



JIPEL

NYU Journal of Intellectual Property
& Entertainment Law

ARTICLES

Encryption and the Press Clause *D. Victoria Baranetsky*

A Framework for Assessing Technology
Hubs in Africa *Jeremy de Beer, Paula Millar,
Jacqueline Mwangi, Victor Nzomo and Isaac Rutenberg*

NOTES

Through the Looking Glass: Photography
and the Idea Expression Dichotomy *Carl Mazurek*

Tomato, Tamatie? Revising the Doctrine of Foreign
Equivalents in American Trademark Law *Thomas Merante*

Fitting Marrakesh into a Consequentialist Copyright
Framework *Neil Yap*

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
& ENTERTAINMENT LAW

VOLUME 6

NUMBER 2

CONTENTS

Preface.....	v
2016-2017 Journal Awards	vi

ARTICLES

Encryption and the Press Clause	179
<i>D. Victoria Baranetsky</i>	
A Framework for Assessing Technology Hubs in Africa.....	237
<i>Jeremy de Beer, Paula Millar, Jacqueline Mwangi, Victor Nzomo, and Isaac Rutenberg</i>	

NOTES

Through the Looking Glass: Photography and the Idea Expression Dichotomy	278
<i>Carl Mazurek</i>	
Tomato, Tamatie? Revising the Doctrine of Foreign Equivalents in American Trademark Law	310
<i>Thomas Merante</i>	
Fitting Marrakesh into a Consequentialist Copyright Framework	351
<i>Neil Yap</i>	

EVENTS

2017 JIPEL Symposium: Careers in IP	383
---	-----



Statement of Purpose

Consistent with its unique development, The New York University Journal of Intellectual Property & Entertainment Law (JIPEL) is a nonpartisan periodical specializing in the analysis of timely and cutting-edge topics in the world of intellectual property and entertainment law. As NYU's first and only online journal, JIPEL also provides an opportunity for discourse through comments from all of its readers. There are no subscriptions, or subscription fees; in keeping with the open-access and free discourse goals of the students responsible for JIPEL's existence, the content is available for free to anyone interested in intellectual property and entertainment law.

The *New York University Journal of Intellectual Property & Entertainment Law* is published two times per year at the New York University School of Law, 139 MacDougal Street, New York, New York, 10012. In keeping with the Journal's open access and free discourse goals subscriptions are free of charge and can be accessed via www.jipel.law.nyu.edu. Inquiries may be made via telephone (212-998-6101) or electronic mail (law.jipel@gmail.com).

The Journal invites authors to submit pieces for publication consideration. Footnotes and citations should follow the rules set forth in the latest edition of *The Bluebook A Uniform System of Citation*. All pieces submitted become the property of the Journal. We review submissions through ExpressO Bepress (<http://law.bepress.com/expresso/>) and through electronic mail (submissions.jipel@gmail.com).

All works copyright © 2017 by the author, except when otherwise expressly indicated. For permission to reprint a piece or any portion thereof, please contact the journal in writing. Except as otherwise provided, the author of each work in this issue has granted permission for copies of that article to be made for classroom use, provided that (1) copies are distributed to students free of cost, (2) the author and the Journal are identified on each copy, and (3) proper notice of copyright is affixed to each copy.

A nonpartisan periodical, the Journal is committed to presenting diverse views on intellectual property and entertainment law. Accordingly, the opinions and affiliations of the authors presented herein do not necessarily reflect those of the Journal members.

The Journal is also available on WESTLAW, LEXIS-NEXIS and HeinOnline.

NEW YORK UNIVERSITY

JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

EDITORIAL BOARD 2016-2017

Editor-In-Chief
THOMAS C. MERANTE

Senior Articles Editors
HILDA LI
INZER NI

Executive Editor
JACKSON C. YATES

Managing Editors
MATTHEW DUNAY
CARL MAZUREK

Senior Notes Editor
JULIAN PYMENTO

Senior Editors
DAVID BAILEY
NAADIA CHOWDHURY
VICTOR GUO
SARAH HIGGINS
SHREYAS KALE
VICTORIA KIM
AMY ROSENTHAL
MARK SANCHEZ
ALEX SCHINDLER

Senior Web Editors
CHRISTINE CHEN
BRECK WILMOT

Staff Editors

NELSON ADAMS
VLADIMIR ALEXANDER
YASMIN AMIN-REIMER
LAURA ARANGO
MANUEL AYARRA
JORDAN CHAFETZ
CYNTHIA CHANG
SOUNG YOUNG CHOE
DEVANSHI DESAI
COREY GOLDSTEIN
CAROLINE HERALD
VINCENT HONRUBIA
WOO TAEK HWANG

KRISTEN IGLESIAS
HYUNJONG JIN
JOEL JUNG
LAUREN KRAKOVYAK
ALEXIS KUO
WILLIAM LEE
MEGAN LEVY
JIE LI
JIQIANG LIN
THOMAS LOY
DENNIS MAI
YANA MANEVICH
ASHLEY MARTIN
ADAM OFMAN

HANIEL OGBURU-OGBONNAYA
CHARLES PESKOWITZ
HANNAH PHAM
JOSHUA POLEVOY
SYDNEY SCHEIN
JINHWA SEO
MANNAT SHARMA
ADAM STEIN
JUN TONG
MILAGROS VILLALOBOS NAVAS
GIA WAKIL
NEIL YAP
ERIC ZWILLING

Faculty Advisors
AMY ADLER
BARTON BEEBE

PREFACE

Our spring issue reminds us of the amazing breadth and influence of intellectual property and innovation law around the globe.

First, Victoria Baranetsky parses the language and history of the United States Constitution's Press Clause to argue for stronger protection of information privacy. We learn that the Founders' own 'encryption' analogues provided a significant philosophical backdrop for the drafting of the Press Clause. The article argues that particularly in a period of heightened surveillance and information sharing, the Press Clause offers vital protection for encrypted information.

Next, Jeremy de Beer, Paula Millar, Jacqueline Mwangi, Victor Nzomo, and Isaac Rutenberg survey Africa's technological terrain and present a framework for the analysis of technology hub impact on the continent. By synthesizing academic literature, practice- and policy-oriented literature, and field experience, the authors offer a robust scheme for promoting technological growth and connectivity.

Three notes advance the global scope of the issue. They invite the reader to consider the role of copyright law in producing works for the visually impaired under the Marrakesh Treaty, the relationship between communicative branding objectives and non-English words in U.S. trademark law, and the fit between copyright law's idea/expression dichotomy and photographic works.

Finally, we include a brief photographic review of *JIPEL*'s 2017 Careers in IP Symposium, which brought together students, faculty, and distinguished practitioners for an engaging discussion covering a wide range of IP career-oriented topics.

I hope that you find this issue engaging and illuminating. On behalf of the 2016-2017 *JIPEL* board, thank you for reading.

Sincerely,

Thomas C. Merante

Editor-in-Chief

NYU Journal of Intellectual Property & Entertainment Law

JOURNAL OF INTELLECTUAL
PROPERTY & ENTERTAINMENT LAW
AWARD WINNERS
2016-2017

OUTSTANDING NOTE

Julian Pymonto

OUTSTANDING 3L EDITOR

Thomas Merante

OUTSTANDING 2L EDITORS

Gia Wakil

Jiqiang Lin

OUTSTANDING SENIOR EDITOR

Naadia Chowdhury

AWARD IN EXCELLENCE

Matthew Dunay

Julian Pymonto



NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 6

SPRING 2017

NUMBER 2

ENCRYPTION AND THE PRESS CLAUSE

D. VICTORIA BARANETSKY*

Almost twenty years ago, a hostile debate over whether government could regulate encryption—later named the Crypto Wars—seized the country. At the center of this debate stirred one simple question: is encryption protected speech? This issue touched all branches of government percolating from Congress, to the President, and eventually to the federal courts. In a waterfall of cases, several United States Court of Appeals appeared to reach a consensus that encryption was protected speech under the First Amendment, and with that the Crypto Wars appeared to be over, until now.

Nearly twenty years later, the Crypto Wars have returned. Following recent mass shootings, law enforcement has once again questioned the legal protection for encryption and tried to implement “backdoor” techniques to access messages sent over encrypted channels. In the case, Apple v. FBI, the agency tried to compel Apple to grant access to the iPhone of a San Bernardino shooter. The case was never decided, but the legal arguments briefed before the court were essentially the same as they were two decades prior. Apple and amici supporting the company argued that encryption was protected speech.

While these arguments remain convincing, circumstances have changed in ways that should be reflected in the legal doctrines that lawyers use. Unlike twenty years ago, today surveillance is ubiquitous, and the need for encryption is no longer felt by a seldom few. Encryption has become necessary for even the most basic exchange of information given that most Americans share “nearly every aspect of

* First Look Fellow at the Reporters Committee for Freedom of the Press, J.D. 2011, Harvard Law School. All reflections and opinions are, of course, my own and do not speak for anyone else or any other entity, including The Reporters Committee for Freedom of the Press. For helpful conversations, comments, and support, I thank Riana Pfefferkorn, David McCraw, Pam Samuelson, Michael Froomkin, Jeremy Dolan, and participants at the Yale Law School Freedom of Expression Conference. Special thanks also to the tireless support of the N.Y.U. Journal of Intellectual Property & Entertainment Law editors and staff.

their lives—from the mundane to the intimate” over the Internet, as stated in a recent Supreme Court opinion.¹

Given these developments, lawyers might consider a new justification under the Press Clause. In addition to the many doctrinal concerns that exist with protection under the Speech Clause, the Press Clause is normatively and descriptively more accurate at protecting encryption as a tool for secure communication without fear of government surveillance. This Article outlines that framework by examining the historical and theoretical transformation of the Press Clause since its inception.

INTRODUCTION.....	181
I. THE PRESS CLAUSE: THE FOUNDERS’ INTERPRETATION.....	185
A. <i>The Framers’ Understanding</i>	185
II. THE PRESS CLAUSE: THE TRADITIONAL DEBATE.....	198
A. <i>Freedom of the Press: An Institutional Protection</i>	198
B. <i>Freedom of the Press: An Individual Protection of Tools</i>	204
III. ENCRYPTION AND THE PRESS CLAUSE	207
A. <i>Short History of Encryption and Government Control</i>	208
B. <i>The Original Crypto Wars: 1960-2000</i>	211
C. <i>Today’s Crypto Wars and Bernstein</i>	223
D. <i>Encryption Under the Press Clause</i>	226
1. <i>Press Clause is Descriptively Accurate at Protecting Encryption as a Tool</i>	227
2. <i>Press Clause Underscores the Structural Role of Encryption</i>	229
3. <i>The Press Clause is Especially Apt at Protecting Encryption Even if Countervailing National Security Exist</i>	231
4. <i>The Press Clause Mirrors Practices of Founding Fathers</i>	232
CONCLUSION	235

¹ Riley v. California, 134 S. Ct. 2473, 2490 (2014).

What is the liberty of the press? Who can give it any definition, which would not leave the utmost latitude for evasion? I hold it to be impracticable; and from this I infer, that its security, whatever fine declarations may be inserted in any constitution respecting it, must altogether depend on public opinion, and on the general spirit of the people and of the government. – Alexander Hamilton, Federalist 54.

The telegraph, telephone, radio, and especially the computer have put everyone on the globe within earshot We think we're whispering, but we're really broadcasting. – Steven Levy.

INTRODUCTION

The First Amendment houses the Press Clause, one of the most important clauses in the United States Constitution. The Amendment states, “Congress shall make no law . . . abridging the freedom of speech, *or of the press*.”² While the neighboring Speech Clause is the dominant provision under which courts have protected encryption, the Press Clause appears to offer an even stronger normative and descriptive rationale for this tool because of the Founders’ intention that the Press Clause be used as a structural protection against government control over information. Indeed, the Founders established the freedom of the press in the newly-formed country to prevent the type of abuse practiced by the British Crown for centuries.³ Analogizing the Founders’ reasoning for the Press Clause to the present-day needs for protecting encryption reveals a novel justification that could be employed in future encryption cases.

Unfortunately, after nearly 100 years of jurisprudence, the Press Clause is still often treated as a supportive afterthought to the Speech Clause. This disregard has largely been borne out of the confusion over what the word “press” means. Traditionally, the “press” has meant the institution of the news media, composed of professional journalists who act as government watchdogs. Many courts and

* First Look Fellow at the Reporters Committee for Freedom of the Press, J.D. 2011, Harvard Law School. All reflections and opinions are, of course, my own and do not speak for anyone else or any other entity, including The Reporters Committee for Freedom of the Press. For helpful conversations, comments, and support, I thank Riana Pfefferkorn, David McCraw, Pam Samuelson, Michael Froomkin, Jeremy Dolan, and participants at the Yale Law School Freedom of Expression Conference. Special thanks also to the tireless support of the N.Y.U. Journal of Intellectual Property & Entertainment Law editors and staff.

² U.S. CONST. amend. I (emphasis added).

³ See, e.g., *Entick v. Carrington*, 19 How. St. Tr. 1029 (1765) (dismissing a general warrant against a dissenting printer); *Wilkes v. Wood*, 19 How. St. Tr. 1153 (1763) (same).

academics have endorsed this press-as-industry approach.⁴ More recently, courts and academics have adopted the “press-as-technology” approach. Under this interpretation, the Press Clause protects individuals’ use of tools, like the printing press, that help create, populate, and distribute information.⁵ Justice Scalia most recently employed this interpretation in dicta of the infamous Supreme Court case *Citizens United*.⁶ There, he diminished the press-as-industry approach, and endorsed the broader interpretation of freedom of the press as a right of all citizens to communicate with one another.⁷ As Professor Michael McConnell has written, this

⁴ See LEE BOLLINGER, *IMAGES OF A FREE PRESS* 20 (1991) (stating that “the government is untrustworthy when it comes to regulating public debate, for it will forever try to recapture its authoritarian powers” and the press is “the public’s representative, its agent, helping stand guard against the atavistic tendencies of the state”); LEONARD W. LEVY, *EMERGENCE OF A FREE PRESS*, at xii (1985) (concluding that the Framers believed press freedom “meant that the press had achieved a special status as an unofficial fourth branch of government, ‘the Fourth Estate’”); Floyd Abrams, *The Press Is Different: Reflections on Justice Stewart and the Autonomous Press*, 7 HOFSTRA L. REV. 563, 592 (1979) (stating that, at times, the press manages to “serve[] as a vigilant protector of the public from its government”); David A. Anderson, *The Press and Democratic Dialogue*, 127 HARV. L. REV. 331, 334 (2014) (arguing that “the press is one of the entities that usefully serve these functions [as a check on government overreach], and is the one the Framers saw fit to recognize”); Vincent Blasi, *The Checking Value in First Amendment Theory*, 1977 AM. B. FOUND. RES. J. 521, 538 (1977); Potter Stewart, *Or of the Press*, 26 HASTINGS L.J. 631, 634 (1975) (“The primary purpose of the constitutional guarantee of a free press was a similar one: to create a fourth institution outside the Government as an additional check on the three official branches.”).

⁵ See *Citizens United v. Fed. Election Comm’n*, 558 U.S. 310, 390-91 n.6 (2010) (Scalia, J., concurring) (“It is passing strange to interpret the phrase ‘the freedom of speech, or of the press’ to mean, not everyone’s right to speak or publish, but rather everyone’s right to speak or the institutional press’s right to publish.”); *First Nat’l Bank of Boston v. Bellotti*, 435 U.S. 765, 800 n.5 (1978) (Burger, C.J., concurring) (arguing that “‘press,’ the word for what was then the sole means of broad dissemination of ideas and news, would be used to describe the freedom to communicate with a large, unseen audience,” even using new technologies that were not known to the Framers); David Lange, *The Speech and Press Clauses*, 23 UCLA L. REV. 77, 90 (1975) (arguing the notion that it is unlikely that the Framers intended to protect modern journalists); Michael W. McConnell, *Reconsidering Citizens United as a Press Clause Case*, 123 YALE L.J. 412, 454 (2013) (describing the freedom of the press as the right to “disseminat[e] . . . opinion or information to the public through media or communications”); Eugene Volokh, *Freedom for the Press as an Industry, or for the Press as a Technology? From the Framing to Today*, 160 U. PA. L. REV. 459, 505 (2012) (“Under the mass-communications-more-protected view, the Free Press Clause provides special protection to all users of the press-as-technology.”).

⁶ See *Citizens United*, 558 U.S. at 342.

⁷ *Id.* at 391 n.6 (Scalia, J., concurring).

interpretation of the Clause protects “the right of any person to use the technology of the press to disseminate opinions.”⁸

At the outset, it is important to state that this Article does not wholly endorse⁹ either interpretation of the Press Clause.¹⁰ Instead, this Article focuses on the more fundamental first order question: what is the *purpose* of the Press Clause? This Article attempts to answer that question by examining the Founders’ intentions through an examination of historical records. Through this analysis, this Article determines that the likely purpose behind the Press Clause was not specifically aimed at individuals or professional journalists.¹¹ Instead, the Clause was targeted at government. The Press Clause was intended to provide structural protection *against* arrant state control over the *flow of information*.¹² Zooming out of the press-as-technology versus press-as-industry debate reveals a broader interpretation of the Press Clause as a constitutional protection against state censorship by securing

⁸ McConnell, *supra* note 5, at 441; *see also* David B. Sentelle, *Freedom of the Press: A Liberty for All or a Privilege for a Few?*, 2014 CATO SUP. CT. REV. 15, 24 (2014) (“The original meaning of ‘the press,’ then, was not limited to an institution called ‘the press.’”).

⁹ Although this article affirms parts the press-as-technology approach, it does not endorse the holding of *Citizens United*. *See Citizens United*, 558 U.S. at 342 (stating that “political speech does not lose First Amendment protection ‘simply because its source is a corporation’” (quoting *Bellotti*, 435 U.S. at 784)). As Dean Robert Post has said, the *Citizens United* Court’s decision is both “fanciful and baffling.” ROBERT C. POST, *CITIZENS DIVIDED: CAMPAIGN FINANCE REFORM AND THE CONSTITUTION* 71 n.* (2014); *see also* Randall P. Bezanson, *Whither Freedom of the Press?*, 97 IOWA L. REV. 1259, 1263 (2012) (“The second issue raised by the *Citizens United* language is that its conclusion that the free-press guarantee of the First Amendment affords no greater or different protection to the press is almost offhanded.”).

¹⁰ Although this article may endorse elements of the press-as-technology interpretation of the Press Clause, it does not intend to make any further comment on the meaning of *Citizens United*. For the author’s other writings on *Citizens United*, *see* Victoria Baranetsky, *The Economic Liberty Approach of the First Amendment: A Story of American Booksellers v. Hudnut*, 47 HARV. C.R.—C.L. L. REV. 169 (2012).

¹¹ *See* Anderson, *supra* note 4, at 334 (noting that determining what “press” under the Press Clause means should “develop incrementally; it is unrealistic to expect its constitutional meaning to emerge full-blown”).

¹² *See* Sonja R. West, *Press Exceptionalism*, 127 HARV. L. REV. 2434, 2436 (2014) (describing the Supreme Court’s different treatment of the Speech and Press Clauses); Sonja R. West, *The ‘Press’ Then & Now*, 77 OH. STATE L. REV. 49, 54, 67 (2016) (stating the “emphasis on the structural function is found in the early documents, which repeatedly hailed press freedom” as “defend[ing] and protect[ing] the people and the republic,” and stating it “was rarely discussed as a matter of individual expressive value” or “means to individual self-fulfillment or self realization”).

channels of communication for individuals.¹³ This is distinguishable from the positive right of an individual *to speak*, as is ensured under the Speech Clause.¹⁴

After establishing this point, this Article then finds that encryption falls squarely within this protection of the Press Clause. Encryption¹⁵—a tool for concealing information—obstructs government from controlling access to certain information.¹⁶ By increasing trust among citizens, encryption enables speakers to freely express themselves unafraid of government retribution.¹⁷ In this way, encryption is like a modern day printing press, increasing individual control over the spread of information to different audiences at different rates without fear of government intrusion.¹⁸

Moving forward, Section I of this Article will provide historical support¹⁹ to assert that the Founder’s ultimate and distinct purpose for including the Press Clause was to build a structural protection against government control over methods of

¹³ This argument is akin to the argument made by Vincent Blasi. *See* Blasi, *supra* note 4, at 606 (“These news sources play a unique role in the checking process because they sometimes have access to inside information relating to the misconduct of public officials—information of the highest possible significance under the checking value.”).

¹⁴ The Speech Clause does not, however, only protect speech in its positive aspect. *See* W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624 (1943) (The right of freedom of thought and of religion as guaranteed by the Constitution against State action includes both the right to speak freely and the right to refrain from speaking at all); *Wooley v. Maynard*, 430 U.S. 705 (1977) (recognizing the same).

¹⁵ Note that there are various definitions of encryption.

¹⁶ *See, e.g.,* Joseph Menn, *Exclusive: Yahoo Secretly Scanned Customer Emails for U.S. Intelligence—Sources*, REUTERS (Oct. 4, 2016), <http://www.reuters.com/article/us-yahoo-nsa-exclusive-idUSKCN1241YT>.

¹⁷ *Cf.* American Civil Liberties Union and Human Rights Watch, *With Liberty to Monitor All*, HUMAN RIGHTS WATCH (July 28, 2014), <https://www.hrw.org/report/2014/07/28/liberty-monitor-all/how-large-scale-us-surveillance-harming-journalism-law-and>.

¹⁸ *See* Susan McGregor, *Digital Security and Source Protection for Journalists*, TOW CENTER FOR DIGITAL JOURNALISM (July 16, 2014), <http://towcenter.org/digital-security-and-source-protection-for-journalists-research-by-susan-mcgregor/>.

¹⁹ *See generally* Patrick J. Charles & Kevin Francis O’Neill, *Saving the Press Clause from Ruin: The Customary Origins of a “Free Press” as Interface to the Present and Future*, 2012 UTAH L. REV. 1691, 1693 n.6, 1703 (discussing the dearth of historical analysis of the Press Clause in the writings of esteemed press scholars like David Anderson, Eugene Volokh, and Sonja West, but also stating that despite such dearth, “[i]nterpreting the Constitution through the events of the American Revolution . . . is crucial to understanding the evolution of eighteenth-century political and constitutional thought”—and that this is particularly important for the press clause because “[u]nlike most Anglo-American rights, the development of a free press stems from customary practice”).

disseminating information.²⁰ Having understood this fundamental point, Section II will then discuss how both interpretations (press-as-industry and press-as-technology) agree on and start from this fundamental point.²¹ Finally, Section III considers how the Press Clause, under either interpretation, more soundly protects encryption than the Speech Clause.²²

I

THE PRESS CLAUSE: THE FOUNDERS' INTERPRETATION

The Press Clause has an essential role within our Constitutional order. As discussed below, many of the Founders believed the Clause to be the most important privilege within the Bill of Rights.²³ Over the past half-century, however, the clause has lost this prestige.²⁴ In many ways this precipitous fall is due to the divide caused by warring interpretations of the meaning of “the press.” Moving past this bifurcation, this Section attempts to uncover a more unified, foundational account of the Clause as a structural protection for citizens to control where and how one may share information. It argues that the Founders crafted the Press Clause to provide protection for physical spaces, such as the printer’s office, one’s own home, and other spheres where information tools could be leveraged without government intrusion, an interpretation with which both sides of the debate might agree. Today, that protection can be broadened to other zones like smartphones, encryption, and similar tools that capture our most private communications.

A. The Framers’ Understanding

The Framers understood freedom of the press to be one of, if not, *the most* vital constitutional protections.²⁵ James Madison, the drafter of the First Amendment, called the liberty of the press “inviolable” and the “choicest privileges of the people.”²⁶ Patrick Henry named “liberty of the press” as one of two “peak

²⁰ See *infra* Part I.

²¹ See *infra* Part II.

²² See *infra* Part III.

²³ JEFFREY A SMITH, PRINTERS AND PRESS FREEDOM: THE IDEOLOGY OF EARLY AMERICAN JOURNALISM 69 (1988) (quoting Madison as stating that “freedom of the press and rights of conscience” are the “choicest privileges of the people,” and that Jefferson said that liberties such as freedom of the press should be protected ‘in all cases’ rather than none”).

²⁴ See David Anderson, *Freedom of the Press*, 80 TEX. L. REV. 429, 448 (2002) (referring to the period between the 1930s and 1960s as “the heyday of the Press Clause in the Supreme Court”).

²⁵ See SMITH, *supra* note 23, at 68 (stating “two of the most populous states, Virginia and New York, demanded a press amendment and were joined by North Carolina, which refused to approve the Constitution until a bill of rights was provided”); see also Blasi, *supra* note 4, at 527.

²⁶ West, *The ‘Press’ Then & Now*, *supra* note 12, at 23 (citing SMITH, *supra* note 23, at 166).

concerns” for the Bill of Rights (along with trial by jury).²⁷ Thomas Jefferson also distinguished having free presses as the chief component of a healthy government.²⁸

Its importance was borne out of its structural role.²⁹ As Justice Stewart explained, the primacy of the Press Clause is because it is “a structural provision” operating “outside the Government as an additional check on the three official branches.”³⁰ Vincent Blasi has expounded that this “checking value” of freedom of the press makes the Clause the most powerful and dangerous constitutional privilege, informing the structure of democracy.³¹ “It is a value grounded in distrust of government,” wrote Lucas Powe, that “assumes a darker side of human nature and holds that those who wield governmental power will be prone to overreaching, and thus that it is essential to provide information for a resisting citizenry.”³² Thus, the freedom of press has long been described as a resistance mechanism against state power.

Both interpretations of the Press Clause have acknowledged its structural role as removing government from the private realm. For instance, press-as-industry academics acknowledge the structural role of the Clause as protecting an entirely separate institution of the Fourth Estate—an independent watchdog of the government.³³ Similarly, press-as-technology academics have expanded the structural protection beyond the news media³⁴ to all individuals’ property where

²⁷ NEIL H. COGAN, *THE COMPLETE BILL OF RIGHTS* 642 (1997) (quoting Patrick Henry naming several protections necessary for the Bill of Rights, but leaving out the freedom of speech).

²⁸ See LEONARD LEVY, *LEGACY OF SUPPRESSION: FREEDOM OF SPEECH AND PRESS IN EARLY AMERICAN HISTORY* 300 (1960) (quoting Thomas Jefferson for his belief that only lying offenders should be prosecuted).

²⁹ See SMITH, *supra* note 23, at 70 (quoting Madison for stating that the Bill of Rights would “prevent abuses of power, would satisfy the public, and would contribute to political stability” and that through the amendments “[the] people shall not be deprived or abridged of their right to speak, to write or to publish their sentiments; and the freedom of the press, [w]as one of the great bulwarks of liberty, shall be viable”); *id.* at vii (stating “[t]he basis of this structural rationale for press freedom lies in the political and intellectual experiences of eighteenth-century America”); see also West, *The ‘Press’ Then & Now*, *supra* note 12, at 67 (stating “the evidence suggests [the framing generation] believed the structural function to be of paramount importance”).

³⁰ Stewart, *supra* note 4, at 634.

³¹ Blasi, *supra* note 4, at 527.

³² L.A. SCOTT POWE, *THE FOURTH ESTATE AND THE CONSTITUTION: FREEDOM ON THE PRESS IN AMERICA* 238 (1992).

³³ See *id.* at 260-61 (noting Burke’s characterization of the press as the “Fourth Estate”).

³⁴ See Volokh, *supra* note 5, at 469 (“It seems unlikely that the Framers would have secured a special right limited to this small industry, an industry that included only part of the major contributors to public debate.”); see also Lange, *supra* note 5, at 90 (arguing that the notion that

channels of information were developed. Under this analysis, the Press Clause protects against government intrusions into homes and offices in which citizens use technology to combat government propaganda.³⁵ It also protects the private tools and technology used in these spaces, such as the printing press and pamphlets.³⁶ Under both approaches, the structural role was intended by the Founders to be particularly protective in circumstances where it could “effectively expose deception in government.”³⁷

These interpretations lean on a long history of writers who experienced oppressive regulations on printing stemming as far back as the sixteenth century in

the Framers intended to protect modern journalists is unpersuasive, in part, because he said that the partisan press of the day “bore little relationship to . . . the press of Hearst and Pulitzer”).

³⁵ See *Entick*, 19 How. St. Tr. 1029; *Wilkes*, 19 How. St. Tr. 1153.

³⁶ See Jasper L. Tran, *Press Clause and 3D Printing*, 14 NW. J. TECH. & INTELL. PROP. 75, 79 (2016) (arguing that 3D printers are “the modern equivalent of the printing press” protected by the First Amendment); see also Volokh, *supra* note 5, at 462 n.10 (“I speak here of communications technologies that today serve the role the printing press did in the 1700s, not just of the printing press as such.”); see also *id.* (citing *First Nat’l Bank of Boston v. Bellotti*, 435 U.S. 765, 800 n.5 (1978) (Burger, C.J., concurring) (stating “[i]t is not strange that ‘press,’ the word for what was then the sole means of broad dissemination of ideas and news, would be used to describe the freedom to communicate with a large, unseen audience,” even using new technologies that were not known to the Framers)); cf. Charles & O’Neill, *supra* note 19, at 1701 (“With only one publishing technology available circa 1791, it is impossible to ascertain how the founding generation viewed the Press Clause as an evolving technological right of the people to employ free speech. Are we to believe the founding generation had the foresight to predict other popular publishing mediums such as radio, television, and the Internet? The answer remains no.”).

³⁷ Andrew Bradford, *Sentiments on the Liberty of the Press*, AM. WKLY. MERCURY (Phila.), Apr. 25, 1734, reprinted in FREEDOM OF THE PRESS FROM ZENGER TO JEFFERSON 38, 41-42 (Leonard W. Levy ed., 1966) (discussing freedom of the Press as the “great Palladium of all our other Liberties” because it is the “Liberty of detecting the wicked and destructive Measures of certain Politicians; of dragging Villany out of its obscure lurking Holes, and exposing it in its full Deformity to open Day; of attacking Wickedness in high Places, of disentangling the intricate Folds of a wicked and corrupt Administration, and pleading freely for a redress of Grievances”); see also *Near v. Minnesota*, 283 U.S. 697, 719-20 (1931) (“Meanwhile, the administration of government has become more complex, the opportunities for malfeasance and corruption have multiplied, crime has grown to most serious proportions, and the danger of its protection by unfaithful officials and of the impairment of the fundamental security of life and property by criminal alliances and official neglect, emphasizes the primary need of a vigilant and courageous press, especially in great cities. The fact that the liberty of the press may be abused by miscreant purveyors of scandal does not make any the less necessary the immunity of the press from previous restraint in dealing with official misconduct.”); *New York Times Co. v. United States*, 403 U.S. 717 (1971) (Black, J. concurring).

Great Britain.³⁸ While many credit eighteenth century thinkers like William Blackstone, Thomas Gordon, and John Trenchard with establishing the Anglo-American origins of freedom of the press, they overlook that these theorists, who heavily influenced the founding generation, borrowed their views from earlier writers who experienced firsthand brutal punishment from the British government for using early private printing presses.³⁹ Perhaps it was because this initial cadre of writers received brutal punishment for sharing information that they were able to highlight the importance of freedom of the press so precisely.

During the seventeenth century, the Crown imposed corporal and even capital punishment on those who used new printing technology of their own accord. Indeed, “[s]hortly after the first printing press arrived at Westminster in 1476, the Crown established a primitive schem[a]” of censorship for printing materials.⁴⁰ At the time, the freedom of the press had not been codified—not in the Magna Carta, nor the 1689 English Bill of Rights, nor the 1701 Act of Settlement.⁴¹ Because of this lacuna, the Tudors were easily able to establish a “whole machinery of censorship and control” that resulted in some of the most violent practices to individuals.⁴² Historian Frederick Siebert traces the rapid rise of regulation over information tools during the reigns of Henry VIII to Elizabeth I, including a patchwork of proclamations, patents, trade regulations, judicial decrees, and Privy Council and parliamentary actions—all penalizing individuals for using printing presses and other tools in their homes.⁴³

Siebert writes that the most punishing edicts occurred under the reign of Elizabeth I (1533-1603), through a regulation called the Star Chambers Decree of

³⁸ See Edward S. Lee, *Freedom of the Press 2.0*, 42 GA. L. REV. 309, 320 (2008) (“The protection for an individual’s use of the printing press—free of intrusive governmental regulation—was a response to the repressive regime of strict regulation of the press that enabled the Crown and later Parliament to control the production of all printed materials in England from the 1500s until the early 1700s.”); see generally MARK ROSE, *AUTHORS AND OWNERS* 12 (1993); FREDERICK S. SIEBERT, *FREEDOM OF THE PRESS IN ENGLAND* 346-92 (1965); Edward A. Bloom, *Neoclassic “Paper Wars” for a Free Press*, 56 MOD. LANGUAGE REV. 481 (1961); Douglas M. Ford, *The Growth of the Freedom of the Press*, 4 ENG. HIST. REV. 1 (1889); Charles & O’Neill, *supra* note 19, at 1703 (“Unlike most Anglo-American rights, the development of a free press stems from customary practice” during the eighteenth century, “the bowels of the print culture itself.”).

³⁹ See Charles & O’Neill, *supra* note 19, at 1703.

⁴⁰ Michael W. Price, *Rethinking Privacy: Fourth Amendment ‘Papers’ And the Third-Party Doctrine*, 8 J. NAT’L SEC. L. POL’Y 247, 251 (2016).

⁴¹ *Id.*

⁴² CYNDIA S. CLEGG, *PRESS CENSORSHIP IN ELIZABETHAN ENGLAND* 3 (1997) (citing SIEBERT, *supra* note 38); SIEBERT, *supra* note 38, at 5.

⁴³ See SIEBERT, *supra* note 38, at 47-63

1586 and the Stationers' Company.⁴⁴ In 1643, Parliament established a consortium of printers, called the Stationers' Company, which was permitted by license to maintain a monopoly on printing in exchange for suppressing licentious remarks about the government.⁴⁵ The Company had complete control over the "right to copy" and release information.⁴⁶ Any person trying to operate outside this system was immediately disciplined. Non-government sanctioned books or pamphlets were marked "seditious" or "offensive" and the Decree made unofficial presses "subject to warrantless searches" that could "result[] in destruction of the nonconforming printing press."⁴⁷ To enforce these rules, government surveillance over printing extended to searches of "people's houses to confiscate illegal presses and materials."⁴⁸

These draconian Elizabethan rules worsened during the Stuart kings of the seventeenth century. In 1662 Parliament, passed the Printing Act which permitted enforcement of the Stationer system by "death or otherwise"⁴⁹ and decreased the number of Stationer printers in England to just twenty.⁵⁰ In 1683, Algernon Sydney, a member of Parliament and an outspoken critic of King Charles of England, was executed for his unpublished writings "found in [the privacy of] his home."⁵¹ Sydney's claim that the writings were never distributed nor intended for publication made little difference to the government.⁵² The mere fact that Sydney challenged the government's regime by privately possessing materials was by itself punishable by death. For the decades that followed, the British government prosecuted individuals like Sydney to obtain complete control over the flow of information.

During that time several important political thinkers publicly criticized the Crown's tyrannical actions in writings that would later come to influence the drafters of the First Amendment. For example, just sixteen months after Parliament passed the Star Chambers Act, poet and political theorist John Milton penned *Areopagitica*, as the author explained, "to deliver the press from restraints with which it was

⁴⁴ See Lee, *supra* note 38, at 315-16, 339-56.

⁴⁵ See SIEBERT, *supra* note 38, at 47-63.

⁴⁶ *Id.*

⁴⁷ Lee, *supra* 38, at 321 (referencing the Star Chamber Ordinance of 1586, at §§ 1-2).

⁴⁸ *Id.* at 315-16, 339-56.

⁴⁹ *Id.*

⁵⁰ See *id.*

⁵¹ *Id.*

⁵² See NANCY C. CORNWELL, FREEDOM OF THE PRESS RIGHTS AND LIBERTIES UNDER THE LAW 24 (2004).

encumbered.”⁵³ The freedom of the press, he argued, had the “power of determining what was true and what was false, what ought to be published and what [ought] to be suppressed.”⁵⁴ Ultimately, Milton argued, truth depended on the people, not the government, controlling what information could be published and disseminated expeditiously.

Milton was certainly not alone.⁵⁵ His contemporaries, similarly influential to the Founders, including Henry Robinson, William Walwyn, Roger Williams, John Lilburne, John Saltmarsh, and John Goodwin, also expounded on the importance of freedom of the press. Robinson, for instance, advocated for the “free trading of truth”, and wrote, referring to the printing press, that “no man can have a natural monopoly of [it].”⁵⁶ Walwyn, a pamphleteer himself and a central figure of the Levellers movement (which advocated for popular sovereignty during the English Civil War) underscored that the press should be free from government “for any man” not just those licensed by government.⁵⁷ In particular, Walwyn wrote that this also meant “freedom for [all printing] materials.”⁵⁸

While Parliament eventually abolished the Star Chamber and the Stationers’ Company, the prohibition against seditious libel remained “alive and well in English common law,” as did the “practice of issuing ‘general warrants’ to search and seize paper.”⁵⁹ At this time three writers—John Wilkes, Father Candor and Junius—would have an especially important role in developing political thought on freedom of the press that would come to influence the Founders that the freedom to use information tools was an important structural protection within the Constitution.⁶⁰ Among these three, the most important was arguably that of English political activist John Wilkes.⁶¹

⁵³ MILTON, THE SECOND DEFENSE OF THE PEOPLE OF ENGLAND (1654) (discussing *Areopagitica*).

⁵⁴ *Id.*

⁵⁵ See BALACHANDRA RAJAN ET AL., MILTON AND THE CLIMATES OF READING: ESSAYS 137 (2006).

⁵⁶ CORNWELL, *supra* note 52, at 24 (quoting Robinson).

⁵⁷ *Id.* (quoting Walwyn). Nearly a century later, Blackstone similarly adapted this understanding as the liberty of “every freeman [having] undoubted right to lay what sentiments he pleases before the public; to forbid this, is to destroy the freedom of the press” 2 WILLIAM BLACKSTONE, COMMENTARIES *151, *152.

⁵⁸ CORNWELL, *supra* note 52, at 24 (quoting Walwyn).

⁵⁹ Price, *supra* note 40, at 252.

⁶⁰ See Blasi, *supra* note 4.

⁶¹ See LEVY, *supra* note 28, at 145-47.

While John Wilkes was more of a radical than a political theorist, his writings on his personal experience with government censorship and retribution are foundational for understanding the distinct privileges provided under freedom of the press.⁶² Wilkes' difficulties with the Crown started in 1762 when he founded a weekly publication called *The North Briton*. In the forty-fifth issue of the journal, Wilkes wrote stinging attacks on George III.⁶³ The comments prompted the King's ministers to issue general warrants for the arrest and search of some 200 persons involved in the publication.⁶⁴ Ultimately, forty-nine people were arrested, including Wilkes himself.⁶⁵ In addition to the arrests, the government also ransacked Wilkes' home; and "[w]hile the search was nominally justified by charges of sedition, it in fact swept much more broadly."⁶⁶ Lord Halifax ordered that, "all must be taken, manuscripts and all."⁶⁷ According to records, messengers "fetched a sack and filled it" with Wilkes' tools and private papers.⁶⁸ His arrest and seizure of belongings was heavily criticized by the public and newspapers.⁶⁹

Wilkes immediately challenged the warrant, turning the investigation into a salvo for the freedom of the press. In front of a crowd at the Court of Common Pleas, he stated that the case would "teach ministers of arbitrary principles, that the liberty of an English subject is not to be sported away with impunity, in this cruel and despotic manner."⁷⁰ The Court appeared to agree, and ruled the general warrant invalid. But this result was insufficient. Immediately after the criminal charges were dropped, Wilkes brought a civil suit for trespass, which he argued would "determine

⁶² See *id.*

⁶³ See John Wilkes, *The North Briton*, No. 45, Apr. 23, 1763 (London: W. Bingley, 1769), reprinted in THOMAS CURSON HANSARD, *THE PARLIAMENTARY HISTORY OF ENGLAND, FROM THE EARLIEST PERIOD TO THE YEAR 1803*, at 1335 (1813) ("A despotic minister will always endeavour to dazzle the prince with high flown ideas of the prerogative and honour of the crown, which the minister will make a parade of firmly maintaining. I wish as much any man in the kingdom to see the honour of the crown maintained in a manner truly becoming royalty. I lament to see it sunk even to prostitution.").

⁶⁴ See *id.*

⁶⁵ See *id.*

⁶⁶ Price, *supra* note 40.

⁶⁷ *Id.*

⁶⁸ *Id.*

⁶⁹ Roger P. Mellen, *John Wilkes and the Constitutional Right to a Free Press in the United States*, 41 JOURNALISM HIST. 2, 9 (2015) (stating the *Boston Post Boy* of June 1763 admonished "the Hand of Authority being lifted up in order to fall heavily on such Political Writers as may displease").

⁷⁰ Jack Lynch, *Wilkes, Liberty, No. 45*, Colonial Williamsburg, Summer 2003, <http://www.history.org/foundation/journal/summer03/wilkes.cfm>.

at once whether English liberty be a reality or a shadow.”⁷¹ In *Wilkes v. Wood*, Wilkes condemned the Crown’s use of general warrants as enabling the “promulgation of our most private concerns, affairs of the most secret personal nature,” signifying “an outrage to the constitution itself” and likening the government’s behavior to the Spanish Inquisition.⁷² Ultimately, Wilkes prevailed. The jury took thirty minutes to decide in favor of Wilkes and awarded him £1000.⁷³

Scholar Arthur Cash calls the case “a momentous shift in the locus of power in government” from the privileged to the masses.⁷⁴ Energized by Wilkes’ victory, other persons arrested by the warrant sued and won against the government in an unprecedented action.⁷⁵ The case shifted power not merely in terms of who got to publish what information, but also uniquely designated privacy as a main concern for cases involving publishers.⁷⁶ In a subsequent case, *Entick v. Carrington*, involving charges brought against the “very seditious” weekly paper known as the *Monitor*, privacy of a political dissenter’s home and papers became the crux of the case. Investigators had “read over, pryed [sic] into, and examined all [of John Entick’s] private papers, books, etc.[.]” a process the Court “compared to racking his body to come at his secret thoughts.”⁷⁷ Lord Camden, deciding the case, found that Entick’s papers were “his dearest property. . . [and w]here private papers are removed and carried away, the secret nature of those goods will be an aggravation of the trespass, and demand more considerable damages in that respect.”⁷⁸

The cases of Wilkes and Entick generated copious political writing in the American colonies on liberty of the press. In particular, it inspired two anonymous writers, The Father of Candor and Junius, who advocated for freedom of the press as a challenge to government intrusion in the private home. In 1764, the Father of Candor, an eminent public figure and legal thinker, wrote that no gentleman “would rest easy in his bed, if he thought, that . . . he was liable not only to be taken up himself, but every secret of his family made subject to the inspection of a whole

⁷¹ *Id.*

⁷² Price, *supra* note 40.

⁷³ See *Wilkes v. Wood*, (1763) 98 Eng. Rep. 489 (C.P.) 490; Lofft 1, 5.

⁷⁴ *Id.*

⁷⁵ See *Money v. Leach*, (1765) 97 Eng. Rep. 1075 (K.B.) 1077; *Huckle v. Money*, (1763) 95 Eng. Rep. 768 (K.B.).

⁷⁶ See *Wilkes v. Wood*, (1763) 98 Eng. Rep. 489 (C.P.) 490; Lofft 1, 5.

⁷⁷ Price, *supra* note 40, at 254; *id.* (stating the intrusion was “directly aimed at [a] political dissenter[] and political papers”).

⁷⁸ Price, *supra* note 40, at 254.

Secretary of State's office."⁷⁹ He continued, "Everybody has some private papers, that he would not on any account have revealed;" giving government such authority would be "inconsistent with every idea of liberty."⁸⁰ Similarly, Junius, a pseudonym for an author of a series of anonymous letters published in the *Public Advertiser* from 1769-1772, wrote, "The liberty of the press is the palladium of all the civil, political, and religious rights of an Englishman," acknowledging freedom of the press as a gateway to all other rights.⁸¹

Despite these writings, similar events occurred in the colonies that would also influence the Founders' ideas on freedom of the press. For instance, in 1735, the Crown prosecuted Peter Zenger, a New York newspaper printer, for libel after Zenger published a newspaper criticizing New York Governor William Cosby.⁸² The case was a focal point for the Founders and the general public. Andrew Hamilton, serving as Zenger's attorney, strongly admonished the Crown for its actions, and gave a rousing closing argument leading to Zenger's victory and cheers from the courtroom spectators.⁸³ In fact, despite "overwhelming evidence against" Zenger, it took the American jury just ten minutes to acquit him, after which the decision was widely praised in the press.⁸⁴

By the late eighteenth century, the importance of freedom of press and its structural protection against the government had become deeply ingrained within the minds of the Founders. "Opposition thought . . . was devoured by the colonists"⁸⁵ and nowhere more central was the idea of rebellion ingrained than in the Press Clause. The drafters of the First Amendment were "men to whom Wilkes and Junius were household words."⁸⁶ Most commonly their discussions were focused on the free press's "power and its essential characteristic of aggressiveness: its endlessly propulsive tendency to expand itself beyond legitimate boundaries."⁸⁷

⁷⁹ LAURA K. DONAHUE, *THE FUTURE OF FOREIGN INTELLIGENCE: PRIVACY AND SURVEILLANCE IN A DIGITAL AGE* 99 (2016) (quoting John Almon).

⁸⁰ *Id.*

⁸¹ TIM O'MALLEY & CLIVE SOLEY, *REGULATING THE PRESS* 21 (2000).

⁸² See Arthur E. Sutherland, Book Review, 77 HARV. L. REV. 787, 787-88 (1964) (reviewing JAMES ALEXANDER, *A BRIEF NARRATIVE OF THE CASE AND TRIAL OF JOHN PETER ZENGER* (1963)).

⁸³ See *id.*

⁸⁴ See Price, *supra* note 40, at 255.

⁸⁵ WILLIAM H. GOETZMANN, *BEYOND THE REVOLUTION: A HISTORY OF AMERICAN THOUGHT FROM PAINE TO PRAGMATISM* 26 (2009).

⁸⁶ ZACHARIAH CHAFFEE, JR., *FREE SPEECH IN THE UNITED STATES* 21 (1969); see also Blasi, *supra* note 4, at 533.

⁸⁷ BERNARD BAILYN, *THE IDEOLOGICAL ORIGINS OF THE AMERICAN REVOLUTION* 56 (1967).

In addition to the writings of Milton and Father Candor, many of the Founders also closely identified with the practical experiences of Algernon Sydney, John Wilkes, John Entick, and Peter Zenger. Thomas Paine, Benjamin Franklin, Thomas Jefferson, and George Washington were all printers or wrote opposition publications during the Revolution.⁸⁸ The Founders recognized that they would have faced similar charges of treason for using the press and other tools, such as ciphers, had the Revolution ended differently. It is of little surprise then, that in drafting the Bill of Rights, the Founders described that the freedom of the press as a central concern.⁸⁹ Its importance was reflected in their letters and writings, where freedom of the press was repeatedly mentioned as a key, if not the most important protection.⁹⁰

The original state constitutions also illustrate the primacy of the structural protection.⁹¹ Each of the thirteen original states protected freedom of the press in their declarations or constitutions (while only one included protection for free speech).⁹² Out of the eleven state constitutions that adopted a protection for freedom of the press, Pennsylvania's is perhaps most emblematic, containing not one, but two press provisions.⁹³ Its second Press Clause, located in the main section of the constitution, titled "the Plan or Frame of Government for Commonwealth or State of Pennsylvania" described press freedom as structurally essential to a healthy government. It read, "[t]he printing presses shall be free to every person who undertakes to examine the proceedings of the legislature, or any part of government."⁹⁴ As Professor Sonja West has noted, "[t]he placement of this right in the part of the document establishing the state's organizational framework further emphasizes that press freedom filled a specific structural function."⁹⁵

In addition, First Amendment draftsman James Madison elaborated an even stronger structural interpretation of the Press Clause in the Virginia Report of 1799-

⁸⁸ MARK A. LAUSE, *SOME DEGREE OF POWER* 25-26 (1991).

⁸⁹ See SMITH, *supra* note 23, at 162 ("Early American journalists and libertarian theorists distrusted state power and continually argued that the press should serve as a check on its use.").

⁹⁰ West, *The 'Press,' Then & Now*, *supra* note 12, at 62 (citing SMITH, *supra* note 23, at 166).

⁹¹ *Ex parte Milligan*, 71 U.S. 2, 37 (1866) (stating the original constitutions were "framed with the most jealous care").

⁹² Seth F. Kreimer, *The Pennsylvania Constitution's Protection of Free Expression*, 5 U. PA. J. CONST. L. 1, 15 (2004).

⁹³ West, *supra* note 12, at 67 n.95 (citing David A. Anderson, *The Origins of the Press Clause*, 30 UCLA L. REV. 455, 489-90 (1983) (stating that Pennsylvania's second Press Clause, is "unmistakable" evidence of "the right to examine government")).

⁹⁴ Pennsylvania Declaration of Rights, 1776, *reprinted in* West, *The 'Press,' Then & Now*, *supra* note 12, at 66.

⁹⁵ *Id.* at 66-67.

1800.⁹⁶ There, Madison wrote freedom of the press was “[t]he essential difference between the British government, and the American Constitutions,” and “that the unconstitutional power exercised over the press . . . ought more than any other to produce universal alarm.”⁹⁷ Madison’s words harken back to the cases of *Wilkes* and *Entick*, and the principle that freedom of the press did not simply protect criticism of government, but also thwarted government intrusions into physical spaces. He wrote, “[t]he freedom of the press under the common law is . . . an exemption from all previous restraint on printed publications, by persons authorized to inspect and prohibit them.”⁹⁸ In essence, inherent to the freedom of the press was broad protection of the press, including certain zones and tools used for the dissemination of information—zones and tools which, if employed, could rival the government’s control of information.⁹⁹

Despite its long history, the Supreme Court did not expound on the meaning of the Clause until the early twentieth century. However, several instances in the early nineteenth century also suggest a tacit recognition by the federal government of the structural right. For example, during the Civil War, Northern publishers referred to as the “Copperhead Press” hotly opposed President Abraham Lincoln.¹⁰⁰ The group, named after a venomous snake, made repeated vitriolic verbal attacks on the President in the media, believing Lincoln had acted beyond the bounds of his constitutional powers. In December 1862, one Copperhead Congressman “boldly introduced a congressional resolution calling for Lincoln’s imprisonment.”¹⁰¹ The Congressman “had a large public following and strong associations with several newspaper editors”¹⁰² and “energetically pushed the envelope in speech after speech, encouraging soldiers to desert and inciting weary crowds, all the while knowing how he enraged official Washington.”¹⁰³ Still, Lincoln persisted to tolerate the press. Such

⁹⁶ JAMES MADISON, REPORT ON THE VIRGINIA RESOLUTIONS (1800), http://presspubs.uchicago.edu/founders/documents/amendI_speechs24.html. Madison prepared this document for the Virginia House of Delegates as a criticism of the federal Alien and Sedition Acts.

⁹⁷ *Id.*

⁹⁸ *Id.*

⁹⁹ *See id.*; *see also* BERNARD SCHWARTZ ET AL., THE BILL OF RIGHTS: A DOCUMENTARY HISTORY 235, 287 (1971) (discussing Madison’s idea, “That the freedom of the press is one of the great bulwarks of liberty, and therefore ought never to be restrained”).

¹⁰⁰ Brandon Johnson, *Oppression in the Defense of Liberty: Abraham Lincoln and Freedom of the Press During the Civil War*, 18 HISTORIA 1, 6 (2009).

¹⁰¹ *Id.*

¹⁰² *Id.* at 7.

¹⁰³ Wyatt Kingseed, *The Fire in the Rear: Clement Vallandigham and the Copperheads*, AM. HIST. MAGAZINE, (Apr. 11, 2016), <http://www.historynet.com/the-fire-in-the-rear-clement-vallandigham-and-the-copperheads.htm>.

patience for publication and tools used against the government without state interference was a clear divergence from the Crown's precedent.

This tolerance for the private press continued as technology changed. During the mid-nineteenth century, for example, the boom of the telegraph created the ability for communication to move with unprecedented celerity over expansive distances, moving it outside the grasp of government oversight.¹⁰⁴ The high-speed printing press at the end of the nineteenth century, also known as the penny press, created a similarly unique ability for publishing much more quickly than ever before.¹⁰⁵ The telegraph and the penny press, combined with the notoriously irresponsible reporting techniques of "yellow journalists," often caused difficulty for the government.¹⁰⁶ "Even so, the high-speed presses were seen as solidly within the First Amendment protection. They did the same thing that an old-fashioned Franklin press did (put ink onto sheets of newspaper), except that they did so much more rapidly."¹⁰⁷ Yet, "nobody in 1888 claimed the high-speed presses were outside the First Amendment."¹⁰⁸

Thus, when the Supreme Court finally interpreted the Press Clause in the early twentieth century,¹⁰⁹ it is not altogether surprising that the Court undergirded the Clause's structural importance. It did so most apparently in the 1931 Supreme Court case, *Near v. Minnesota*,¹¹⁰ later called by First Amendment scholar Anthony Lewis the "Court's first great press case".¹¹¹ Plaintiff Jay M. Near was the owner of a local

¹⁰⁴ AMERICAN CIVIL LIBERTIES UNION, *ACLU SUBMISSION TO THE SPECIAL RAPPORTEUR ON THE PROMOTION AND PROTECTION OF THE RIGHT TO FREEDOM OF EXPRESSION AND OPINION* 15 n.68 (2015), <https://www.aclu.org/other/aclu-submission-special-rapporteur-encryption-and-anonymity> [hereinafter *ACLU Submission*] (citing both SIMON SINGH, *THE CODE BOOK* 61, 79 (1999) ("In the century following the invention of the telegraph in 1844, forty-four new commercial ciphers were patented by Americans for both commercial and private uses.") and STEVEN LEVY, *CRYPTO* 16 (2001) ("Just as the invention of the telegraph upped the cryptographic ante by moving messages thousands of miles in the open, presenting a ripe opportunity for eavesdroppers of every stripe, the computer age would be moving billions of messages previously committed to paper into the realm of bits.")).

¹⁰⁵ See LOREN COLEMAN, *THE COPYCAT EFFECT: HOW THE MEDIA AND POPULAR CULTURE TRIGGER THE MAYHEM IN TOMORROW'S HEADLINES* 135-37 (2004).

¹⁰⁶ *Id.* For example, the "penny press" is often blamed for the harrowing violence of the late nineteenth century, such as that period's surge in copycat killings.

¹⁰⁷ David B. Kopel, *First Amendment Guide*, 81 TENN. L. REV. 417, 455 (2014).

¹⁰⁸ *Id.*

¹⁰⁹ See *Schenck v. United States*, 249 U.S. 47, 52 (1919) (being the Court's first case to discuss freedom of the press (and speech) and creating the "clear and present danger" test).

¹¹⁰ 283 U.S. 697 (1931).

¹¹¹ ANTHONY LEWIS, *MAKE NO LAW: THE SULLIVAN CASE AND THE FIRST AMENDMENT* 90 (1991).

Minneapolis newspaper that had been enjoined under a state statute for publication of “malicious, scandalous and defamatory” comments.¹¹² In the landmark decision,¹¹³ relying on the Press Clause, the Court struck down the Minnesota law, removed the injunction, and cemented prior restraint as one of the most punitive state measures only to be employed in the most extreme circumstances.¹¹⁴ Although the Speech Clause played a role in the Court’s decision, it is often believed that absent the Press Clause, the Supreme Court would have upheld the prior restraint.¹¹⁵

Over the next century, as academics battled over the meaning of the word “press,” the power of the Clause only grew stronger, reaching its pinnacle in American society in the 1960s, in what legal scholar David Anderson refers to as the “heyday of the Press Clause.”¹¹⁶ *Near*, standing on the shoulders of hundreds of years of “free press” case development (from Wilkes to Zenger) had planted the seeds for watershed press cases, most important *New York Times Co. v. United States*.¹¹⁷ Before diving into this jurisprudence, however, it is imperative to understand the two interpretations of the Press Clause, and how both meanings are incomplete without the other. To understand the full protection intended by the Press Clause, we must marry elements of both.¹¹⁸

¹¹² *Near*, 283 U.S. at 697. In that case, a county prosecutor sought to enjoin a local newspaper in Minneapolis that accused local police for cooperating with a ring of “Jewish gangster[s]” involved in a string of crimes. *Id.* at 704.

¹¹³ *Near*, 283 U.S. at 704.

¹¹⁴ *Id.* at 713.

¹¹⁵ In large part, the heavy lifting of the case was done by the Press Clause to strike down the government’s seemingly reasonable law where the newspaper had been notoriously inflammatory and contained racist remarks critical of government. The law indeed only required that the reporters instill ethical journalistic codes in reporting truthful material. But ultimately the Court held the Press Clause protected the use of private presses and decisions to control information, especially where claims of government legitimacy were at issue. Anderson, *supra* note 24, at 458. While the Court has never given the Press Clause independent significance, neither has it foreclosed the possibility of its additional and separate importance. Chief Justice Burger is the only member of the Court who has expressed hostility toward the prospect of specific constitutional protection for the press, and even he concedes that the question is still open.

¹¹⁶ Anderson, *supra* note 24, at 448.

¹¹⁷ 403 U.S. 713 (1971) (per curiam).

¹¹⁸ Charles & O’Neill, *supra* note 19, at 1703-05 (discussing the importance of history and the various theories that helped to create the Press Clause as it developed)

II

THE PRESS CLAUSE: THE TRADITIONAL DEBATE

Over the past century, two theories of the Press Clause have developed. The two theories diverge in how they define the word “press”.¹¹⁹ A resurgence of this debate has percolated in recent years.¹²⁰ Revisiting foundational texts reveals that the Press Clause was likely meant to protect against government intrusion into private zones as a check on government control of information.

A. *Freedom of the Press: An Institutional Protection*

In the 1970s, a series of legal scholars, including David Anderson,¹²¹ Floyd Abrams,¹²² Edwin C. Baker,¹²³ and Vincent Blasi¹²⁴ (and more recently Sonja West¹²⁵) argued that the Press Clause was established to protect the institutional press, often referred to as the “Fourth Estate.”¹²⁶ According to this interpretation, members of the news media are afforded special protections under the Press Clause because of their contribution to democracy, acting as watchdogs over government.¹²⁷ Just as the judiciary, executive, and legislative branches of government act as a check on one another, the press—the Fourth Estate—serves as a check on the whole of government.¹²⁸ This position holds that without the checking value of the press,

¹¹⁹ See Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2302 (2014) (“The word ‘press’ has the dual signification of an institution for creating and distributing content and a technology for creating and distributing content. At the Founding it referred to the freedom to use the key mass communication technology of the day—the printing press.”).

¹²⁰ The resurgence followed discussion over the Press Clause in dicta of *Citizens United*, 558 U.S. 310 (2010).

¹²¹ Anderson, *supra* note 24.

¹²² Floyd Abrams, *The Press Is Different: Reflections on Justice Stewart and the Autonomous Press*, 7 HOFSTRA L. REV. 563, 580 (1979).

¹²³ C. Edwin Baker, *The Independent Significance of the Press Clause Under Existing Law*, 35 HOFSTRA L. REV. 955, 956 (2007).

¹²⁴ Blasi, *supra* note 4.

¹²⁵ West, *The ‘Press,’ Then & Now*, *supra* note 12.

¹²⁶ Stewart, *supra* note 4, at 634 (“The primary purpose of the constitutional guarantee of a free press was a similar one: to create a fourth institution outside the Government as an additional check on the three official branches.”).

¹²⁷ TIMOTHY E. COOK, *FREEING THE PRESSES: THE FIRST AMENDMENT IN ACTION* 7 (2005) (quoting Anderson, *supra* note 93, at 465 (noting that the first section “values the press as a public forum open to all [while] the second highlights the watchdog function”)).

¹²⁸ See generally Blasi, *supra* note 3 (examining the idea that free expression has a “checking value” against the abuse of government power).

corruption within government would likely ensue. Therefore, the Press Clause affords protection to members of news media to safeguard democracy.

This interpretation was not cabined to academia. Members of the Supreme Court, including Justices Stewart,¹²⁹ Powell,¹³⁰ and Douglas,¹³¹ vociferously supported this position during a twenty-year period. Starting in 1964, with *New York Times Co. v. Sullivan*,¹³² the Court established a preference for the press-as-industry approach by establishing the actual malice standard for libel actions.¹³³ The case created a qualified protection for the news media arising from their structural role as a check on government.¹³⁴ Quoting Madison, the Court asserted that in every state in the Union, “the press has exerted a freedom in canvassing the merits and measures of public men, of every description”¹³⁵ and that government ought not attempt to “controul [sic] the freedom of the press.”¹³⁶ Because of quotes like these, *Sullivan* has become known as one of the most important First Amendment cases to have ever been decided.¹³⁷ Although the case did not ultimately rely on the Press Clause, “it create[d] a bedrock of press-supportive dicta on the way to a press-prevailing . . . conclusion.”¹³⁸ In other words, *Sullivan* acted a building block for other cases that

¹²⁹ Stewart, *supra* note 4, at 634.

¹³⁰ See *Saxbe v. Wash. Post Co.*, 417 U.S. 843, 862-63 (1974) (Powell, J., dissenting).

¹³¹ See *Branzburg v. Hayes*, 408 U.S. 665, 721 (1972) (Douglas, J., dissenting).

¹³² 376 U.S. 254 (1964).

¹³³ If the plaintiff in a defamation suit is a public figure, then to find the defendant guilty under the actual malice standard, the plaintiff must prove the defendant made the publication either (i) knowing it was false, or (ii) with reckless disregard. At issue in *Sullivan* was an advertisement that contained certain factual inaccuracies about a Southern public official who had clashed with civil rights demonstrators. See *id.*

¹³⁴ Stewart, *supra* note 4, at 635.

¹³⁵ *Sullivan*, 376 U.S. at 275.

¹³⁶ *Id.* at 277 (citing an 1804 Letter to Abigail Adams as quoted in *Dennis v. United States*, 341 U.S. 494, 522 n.4 (1951) (Frankfurter, J., concurring)).

¹³⁷ Harry Kalven, Jr., *The New York Times Case: A Note on “The Central Meaning of the First Amendment”*, 1964 SUP. CT. REV. 191, 193-94 (1964).

¹³⁸ RonNell Andersen Jones, *The Dangers of Press Clause Dicta*, 48 GA. L. REV. 705, 851 (2014).

established protections for journalists, such as the “reporter’s privilege,”¹³⁹ disclosure of intercepted information,¹⁴⁰ and “access.”¹⁴¹

Perhaps more than *Sullivan*, however, *New York Times Co. v. United States* (the “*Pentagon Papers*” case) is often identified as the most influential press-as-industry case.¹⁴² In *Pentagon Papers*, the Court rejected the government’s injunction against the *New York Times* and the *Washington Post* for trying to publish the then-classified Pentagon Papers, despite the government’s arguments that publication of confidential material would disrupt national security.¹⁴³ The watershed case was revolutionary for restricting government power. It illuminated that the Press Clause could thwart government from interfering in private decisions about information, even in circumstances involving national security, in order to benefit of democracy. Many credit the development of this reasoning to Max Frankel, then-chief of the *New York Times*’ Washington bureau, who submitted a now famous affidavit in the district court. Frankel wrote:

Without the use of “secrets” that I shall attempt to explain in this affidavit, there could be no adequate diplomatic, military and political reporting of the kind our people take for granted, either abroad or in Washington and there could be no mature system of communication between the Government and the people

In the field of foreign affairs, only rarely does our Government give full public information to the press for the direct purpose of simply informing the people. For the most part, the press obtains significant information bearing on foreign policy only because it has managed to make itself a party to confidential materials of value in transmitting these materials from government to other branches and offices of government as well as to the public at large. This is why the press has

¹³⁹ See, e.g., *Branzburg v. Hayes*, 408 U.S. 665, 698 (1972) (Douglas, J., dissenting).. Although the Court denied the claims in *Branzburg*, many states have provided a reporter’s privilege through legislation. *Id.*

¹⁴⁰ *Bartnicki v. Vopper*, 53 U.S. 514, 527-28 (2001) (citing *N.Y. Times Co. v. Sullivan*, 376 U.S. 254 (1964)).

¹⁴¹ David Lange, *The Role of the Access Doctrine in the Regulation of the Mass Media: A Critical Review and Assessment*, 52 N.C. L. REV. 1, 8-34 (1973); see, e.g., *Miami Herald Pub. Co. v. Tornillo*, 418 U.S. 241 (1974); *CBS v. Democratic Nat’l Comm.*, 412 U.S. 94 (1973); see generally Jerome A. Barron, *Access to the Press—A New First Amendment Right*, 80 HARV. L. REV. 1641 (1967).

¹⁴² 403 U.S. 713 (1971).

¹⁴³ *Id.*

been wisely and correctly called The Fourth Branch of Government . .

..

But for the vast majority of “secrets,” there has developed between the Government and the press (and Congress) a rather simple rule of thumb: the Government hides what it can, pleading necessity as long as it can, and the press pries out what it can, pleading a need and right to know. Each side in this “game” regularly “wins” and “loses” a round or two. Each fights with the weapons at its command. When the Government loses a secret or two, it simply adjusts to a new reality. When the press loses a quest or two, it simply reports (or misreports) as best it can. Or so it has been, until this moment.¹⁴⁴

Never before had the need for a structural division between the press and government as well as the struggle over information been made so concrete. Soon after, the idea that journalists should be protected from government became commonplace. Just three years after *Pentagon Papers*, Justice Stewart presented a popular convocation speech at Yale Law School that buttressed this perspective.¹⁴⁵ Stewart wrote that press rights in the First Amendment were “no constitutional accident, but an acknowledgment of the critical role played by the press in American society,”¹⁴⁶ and that “so far as the Constitution goes, the autonomous press may publish what it knows, and may seek to learn what it can.”¹⁴⁷

While many legal scholars and litigators have held this point of view, “[t]he Court has not yet squarely resolved whether the Press Clause confers upon the ‘institutional press’ any freedom from government restraint not enjoyed by all others.”¹⁴⁸ In fact, the Supreme Court has never recognized any constitutional rights belonging exclusively to the media.¹⁴⁹ Justice Brennan, for instance, wrote that “in

¹⁴⁴ Affidavit of Max Frankel paras. 3, 17, Editor, N.Y. Times, *N.Y. Times Co. v. United States*, 403 U.S. 713 (1971), <http://www.pbs.org/wgbh/pages/frontline/newswar/part1/frankel.html> [hereinafter *Frankel Affidavit*].

¹⁴⁵ Stewart, *supra* note 4, at 634.

¹⁴⁶ *Houchins v. KQED, Inc.*, 438 U.S. 1, 17 (1978) (Stewart, J., concurring); *see also* Stewart, *supra* note 4, at 634.

¹⁴⁷ Stewart, *supra* note 4, at 634.

¹⁴⁸ *First Nat’l Bank of Boston v. Bellotti*, 435 U.S. 765, 798 (1978) (Burger, C.J., concurring)

¹⁴⁹ *Richmond Newspapers, Inc. v. Va.*, 448 U.S. 555, 573 (1980) (plurality opinion); Baker, *supra* note 123, at 958-59 (arguing that the existing Court does not recognize special privileges for the press, and holding that in criminal trials, “media representatives enjoy the same right of access as the public”); *see* *Branzburg v. Hayes*, 408 U.S. 665, 704 (1972) (“Freedom of the press is a ‘fundamental personal right’ which ‘is not confined to newspapers and periodicals.’” (quoting *Lovell v. Griffin*, 303 U.S. 444, 450, 452 (1938))); *see* Anderson, *supra* note 24, at 432 (“Nonconstitutional sources of special protection for the press are far more numerous.”).

the context of defamation law, the rights of the institutional media are no greater and no less than those enjoyed by other[s].”¹⁵⁰ In scholarship, the right has always been described as an *individual* right, not an *institutional* right.¹⁵¹

In addition, academics have pointed out that creating a special privilege for the media would contradict the fact that no part of the Constitution protects members of the news media. For example, despite popular beliefs, journalists are not protected from government subpoenas, search warrants, or work-product requests.¹⁵² Similarly, “if a reporter commits a minor tort such as a trespass, minor deception or breach of loyalty, no judicial consideration is given to the fact that she was engaged in news reporting.”¹⁵³ Additionally, the “definitional problem,”¹⁵⁴ in other words, the problem of defining *who would* constitute the press, particularly in today’s world where citizen journalists, bloggers, and journalism students have all taken on the title of reporter, also complicates the question of how the Press Clause could or should provide specific protections for one particular profession of individuals.¹⁵⁵

Despite these obstacles, this understanding of the Press Clause acknowledges a fundamental characteristic of the original understanding of the Press Clause: that the press (whatever it may be) plays a *structural* role in our democracy.¹⁵⁶ As David Anderson argues, the Press Clause provides “an additional check on official power” because it provides a barrier between the “press” and “the three official branches of government.”¹⁵⁷ In essence, the press can necessarily enable the free flow of information and act as a check on a tyrannical government because it has no part of it.¹⁵⁸ As Vincent Blasi explains, Madison’s view of the Press Clause was that it

¹⁵⁰ *Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc.*, 472 U.S. 749, 784 (1985).

¹⁵¹ *See generally* Volokh, *supra* note 5; *see also* BLACKSTONE, *supra* note 57.

¹⁵² Sonja R. West, *Awakening the Press Clause*, 58 UCLA L. REV. 1025, 1029 (2011).

¹⁵³ *Id.*

¹⁵⁴ *Id.*

¹⁵⁵ As the Supreme Court has noted, “[thanks to] the advent of the Internet and the decline of print and broadcast media . . . the line between the media and others who wish to comment on political and social issues becomes far more blurred.” *Citizens United v. Fed. Election Comm’n*, 558 U.S. 310, 352 (2010); *see also* SCOTT GANT, *WE’RE ALL JOURNALISTS NOW* 6 (2007).

¹⁵⁶ Paul Horwitz, *Churches as First Amendment Institutions: Of Sovereignty and Spheres*, 44 HARV. C.R.–C.L. L. REV. 79, 87 (2009).

¹⁵⁷ Anderson, *supra* note 24, at 460 (stating that while the Founders did not seem to argue that the checking power of the Press Clause was written to “protect editorial autonomy” or that the “government may not interfere with journalistic decisions,” they did believe that the Press Clause provided an important structural role to check government intrusion into personal realms where tools like the press, computers, or code could be used to share certain information).

¹⁵⁸ William W. Van Alstyne, *First Amendment and the Free Press: A Comment on Some New Trends and Some Old Theories*, 9 HOFSTRA L. REV. 1, 19 (1980).

served as a structural barrier and protected individuals from state action.¹⁵⁹ Through the space provided by this barrier, people could judge and challenge the actions of their rulers.¹⁶⁰

Alexander Bickel echoed this position. Bickel “believed the press was a constitutionally recognized countervailing power to the official branches of government.”¹⁶¹ Although Bickel did not restrict the Press Clause to “institutionalized print and electronic media, he believed that the first amendment recognized the press—as an entity outside of and in competition with the government.”¹⁶² Bickel called this necessary but contentious relationship between government and press the “disorderly situation.”¹⁶³ In this relationship, both entities had the right and obligation to keep information and materials secret from one another.¹⁶⁴ As other scholars have commented, Bickel found a parallel between James Madison’s theories regarding “the separation of powers” and Madison’s view of freedom of the press.¹⁶⁵ Like the separation of powers principle, the Press Clause created a structural protection that shielded against government intrusion into the private realm, particularly when dealing with distribution of information.¹⁶⁶

Although the Court has never elaborated on this structural point, certain Justices on the Court have remarked on it. As if mimicking the words of Max Frankel, Justice Stewart’s 1974 Yale Law School convocation speech buttressed this perspective.¹⁶⁷ Stewart wrote that press rights in the First Amendment were “no constitutional accident, but an acknowledgment of the critical role played by the

¹⁵⁹ Blasi, *supra* note 4, at 538.

¹⁶⁰ *Id.*

¹⁶¹ John Nowak, *Using the Press Clause to Limit Government Speech* 30, ARIZ. L. REV. 1, 13 (1988) (citing ALEXANDER M. BICKEL, *THE MORALITY OF CONSENT* (1975)).

¹⁶² *Id.*

¹⁶³ See David McCraw & Stephen Gikow, *The End to an Unspoken Bargain? National Security and Leaks in a Post-Pentagon Papers World*, 48 HARV. C.R.–C.L. L. REV. 473 (2013) (citing ALEXANDER M. BICKEL, *THE MORALITY OF CONSENT* (1975)) (“It is a disorderly situation surely. But if we ordered it we would have to sacrifice one of two contending values—privacy or public discourse—which are ultimately irreconcilable.”).

¹⁶⁴ *Id.*

¹⁶⁵ See Nowak, *supra* note 162, at 12.

¹⁶⁶ Sonja West, *The Stealth Press Clause*, 48 GA L. REV. 729, 753 (2014) (In *Mills v. Alabama*, the Court specifically enumerated the press as providing this protection because it “serves[] as powerful antidote to any abuses of power by governmental officials and as a constitutionally chosen means for keeping officials elected by the people responsible to all the people whom they were selected to serve.”) (citing *Mills v. Alabama*, 384 U.S. 214, 219 (1966)); see also Blasi, *supra* note 4, at 538 (“the generation of Americans which enacted the First Amendment built its whole philosophy of freedom of the press around the checking value”).

¹⁶⁷ Stewart, *supra* note 4, at 634.

press in American society,”¹⁶⁸ and that “so far as the Constitution goes, the autonomous press may publish what it knows, and may seek to learn what it can.”¹⁶⁹ Similarly, Justice Douglas characterized the freedom of the press as part of a constitutional scheme (also consisting of the separation of powers and an independent judiciary) designed “to take Government off the backs of people.”¹⁷⁰ Justice Douglas explained that the right to have the government “off your back” meant the protection of a zone where one could make certain private choices without state interference—such as deciding what information to publish and what to keep private from government intrusion. As explained more below, while the press-as-technology perspective does not highlight the structural importance of the press, it stresses the importance of the individual. Combining, these approaches therefore presents an opportunity to uphold the Founders’ true intention.¹⁷¹

B. Freedom of the Press: An Individual Protection of Tools

While the press-as-industry approach was the popular stance among legal scholars for most of the twentieth century, following the decision of *Citizens United*¹⁷² the press-as-technology approach gained momentum. In fact, some claim the press-as-technology approach has now been adopted as the “prevailing” approach.¹⁷³ In the 2010 decision, Justice Scalia engaged in debate with Justice Stevens over the two interpretations.¹⁷⁴ Arguing against the press-as-technology approach, Justice Stevens asserted that the Court was wrongly focused on the Speech Clause and should turn to a press-as-institution analysis under the Press Clause.¹⁷⁵

¹⁶⁸ *Houchins v. KQED, Inc.*, 438 U.S. 1, 17 (1978) (Stewart, J., concurring); *see also* Stewart, *supra* note 4, at 634.

¹⁶⁹ Stewart, *supra* note 4, at 634.

¹⁷⁰ *Columbia Broadcasting Sys., Inc. v. Democratic Nat’l Comm.*, 412 U.S. 94, 162 (1973) (Douglas, J., concurring); *see also* Dorothy Glancy, *Getting Government off the Backs of People: The Right of Privacy and Freedom of Expression in the Opinions of Justice William O. Douglas*, 21 SANTA CLARA L. REV. 1047, 1049 (1981) (citing *Pub. Util. Comm’n. v. Pollak*, 343 U.S. 451, 467 (1952) (Douglas, J., dissenting)).

¹⁷¹ *See* Nowak, *supra* note 162, at 13 (discussing the press as a “structural protection for individuals’ physical ability to control the flow of information through technology” or as the “independent press, free of direct government censorship”).

¹⁷² 558 U.S. 310 (2010).

¹⁷³ West, *The ‘Press,’ Then & Now*, *supra* note 12, at 59 (stating Volokh’s interpretation appears to reflect the now-prevailing view of press freedom).

¹⁷⁴ In the opinion, Justices Scalia and Stevens argued over the point of whether newspapers had a special protection under the Clause. *Citizens United v. Fed. Election Comm’n*, 558 U.S. 310, 390 (2010).

¹⁷⁵ *Id.* at 431 n.57 (Stevens, J., concurring in part and dissenting in part) (describing a Press Clause which shows “why one type of corporation, those that are part of the press, might be able to claim special First Amendment status”).

Justice Scalia rebuked this comment, arguing that Justice Steven's perspective "boggles the mind," and that the Press Clause referred simply to "everyone's right" to publish, not "the institutional press's right to publish."¹⁷⁶ While *Citizens United* spurred much more heated debate about the protection for corporations under the Speech Clause,¹⁷⁷ the opinion also inspired an article by Professor Eugene Volokh that would clarify the press-as-technology approach as the prevailing view under the Press Clause.¹⁷⁸

In 2012, Professor Volokh picked up on the argument where Justice Scalia left off,¹⁷⁹ arguing, "the purpose of the Constitution was not to erect the press into a privileged institution but to protect *all persons* in their right to print what they will as to utter it."¹⁸⁰ Volokh asserts that freedom of press is a general protection for individuals to use technology like the printing press,¹⁸¹ and not a limited protection for the institutional press.¹⁸² To demonstrate this point, he examines several nineteenth-century political writers, as well as twelve American and three British cases from between 1784 and 1840, in which freedom of press was used to protect individuals who were not professional journalists.¹⁸³ He references Francisco Ludlow Holt and his 1812 work, *The Law of Libel*,¹⁸⁴ which broadly defined "[t]he liberty of the press as the personal liberty of the writer to express his thoughts in the more [im]proved way invented by human ingenuity in the form of the press."¹⁸⁵

Professor Volokh offers further evidence to argue that to the Founders, the "press" was nothing more than a tool. He cites James Madison, who spoke of the

¹⁷⁶ *Id.* at 390 n.6 (Scalia, J., concurring).

¹⁷⁷ Indeed, *Citizens United* was a Speech Clause case. See RANDALL P. BEZANSON, TOO MUCH FREE SPEECH? 40 (2012) (noting that the Press Clause analysis was neither briefed nor argued in *Citizens United*).

¹⁷⁸ See generally Volokh, *supra* note 5.

¹⁷⁹ *Id.* at 463.

¹⁸⁰ *Pennekamp v. Florida*, 328 U.S. 331, 364 (1946) (Frankfurter, J., concurring).

¹⁸¹ Volokh points to early cases, treatises, and state constitutions to suggest the "liberty of the press" is a mere addendum to freedom of speech, and simply provides that "every citizen" may freely speak, write, and publish his sentiments. Volokh, *supra* note 5, at 466-68.

¹⁸² "Professor Volokh argues that 'it was not until the 1970s that some courts extended special protections under the Press Clause to the institutional press and these decisions remained a minority.'" McConnell, *supra* note 5, at 431 (citing Volokh, *supra* note 5, at 522-23).

¹⁸³ See Volokh, *supra* note 5.

¹⁸⁴ *Id.* at 471.

¹⁸⁵ *Id.*

“use of the press” in his 1800 Report on the Virginia Resolution.¹⁸⁶ Similarly, he quotes the Massachusetts response to the Virginia Resolutions and St. George Tucker’s influential work that also described the “use of the press.”¹⁸⁷ Indeed, other writers of that time, such as William Rawle likewise characterized “[t]he press” as a “vehicle” of speech and stated “[t]he art of printing illuminates the world, by a rapid dissemination of what would otherwise be slowly communicated and partially understood.”¹⁸⁸ Citing these sources in a straightforward textual reading, Volokh underscores the importance of these tools and the understanding of the Press Clause as a broad protection for all individuals who use them.

Although Volokh convincingly highlights the Press Clause as a general protection for individuals and the private use of technology, his article seems to mute the structural importance of the Clause.¹⁸⁹ Arguing that no distinction exists between the Speech Clause and Press Clause, he finds that the use of the printing press was just another form of speech.¹⁹⁰ But even Volokh concedes that the freedom of the press posed separate “dangers that ordinary ‘speech’ did not.”¹⁹¹ He admits that historically the freedom of the press was considered to be “especially dangerous” and in this way distinct from speech.¹⁹² Volokh also cites to Madison,¹⁹³ who

¹⁸⁶ *Id.* at 473 n.47 (emphasis added); *see also id.* (“The inclusion of the word ‘use’ in . . . ‘freedom in the use of the press’ makes it unmistakably clear that Madison . . . w[as] referring to the machine of the printing press.”).

¹⁸⁷ *Id.* at 473 (stating that “freedom of the press is a security for the rational use, and not the abuse of the press”). Additionally, in St. George Tucker’s influential 1803 work, he discussed the freedom of the press and spoke of “[w]hoever makes use of the press as the *vehicle* of his sentiments on any subjects.” St. George Tucker, *Of the Freedom of Speech and Press* (1803), <http://lonang.com/library/reference/tucker-blackstone-notes-reference/tuck-2g/> (emphasis added).

¹⁸⁸ Volokh, *supra* note 5, at 476 (citing WILLIAM RAWLE, A VIEW OF THE CONSTITUTION OF THE UNITED STATES OF AMERICA 119 (Philadelphia, H.C. Carey & I. Lea 1825)).

¹⁸⁹ Volokh cites Justice Scalia’s argument in *Citizens United* that the shared words “‘freedom of’ in the phrase the ‘freedom of speech, or of the press’ are most reasonably understood as playing the same role for both ‘speech’ and ‘press.’” Volokh, *supra* note 5, at 472 n.46. Volokh also tries to marry the freedom of speech clause with freedom of the press by stating that “over the last several decades, the phrase ‘freedom of speech’ has often been used to mean “freedom of expression” and to encompass all means of communication . . . [including] radio, films, television, and the Internet.” *Id.* at 477.

¹⁹⁰ *Id.* at 505-06. Volokh writes that a discussion about the independence of the Press Clause is “outside the scope of this Article” but throughout cites several cases suggesting its indivisibility from the Speech Clause. *Id.* at 506 n.214 (citing to cases in Section I.E of his own article for evidence that the Speech Clause is indivisible from Freedom of Press.).

¹⁹¹ Volokh, *supra* note 5, at 476.

¹⁹² *Id.*

¹⁹³ As stated previously, Madison clearly thought the Press Clause provided a distinct structural protection. James Madison, Report on the Virginia Resolutions to the House Delegates (1880), reprinted in 4 DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION 546, 570 (Jonathan Elliot ed., Philadelphia, J.B. Lippincott Co. 2d ed. 1891).

undoubtedly understood the Press Clause to be a structural protection, further undermining this perspective. Indeed, Professor Volokh nods at this conclusion when he cites to David Anderson, the foremost scholar for the institutional press perspective.¹⁹⁴

This short but meaningful acknowledgement by Volokh seems to suggest that while both the press-as-technology and the press-as-speech perspectives have often been pitted against one another, they nevertheless overlap. Although neither exactly agrees on *who* specifically is protected by the Press Clause (a narrow set of professional journalists as opposed to anyone utilizing a press tool), both approaches would likely agree that the Clause is a protection for the use of tools and private spaces that enable the public to control the flow of information. With that broader perspective, we can move forward and see how encryption falls under this definition. If the Press Clause is ultimately a simple check on government by allowing the public to control the flow of information, then encryption seems to fall squarely within that definition.

III

ENCRYPTION AND THE PRESS CLAUSE

Encryption is the mathematical “process of converting messages, information, or data into a form unreadable by anyone except the intended recipient.”¹⁹⁵ Through this process, a plaintext message is paired with a randomly generated key, and both are scrambled until all that is left is an undecipherable message, called ciphertext.¹⁹⁶ Although this description makes encryption seem like a modern invention, the tool is “as old as communication itself.”¹⁹⁷ Based on the Greek words *kryptos*, meaning

Similarly, Francis Holt understood speech and press as distinct. FRANCIS LUDLOW HOLT, *THE LAW OF LIBEL* 38-39 (1812). Holt writes that “with a very few exceptions, whatever any one has a right both to think and to speak, he has *likewise* a consequential right to print and to publish.” *Id.* at 38 (emphasis added). Holt uses the word “likewise” to distinguish separate right to publish and control the actual printing and spreading of information through technology. *Id.*

¹⁹⁴ He quotes Anderson to state that “the existence of a Press Clause may have been crucial” to the Supreme Court’s original First Amendment cases. Volokh, *supra* note 5, at 477 n.67 (citing Anderson, *supra* note 93, at 458).

¹⁹⁵ SANS Institute, *History of Encryption* (2001), <http://www.sans.org/reading-room/whitepapers/vpns/history-encryption-730> (last visited Feb. 27, 2017).

¹⁹⁶ Micah Lee, *Encryption Works: How to Protect Your Privacy in the Age of NSA Surveillance*, FREEDOM OF THE PRESS FOUND. (July 2, 2013), <https://freedom.press/encryption-works>; see also John J Browder, *Encryption Source Code and the First Amendment*, 40 JURIMETRICS J. 431, 431 n.2 (2000).

¹⁹⁷ Wendy McElroy, *Thomas Jefferson Used Encryption*, INFORMATION LIBERATION (Sept. 1, 2012), <http://www.informationliberation.com/?id=40786>.

hidden or secret, and *graphia*, meaning writing—encryption is inviolably intertwined with all technological communication—dating back to pen and invisible ink.¹⁹⁸ Throughout history, encryption’s purpose has been to keep information secret rather than to make it public, but it is just as much a tool for individual control over the flow of information as the printing press.¹⁹⁹ This section will detail the legal justifications previously used to protect encryption under the Speech Clause, and will conclude that while the Speech Clause is an appropriate home for protection, the Press Clause acts as an additional and descriptively powerful rationale for protecting encryption.

A. *A Short History of Encryption and Government Control*

Control over information is inherent to power.²⁰⁰ To accomplish this, governments have historically held tight grips over all technology that is capable both of spreading as well as censoring messages. Inevitably, governments have long imposed controls over encryption in order to manage the flow of information.²⁰¹ Dating as far back to the Roman state, Julius Caesar safely guarded his encryption method (shifting the Roman alphabet three places) to send covert messages to his military.²⁰² At the time of the Renaissance, many European countries developed

¹⁹⁸ See generally JOHN A. NAGY, *INVISIBLE INK: SPYCRAFT OF THE AMERICAN REVOLUTION* (2011).

¹⁹⁹ DAVID KAYE, REPORT OF THE SPECIAL RAPPORTEUR ON PROMOTION AND PROTECTION OF THE RIGHT TO FREEDOM OF EXPRESSION AND OPINION A/HRC/29/32 (May 22, 2015) (“Drawing from research on international and national norms and jurisprudence, and the input of States and civil society, the report concludes that encryption and anonymity enable individuals to exercise their rights to freedom of opinion and expression in the digital age and, as such, deserve strong protection.”); *id.* at 5 (“Where States impose unlawful censorship through filtering and other technologies, the use of encryption and anonymity may empower individuals to circumvent barriers and access information and ideas without the intrusion of authorities. Journalists, researchers, lawyers and civil society rely on encryption and anonymity to shield themselves (and their sources, clients and partners) from surveillance and harassment.”).

²⁰⁰ See MICHEL FOUCAULT, *TRUTH AND POWER*, in *POWER/KNOWLEDGE: SELECTED INTERVIEWS AND OTHER WRITINGS, 1972-1977*, at 109, 115 (Colin Gordon ed., Colin Gordon et al. trans., 1980). Panopticism is defined as State power over citizens in the form of continuous supervision. See also Michel Foucault, *Truth and Juridical Forms*, in *POWER: ESSENTIAL WORKS OF FOUCAULT 1954-1984*, at 1, 70 (James D. Faubion ed., Robert Hurley et al. trans., The New Press 1994).

²⁰¹ See Jeffrey L. Vagle, *Furtive Encryption: Power, Trust, and the Constitutional Cost of Collective Surveillance*, 90 INDIANA L. J. 101, 106-07 (2015) (stating that over the past 4000 years encryption, in different forms has been used by priests, emperors, diplomats, generals, spies, insurgents, dissidents, criminals, prisoners, and even lovers).

²⁰² Jason Kerben, Comment, *The Dilemma for Future Communication Technologies: How to Constitutionally Dress the Crypto-Genie*, 5 COMM. LAW CONSPECTUS 125, 125 (1997).

secret codebreaking bureaus, called black chambers, that were kept secret from the public.²⁰³ During World War II, cryptography became a focal point of the struggle over power when German intelligence began developing tools, like the Enigma machine, and the U.S. later decrypted Enigma and covertly employed it to win the War.²⁰⁴ By the 1960s, government control and secrecy around encryption reached new heights of paranoia in the midst of the Cold War²⁰⁵—that is, until today.²⁰⁶

Recent attempts within the United States have caused national alarm around encryption. On February 16, 2016, a federal magistrate judge in the U.S. District Court for the Central District of California issued an order requiring Apple, Inc. to assist the Federal Bureau of Investigation (FBI) in obtaining data from an iPhone belonging to one of the members involved in the 2015 shooting in San Bernardino, California.²⁰⁷ After Apple resisted the government's order, but before the court issued an opinion, the case was resolved when FBI agents accessed the data through other methods.²⁰⁸ This case instigated heated rhetoric that reignited a decades-old debate about government control over encryption methods.²⁰⁹ Few realize that the debate stems back to the country's origin.

The history of the United States government siphoning encryption from the hands of private innovators dates back to the Founding Era.²¹⁰ From the time of the

²⁰³ Throughout history, this practice continued to be mechanized in government institutions. Indeed, by the eighteenth century, the British Royal Mail was so commonly trying to break private and diplomatic ciphers that in 1720 it began to operate a specific system for deciphering mail. See John A. Fraser III, *The Use of Encrypted, Coded and Secret Communications is an 'Ancient Liberty' Protected by the United States Constitution*, 2 VA. J.L. & TECH. 1, 19 (1997).

²⁰⁴ Enigma's code was infamously broken by the Allies, enabling them to locate and sink many German U-boats and contributing to their ultimate success in the war. Thinh Nguyen, *Cryptography, Export Controls, and the First Amendment in Bernstein v. United States Department of State*, 10 HARV. J.L. & TECH. 667, 668 (1997); John Duong, *Intersection of the Fourth and Fifth Amendments in the Context of Encrypted Personal Data at the Border*, 2 DREXEL L. REV. 313, 324 (2009).

²⁰⁵ Cf. Delaney Hall, Episode 208: *Vox Ex Machina*, 99% INVISIBLE (Apr. 12, 2016), <http://99percentinvisible.org/episode/vox-ex-machina/>.

²⁰⁶ See COMMITTEE TO STUDY NATIONAL CRYPTOGRAPHY POLICY, NATIONAL RESEARCH COUNCIL, CRYPTOGRAPHY'S ROLE IN SECURING THE INFORMATION SOCIETY 96-97 (1996).

²⁰⁷ HOUSE JUDICIARY COMMITTEE, ENCRYPTION WORKING GROUP YEAR END REPORT (Dec. 20, 2016), http://energycommerce.house.gov/sites/republicans.energycommerce.house.gov/files/documents/114/analysis/20161219EWGFINALReport_0.pdf.

²⁰⁸ *Id.*

²⁰⁹ *Id.*

²¹⁰ Governments have always encouraged private design because innovation, particularly in the creation of novel cryptographic systems, historically comes primarily from amateurs. RICHARD A. MOLLIN, CODES: THE GUIDE TO SECRECY FROM ANCIENT TO MODERN TIMES 101 (2005); see

Revolution through the early years of the United States, Benjamin Franklin, Thomas Jefferson, James Madison and a covey of other Founders used and even invented various codes, signals, and ciphers for their private communications.²¹¹ Initially used to covertly send messages during the Revolution, these tools were later used by the Founders in their public capacities.²¹² Franklin and Jefferson, who both invented their own encryption tools for private correspondence, later used these tools as French Ambassador and Secretary of State, respectively.²¹³ In fact, Thomas Jefferson used a cipher to encrypt a message to James Madison in August of 1789 to strengthen the language in the First Amendment.²¹⁴ Many civilians, including businesspeople and revolutionaries, also used encryption at that time to secure their communications “[a]nd no one questioned their right to do so—no matter the context or rationale behind their use of encryption.”²¹⁵

As the country grew, what was once a tacit usurpation by the government of private encryption quickly developed into more intense restraint.²¹⁶ Much like the Elizabethan regulation over the printing press, by the twentieth century, the federal government began enacting a series of laws, codes, and even patent regulations to control encryption. In fact, between 1874 and 1928 the United States Patent & Trademark Office only granted and published 105 patents on cryptological devices.²¹⁷ By the early twentieth century, the United States government also began

also Danielle Kehl, Andi Wilson & Kevin Bankston, *Doomed to Repeat History: Lessons from the Crypto Wars of the 1990s*, NEW AMERICA (June 2, 2015), <https://www.newamerica.org/cybersecurity-initiative/policy-papers/doomed-to-repeat-history-lessons-from-the-crypto-wars-of-the-1990s/>.

²¹¹ Fraser, *supra* note 204, at 24-33.

²¹² *Id.*

²¹³ *Id.*

²¹⁴ Sarah Elizabeth Adler, *Encryption for All: Why This American Tradition Must be Upheld*, CAL. MAG., Jan. 31, 2017.

²¹⁵ Seth Schoen & Jamie Williams, *Crypto is for Everyone—History Proves It*, ELEC. FRONTIER FOUND. (Oct. 30, 2015), <https://www.eff.org/deeplinks/2015/10/crypto-everyone-and-american-history>.

²¹⁶ Norman Andrew Crain, *Bernstein, Karn, and Junger: Constitutional Challenges to Cryptography*, 50 AL. L. REV. 869, 874 (“The NSA has continuously attempted to control the development and expansion of cryptography in the private sector because it views the technology as a threat to national security.”) (citing Laura M. Pilkington, *First and Fifth Amendment Challenges to Export Controls on Encryption: Bernstein and Karn*, 37 SANTA CLARA L. REV. 159, 162)).

²¹⁷ Fraser, *supra* note 204, at 59 (“The Inventions Secrecy Act authorizes the Commissioner of Patents to refuse to issue patent secrecy orders, but it has not been effective in preventing the public dissemination of a number of strong, unpatented encryption products, and is not a mainstay of federal attempts to control encryption.”).

poaching private cryptographers to work in government intelligence.²¹⁸ In addition, the government began restricting innovation in less visible ways.²¹⁹ As one commentator said, the U.S. government attempted to “control the development and expansion of cryptography in the private sector [by slowing] the growth and dissemination of cryptography by controlling public funding, patent publications, and presentation of scientific papers at academic conference[s].”²²⁰ In essence, this created a ‘Jekyll and Hyde’ approach where the government’s dependency on private innovation required careful facilitation of encryption for government purposes, while simultaneously choking back private use of encryption tools.

B. The Original Crypto Wars: 1960-2000

After a century of schizophrenic give-and-take behavior, the tension between government and private control over encryption finally began to boil over into what is often now described as the Crypto Wars. In the 1970s, a perfect storm began to brew when the paranoid Cold War government and its equally suspicious citizenry were introduced to a new encryption technology that could easily be employed. In particular, the invention of public key cryptography created in 1976 by Whitfield Diffie and Martin Hellman created the landing site for the government’s fight for control.²²¹ Now called the “Diffie Helman key exchange,” the invention was unique because, unlike previous cryptography, this approach allowed two or more parties to communicate securely “even if they had never previously met.”²²² This discovery was so profound that it “laid the foundation for a number of innovations in secure communications over the next 40 years.”²²³

²¹⁸ For example, the government hired William Frederick Friedman, later honored as the father of U.S. cryptanalysis, from a private lab. Friedman was named the head of the U.S. Army’s Signal Intelligence and later the NSA’s chief cryptologist.

²¹⁹ Duong, *supra* note 205, at 325 (citing Press Release, Philip Bulman, Nat’l Inst. Of Standards & Tech., Commerce Department Announces Winner of Global Information Security Competition (Oct. 2, 2000), <http://csrc.nist.gov/archive/aes/index.htm-l> (follow “AES Press Release” hyperlink)). Perhaps the most dramatic example is when the U.S. government, through the National Institute of Standards and Technology, sponsored an open contest in 1997 to develop a new encryption standard. The winning entry was a cipher created by two Belgian scientists. The NSA eventually adopted the tool for classifying highly sensitive materials. *Id.*

²²⁰ Crain, *supra* note 217, at 874 (citing Pilkington, *supra* note 217, at 162).

²²¹ Kehl et al., *supra* note 211, at 2 (“Everything changed with the invention of public key cryptography in 1976.”).

²²² *Id.* at 2-3.

²²³ *Id.* at 3.

In response to this new technology, the federal government began exercising fierce restrictions on private use of encryption in all branches of government.²²⁴ First, Congress began developing laws that made possession of encryption a crime itself, thwarting private industry from exporting certain software.²²⁵ Then, in 1983, President Reagan signed the National Security Decision Directive, an Executive Order that gave the NSA authority to control all private sector technology that could reasonably affect national security.²²⁶ Coextensive with this directive, the agency decided by 1987 to no longer guarantee the security of its own encryption tools that it offered to the public,²²⁷ and “developed a policy of opposition to private cryptographic research.”²²⁸ In the government’s eyes, private innovation had become a threat to “government codes and intelligence gathering.”²²⁹

The Crypto Wars came into full swing during the next decade when individuals gained even more access to new technology.²³⁰ The 1990s were an especially tumultuous period because “the balance of encryption technology had effectively shifted back to the citizen,” making the government incapable of asserting its control over information.²³¹ Although corporations had previously driven much of the private cryptographic development, by the 1990s individuals were easily able to purchase encryption tools²³² and began to develop “ever more secure methods of encryption” that could “frustrate interception attempts by even the most sophisticated government agencies.”²³³ Perhaps the best example of this

²²⁴ *Id.*

²²⁵ Facing these rules, companies such as IBM and Digital Equipment Corporation questioned whether they could export hardware and software with strong encryption without government punishment. HAROLD ABELSON ET AL., KEYS UNDER DOORMATS: MANDATING INSECURITY BY REQUIRING GOVERNMENT ACCESS TO ALL DATA AND COMMUNICATIONS (2015), <https://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf>.

²²⁶ The Directive was withdrawn in 1987.

²²⁷ This policy stood in sharp juxtaposition to ten years prior, in 1977, when the National Bureau of Standards developed with help from the NSA an IBM-made encryption chip, known as DES, for commercial use.

²²⁸ Fraser, *supra* note 204, at 63. During that time, debates also brewed over whether the NSA or the National Institute of Standards and Technology (NIST) would control the development of cryptographic standards for the non-national security side of the government (NIST had been given authority under the 1987 Computer Security Act). ABELSON ET AL., *supra* note 226, at 8.

²²⁹ Fraser, *supra* note 204, at 63.

²³⁰ Kehl et al., *supra* note 211, at 3. “[P]olicymakers and advocates fiercely debated the tradeoffs related to the proliferation of encryption technology both in the United States and overseas.” *Id.* at 1.

²³¹ Fraser, *supra* note 204, at 65.

²³² Kehl et al., *supra* note 211.

²³³ Fraser, *supra* note 204, at 65.

type of encryption tool was Pretty Good Privacy (“PGP”), released publicly in 1991, for the encryption of email and files. As explained by its designer Philip Zimmerman, “until recently, if the government wanted to violate the privacy of ordinary citizens, they had to expend a certain amount of expense and labor to intercept and steam open and read paper mail”; with PGP and other similar tools, however, citizens could “take privacy into their own hands.”²³⁴

While citizens saw this individual control as a victory, the government excoriated this shift as the “going dark” problem.²³⁵ The “going dark” problem referred to “advancements in technology that le[ft] law enforcement and the national security community unable to obtain certain forms of evidence.”²³⁶ To frustrate this private hold over information, the U.S. government launched three large-scale attacks that have now become iconic events of the Crypto Wars.²³⁷ The first occurred in 1993 when the government compromised the DES key, a private use encryption tool, to the point of total inefficacy by allowing a “backdoor” or alternative method of access for law enforcement.²³⁸ The second followed when the Clinton Administration proposed a new device, called the Clipper Chip. The Clipper Chip infamously sought to require all encryption systems to retain a “key escrow,” or a third party copy of the keys that could be handed over to law enforcement to decrypt a message, which inherently compromised the technology. This proposal created a “groundswell of opposition from privacy advocates, industry representatives, and prominent politicians.”²³⁹

While the Clipper Chip debacle ultimately failed after immense public outcry, law enforcement persisted with yet a third tactic, which would spur the biggest fight yet: legislation. The idea for the legislation occurred after telephone companies switched from analog to digital, making it impossible for law enforcement to continue wiretapping calls.²⁴⁰ The FBI complained that this switch amounted to yet another example of the “going dark” problem, and firmly held that the agency

²³⁴ Kehl et al., *supra* note 211, at 3.

²³⁵ Federal Bureau of Investigation, *Going Dark*, <https://www.fbi.gov/services/operational-technology/going-dark> (last visited Feb. 26, 2017) [hereinafter *Going Dark*].

²³⁶ HOUSE JUDICIARY COMMITTEE, *supra* note 211, at 2.

²³⁷ Jill M. Ryan, *Freedom to Speak Unintelligibly: The First Amendment Implications of Government-Controlled Encryption*, 4 WM. & MARY BILL RTS. J. 1165, 1174-89 (1996) (reviewing actions of the federal government in recent years and concluding that the goal is the control of encryption).

²³⁸ Kehl et al., *supra* note 211.

²³⁹ *Id.* at 1, 9 (stating Congress held hearings on the merits of key escrow).

²⁴⁰ Ahmed Ghappour, *Searching Places Unknown: Law Enforcement Jurisdiction on the Dark Web*, STAN. L. REV. 10, 22 (2017) (citation omitted).

needed to preserve its access to telephone and Internet data.²⁴¹ After fierce debate between privacy advocates and law enforcement, both sides reached a compromise with the passage of the Communications Assistance for Law Enforcement Act (“CALEA”) in 1994.²⁴² CALEA required telecommunications carriers to make their equipment available to federal agencies to conduct surveillance,²⁴³ but a key concession was made for encryption. CALEA included an explicit exception so that providers are not compelled to decrypt customer communications.²⁴⁴

Because the government was unable to impose backdoors through CALEA, it instead created another, more aggressive, program known as the International Traffic

²⁴¹ *Going Dark*, *supra* note 236.

²⁴² Communications Assistance for Law Enforcement Act, Pub. L. No. 103-414, 108 Stat. 4279 (1994).

²⁴³ Albert Gidari, *More CALEA and Why It Trumps the FBI’s All Writs Act Order*, THE CENTER FOR INTERNET AND SOCIETY (Feb. 22, 2016), <http://cyberlaw.stanford.edu/blog/2016/02/more-calea-and-why-it-trumps-fbis-all-writs-act-order> (“The law was passed in 1994 in response to law enforcement’s concerns that it was then ‘going dark’ with the advent of digital telephony and the Internet. The Director of the FBI at that time testified that CALEA was necessary to preserve the capabilities it always had to intercept *all* communications. But in the end, CALEA was a compromise, giving law enforcement a narrowly focused capability to carry out lawfully authorized surveillance on public switched and cellular networks, but imposing certain privacy protections and limitations on law enforcement’s ability to ‘imped[e] the development of new communications services and technologies.’ In short, the FBI did not get a future-proof legislative mandate to gain access to evidence it all new technologies or the ability to block introduction of secure technologies . . . Congress also determined that carriers would have no responsibility to decrypt encrypted communications unless the carrier provided the encryption and could in fact decrypt it . . . So while CALEA provided law enforcement with some surveillance capabilities on phone networks (which the Federal Communications Commission later extended to broadband Internet access and two-way Voice over IP), it precluded the government from requiring ‘any specific design of equipment, facilities, services, features, or system configurations to be adopted by any manufacturer of telecommunications equipment.’”). CALEA contains a major carve-out for encryption that was, like CALEA’s other exemptions, the result of a very hard-fought battle during the drafting of CALEA, and a major win for the cryptography and civlib communities that provided a lot of input and expert testimony in the drafting process. *See* 47 U.S.C. 1002(b)(3).

²⁴⁴ Kehl et al., *supra* note 211, at 8 (47 U.S.C. § 1002(b)(3) explicitly states that a “telecommunications carrier shall not be responsible for decrypting, or ensuring the government’s ability to decrypt, any communication encrypted by a subscriber or customer, unless the encryption was provided by the carrier and the carrier possesses the information necessary to decrypt communication.” The legislative history further explains, “nothing in this paragraph would prohibit a carrier from deploying an encryption service for which it does not retain the ability to decrypt communications for law enforcement access.”). *See also* Steven M. Bellovin et al., *Lawful Hacking: Using Existing Vulnerabilities for Wiretapping on the Internet*, 12 NW. J. TECH. & INTELL. PROP. 8, 13 (2014).

in Arms Regulations (“ITAR”).²⁴⁵ Under ITAR, the President was authorized to include certain articles on the United States Munitions List.²⁴⁶ In order to export an item on the list, an individual would be required to provide further approval, such as obtaining a license from the government, akin to the seventeenth-century Stationer system in England.²⁴⁷ Conveniently, the regulation defined “export” as including any process of making cryptography public, including downloading or causing the downloading of software to the Internet. Therefore, not only was encryption code being flagged as a type of munition under the statute requiring government approval,²⁴⁸ but also the regulation defined “export” as broadly as encompassing sharing the code on the Internet.²⁴⁹ This categorization subjected encryption to “more stringent export regulations” than any other type of software.²⁵⁰

ITAR would eventually become the fatal blow to the government’s position during the Crypto Wars, winning disfavor not only with the public but also within the courts. A major criticism of ITAR was that it created a massive impediment for cryptographers, students, and scientists hoping to share cryptographic code with one another for teaching, edification, and further academic development. In other words, the government’s regulation of encryption was limiting free speech, a point underscored in the landmark encryption case, *United States v. Bernstein*. At the center of the controversy stood Daniel J. Bernstein, a graduate student at the University of California, who wrote and tried to publish code “for a zero delay private-key stream encryption based upon a one-way hash function,”²⁵¹ an

²⁴⁵ 22 C.F.R. §§ 120.1-130.17 (1999). The ITAR regulations are adopted under the authority of the Arms Export Control Act, 22 U.S.C. §§ 2751-2796(d) (1994).

²⁴⁶ 22 U.S.C. § 27786(a)(1) (1996).

²⁴⁷ Ryan Fox, *Old Law and New Technology: The Problem of Computer Code and the First Amendment*, 49 UCLA L. REV. 871, 887 (2001).

²⁴⁸ Encryption Items Transferred from the U.S. Munitions List to the Commerce Control List, 61 Fed. Reg. 68, 572-73 (Dec. 30, 1996) (codified in scattered sections of 15 C.F.R. pts. 730774).

²⁴⁹ See 15 C.F.R. § 734.2(b)(9)(i)(B)(ii).

²⁵⁰ *Junger v. Daley*, 8 F. Supp. 2d 708, 711 (N.D. Ohio 1998), *rev’d*, 209 F.3d 481 (6th Cir. 2000).

²⁵¹ *Bernstein v. Dep’t of Justice*, 176 F.3d 1132, 1136 (9th Cir.), *vacated for rehearing en banc*, 192 F.3d 1308 (1999). In Judge Betty Fletcher’s analysis, she described that Bernstein’s “one-way hash function could be employed as the heart of an encryption method. The Snuffle source code . . . was meant as an expression of how this might be accomplished . . . Snuffle was also intended, in part, as a political expression. Bernstein discovered that the . . . regulations controlled encryption exports, but not one-way hash functions. Because he believed that an encryption system could easily be fashioned from any of a number of publicly-available one-way hash functions, he viewed the distinction . . . as absurd. To illustrate his point, Bernstein developed Snuffle, which is encryption system built around a one-way hash function.” 176 F.3d at 1140 nn.11-12.

encryption program he called “*Snuffle*.”²⁵² Put simply, Bernstein wanted to publish a paper containing his mathematical equations.²⁵³ To move forward with his plan, Bernstein contacted the State Department to determine whether his code qualified as munitions under ITAR.²⁵⁴ The Department wrote back, concluding that Bernstein had to “register as an arms dealer, [and] apply for an export license,” and warned that even then a license was unlikely.²⁵⁵ Upon receiving this response, Bernstein sued for injunctive and declaratory relief in the District Court for the Northern District of California.²⁵⁶

On May 6, 1999, after four years and one regulatory change, the Ninth Circuit issued a ruling in favor of Bernstein. The court held that encryption code was more like speech than conduct,²⁵⁷ and that the requirement for a license amounted to an invalid prior restraint.²⁵⁸ In making this determination, the court admitted, “[t]he more difficult issue arises” in determining whether encryption exhibits “a close enough nexus to expression”²⁵⁹ or unprotected conduct as in *United States v. O’Brien*, in which the Supreme Court held that burning a draft card was unprotected conduct.²⁶⁰ The difficulty in making this determination was that encryption code seemed to be actively related to technology, which the court considered distinct from

²⁵² *Bernstein*, 176 F.3d at 1136 n.1.

²⁵³ Browder, *supra* note 197, at 437.

²⁵⁴ *Id.*

²⁵⁵ Alison Dame-Boyle, *EFF at 25: Remembering the Case that Established Code as Speech*, ELEC. FRONTIER FOUND. (Apr. 16, 2015), <https://www.eff.org/deeplinks/2015/04/remembering-case-established-code-speech>.

²⁵⁶ *Bernstein v. U.S. Dep’t of State*, 922 F. Supp. 1426, 1430 (N.D. Cal. 1996). He also filed an administrative appeal with respect to the department’s jurisdiction determination. His theories for relief were expansive, claiming that the regulations were “a content-based infringement on speech, act as an unconstitutional prior restraint on speech, are vague and overbroad...infringe the rights of association and equal protection” and “that the actions of defendants are arbitrary and capricious and constitute an abuse of discretion under the Administrative Procedure Act.” *Id.* at 1430-31.

²⁵⁷ *Bernstein v. Dep’t of Justice*, 176 F.3d 1132, 1141 (9th Cir.). The *Bernstein* decision generated three separate opinions from the three-judge panel. In the majority opinion, the judges appeared more convinced by the possible negative outcomes not wanting to chill scientific speech, rather than being convinced that code is speech. *See id.* (stating that code had sufficiently expressive characteristics to be speech and emphasizing that a contrary result would chill speech related to cryptography.)

²⁵⁸ *Id.* at 1144-45.

²⁵⁹ The decision generated three district court opinions and, consequently, was appealed to the Ninth Circuit. The Ninth Circuit initially ruled in favor of Bernstein but ultimately the full Ninth Circuit granted the government a rehearing en banc and withdrew the panel’s opinion. 192 F.3d 1308 (9th Cir. 1999). For a fuller history of the case, *see* Browder, *supra* note 197, at 437-40.

²⁶⁰ *United States v. O’Brien*, 391 U.S. 367 (1968).

speech. In fact, the court abstained from deciding whether object code (instructions created by a computer)²⁶¹ amounted to protected speech.²⁶² Ultimately, despite this difficulty, the Ninth Circuit decided that “cryptographers use source code to express their scientific ideas in much the same way that mathematicians use equations or economists use graphs . . . [because of this]” it held that ITAR amounted to a prior restraint on speech.”²⁶³

Soon after *Bernstein*, other Circuit Courts in the United States Courts of Appeals decided similarly, finding that cryptography was protected under the Speech Clause of the First Amendment.²⁶⁴ These decisions did not merely adopt *Bernstein*’s holding, they greatly elaborated on it, making ever more florid arguments to assert that code is speech. For example, the Second Circuit analogized code to “[m]athematical formulae and musical scores” because all three are “symbolic notations not comprehensible to the uninitiated.”²⁶⁵ The Sixth Circuit similarly wrote that code, like a musical score, is “an expressive means for the exchange of information and ideas.”²⁶⁶ Citing the Supreme Court, the Sixth Circuit

²⁶¹ Fox, *supra* note 248, at 876-77 (distinguishing source code as programming steps that are created by a programmer with a text editor in languages such as C, C++, LISP, Java, HTML and XML that can be easily read by professionals consisting of statements that demonstrate the exact steps a computer should be taking, from object code which is a “sequence of instructions that the processor can understand but that is difficult for a human to read or modify”). *See also Bernstein*, 176 F.3d at 1141 n.15; *Universal City Studios, Inc. v. Reimerdes*, 111 F. Supp. 2d 294, 306 (S.D.N.Y. 2000), *aff’d sub nom.*, *Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001). *But see Lee Tien, Publishing Software as a Speech Act*, 15 BERKELEY TECH. L.J. 629, 633 n.13 (2000) (stating that distinguishing source code and object code “for First Amendment purposes appears largely pointless”).

²⁶² Judge Bright’s separate concurrence stated that “encryption source code also has the functional purpose of controlling computers and in that regard does not command protection under the First Amendment.” *Bernstein*, 176 F.3d at 1147 (concurring opinion). Thus, the Circuit did not express an opinion on whether object code manifests a “close enough nexus to expression” to warrant application of the prior restraint doctrine. However, it should be noted that *Bernstein*’s Snuffle did not involve object code, nor does the record contain any information regarding expressive uses of object code in the field of cryptography.

²⁶³ *Bernstein v. Dep’t of Justice*, 176 F.3d 1132, 1141 (9th Cir.),. Government efforts to control encryption thus may well implicate not only the First Amendment rights of cryptographers intent on pushing the boundaries of their science, but also the constitutional rights of each of us as potential recipients of encryption’s bounty. Viewed from this perspective, the government’s efforts to retard progress in cryptography may implicate the Fourth Amendment, as well as the right to speak anonymously. *See McIntyre v. Ohio Elections Comm’n*, 115 S. Ct. 1511, 1524 (1995).

²⁶⁴ *See e.g., Corley*, 273 F.3d at 449 (2d Cir. 2001); *Junger v. Daley*, 209 F.3d 481, 485 (6th Cir. 2000); *321 Studios v. Metro Goldwyn Mayer Studios, Inc.*, 307 F. Supp. 2d 1085, 1099-1100 (N.D. Cal. 2004); *United States v. Elcom Ltd.*, 203 F. Supp. 2d 1111, 1126 (N.D. Cal. 2002).

²⁶⁵ *Corley*, 273 F.3d at 445-46.

²⁶⁶ *Junger*, 209 F.3d at 484.

stated that code mirrored the expression found in “the paintings of Jackson Pollock, [the] music of Arnold Schoenberg and [in] Lewis Carroll’s *Jabberwocky*.”²⁶⁷ Code was soon being compared to the greatest artistic expressions of all time. With this judicial consensus the Crypto Wars came to an end.

In the wake of *Bernstein*, however, some lawyers and academics appeared to disagree. A number of articles were published describing various holes in the Speech Clause justification. Some argued that the Speech Clause was not enough to shield encryption from protection because “[j]ust as there are certain classes of ‘traditional’ speech that lack standard First Amendment protection (like defamation, obscenity, or threats), there is code so dangerous that it cannot be allowed at all.”²⁶⁸ Viruses, such as ‘Michelangelo’ in the early 1990s, could be so destructive that they would erase all data on the computer’s hard drive, and spread unannounced from an unknown origin.²⁶⁹ Similarly, encryption could be described as so dangerous that it would not warrant protection.

Others argued that the Speech Clause justification for encryption produced inconsistent results²⁷⁰ because of the “desire to peg software as either all speech or not speech, which is overly simplistic.”²⁷¹ This pithy argument that all “code is speech” harkened back to the defunct *Chaplinsky* doctrine. In *Chaplinsky*, the Court established that certain categories of speech—such as hate speech, commercial speech and pornography—are outside of the Speech Clause’s protection.²⁷² Since the 1970s, however, courts have largely eviscerated these silos of speech.²⁷³ Expanding on these already lingering absolutist categories of speech today would go against the trend of First Amendment protection established over the last thirty years.²⁷⁴

²⁶⁷ *Id.* (citing *Hurley v. Irish-American Gay, Lesbian and Bisexual Group*, 515 U.S. 557, 569 (1995)); *see also* *Fox*, *supra* note 248, at 879; *Bernstein*, 176 F.3d at 1141. In essence, the argument was that program code was language, a “formal system of expression” with a set vocabulary, syntax rules, and assignment of meaning.

²⁶⁸ *Fox*, *supra* note 248, at 882.

²⁶⁹ *Id.*

²⁷⁰ *Id.* at 904 (citing Robert Post, *Recuperating First Amendment Doctrine*, 47 STAN. L. REV. 1249, 1250-51 (stating that the contemporary First Amendment doctrine is full of “internal incoherence . . . [and] its distressing failure to facilitate constructive judicial engagement with significant contemporary social issues connected with freedom of speech”)).

²⁷¹ *Id.* at 904.

²⁷² *Chaplinsky v. New Hampshire*, 315 U.S. 568, 571-72 (1942).

²⁷³ *See generally* Baranetsky, *supra* note 10; *cf.* *New York v. Ferber*, 258 U.S. 747 (1982) (holding that the First Amendment right to free speech did not preclude states from banning the sale of sexually explicit material depicting children, even if not obscene).

²⁷⁴ Tien, *supra* note 262, at 652.

Yet as many academics noted,²⁷⁵ the greatest difficulty with protecting encryption under the Speech Clause is that under this analysis courts must distinguish between encryption's expressive value and its conduct, which is not protected under the Clause.²⁷⁶ This speech/conduct doctrine was first established in *United States v. O'Brien*, where the Supreme Court found that burning a draft card was not protected speech because the action was not pure speech, but rather speech mixed with conduct.²⁷⁷ Applying an intermediate level of scrutiny test,²⁷⁸ the Court concluded that it could not "[a]ccept the view that an apparently limitless variety of conduct can be labeled 'speech' whenever the person engaging in the conduct intends thereby to express an idea."²⁷⁹

To overcome this speech/conduct hurdle, many Courts discussed encryption as merely "functional." In *Junger*, for example, the Sixth Circuit wrote that source code has both "an expressive feature and a functional feature."²⁸⁰ Similarly, the *Corley* court wrote, "the fact that a program has the capacity to direct the functioning of a computer does not mean that it lacks the additional capacity to convey

²⁷⁵ Much of these discussions were based on academics writing at the time. See, e.g., Katherine A. Moerke, Note, *Free Speech to a Machine? Encryption Software Source Code Is Not Constitutionally Protected "Speech" Under the First Amendment*, 84 MINN. L. REV. 1007 (2000); Yvonne C. Ocrant, Comment, *A Constitutional Challenge to Encryption Export Regulations: Software Is Speechless*, 48 DEPAUL L. REV. 503 (1998).

²⁷⁶ Nguyen, *supra* note 205, at 675 n.69; see LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW § 12-7 (2d ed. 1988) (noting that the speech/conduct distinction was created in labor picketing cases such as *Thornhill v. Alabama*, 310 U.S. 88 (1940) and that the dichotomy is too simple, consistent or meaningful); Cass R. Sunstein, *Words, Conduct, Caste*, 60 U. CHI. L. REV. 795 (1993) (asserting that protection of speech should be determined based on its goals of fostering democracy and equality, not on the speech/conduct distinction). See also Stephanie M. Kaufman, *The Speech/Conduct Distinction and First Amendment Protection of Begging in Subways*, 79 GEO. L.J. 1803 (1990); Paul Reidinger, *The Expressionists: When Is Conduct Speech?*, 76 A.B.A. J. 90 (1990) (surveying recent court decisions implementing the speech/conduct distinction); Sally A. Specht, *The Wavering, Unpredictable Line Between "Speech" and Conduct*, 40 WASH. U. J. URB. & CONTEMP. L. 173 (1991); Aviva O. Wertbeimer, *The First Amendment Distinction Between Conduct and Content*, 63 FORDHAM L. REV. 793 (1994).

²⁷⁷ 391 U.S. 367, 377 (1968).

²⁷⁸ The Court in *O'Brien* created a three-part test, requiring the government to merely show that the regulation (1) furthered an important or substantial government interest, (2) was unrelated to the suppression of free speech, and (3) was no greater than necessary to further the non-speech interest. *Id.* at 369, 377.

²⁷⁹ *Id.* at 376. The Court upheld a law providing that any person who "forges, alters or knowingly destroys, knowingly mutilates, or in any manner changes" his draft certificate has committed a crime. *Id.* at 370. The regulation was brought before the Court when David O'Brien burned a copy of his draft card as a sign of "public protest." *Id.* at 369.

²⁸⁰ *Junger v. Daley*, 209 F.3d 481, 484 (6th Cir. 2000).

information.”²⁸¹ Speech that is merely functional, unlike conduct, warrants protection according to First Amendment doctrine.²⁸² For example, how-to manuals, blueprints, and even guides for committing murder have all received First Amendment protection.²⁸³ Employing this analysis, courts were able to describe encryption’s practical effect without having to exclude it from protection under the Speech Clause.²⁸⁴

However, this analogy was hard pressed.²⁸⁵ Encryption often does more than a blueprint or a score of music; it often behaves like the piano, as well as the notes. As other commentators in the wake of *Bernstein* mentioned, the Speech Clause is “ill-suited to the realities of computer technology because software inseparably incorporates elements of both expression and function.”²⁸⁶ This is precisely the hiccup that the District Court in *Junger* pointed to when it stated, “source code is by

²⁸¹ *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 447-48 (2d Cir. 2001) (“A recipe is no less ‘speech’ because it calls for the use of an oven, and a musical score is no less ‘speech’ because it specifies performance on an electric guitar. Arguably distinguishing computer programs from conventional language instructions is the fact that programs are executable on a computer. But the fact that a program has the capacity to direct the functioning of a computer does not mean that it lacks the additional capacity to convey information, and it is the conveying of information that renders instructions ‘speech’ for purposes of the First Amendment. The information conveyed by most ‘instructions’ is how to perform a task.”).

²⁸² Writing an amicus brief in the *Bernstein* case, several professors tried to distinguish that “[p]rogram code conveys information both from and to humans; humans themselves write the program code. Such code is admittedly characterized by functionality, but this alone fails to distinguish it from other languages.” Brief for Bernstein as Amici Curiae Supporting Appellee at 2, *Bernstein*, 192 F.3d 1308 (9th Cir. 1999) (No. 97-16686). Similarly, the Court in *Junger* wrote, “[t]he fact that a medium of expression has a functional capacity should not preclude constitutional protection.” 209 F.3d at 484-85. See also *Bernstein*, 922 F. Supp. at 1435-36 (stating that the functional nature of source code is “immaterial” to the free speech analysis).

²⁸³ See *United States v. The Progressive, Inc.*, 467 F. Supp. 990 (W.D. Wis. 1979) (finding that a set of instructions for how to build a nuclear bomb was speech and stating that prior restraints are unconstitutional); *Rice v. Paladin Enterprises, Inc.*, 940 F. Supp. 836 (S.D. Md. 1996). These informational products, like computer program code, convey both expressive and functional content; in many cases, the functional content of such speech far exceeds its creative expression. However, First Amendment jurisprudence does not support diminished protection for speech based on its functionality. See, e.g., *id.*

²⁸⁴ As Lee Tien has suggested, however, functionality is really a “proxy for effect or harm” because there are strong causation issues inherent in the relationship between the code’s function and the harm caused by its use. Tien, *supra* note 262, at 695.

²⁸⁵ Fox, *supra* note 248, at 876 (“[T]his binary ‘is it speech or not’ question, and the doctrine upon which the functionality questions are based, are fundamentally flawed discussions that place computer code in a tenuous position from the start.”).

²⁸⁶ Nguyen, *supra* note 205, at 675-76. See also Fox, *supra* note 248, at 873; Tien, *supra* note 262, at 652.

design functional: it is created and, if allowed, exported to do a specified task, not to communicate ideas.”²⁸⁷ Unlike other software, encryption code “actually do[es] the function of encrypting data.”²⁸⁸ With this thought in mind, the *Bernstein* court reserved judgment on object code—“the pure instructional data that is created to run directly on a computer’s processor”—because object code is even more distinctly part of conduct.²⁸⁹ To make matters more complicated, it is this element of encryption—its act of encrypting—that government is wary of because it undermines government power²⁹⁰ and can have deleterious secondary effects by helping nefarious minds communicate.²⁹¹ Therefore, protection under the Speech Clause is limited.

While most courts dodged the conduct analysis,²⁹² one court did not. In *Karn v. United States Department of State*,²⁹³ the District Court for the District of Columbia applied the conduct analysis from *O’Brien* to reach an unfavorable decision with regard to encryption. In 1994, Philip Karn, like Bernstein, submitted a request to the Department of State to determine whether a floppy disk containing code from a book titled *Applied Cryptography* would be subject to ITAR.²⁹⁴ The

²⁸⁷ *Junger v. Daley*, 8 F. Supp. 2d 708, 717 (N.D. Ohio 1998).

²⁸⁸ *Id.* at 720. The lower court continued that although “encryption source code is occasionally expressive,” it was not “sufficiently communicative” because “encryption software is especially functional rather than expressive” so “its export is not protected conduct under the First Amendment.” *Id.* at 715-16.

²⁸⁹ Fox, *supra* note 248, at 880-81 (describing object code as looking to a human like “nothing more than a string of seemingly meaningless numbers” in contrast to source code, which is more meaningful than the English language in certain circumstances, but still arguing that “the idea of removing First Amendment protection directly on determination of the particular type of code at issue, while an easy test to perform, needs to be replaced” because “there are some pieces of object code that by their nature or their use implicate values served by the First Amendment.”).

²⁹⁰ *Id.* at 881 (citing LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 5 (1999) (stating that the values in different codes “can even be political in nature; code can represent and change power struggles in cyberspace”)).

²⁹¹ While cryptography can protect the privacy of journalists, bank records and other transactions, other injurious uses of cryptography do exist, such as crime and terrorism. “The ineluctable fact is that the government’s interest in preventing technology from falling into the hands of terrorists and criminals is not significantly advanced by curtailing the publication of source code.” Browder, *supra* note 197, at 443.

²⁹² In *Junger*, the court wrote about function, rather than conduct to sidestep this problem. *Junger v. Daley*, 209 F.3d 481, 485 (6th Cir. 2000).

²⁹³ *Karn v. United States Dep’t of State*, 925 F. Supp. 1 (D.D.C. 1996), *remanded*, 107 F.3d 923 (D.C. Cir. 1997).

²⁹⁴ *Id.* at 3. In fact, Karn submitted an application for the book as well as a diskette containing the same exact information. The government determined the book was not subject to the jurisdiction of the State Department under ITAR, but that the diskette was. See Kerben, *supra* note

government took the position that the code was subject to the regulation as a defense article and, therefore required a license.²⁹⁵ Like Bernstein, Karn filed suit against the government, arguing that the government's regulation of the transport of the diskette violated his First Amendment rights.²⁹⁶

By applying the *O'Brien* test,²⁹⁷ the District Court in *Karn* awarded summary judgment to the government and dismissed the case.²⁹⁸ To reach this conclusion, the lower court reasoned ITAR was content-neutral because the government merely intended to stop encryption which threatened national security.²⁹⁹ Although the case was appealed, the Circuit Court remanded it to back to the lower court for rehearing because of a change in law.³⁰⁰ To this day, *Karn* stands in opposition to *Bernstein*, for the proposition that "so long as judges continue to see functional aspects [of encryption] as being nonexpressive conduct," *O'Brien* will likely make protection of encryption under the Speech Clause difficult.³⁰¹

Although both perspectives of the speech versus conduct debate are convincing, neither is more correct than the other. Code is at the same time speech *and* conduct. Trying to create a speech/nonspeech division through the *O'Brien* analysis is a fallacy, as John Hart Ely explained in a canonical *Harvard Law Review* article.³⁰² Discussing *O'Brien*, Ely noted that just as "burning a draft card to express opposition to the draft is an undifferentiated whole, 100% action and 100% expression encryption involves no conduct that is not at the same time

203, at 152 n.185 (explaining Karn was interested in showing the incongruities of the law and so he tested the law by applying for the declaratory judgment on the book to compare it with the information later tested on the diskette).

²⁹⁵ *Karn*, 925 F. Supp. at 4.

²⁹⁶ *Id.*

²⁹⁷ *Id.* at 10.

²⁹⁸ *Id.* at 14.

²⁹⁹ *Id.* at 10 (stating encryption made it "easier for foreign intelligence targets to deny the United States Government access to information vital to national security interests").

³⁰⁰ *Id.*

³⁰¹ Fox, *supra* note 248, at 886 (stating the test articulated in *Spence* will "also erect further barriers if one does not understand exactly to whom the 'speaker' of the code is speaking").

³⁰² See John Hart Ely, *Flag Desecration: A Case Study in the Roles of Categorization and Balancing in First Amendment Analysis*, 88 HARV. L. REV. 1482, 1495 (1975) ("[B]urning a draft card to express opposition to the draft is an undifferentiated whole, 100% action and 100% expression. It involves no conduct that is not at the same time communication, and no communication that does not result from conduct. Attempts to determine which element 'predominates' will therefore inevitably degenerate into question-begging judgments about whether the activity should be protected.").

communication”³⁰³ because the same structure used to express idea can be “transformed into an instrument of conducts.”³⁰⁴ An apt constitutional justification, therefore, would recognize this aspect of code as both expressive speech and as enabling conduct.³⁰⁵ For this reason, even though encryption retains its constitutional protection under the Speech Clause,³⁰⁶ a helpful alternative method of protection would acknowledge this dual nature of encryption.³⁰⁷

C. *Today’s Crypto Wars and Bernstein*

In the decades following *Bernstein*, the climate quickly shifted, and many proclaimed that the Crypto Wars were dead.³⁰⁸ *Bernstein* had altered the tide for cryptographers by securing protection for encryption under the Speech Clause.³⁰⁹ In addition to the positive jurisprudence, encryption soon became ubiquitous. No longer were a limited number of government officers and a few technical specialists the only individuals with access to encryption technology.³¹⁰ Instead, encryption began appearing in various household devices (such as DVDs, Blu-ray players, wireless routers), as well as everyday online billing services, bank transactions, HTTPS, and commonplace operating systems (such as Microsoft Windows).³¹¹ Given its new omnipresence, encryption became seemingly impossible to regulate.

For many years, in this stable climate, encryption ushered in a new level of free expression. As the American Civil Liberties Union stated in a report to the UN

³⁰³ *Id.*

³⁰⁴ Fox, *supra* note 248, at 907.

³⁰⁵ See *id.* at 873 (“Simply put, code is the instructions people write to tell computers what to do. Computers operate by executing those instructions, thus code is fundamental to all computer technology.”).

³⁰⁶ See *Junger v. Daley*, 209 F.3d 481, 485 (6th Cir. 2000).. The strength for protection of encryption under the Speech Clause is still somewhat debatable. In addition to *Bernstein* being removed by the panel, the Circuit in *Junger* did not explicitly state whether encryption should receive First Amendment protection, but merely that the district court too quickly accepted the government’s assertions for the need of export regulations. See Fox, *supra* note 248, at 876 (“[T]his binary ‘is it speech or not’ question, and the doctrine upon which the functionality questions are based, are fundamentally flawed discussions that place computer code in a tenuous position from the start.”).

³⁰⁷ Cf. TRIBE, *supra* note 277, at 978 n.2 (“It would . . . be wrong to conclude that some form of first amendment scrutiny is triggered whenever government does anything that happens to reduce the flow of information or ideas.”).

³⁰⁸ See ABELSON et al., *supra* note 226, at 8.

³⁰⁹ See *id.* This statement is made with the caveat of the *Karn* ruling.

³¹⁰ See Pilkington, *supra* note 217, at 168 (discussing how private encryption technology has steadily grown with the advancement of computer technology).

³¹¹ Duong, *supra* note 207, at 326.

Human Rights Council, “encryption and anonymity are the only effective safeguards against [the government’s] assault on freedom of expression and association.”³¹² Similarly, a report by David Kaye, UN Special Rapporteur for Freedom of Expression and Opinion, stated that “[e]ncryption and anonymity provide individuals and groups with a zone of privacy online to hold opinions and exercise freedom of expression without arbitrary and unlawful interference or attacks[;] encryption and anonymity are protected because of the critical role they can play in securing those rights.”³¹³ Another report went on to even credit encryption with having improved the economy.³¹⁴

Recently, the consensus that encryption is good for democracy has once more been called into question.³¹⁵ Following the revelations made by Edward Snowden in June 2013 that the NSA maintains a pervasive surveillance program³¹⁶ civic society organizations, technology companies, journalists, and members of the public came forward once more to embrace encryption as a method to resisting government overreach. This coalition of groups highlighted encryption as crucial to ensuring secure messaging free from government intrusion. However, government officials as well as law enforcement revived their critique and warned against the going dark effect, using recent terrorist attacks as a rallying cry to regulate encryption.

In 2010, for instance, the FBI advocated that Congress legislate that all communications systems create backdoors for law enforcement.³¹⁷ Immediate pushback from technology companies, privacy advocates, and the public caused the initiative to quickly fade, but the FBI returned to Congress in 2013, urging it to extend CALEA to online companies that encrypt their messaging services.³¹⁸ Yet again, the FBI’s urgings failed. After the attacks in Paris and San Bernardino,

³¹² *ACLU Submission*, *supra* note 105.

³¹³ KAYE, *supra* note 200, at 3 (“Encryption and anonymity, today’s leading vehicles for online security, provide individuals with a means to protect their privacy, empowering them to browse, read, develop and share opinions and information without interference and enabling journalists, civil society organizations, members of ethnic or religious groups, those persecuted because of their sexual orientation or gender identity, activists, scholars, artists and others to exercise the rights to freedom of opinion and expression.”).

³¹⁴ Kehl et al., *supra* note 211, at 20 (“[o]ver the past fifteen years, a virtuous cycle between strong encryption, economic growth, and support for free expression online has evolved”).

³¹⁵ *See id.* at 21.

³¹⁶ *See id.*

³¹⁷ *See* Charlie Savage, *U.S. Tried to Make it Easier to Wiretap the Internet*, N.Y. TIMES (Sept. 27, 2010), http://www.nytimes.com/2010/09/27/us/27wiretap.html?_r=2&; *see also* Ryan Singel, *FBI Drive for Encryption Backdoors is Déjà vu for Security Experts*, WIRED (Sept. 27, 2010), <https://www.wired.com/2010/09/fbi-backdoors-2/>.

³¹⁸ *Id.*

however, law enforcement's commitment became unwavering. In 2014, FBI Director James Comey proclaimed that encryption was a serious threat to national security.³¹⁹ Soon after, the agency filed suit against Apple and Congress introduced the Burr-Feinstein anti-encryption bill.³²⁰

Those in favor of encryption responded with many of the same arguments made in the 1990s, as well as some new ones.³²¹ In *Apple v. FBI*, for instance, the government argued that according to an antiquated law called the All Writs Act, Apple was required to comply with the FBI's request and unlock the iOS (Apple's iPhone operating system) of a San Bernardino shooter.³²² In response, Apple fell back on the stalwart recitation of speech cases starting with *Bernstein*. In its motion to vacate, the company wrote, "[t]he government asks this Court to command Apple to write software that will neutralize safety features . . . [but u]nder well-settled law, computer code is treated as speech within the meaning of the First Amendment

³¹⁹ James B. Comey, *Going Dark: Are Technology, Privacy, and Public Safety on a Collision Course?*, Speech at the Brookings Institution (Oct. 2014), <https://www.fbi.gov/news/speeches/going-dark-are-technology-privacy-and-public-safety-on-a-collision-course> ("Cyber adversaries will exploit any vulnerability they find. But it makes more sense to address any security risks by developing intercept solutions during the design phase, rather than resorting to a patchwork solution when law enforcement comes knocking after the fact. And with sophisticated encryption, there might be no solution, leaving the government at a dead end—all in the name of privacy and network security.").

³²⁰ See Compliance with Court Orders Act of 2016 (discussion draft), 114th Cong. (2d Sess. 2016), <http://www.burr.senate.gov/imo/media/doc/BAG16460.pdf> (last accessed Mar. 26, 2017). Ultimately, the bill was not passed. Rainey Reitman, *Security Win: Burr-Feinstein Proposal Declared "Dead" for This Year*, ELEC. FRONTIER FOUND. (May 27, 2016), <https://www EFF.org/deeplinks/2016/05/win-one-security-burr-feinstein-proposal-declared-dead-year>. Recent developments following the election of President Donald Trump and his new Administration have caused many anticipate that the Administration will take a strong stance against encryption, especially because of heavy leaks from the White House. Chris Kanaracus and Steve Wilson, *Expect Renewed Push for Encryption Backdoors from Trump Administration*, ZDNET (Jan. 26, 2017, 2:39 PM), <http://www.zdnet.com/article/expect-renewed-push-for-encryption-backdoors-from-trump-administration/>.

³²¹ Kehl et al., *supra* note 211, at 21 (citing Statement of Kevin S. Bankston, Policy Director of New America's Cybersecurity Initiative, *Hearing on Encryption Technology and U.S. Policy Responses, Before the US House of Representatives Subcommittee on Information Technology of the Committee on Oversight and Government Reform*, (Apr. 29, 2015), https://static.newamerica.org/attachments/2982-at-crypto-hearing-best-arguments-against-backdoor-mandates-come-from-members-of-congress-themselves/Bankston_Written_Testimony.5876d326c5fc4e0cbd17b59e8d53384f.pdf).

³²² 28 U.S.C. § 1651(a) (stating federal courts may "issue all writs necessary or appropriate in aid of their respective jurisdictions and agreeable to the usages and principles of law").

(citing *Bernstein*, *Corley*, *Junger*, and *MGM*).³²³ At least three other amicus briefs supporting Apple's position also cited *Bernstein*. The case also began appearing in news articles around the country discussing the government's position.³²⁴

Although the court in *Apple* never reached a decision because the government found an alternate method to access the phone,³²⁵ the case solidified *Bernstein* as the predominant legal approach for encryption advocates, and, moreover, that the Crypto Wars were back. In this second epoch of the Crypto Wars, this seemingly familiar debate has an "added urgency" that requires consideration of new legal analysis to take into account the development of the surveillance state and the growing dependency on technology and the Internet.³²⁶ While some commentators have posed important new economic and policy arguments for protecting encryption, this Article suggests the Press Clause as novel legal justification for encryption.³²⁷

D. Encryption under the Press Clause

In addition to the *Bernstein* line of cases, the Press Clause should be utilized as a legal justification for protecting encryption because it highlights certain descriptive qualities of encryption that the Speech Clause does not encapsulate. For instance, the Press Clause specifically protects the use of tools that moderate the flow of information even if they might appear to threaten national security. In addition to its descriptive power, the Press Clause also has normative force, as it underscores the tool's structural importance for checking government power in this "golden age of surveillance."³²⁸ In this way, the Press Clause possibly promises to restore the delicate balance between government and individual power to "something closer to what the Founding Fathers intended."³²⁹

³²³ See Apple's Motion to Vacate, dated 32 February 25, 2016. ED No. CM 16-10 (SP), <https://www.justsecurity.org/wp-content/uploads/2016/02/apple.motiontovacate.pdf>.

³²⁴ See Eric Geller, *The Rise of the New Crypto War*, DAILY DOT (July 10, 2015), <http://www.dailydot.com/politics/encryption-crypto-war-james-comey-fbi-privacy/>.

³²⁵ The court never reached the First Amendment question because the government dropped the case before the court issued an opinion.

³²⁶ ABELSON, *supra* note 226, at 10.

³²⁷ *Encryption Technology and Possible U.S. Policy Responses: Hearing Before the U.S. H.R. Subcomm. on Info. Tech. of the Comm. on Oversight and Gov't Reform* (2015) (statement of Kevin S. Bankston) at 11, <http://oversight.house.gov/wp-content/uploads/2015/04/4-29-2015-IT-Subcommittee-Hearing-on-Encryption-Bankst on.pdf>.

³²⁸ *Id.*

³²⁹ *Id.*

1. *Press Clause is Descriptively Accurate at Protecting Encryption as a Tool*

While the Speech Clause is effective at protecting encryption³³⁰ the Press Clause provides a significantly more descriptive justification because of the tool's similarities between to the printing press.³³¹ To begin, "[a] printing press is a device for applying pressure to an inked surface resting upon a print medium (such as paper or cloth), thereby transferring the ink."³³² More generally, however, the invention of the printing press was revolutionary because it allowed individuals to control the spread of information by enabling speedy, copious, and selective publication. This ability for citizen circulation of information, or the "democratization of knowledge" led to a substantial increase in human knowledge that spawned a fundamental change in human society, later named the Printing Revolution, that started soon after Gutenberg's invention in 1440.³³³

Encryption, like the printing press, is a modern tool that not only controls the flow of information, but also protects individuals from government interference. Its code works as a tool or "technology"³³⁴ by "encoding of information called "plaintext," into unreadable form, termed "ciphertext."³³⁵ The tool commonly allows academics, journalists, activists, and others to communicate with each other freely by "prevent[ing] anyone other than the user or intended recipient from reading private information."³³⁶ As a recent UN report has described, encryption tools are "protected because of the critical role they can play in securing" the ability for

³³⁰ This is especially the case given the Speech Clause's reputation as the "lodestar of our Constitution." *W. Va. State Bd. of Educ. v. Barnette*, 319 U.S. 624, 639 (1943). In recent years, particularly following *Citizens United*, the Clause has become very effective at providing protection for actions that go beyond traditional notions of speech.

³³¹ See generally WARREN CHAPPELL & ROBERT BRINGHUSRT, *A SHORT HISTORY OF THE PRINTED WORD* (2000) (discussing the history of printing).

³³² Wikipedia, *Printing Press*, https://en.wikipedia/wiki/Printing_press (last visited Feb. 25, 2017).

³³³ *Id.*

³³⁴ *Hearing before Senate Judiciary Comm.'s Subcomm. on Constitution, Federalism and Property Rights* (1998) (statement of Cindy Cohn, Lead Counsel in *Bernstein*), https://www.eff.org/files/filenode/bernstein/19980317_cohn_testimony.html ("From a legal standpoint, the *Bernstein* case is not complex, nor does it break any dramatic new ground. It simply asks the courts to recognize that the First Amendment extends to science on the Internet, just as it does to science on paper and in the classroom. For it is this scientific freedom which has allowed us to even have an Internet, as well as the many other technologies which we enjoy today.").

³³⁵ Duong, *supra* note 205, at 324.

³³⁶ *Id.*

individuals to control the exchange of knowledge “without arbitrary and unlawful interference or attacks.”³³⁷

Seemingly, one difficulty in drawing the analogy between encryption and the printing press arises from the tools’ two different ways of promoting information. While the printing press appears to make information more easily visible to the public and encryption appears to conceal it,³³⁸ that characterization is overly simplistic. Like encryption, the printing press has historically helped publishers review and edit information before selectively deciding what information to release into the public domain, and what information to keep private. John Milton’s own description of the printing press highlighted its significance as the “power of determining . . . what ought to be published and what [ought] to be *suppressed*.”³³⁹ When the Crown confiscated John Wilkes’ materials, for example, many of the seized documents he had printed on his press were unpublished materials he had editorially decided to keep private, and not to share with the public.³⁴⁰ Similarly, Max Frankel described in his aforementioned affidavit how in the 1970s, newspapers used printing presses to selectively publish certain stories for public consumption.³⁴¹ In the modern day, publications continue to use printing tools, such as websites, to choose what to share and what to censor. For example, only key texts from the Snowden leaks and the Panama Papers were published from the thousands of pages made available to journalists.³⁴² Editors choose to elevate certain points but not publishing others.

Encryption has largely the same effect, especially given the copious amount of information in our society. Today the public domain is oversaturated with data. With the click of a button, news articles, documents, and other information are easily made public in seconds. By encrypting certain information and removing those messages from the public domain, individuals are able to highlight the information

³³⁷ KAYE, *supra* note 200, at 7.

³³⁸ A. Michael Froomkin, *The Metaphor is the Key: Cryptography, the Clipper Chip, and the Constitution*, 143 U. PA. L. REV. 709, 712 (1995) (“Without the ability to keep secrets, individuals lose the capacity to distinguish themselves from others, to maintain independent lives, to be complete and autonomous persons . . . This does not mean that a person actually has to keep secrets to be autonomous, just that she must possess the *ability* to do so. The ability to keep secrets implies the ability to disclose secrets selectively, and so the capacity for selective disclosure at one’s own discretion is important to individual autonomy as well.”).

³³⁹ MILTON, *supra* note 53 (emphasis added).

³⁴⁰ See Mellen, *supra* note 69.

³⁴¹ See Frankel Affidavit, *supra* note 145.

³⁴² Frank Jordans, *German Newspaper that First Obtained the Panama Papers Says It Won’t Publish All the Files*, BUSINESS INSIDER (Apr. 7, 2016), <http://www.businessinsider.com/some-panama-papers-files-wont-be-published-2016-4>.

which *is* public in the growing sea of data. By sending documents over encrypted channels, for example, and then publishing particular sections, rather than merely dumping the data on the Internet, encryption allows individuals to bring to the surface only those points they wish to share with readers. As Professor Froomkin discusses in his seminal article about encryption, *The Metaphor is Key*, “[t]he ability to keep secrets implies the ability to disclose secrets selectively.”³⁴³ For these reasons, just as the printing press before it, encryption has become an effective, powerful, and widespread tool for information sharing, used in “commercial, political, and personal life in a surprising number of ways.”³⁴⁴

2. *Press Clause Underscores the Structural Role of Encryption*

Practically, encryption is used to ensure a zone of privacy for individuals where one can share information securely. Encryption has become imperative for providing these spaces as technological realms are increasingly surveilled by governments or hacked by private parties. Indeed, since the Snowden revelations, journalists have increasingly turned to in-person communication to lower the risk of interference. As Professor Froomkin explains, encryption “is important to individual autonomy” in a landscape where spheres of privacy can easily be intruded on by government and other actors.³⁴⁵

It is this structural aspect of encryption that poses a threat to government much more than its speech characteristics. Law enforcement is concerned over the physical barrier encryption creates—the going dark effect—not its written code. Therefore, just like the British Crown’s “despotick” regulations over the private use of the printing press in the eighteenth century, today members of the United States government have tried to impose restraints on private forms of encryption rather than limiting the writing of code itself.³⁴⁶ During the course of the *Apple* litigation, *amici* explained this tension by stating,³⁴⁷ “[t]his case is about giving the government the

³⁴³ Froomkin, *supra* note 339, at 712.

³⁴⁴ *Id.* at 718.

³⁴⁵ Froomkin, *supra* note 339, at 713.

³⁴⁶ As explained above, many instances of the government regulating encryption in the early 1990s made individuals concerned about their safety. For example, in 1993, Philip Zimmermann, the inventor of PGP, was visited by U.S. Customs Service agents “who were investigating a complaint from RSA Data Security alleging the theft and international shipment of their intellectual property. The seed of this . . . inquiry quickly bloomed into an investigation of possible ITAR violations by a U.S. Attorney. For years, Zimmermann remained under an investigatory cloud but was never indicted.” Vagle, *supra* note 202, at 114.

³⁴⁷ Laura Hautala, *Cybersecurity Pros Fear Government Overreach in iPhone Fight*, CNET.COM, Mar. 3, 2016, <https://www.cnet.com/news/apple-versus-fbi-iphone-encryption-cybersecurity-experts-fear-government-overreach/>; Steve Tobak, *Why Apple’s Tim Cook Is Right*

power to conscript technology providers[,] to create new versions of their products intended solely to defeat the security features designed to safeguard their users.”³⁴⁸ After the government backed down from the case, FBI director James Comey did not relent.³⁴⁹ Comey warned “the balance we have long struck is fundamentally challenged and changed,”³⁵⁰ and once more asserted going dark effect arguments.³⁵¹ More recently, incidents from the Trump Administration have shown similar dislike for encrypted messaging,³⁵² especially as government leaks increase.

Unlike the Speech Clause, the Press Clause was created to address this precise tension and strike a balance in favor of the citizen. The Framers’ intended the Clause to create a Constitutional protection for enabling unencumbered spaces of *private* communication, free from government surveillance and retribution. Or as Justice Douglas stated, the Clause was designed “to take Government off the backs of

on *Encryption*, FOXBUSINESS.COM, Feb. 19, 2016, <http://www.foxbusiness.com/features/2016/02/19/why-apples-tim-cook-is-right-on-encryption.html>; John Eden, *Why Apple is Right to Resist the FBI*, TECHDIRT.COM, Mar. 13, 2016, <https://techcrunch.com/2016/03/13/why-apple-is-right-to-resist-the-fbi/>.

³⁴⁸ Brief for Center for Democracy & Technology as Amicus Curiae, Motion to Vacate and In Opposition to Government’s Motion to Compel Assistance (2016) (No. CM 16–10 (SP)), http://images.apple.com/pr/pdf/Center_for_Democracy_and_Technology.pdf.

³⁴⁹ Kevin Johnson, *FBI Director Reflects on Apple Dispute*, USA TODAY (Apr. 12, 2016, 6:36 PM), <http://usatoday.com/story/news/politics/2016/04/12/james-comey-fbi-apple/82940460/>. Comey stated to the public that “Apple is not a demon” and that he hoped the “people don’t perceive the FBI as a demon.” *Id.*

³⁵⁰ *Id.*

³⁵¹ Eric Tucker, *Comey: FBI Wants ‘Adult Conversation’ on Device Encryption*, ASSOCIATED PRESS (Aug. 20, 2016, 9:10 PM), <http://bigstory.ap.org/article/7d57f576e3f74b6ca4cd3436fbeb160/comey-fbi-wants-adult-conversation-device-encryption> (quoting Comey stating, “Widespread encryption built into smartphones is ‘making more and more of the room that we are charged to investigate dark.’”); Tom Winter, Tracy O’Connor and Pete Williams, *Comey: FBI Couldn’t Access Hundreds of Devices Because of Encryption*, NBC NEWS (Mar. 8, 2017), <http://www.nbcnews.com/news/us-news/comey-fbi-couldn-t-access-hundreds-devices-because-encryption-n730646> (quoting Comey saying, “Picture a room. There’s always been a corner of the room that was dark to us, where nation-states and the most sophisticated criminals would find ways to encrypt their data.”).

³⁵² After the Trump Administration launched an investigation into the increased leaks coming from the White House, the White House summoned their staff to force deletion of encryption apps such as Confide and Signal from staffers’ phones. Cory Doctorow, *Trump vs. Leaks: Spicer’s Staff Forced to Undergo “Phone Searches” and Delete Privacy Apps*, BOINGBOING (Feb. 27, 2017), <http://boingboing.net/2017/02/27/trump-vs-leaks-spicers-staf.html>; Lily Hay Newman, *Encryption Apps Help White House Staffers Leak-And Maybe Break the Law*, WIRED (Feb. 15, 2017), <https://www.wired.com/2017/02/white-house-encryption-confide-app/>

people.³⁵³ Just as government imposing prior restraints on newspapers for publishing confidential ideas is deemed incorrect under the Press Clause,³⁵⁴ “law enforcement and intelligence agencies [should not] have the capability to penetrate [citizens’] secrets at will.”³⁵⁵

3. The Press Clause is Especially Apt at Protecting Encryption Even if Countervailing National Security Exist

The Press Clause is especially potent because its protections apply even where countervailing national security interests exist. As previously explained, this tenet was established in *New York Times Co. v. United States*. There, the Court dismissed the government’s argument that national security concerns alone could justify thwarting a newspaper from publication. Indeed, “[e]ven though the case has nine separate written opinions, the majority of [] Justices [still] found that a national security interest, without more, was too amorphous a rationale to abrogate the protections of the First Amendment” and the Press Clause in particular.³⁵⁶

Justice Black’s opinion in *New York Times* most clearly explains this rationale. “[T]he press must be left free[,]” no matter what, he argued, because the press exists “[to] bare the secrets of government.”³⁵⁷ The Press Clause was intended “to serve the governed, not the governors” Black wrote.³⁵⁸ The idea behind this absolutism, Justice Black explained, originated from the belief that the greater the threat to national security, the greater the need to preserve the right to a free press because “security” is often used by government as a decoy, “a broad, vague generality.”³⁵⁹ Its “contours,” he wrote, “should not be invoked to abrogate the fundamental law embodied in the First Amendment.”³⁶⁰ And “[i]f the government had the inherent power to halt the free flow of information and ideas based on national security interests alone, it could wipe out the First Amendment and destroy the fundamental liberty of the very people the government hopes to make secure.”³⁶¹

Although Justice Black’s position appears extreme, he was not alone in his reasoning. Justice Stewart, for instance, stated that despite the government’s claims,

³⁵³ *Columbia Broadcasting Sys., Inc. v. Democratic Nat’l Comm.*, 412 U.S. 94, 162 (1973) (Douglas, J., concurring); *see also* Glancy, *supra* note 171, at 1049.

³⁵⁴ Froomkin, *supra* note 339, at 712; *see New York Times Co.*, 403 U.S. at 713.

³⁵⁵ Froomkin, *supra* note 339, at 713.

³⁵⁶ Crain, *supra* note 217, at 894.

³⁵⁷ *Id.*

³⁵⁸ *New York Times Co. v. United States*, 403 U.S. 717 (1971) (Black, J., concurring).

³⁵⁹ *Id.*

³⁶⁰ *Id.*

³⁶¹ Crain, *supra* note 217, at 894.

national security alone could not justify halting publication. Justice Stewart explained that if publishing the Pentagon Papers would not “result in direct, immediate, and irreparable damage to our Nation or its people” then prior restraint was not justified.³⁶² This standard set a high bar for government restraint. Collectively, “[t]he Supreme Court decision in the Pentagon Papers case was a clear defeat for government claims of national security and an equally clear victory for freedom of the press.”³⁶³ Similarly, only in very few cases where “direct, immediate, and irreparable damage to our Nation or its people” could result from encryption, would the tool not be protected.

4. *The Press Clause Mirrors Practices of Founding Fathers*

Although there is no record of encryption being intentionally protected as a “tool” under the Press Clause or elsewhere in the founding documents, we may also consider the practices of the Founding Fathers to determine whether they may have considered protecting it.

Just as many of the Founders have been cited for their uses of the printing press,³⁶⁴ examining the practices of the Founders from the time of the Revolution to the adoption of the United States Constitution reveals many instances where the Founders used secret communication tools similar to encryption. For instance, the “one if by land, two if by sea” code employed by Paul Revere warned of British forces by displaying lanterns during the battles at Lexington and Concord. In addition to surreptitious codes and signals, the Founders also used ciphers. As John Fraser wrote, “[f]rom the beginnings of the American Revolution in 1775 until the adoption of the United States Constitution, Americans used codes, ciphers and other secret writings to foment, support, and carry to completion a rebellion against the British government.”³⁶⁵ Fraser explains, “America was born of revolutionary conspiracy . . . as rebels and conspirators, the young nation’s leaders . . . turned to codes and ciphers in an effort to preserve the confidentiality of their communications.”³⁶⁶

³⁶² *New York Times Co.*, 403 U.S. at 730 (Stewart, J., concurring).

³⁶³ Jerome A. Barron, *The Pentagon Papers Case and the Wikileaks Controversy: National Security and the First Amendment*, 1 WAKE FOREST J. OF LAW & POL’Y 47, 49 (2011).

³⁶⁴ West, *The ‘Press’ Then & Now*, *supra* note 12, at 79.

³⁶⁵ Fraser, *supra* note 204, at 21.

³⁶⁶ *Id.* at 21 n.63 (citing RALPH E. WEBER, MASKED DISPATCHES CRYPTOGRAMS AND CRYPTOLOGY IN AMERICAN HISTORY 1775-1900, at 5-6 n.6. (1993)). These practices were fueled by the Crown’s habit of opening and reading private mail. Similarly, the Continental Congress ordered its Committee handling foreign correspondence to use “cyphers.” For instance, the

These tools were in no way reserved for select communications. Many of the Founders used encryption in their daily life, including Thomas Jefferson, John Adams,³⁶⁷ James Madison, John Jay, and Benjamin Harrison.³⁶⁸ John Adams famously used a cipher to write letters to his wife, Abigail Adams, as well as in his official business correspondence.³⁶⁹ Madison used ciphers for private correspondence, as well as to discuss actions with fellow actors in the Revolution, including Thomas Jefferson. In fact, “[p]rior to the adoption of the Bill of Rights, Madison and Jefferson . . . used a 1700-word code for confidential discussion of sensitive personal and political issues.”³⁷⁰ In a letter written to Madison dated August 28, 1789, Jefferson’s “comments on the proposed First Amendment is partially enciphered.”³⁷¹ These comments would later become the text of the First Amendment.³⁷²

The Founders did not merely use these tools, but developed them themselves. Thomas Jefferson and Benjamin Franklin, perhaps best known for their use of the printing press, are lesser known for their invention of encryption tools.³⁷³ In 1764, Jefferson “suggested to John Page the use of a hundred-year-old English text (Shelton’s *Tachygraphia*) to encode their letters.”³⁷⁴ Jefferson also invented a wheel cipher consisting of twenty-six cylindrical wooden pieces on an iron spindle.³⁷⁵

Committee of Secret Correspondence, which later rebuked the Stamp Act of 1765 became a powerful group opposing the Crown. *Id.*

³⁶⁷ *Id.* at 23 (stating John and Abigail Adams, his wife, used a cipher provided by James Lovell for family correspondence while John Adams was away from home).

³⁶⁸ *ACLU Submission*, *supra* note 105, at 14.

³⁶⁹ See generally DAVID KAHN, *THE CODE-BREAKERS: THE COMPREHENSIVE HISTORY OF SECRET COMMUNICATION FROM ANCIENT TIMES TO THE INTERNET* 181 (1996); WEBER, *supra* note 367.

³⁷⁰ Fraser, *supra* note 204, at 44.

³⁷¹ *Id.* at 43.

³⁷² *Id.*

³⁷³ *Id.* at 70 (“By protecting his communications and raising a shield of privacy around his intentions and statements, Jefferson protected himself against government and private interlopers. Relative to the government’s abilities in 1796 (which were comparatively weak) Jefferson was able to maintain control over the message and the audience. It is simply implausible to suggest that Jefferson, Madison, Washington, Adams, Monroe *et al.* would have been willing to surrender the protection against government that ciphers provided, because the government found it expensive or impossible to crack the codes.”); see also ANN M. LUCAS, THOMAS JEFFERSON FOUNDATION, THOMAS JEFFERSON’S WHEEL CIPHER (Sept. 1995).

³⁷⁴ Fraser, *supra* note 204, at 20.

³⁷⁵ The cipher was subsequently repurposed for government use during World War II.

Similarly, Benjamin Franklin, a printer,³⁷⁶ also created a variety of encryption tools in his print shop. In 1724, Franklin moved to London to work as a printer in the firm of John Watts before returning to Philadelphia in 1726. By 1730, Franklin had set up his own printing business and published a newspaper. On that same press, Franklin printed one of the earliest American texts on the uses of codes, ciphers, and secret writing.³⁷⁷ Subsequently, Franklin must have become even more interested in encryption as he began using a myriad of codes and signals himself.³⁷⁸ For instance, Franklin's international correspondence on behalf of the Continental Congress was often done through cipher. By 1781, Franklin had invented a "homophonic substitution cypher" while representing the United States in Paris.

Relatedly, many of the Founders also published anonymously. "[A]nonymity" for the Framers and their contemporaries "was the deciding factor between whether their writings would produce a social exchange or a personal beating."³⁷⁹ The Founders followed the example of their predecessors, such as Junius and Father Candor who also wrote anonymously. Indeed, some of the most canonical essays that shaped the history of the republic, such as the *Federalist Papers* "were published under fictitious names, such as Publius, Americanus, and Caesar."³⁸⁰

The Supreme Court recently underscored the value of anonymous speech in *McIntyre v. Ohio Elections Committee*.³⁸¹ The case involved Margaret McIntyre, a citizen of Ohio who had distributed leaflets—some of which were not signed—at several public meetings. Under an Ohio ordinance, McIntyre was found guilty for anonymous pamphleteering. McIntyre appealed. Reviewing the case, the Supreme Court briefly canvassed the history of anonymous literature from the time of the founding, including the *Federalist Papers* and other American Revolution-era writings. Ultimately, the Court found "[u]nder our Constitution, anonymous pamphleteering is not a pernicious, fraudulent practice, but an honorable tradition of advocacy and dissent."³⁸² "Anonymity" it continued, "is a shield from the tyranny of

³⁷⁶ See Franklin Common Press, SMITHSONIAN, http://americanhistory.si.edu/collections/search/object/nmah_882271.

³⁷⁷ *Id.*

³⁷⁸ *Id.*

³⁷⁹ *ACLU Submission*, *supra* note 105, at 15 (citing Jonathan Turley, *Registering Publius: The Supreme Court and the Right to Anonymity*, 2002 CATO SUP. CT. REV. 57, 59 (2002)).

³⁸⁰ *Id.*

³⁸¹ *McIntyre v. Ohio Elections Comm'n.*, 514 U.S. 334 (1995); *see also* Lee Tien, *Who's Afraid of Anonymous Speech? McIntyre and the Internet*, 75 OR. L. REV. 117 (1996); Richard K. Norton, *McIntyre v. Ohio Elections Commission: Defining the Right to Engage In Anonymous Political Speech*, 74 N.C. L. REV. 553 (1996); Erika King, Comment, *Anonymous Campaign Literature and the First Amendment*, 21 N.C. CENT. L.J. 144 (1995).

³⁸² *McIntyre*, 514 U.S. at 357.

the majority.”³⁸³ Recognizing that secret messages “sometimes have unpalatable consequences,” the Court still stated that in general, “our society accords greater weight to the value of free speech than to the dangers of its misuse.”³⁸⁴

Although the majority’s opinion in *McIntyre* is on its own a strong support for anonymous speech, Justice Thomas’ concurring opinion ties the Founder’s protection of anonymous speech to encryption and the Press Clause. Finding there is no record of “discussions of anonymous political expression either in the First Congress or in the state ratifying conventions,” Justice Thomas turns his attention to the Founders’ own behavior.³⁸⁵ He observes the Founders’ “universal practice of publishing anonymous articles and pamphlets, [indicating] that the Framers shared the belief that such activity was firmly part of the freedom of the press.”³⁸⁶ Justice Thomas proceeds to mention Thomas Jefferson’s use of encryption “was a partial response to a broad need for secrecy.”³⁸⁷ Highlighting the Founder’s use of tools as a necessary behavior of secrecy, Justice Thomas opens yet another channel for protection under the Press Clause by relying on the Founders’ own ubiquitous practices.

CONCLUSION

The question of whether encryption should be protected under the Constitution is arguably more pertinent than ever before. Moving forward, we should consider the normative and descriptive strengths of the Press Clause. While Speech Clause holds well-accepted precedent, the speech doctrine is both normatively and descriptively complicated. Although the Press Clause presents its own challenges, unlike the Speech Clause, it highlights the fundamental point that encryption is an *instrument* used to empower citizens to candidly share information. Like the printing press, the computer, and the Internet, modern day encryption is a tool that enables robust and trusted conversation. This is especially true as increasing government surveillance intrudes on more private squares of communication. The Press Clause highlights that encryption enables channels for private and unseen communication that are altogether quickly evaporating from our saturated ecosystem of information. In this modern world of data, tools like encryption keep certain channels of conversation outside the view of government and other private

³⁸³ *Id.*

³⁸⁴ *Id.* at 334.

³⁸⁵ *Id.* at 360 (Thomas, J., concurring).

³⁸⁶ *Id.*

³⁸⁷ *Id.* at 367.

individuals who threaten our ability to develop thoughts and ideas, and the Press Clause stands as a barrier to such intrusion.

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 6

SPRING 2017

NUMBER 2

A FRAMEWORK FOR ASSESSING TECHNOLOGY HUBS
IN AFRICA

JEREMY DE BEER, PAULA MILLAR, JACQUELENE MWANGI, VICTOR NZOMO, AND
ISAAC RUTENBERG *

This article explains the importance of technology hubs as drivers of innovation, social change, and economic opportunity within and beyond the African continent. It includes a thorough review and synthesis of findings from multi-disciplinary literature, and integrates insights from qualitative data gathered via interviews and fieldwork. It identifies three archetypes of hubs—clusters, companies, and countries—and discusses examples of each archetype using Kenya as a case study. The article then discusses potential collaboration, conflicts, and competition among these archetypes of hubs, and concludes with recommendations for future researchers.

* Jeremy de Beer, B.Comm., J.D., B.C.L. (Oxon), is a Full Professor at the Centre for Law, Technology and Society, Faculty of Law, University of Ottawa, Senior Research Associate at the IP Unit, University of Cape Town, and a co-founder and director of the Open African Innovation Research Network, Open AIR. He is online at www.jeremydebeer.com.

* Paula Millar, B.A. (Hons.), J.D., is a business and technology lawyer at Blackhawk Network, a publicly traded, financial technology company based in Pleasanton, California. paula.millar@bhnetwork.com.

* Jacqueline Mwangi, LL.B. (KU) is a Research Assistant at the Centre for Intellectual

INTRODUCTION.....	239
I. SYNTHESIS OF EXISTING RESEARCH	243
II. ARCHETYPES OF AFRICAN TECHNOLOGY HUBS	247
A. <i>Cluster Hubs</i>	250
1. <i>Characteristics of a Cluster Hub</i>	254
B. <i>Company Hubs</i>	254
1. <i>Characteristics of a Company Hub</i>	260
C. <i>Country Hubs</i>	262
1. <i>Vision 2030, a New Constitution, and the ICT Masterplan</i>	263
2. <i>Characteristics of a Country Hub</i>	266
III. DISCUSSION/ANALYSIS.....	268
A. <i>Hubs as Complementary and Collaborative</i>	269
1. <i>Company Hubs and Cluster Hubs</i>	269
2. <i>Company Hubs/Cluster Hubs and Country Hubs</i>	270
B. <i>Hubs in Conflict</i>	271
C. <i>Hubs in Competition</i>	274
CONCLUSION	276

Property and Information Technology Law (CIPIT), Strathmore Law School and Member of Open AIR New and Emerging Researchers Group. jwmwangi@strathmore.edu.

* Victor Nzomo, LL.B. (UNISA), LL.M. (UON), Ph.D. (Cand.) (UCT), is a Research Fellow at the Centre for Intellectual Property and Information Technology (CIPIT), Strathmore University Law School and Member of Open AIR New and Emerging Researchers Group. vnzomo@strathmore.edu.

* Isaac Rutenberg, BSc (CSM), Ph.D. (CIT), J.D (SCU), is Senior Lecturer and Director of the Centre for Intellectual Property and Information Technology (CIPIT) at Strathmore University Law School in Nairobi, Kenya. Holding a doctorate in Chemistry from California Institute of Technology and a law degree from Santa Clara University, he is among East Africa's leading intellectual property lawyers. He is also registered to practice law in California and before the United States Patent and Trademark Office.

* This research was supported by Open AIR, www.OpenAIR.org.za, with funding from the International Development Research Centre (IDRC), the Department of Foreign and International Development (DFID), and the Social Sciences and Humanities Research Council (SSHRC).

INTRODUCTION

Africa's startup revolution has spurred a proliferation of hubs, accelerators, incubators, and co-creation spaces across the continent. When AfriLabs, the leading network of African technology hubs, was founded in 2011, its membership included five incubators in four countries.¹ By 2016, the number of active hubs on the continent had grown to 314.² While half of these hubs are located in five countries (South Africa, Kenya, Nigeria, Egypt, and Morocco), nearly every other African country is home to at least one or two hubs.³

The impressive growth and perceived potential of Africa's technology hubs has garnered global interest from news outlets like the *Economist*⁴ and international agencies like the World Bank.⁵ The global Managing Director of McKinsey recently explained to the World Economic Forum the significance of "rapidly accelerating technological change that can unlock growth and leapfrog the limitations and costs of physical infrastructure in important areas of economic life."⁶ Demographic facts alone suggest that in twenty years, Africa will add more to the world's workforce than the rest of the world combined,⁷ and by 2050 one-quarter of earth's population will be African.⁸ Attention paid to Africa's

¹ Erik Hersman, *Afrilabs Provide a Model for African Innovation, Collaboration*, MEMEBURN (Feb. 8, 2011), <http://memeburn.com/2011/02/afrilabs-provide-a-model-for-african-innovation-collaboration/>.

² Victor Du Boucher, *A Few Things We Learned About Tech Hubs in Africa and Asia*, GSMA (Aug. 5, 2016), <http://www.gsma.com/mobilefordevelopment/programme/ecosystem-accelerator/things-learned-tech-hubs-africa-asia/>. The precise number is subject to some amount of variation, resulting from differences in definitions and methodologies employed. *See infra*, Part II.

³ *Id.*

⁴ *Homes for Africa's Tech Entrepreneurs: Tech Hubs Are Expanding Fast Across Africa*, ECONOMIST, Mar. 12, 2015, <http://www.economist.com/news/middle-east-and-africa/21646216-tech-hubs-are-expanding-fast-across-africa-homes-africas-tech-entrepreneurs>.

⁵ Tim Kelly, *Tech Hubs Across Africa: Which Will Be the Legacy-Makers?*, WORLD BANK GROUP (Apr. 30, 2014), <https://blogs.worldbank.org/ic4d/tech-hubs-across-africa-which-will-be-legacy-makers/>.

⁶ Dominic Barton and Acha Leke, *3 Reasons Things Are Looking up for African Economies*, WORLD ECON. FORUM (May 5, 2016), <https://www.weforum.org/agenda/2016/05/what-s-the-future-of-economic-growth-in-africa>.

⁷ Rene Vollgraaf, *Africa to Add More to Workforce in 2035 Than World Combined*, BLOOMBERG (Apr. 28, 2015), <https://www.bloomberg.com/news/articles/2015-04-28/africa-s-labor-force-newcomers-to-exceed-world-by-2035-imf-says>.

⁸ *World Population Prospects*, UNITED NATIONS (2015) <https://esa.un.org/unpd/wpp/>

technology hubs has also increased as a result of events like Facebook founder and CEO Mark Zuckerberg's first-time visit to hubs in Nigeria and Kenya, when he remarked: "The future will be built in Africa."⁹

The emergence of Africa's technology hubs is of crucial importance for those living within the continent, as the trend represents an opportunity for home-grown entrepreneurship devising local solutions to socio-economic problems and propelling Africa's innovation revolution. Africa's technology hubs are also important for the world outside of the continent, as they are a locus for partnership opportunities. On the margins of the most recent United States-Africa Business Forum, President Obama's former Senior Director for Africa at the White House suggested that one of the keys to understanding the continent's "profound impact on world markets" in coming decades is better education about the state of affairs in Africa.¹⁰ Looking toward the future, technology hubs may indeed be among the most significant drivers of global engagement with Africa's entrepreneurs.¹¹

There is, however, very little research on this important new phenomenon. Blog postings, white papers, and consultancy reports remain the predominant source of detailed information. Much of the terminology associated with Africa's high tech hubs is adapted from Silicon Valley, including the moniker "Silicon Savannah." This language has been popularized in media writing on technology (e.g., *TechCrunch*¹² and *Wired*¹³) and finance (e.g., *Bloomberg*,¹⁴ *Financial*

Publications/Files/Key_Findings_WPP_2015.pdf.

⁹ Toby Shapshak, *Africa Will Build the Future Says Zuckerberg, Visits Kenya on First African Trip*, FORBES (Sept. 1, 2016), <http://www.forbes.com/sites/tobyshapshak/2016/09/01/africa-will-build-the-future-says-zuckerberg-visits-kenya-on-first-african-trip/#67a838595214>.

¹⁰ Grant T. Harris, *U.S. Investors Are Missing Out on Africa's Economic Growth*, FORBES (Sept. 20, 2016), <http://www.forbes.com/sites/realspin/2016/09/20/bedwetting-u-s-capital-missing-out-on-african-growth/2/#6d6795343b8f>.

¹¹ Shirin Elahi et al., *Knowledge and Innovation in Africa: Scenarios for the Future*, OPEN AIR 63 (2013), <http://www.openair.org.za/images/Knowledge-Innovation-Africa-Scenarios-for-Future.pdf>.

¹² Jake Bright & Aubrey Hruby, *The Rise of Silicon Savannah and Africa's Tech Movement*, TECHCRUNCH (July 23, 2015), <https://techcrunch.com/2015/07/23/the-rise-of-silicon-savannah-and-africas-tech-movement/>. These authors have also published a book on the topic. JAKE BRIGHT & AUBREY HRUBY, *THE NEXT AFRICA: AN EMERGING CONTINENT BECOMES A GLOBAL POWERHOUSE* (2015).

¹³ Ian Steadman, *Kenya Breaks Ground on Its 'Silicon Savannah' City Project*, WIRED (Jan. 25, 2013), <http://www.wired.co.uk/article/kenya-silicon-savannah>.

¹⁴ *Silicon Savannah: Kenya's Billion-Dollar Tech Bet*, BLOOMBERG, Apr. 18, 2016,

Times,¹⁵ and *Harvard Business Review*¹⁶), and by influential writers like Melinda Gates in the *New York Times*.¹⁷ As explained in this article, however, our on-the-ground research shows that African technology hubs are more complex than such catchphrases would suggest.

African scholars in economics and innovation studies have started to identify and analyze technology hubs, laying the groundwork for nuanced analysis of legal and policy implications.¹⁸ But besides that work, most research about Africa's technology hubs has focused on highlighting success stories rather than critical analysis. Only recently have some critics begun to question technology hubs and their purposes, outputs, and sustainability.¹⁹ However, both qualitative and quantitative data respecting the operations and impact of the continent's hubs are limited. Even if such data existed, there has been no clear terminology and research framework with which to analyze the incredible and rapid growth of Africa's technology hubs.

This article is among the first to offer a framework for systematically describing and critically assessing the emergence of high technology hubs throughout Africa. It is also the first article to explain the legal and policy implications of Africa's innovation revolution for those both within and outside of the continent. This article will, therefore, open up new opportunities for researchers of African innovation to conduct further empirical studies in years to

<https://www.bloomberg.com/news/videos/b/fa73fc02-c511-4824-806d-5656acdfae7c>.

¹⁵ Sally Davies, *Silicon Savannah Springs to Life*, FIN. TIMES, Nov. 4, 2014, <https://www.ft.com/content/83b0b4f4-5fa5-11e4-986c-00144feabdc0>; David Pilling, *Kenya's Start to Roam Silicon Savannah*, FIN. TIMES, Apr. 27, 2016, <https://www.ft.com/content/1cda231c-0bdb-11e6-9456-444ab5211a2f>.

¹⁶ Jonathan Berman, *Meet the Tech Companies Creating Opportunity in Africa*, HARV. BUS. REV., Apr. 12, 2016, <https://hbr.org/2016/04/meet-the-tech-companies-creating-opportunity-in-africa>.

¹⁷ Melinda Gates, *Technology Can Make a Better World, If We Want It to*, N.Y. TIMES, July 22, 2015, <http://www.nytimes.com/roomfordebate/2015/07/22/is-silicon-valley-saving-the-world-or-just-making-money/technology-can-make-a-better-world-if-we-want-it-to>.

¹⁸ Neville Raymond Comins & Erika Kraemer-Mbula, *Innovation Hubs in Southern Africa*, in INNOVATION AFRICA: EMERGING HUBS OF EXCELLENCE 37 (Olugbenga Adesida, Geci Karuri-Sebina & Joao Resende-Santos eds., 2016).

¹⁹ See, e.g., Tayo Akinyemi, *All Hands on Deck: Why Hub Sustainability is a Complex Challenge that Requires a Systemic Solution*, MEDIUM (Nov. 10, 2014), <https://medium.com/Temitayo/all-hands-on-deck-why-hub-sustainability-is-a-complex-challenge-that-requires-a-systemic-solution-237012240686#.wb1nov4b2>.

come.

To address the current global knowledge gap on Africa's technology hubs, we undertook a thorough review of all relevant literature. Our literature review used a two-step method. First, we reviewed formal academic scholarship, considering a range of academic publications, including journals, conference papers, dissertations, and theses. Second, we examined practice- and policy-oriented literature, including articles, blog posts, and reports. Both steps employed similar search techniques and search terms.²⁰

From our review, we identified three archetypes of "hubs" described in the literature. Based on several variables—including the hubs' size and scope, administrative structure, business practices, funding, participant demography, and sustainability—we created a taxonomy that can be used to guide future research and data collection in this area.

We then supplemented our literature review and framework development with qualitative data gathered via semi-structured interviews in one particular country: Kenya. The purpose of the interviews was to gather primary data on the nature of Kenya's high technology hubs, thereby establishing the country as a case study from which to test and refine our taxonomy. Contributing authors conducted interviews at various hubs, accelerators, and incubators in Nairobi between November 2015 and August 2016.

We present our research findings as follows: Part I of the article synthesizes the current state of knowledge about high technology hubs by presenting the results of our literature review. Part II of the article describes a new analytical framework distilled from our research. Part III contains discussion and analysis, informed by our qualitative data. Finally, Part IV offers conclusions and recommendations for future research.

²⁰ We were guided by leading sources on literature review methods: see, for example, ARLENE FINK, *CONDUCTING RESEARCH LITERATURE REVIEWS: FROM THE INTERNET TO PAPER* (4th ed. 2014). We used Boolean logic techniques to search leading international databases and internet research repositories with the terms AND/OR/NOT to combine or limit results. Search terms included variations of the term "hub," including "incubator," "accelerator," and "co-creation space." We also used bibliographic analysis and citation tracing to work backwards from known sources, thus thoroughly identifying relevant earlier research. After our initial research revealed that the most relevant grey literature was very recent, our methods were innovatively adapted to concentrate on material generated through Google Alerts and Twitter lists.

I SYNTHESIS OF EXISTING RESEARCH

Existing research suggests that hubs can take many forms, yet most operate as some combination of a workspace, Internet café, coffee shop, training center, incubator, accelerator, event venue, and/or makerspace.²¹ While there is great diversity hub-to-hub with respect to structure, amenities, membership, and other factors, the general consensus is that hubs serve as a meeting place for a community.²² For instance, hubs seek to support knowledge sharing and inspire creativity by connecting like-minded people with skilled outsiders through mentorship and networking opportunities.²³ Overall, hubs exist primarily to enable and support entrepreneurship and innovation, more than to create or implement them.²⁴

For example, in 2010, iHub emerged as a “space for the tech community in Nairobi to gather, to call home and build connections to each other and work on ideas from.”²⁵ iHub co-founder Erik Hersman explained that, at least as of 2012, the right environment for a hub could only be found in five or six African cities: Nairobi, Lagos, Accra, Cape Town, Cairo, and possibly Dakar.²⁶ The success of these ventures depends upon the proper combination of “location, talent, policies, entrepreneurial culture, infrastructure, and money.”²⁷

For Hersman, the biggest challenge facing Africa’s technology scene was the absence of angel investors and seed capital. The dearth of investment, according to Hersman, has three causes: local investors who prefer traditional and straightforward investments in real estate, local investors who cannot comprehend the software space, and international investors who fail to understand the local

²¹ Tayo Akinyemi, *Decoding #hubsustainability: Confronting the Critically Important yet Painfully Obvious*, AFRILABS (Mar. 30, 2015), <http://www.afrilabs.com/2015/03/30/decoding-hubsustainability-confronting-the-critically-important-yet-painfully-obvious/>.

²² Nicolas Friederici, *What Is a Tech Innovation Hub Anyway?* OXFORD INTERNET INST. (Sept. 16, 2014), <http://cii.oii.ox.ac.uk/2014/09/16/what-is-a-tech-innovation-hub-anyway/>.

²³ *Id.*

²⁴ *Id.*

²⁵ Erik Hersman, *iHub: The Next Chapter*, iHUB (Mar. 11, 2016), <http://ihub.co.ke/blogs/26994/>.

²⁶ Erik Hersman, *Mobilizing Tech Entrepreneurs in Africa: Innovations Case Narrative: iHub*, 7 INNOVATIONS: TECH., GOVERNANCE, GLOBALIZATION 59, 60 (2012).

²⁷ *Id.* at 60.

situation.²⁸ Despite this, Hersman attests that the strength of African technology initiatives stems from the fact that such ideas could not come from the West.²⁹ The ideas are Africa-specific solutions to Africa-specific problems, attuned to local constraints and derived from cultural idiosyncrasies. A recent case study of iHub found that hubs also serve as links and catalyzers, affording entrepreneurs access to opportunities beyond their immediate locales and enabling integration into global production networks.³⁰

Indeed, the prevailing sentiment is that hubs and startups play an important role in generating employment, creating wealth, and improving Africans' quality of life. Further, a 2015 report by VC4Africa revealed that startups that participate in an incubator or accelerator program, or are selected for a "pitch event," obtain on average \$126,090 in external investment—a figure twenty-three percent higher than their counterparts who do not engage in such activities.³¹

For years, literature on African hubs adopted an optimistic tone, epitomized in such popular catchphrases as "Africa Rising,"³² "Lions go digital,"³³ and "Africa's Silicon Savannah."³⁴ The optimism was not entirely unfounded: In 2000, the entire African continent had less connectivity than Luxembourg;³⁵ today, new undersea fiber optic cables and improved inland connectivity have dramatically increased Internet penetration rates and the widespread accessibility of inexpensive smartphones has allowed Africans to connect online like never before.³⁶ This

²⁸ *Id.* at 59, 64.

²⁹ *Id.* at 59, 67.

³⁰ Mie Hvas, *Tech Hubs in Africa and Their Ability to Act as Catalysers for Integration into Global Production Networks: A Case Study of the iHub in Nairobi, Kenya* 1, 71 (June 29, 2015) (unpublished Masters thesis, Copenhagen Business School), <http://studenttheses.cbs.dk/handle/10417/5760>.

³¹ VENTURE CAPITAL FOR AFRICA, 2015 VENTURE FINANCE IN AFRICA: THE PROGRESS OF EARLY-STAGE HIGH-POTENTIAL GROWTH COMPANIES, VENTURE CAPITAL FOR AFRICA (2015), <https://vc4africa.biz/assets/pdf/Summary-VC4Africa-2015-Report.pdf/>.

³² *Africa Rising*, *ECONOMIST*, Dec. 3, 2011, <http://www.economist.com/node/21541015>.

³³ James Manyika et al., *Lions Go Digital: The Internet's Transformative Potential in Africa*, MCKINSEY & COMPANY (Nov. 2013), <http://www.mckinsey.com/industries/high-tech/our-insights/lions-go-digital-the-internets-transformative-potential-in-africa>.

³⁴ *Id.* at 57.

³⁵ Eleanor Whitehead, *WEF Africa 2012: How Technology Changes Africa's Development Prospects*, THIS IS AFRICA (May 1, 2012), <http://www.thisisafricaonline.com/Reports/WEF-Africa-2012-How-technology-changes-Africa-s-development-prospects/>.

³⁶ Elahi et al., *supra* note 11, at 63.

unprecedented ease of access has spurred a wave of innovation from the likes of entrepreneurs and corporations, resulting in the rise of thousands of technology-based ventures.

Critical voices, however, began emerging regarding concerns over the sustainability of Silicon Cape's startup bubble.³⁷ Other developments fueling this new sentiment included the transition of C4DLab (a startup incubation hub at the University of Nairobi) to a lean model, until it reaches a break-even point,³⁸ and the fact that another accelerator, known as "88mph," pulled out of Kenya and had limited success in South Africa.³⁹ One recent series of blog posts also indicated the fading dominance of iHub in Kenya's maturing technology scene.⁴⁰

Criticism appears to be directed at three main issues: the challenge of measuring the success of African hubs and similar entities, the sustainability of the continent's hubs and its startup revolution, and concerns regarding the reigning incubation model.⁴¹

There is limited academic scholarship on hubs and their processes. While comparative studies form the bulk of research on this topic, a lack of regional or country-specific data and inconsistent nomenclature complicate understanding

³⁷ E.g., Mark Ashton, *Silicon Cape Start-up Bubble Will Burst*, MYBROADBAND (Oct. 26, 2014), <http://mybroadband.co.za/news/business/112729-silicon-cape-start-up-bubble-will-burst.html/>.

³⁸ Tom Jackson, *Nairobi's C4DLab to Adopt Lean Model*, DISRUPT AFRICA (Mar. 11, 2015), <http://disrupt-africa.com/2015/03/nairobis-c4dlab-adopt-lean-model/>.

³⁹ Drazen Jorgic, *Kenya's Technology Push Leaves Investors Cold*, REUTERS, Dec. 31, 2014, <http://www.reuters.com/article/kenya-tech-idUSL6N0UE15920141231/>.

⁴⁰ Wilfred Mutua Mworio, *The Decentralization of Nairobi's Tech Startup and Innovation Ecosystem*, AFRINNOVATOR: FOUNDER'S BLOG (Jul. 20, 2016), <https://founder.afrinnovator.com/the-decentralization-of-nairobis-tech-startup-and-innovation-ecosystem-d2e57ba95224#.bjcomOds8>.

⁴¹ See Malik Fal, *Accelerating Entrepreneurship in Africa*, 8 INNOVATIONS: TECH., GOVERNANCE, GLOBALIZATION 149, 160–63 (2013); see also Randall Kempner, *Incubators Are Popping Up Like Wildflowers...But Do They Actually Work?*, 8 INNOVATIONS: TECH., GOVERNANCE, GLOBALIZATION, 3, 4 (2013); Mark Essien, *Startup Incubators in Africa and Why They Don't Work*, MARKESSIEN.COM (Feb. 28, 2015), <http://markessien.com/startup-incubators-in-africa-and-why-they-dont-work/>; Morgan McClain-McKinney, *Who Needs Incubators? A Lesson from Young Entrepreneurs in Sub-Saharan Africa*, VENTURES AFRICA (Feb. 14, 2015), <http://venturesafrica.com/who-needs-incubators-a-lesson-from-young-entrepreneurs-in-sub-saharan-africa/>.

these works as part of a larger narrative.⁴²

New academic research suggests that hubs are necessary components of the knowledge-based economic and social development emerging across Southern Africa.⁴³ Such hubs should serve as intermediaries connecting players across the public and private sectors, fostering meaningful relationships and inspiring collaboration.⁴⁴ Rather than importing a universal design, each hub should reflect the local ecosystem and cater to its particular needs.⁴⁵

In sum, hubs have been a topic of extensive discussion in recent years. Academics and reporters, both on and off the African continent, have written with optimism about the power of the hub to solve the myriad of development challenges. On the other hand, some have suggested that hubs are overhyped and under-delivering, and are unlikely to have a substantial long-term impact on economic growth.

The shortage of research-based evidence to resolve these debates stems partly from the topic's newness and fluidity, and from the fact that it may take several years before a startup or hub succeeds or fails. Africa's first hub, Ghana's Meltwater Entrepreneurial School of Technology, emerged in 2008 and startup accelerators only emerged on the continent following the 2012 launch of 88mph in Kenya.⁴⁶ While some research suggests it may take anywhere between three and forty months to reach a break-even point, one article notes it may take up to five years to build a business.⁴⁷ Tayo Akinyemi, AfriLabs' first Executive Director, advises new technology hubs to assume that it will take "at least three years,

⁴² Lindile Ndabeni, *The Contribution of Business Incubators and Technology Stations to Small Enterprise Development in South Africa*, 25 DEV. S. AFRICA, 259 (2008) (examining the contribution of small business incubators in South Africa); see also Robert Yawson, *The Place of Technology Based Incubators in Ghana's Economic Development* (Jun. 11, 2009) (unpublished manuscript), https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=1418103.

⁴³ Comins & Kraemer-Mbula, *supra* note 18, at 96.

⁴⁴ *Id.* at 97.

⁴⁵ *Id.* at 39, 96.

⁴⁶ Tom Jackson, *Africa Tech Trends: The Age of Start-up Accelerators*, HOW WE MADE IT IN AFRICA (Feb. 16, 2015), <http://www.howwemadeitinafrica.com/africa-tech-trends-the-age-of-start-up-accelerators/>.

⁴⁷ Charles Rambo, *Time Required to Break-Even for Small and Medium Enterprises: Evidence from Kenya*, 6 INT'L J. OF MGMT. & MARKETING RES. 81 (2013); see also Kelly Berold, *Accelerating Forward with Kresten Buch*, FINWEEK, Oct. 9, 2014, <http://connection.ebscohost.com/c/articles/95842436/accelerating-forward-kresten-buch/>.

probably more” for a hub to become sustainable.⁴⁸ Unsurprisingly, reliable observations and analysis of such phenomena lag several years behind.

At this time, the development of a robust framework characterizing the continent’s hubs and their operations is prudent. Indeed, empirical research on Africa’s high technology hubs will likely be stalled until an adequate research framework exists. The next section of this article introduces such a framework.

II

ARCHETYPES OF AFRICAN TECHNOLOGY HUBS

Many writings about Africa’s technology hubs adopt inconsistent terminology. That is not unexpected since there are dramatic hub-to-hub differences in terms of facilities, services, clientele, and organizational structure, among other factors. The physical and cultural environment of each hub is unique. In effect, generalizations of hubs are difficult to substantiate and are qualified with many exceptions. It helps, therefore, to create an adaptable framework for characterizing the hubs and their various aspects of operation.

A starting point for our framework can be found in research on cluster theory, adapted from the fields of economic geography and strategic management. Derived from the ideas of Alfred Marshall,⁴⁹ cluster theory posits that advantages exist in the geographic concentration of enterprises that compete and cooperate in a specific sector.⁵⁰ Cluster participation affords certain benefits to members of a regional network, including opportunities for coordination, mutual improvement, and productivity growth.⁵¹

This theory has been used to explain the relative successes and failures of high technology clusters in California’s Silicon Valley and the Greater-Boston corridor known as Route 128, both regional network-based industrial systems.⁵² The concepts of clusters and urban agglomeration are potentially applicable in the

⁴⁸ Akinyemi, *supra* note 21.

⁴⁹ ALFRED MARSHALL, *PRINCIPLES OF ECONOMICS* (8th ed. 1890).

⁵⁰ MICHAEL PORTER, *THE COMPETITIVE ADVANTAGE OF NATIONS* (1990).

⁵¹ Michael Porter, *Location, Competition and Economic Development: Local Clusters in a Global Economy*, 14 *ECON. DEV. Q.* 15, 18-19 (2000), http://home.furb.br/wilhelm/COMPETIV/Porter_Cluster3.doc/.

⁵² ANNALEE SAXENIAN, *REGIONAL ADVANTAGE: CULTURE AND COMPETITION IN SILICON VALLEY AND ROUTE 128* (1996).

context of African high technology hubs too.⁵³

Advantages can be based on geography, human capital, natural resources, or a range of other economic and environmental factors. Knowing why clusters emerged or why they were successful, however, does not explain why they developed into certain kinds of high technology clusters seen or described in the literature in Africa.

One study on cluster-based growth in Africa divided cluster development on the continent into two broad categories: (i) the spontaneous congregation of enterprises and supporting entities in a geographic area, and (ii) clusters intentionally induced by policymaking or ‘constructed’ (e.g. “technopoles” and industrial parks).⁵⁴ This study focused mainly on the first category of clusters, with case studies drawn from clusters in sectors such as agriculture, metalworking, and manufacturing.⁵⁵

Attempts to categorize Africa’s technology hubs have been met with minimal success. According to the 2016 World Bank Development Report (“the World Bank Report”), hubs in Africa can be divided into four main categories: government-led hubs, civil society-led hubs, academic institution-led hubs, and hybrid hubs.⁵⁶ Based on this categorization, Kenya is reportedly home to sixteen hubs: ten civil society-led hubs (e.g., iHub), four academic institution-led hubs (University of Nairobi and Strathmore University account for two hubs each) and two hybrid hubs (m:Lab East Africa and GrowthHub).⁵⁷ No hubs in Kenya are government-led, according to the World Bank mapping.

Several commentators and researchers have confirmed, however, that the World Bank’s mapping of hubs is already outdated, and that there are twenty-seven

⁵³ INDUSTRIAL CLUSTERS AND INNOVATION SYSTEMS IN AFRICA (Banji Oyelaran-Oyeyinka & Dorothy McCormick eds., 2007).

⁵⁴ Douglas Zihua Zeng, *Knowledge, Technology, and Cluster-Based Growth in Africa: Findings from 11 Case Studies of Enterprise Clusters in Africa*, in KNOWLEDGE, TECHNOLOGY, AND CLUSTER-BASED GROWTH IN AFRICA 1, 2 (Douglas Zihua Zeng ed., 2008).

⁵⁵ *Id.*

⁵⁶ WORLD BANK GROUP, WORLD DEVELOPMENT REPORT 2016: DIGITAL DIVIDENDS 230 (2016).

⁵⁷ Rachel Firestone, *The Importance of Mapping Tech Hubs in Africa, and beyond*, WORLD BANK GROUP (Aug. 08, 2016), <http://blogs.worldbank.org/ic4d/importance-mapping-tech-hubs-africa-and-beyond/>.

active hubs in Kenya.⁵⁸ Based on our research, the number of Kenyan hubs had at least doubled in roughly one year after the World Bank's assessment. Additional organizations that could be classified as tech hubs under the World Bank's categorization include: FabLab Kivuli, PAWA254, Ubunifu Hub, the Foundry Africa, KICTANET, Sinapis, Nest, Savannah Fund, Swahili Box, mtHub, Chandaria Business Innovation and Incubation Centre, Kenya Climate Innovation Centre, Matili Technology Hub (mtHub), Konza Technopolis, SwahiliPot, and MakersHub.

Another study identified three types of hubs in Southern Africa: traditional science parks, activity-based innovation centers, and co-creation hubs.⁵⁹ While path-breaking, that study proposed a typology specific to Southern Africa which may apply differently to Eastern Africa or other regions of the continent. One reason is the different approaches to establishing innovation hubs. For instance, unlike those in Southern Africa, countries in Eastern Africa like Kenya, Ethiopia, Uganda, Tanzania, and Rwanda seemingly have no government-led hubs that could be categorized as traditional science parks.⁶⁰

Other categorization frameworks may miss key factors that distinguish the various hubs, their operations, and their impacts. Further, they may not account for tech-focused business initiatives such as South Africa's 'Silicon Cape.'⁶¹ As such, this paper proposes a somewhat more general framework for assessing hubs that might be applied in diverse African contexts.

⁵⁸ Du Boucher, *supra* note 2; Gabriella Mulligan, *Niche Tech Hubs to Boom Across Africa in 2016*, DISRUPT AFRICA (Mar. 7, 2016), <http://disrupt-africa.com/2016/03/niche-tech-hubs-to-boom-across-africa-in-2016/>.

⁵⁹ Comins & Kraemer-Mbula, *supra* note 18, at 40.

⁶⁰ WORLD BANK GROUP, *supra* note 56, at 230.

⁶¹ See generally Martin Carstens, *6 Reasons Why SA's Western Cape Deserves to Be Called Silicon Cape*, VENTUREBURN (Apr. 30, 2013), <http://ventureburn.com/2013/04/6-reasons-why-sas-western-cape-deserves-to-be-called-silicon-cape/>; Berenice Magistretti, *Is Cape Town The New Silicon Valley?*, SEEDSTARS WORLD (Jul. 13, 2015), <http://www.seedstarsworld.com/blog/2015/07/is-cape-town-the-new-silicon-valley/>; Jessica Morris, *Silicon Savannah: Is Africa Tipped for a Tech Take-off?*, CNBC (Jan. 7, 2014, 3:13 AM), <http://www.cnbc.com/2014/01/07/silicon-savannah-is-africa-tipped-for-a-tech-take-off.html>; Simon Mundy, *South Africa: Silicon Cape?*, FIN. TIMES: BEYOND BRICS (Jan. 12, 2011, 5:34 PM), <http://blogs.ft.com/beyond-brics/2011/01/12/south-africa-silicon-cape/>; Justin Spratt, *Can Cape Town become SA's Silicon Valley?*, TECH CENTRAL (Oct. 5, 2009), <https://www.techcentral.co.za/can-cape-town-become-sas-silicon-valley/10348/>.

We have identified and defined three archetypes of “hubs”: cluster hubs, company hubs, and country hubs.

First, a “cluster hub” refers to a cluster of distinct entities that are located physically close to one another, and therefore tend to promote regular and intimate interactions. Here, established cluster/agglomeration theory is most applicable. Second, a “company hub” describes a hub as a specific entity, attracting and defining its own community, and interacting with the outside world in a manner similar to a company. This is how “hubs” are most commonly referenced and understood in the practice- and policy-oriented literature we reviewed. Third, a “country hub” tier reflects a more macro view of a hub, where an entire country or region advertises itself as a progressive hub, and government policies guide the actions of the country or region. This latter terminology is most closely aligned with national and regional policy instruments.

Even within this taxonomy, of course, there can be substantial variations in hub structure, operation, and overlap of the entities existing within the tiers. Notwithstanding such complications, the activities of individual actors are given helpful context when hubs are identified and analyzed within this three-tiered framework.

A. Cluster Hubs

The quintessential cluster hub in Kenya is found along Ngong (pronounced ‘gong’) Road, a major roadway connecting Nairobi’s Central Business District to Ngong Town in the western outskirts of Nairobi.⁶² Since 2011, a specific four-kilometer stretch of Ngong Road has become the nerve center of Kenya’s tech boom.⁶³ Currently, eight of the sixteen Kenyan hubs mapped by the World Bank are located on or around this particular section of Ngong Road.⁶⁴ The cluster of hubs, startups, and venture capital firms found within this four-kilometer stretch has been referred to as “Silicon Savannah,”⁶⁵ “Silicon Avenue,”⁶⁶ “Nairobi’s now

⁶² See Dan Evans, *Welcome to the Silicon Savannah - Researching Africa’s Startup Ecosystems*, VENTURE CAPITAL FOR AFRICA (Apr. 18, 2014), <https://vc4a.com/blog/2014/04/18/welcome-to-the-silicon-savannah-researching-africas-startup-ecosystems/>.

⁶³ See *Kenya’s Technology Start-up Scene Is About to Take off*, ECONOMIST, Aug. 25, 2012, <http://www.economist.com/node/21560912>; see also Kenneth Griffith, *The Silicon Savannah - An Insider’s View*, LINKEDIN (Jul. 8, 2014), <https://www.linkedin.com/pulse/20140708181255-31799348-the-silicon-savannah-an-insider-s-view/>.

⁶⁴ Firestone, *supra* note 57, at 229.

⁶⁵ Matthew Hussey, *Silicon Savannah - How Start-ups in Africa are Taking on Some of*

African IT—synonymous Ngong Road,”⁶⁷ “tech industry centre,”⁶⁸ “Ngong Road corridor,”⁶⁹ and “entrepreneurship row.”⁷⁰ Due to the numerous and substantial differences from Silicon Valley, we advocate for alternatives to the term “Silicon Savannah” when describing any part of the Nairobi tech scene. Our particular recommendation is the “Digital Savannah.”

The pioneering and now anchoring entity along Ngong Road is one particular company hub called iHub (also known as Nairobi’s Innovation Hub), founded in March 2010 by technopreneurs previously associated with the not-for-profit startup Ushahidi.⁷¹ Following Kenya’s disputed 2007 presidential election, Ushahidi developed a crowd-sourced mapping tool that garnered international attention.⁷² The founding of iHub along Ngong Road, as a space for the emerging technology community, followed from Ushahidi’s success and popularity.⁷³

For years, iHub occupied the top floor of a four-story building on Ngong Road, called the Bishop Magua Centre.⁷⁴ International donors and Ushahidi

Humanity's Biggest Challenges, HUFFINGTON POST, Jan. 6, 2015, http://www.huffingtonpost.co.uk/matthew-hussey/african-startups-take-on-challenges_b_6416676.html; Guy Alexander, *Kenya's Tech Visionaries Lead the Way*, GUARDIAN, Sept. 15, 2013, <https://www.theguardian.com/world/2013/sep/15/kenya-technology-visionaries/>; Stephanie Findlay, *Welcome to the 'Silicon Savannah'*, CANADIAN BUS., Jul. 10, 2013, <http://www.canadianbusiness.com/global-report/global-report-welcome-to-the-silicon-savannah/>.

⁶⁶ Lillian Nduati, *The Value of a Strong Tech Eco-system*, STARTUP ACADEMY (Dec. 2, 2013), <http://www.startupacademy.co.ke/blog/the-value-of-a-strong-tech-eco-system/>.

⁶⁷ Bright & Hruby, *supra* note 12.

⁶⁸ *Building Africa's Silicon Savannah*, VENTURES AFRICA (Sep. 25, 2013), <http://venturesafrica.com/konza-techno-city-africas-silicon-savannah/>.

⁶⁹ Jonathan W. Rosen, *Kenya Tries to Build Its Silicon Valley*, MIT TECH. REV. (Nov. 23, 2015), <https://www.technologyreview.com/s/543406/kenya-tries-to-build-its-silicon-valley/>.

⁷⁰ Jason Loughnane, *GF Fellow Jason Loughnane: Nairobi's Tech Row*, BANKERS WITHOUT BORDERS, <http://www.bankerswithoutborders.com/blog/gf-fellow-jason-loughnane-nairobis-tech-row/> (last visited Mar. 2, 2017).

⁷¹ WORLD BANK GROUP, *supra* note 56, at 229; *see also* Dena Levitz, *Kenya's Mobile Payments 'Secret Sauce' is Turning Nairobi into the Silicon Savannah*, 1776 (Jan. 20, 2015), <https://www.1776.vc/insights/kenyas-mobile-payments-secret-sauce-is-turning-nairobi-into-the-silicon-savannah/>.

⁷² MARISSA DROUILLARD ET AL., GSMA, DIGITAL ENTREPRENEURSHIP IN KENYA 2014, GSMA 24 (2014), <http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/02/Digital-Entrepreneurship-in-Kenya-2014.pdf>.

⁷³ *Id.*

⁷⁴ *See* Erik Hersman, *iHub: Nairobi's Tech Innovation Hub is Here!*, WHITE AFRICAN (Jan.

provided initial funding for the iHub space. The launch of iHub is considered by many as a key factor that positioned and transformed Kenya's capital, Nairobi, into a center for startups, technology and innovation.⁷⁵ According to a GSM (Groupe Speciale Media) Association (GSMA) study, Kenyan entrepreneurs outside Nairobi (e.g., those located in Mombasa or Kisumu) tend to relocate closer to the Bishop Magua Centre and Ngong Road once their startups gain traction.⁷⁶

A 2014 report by the Vodafone Institute for Society and Communications⁷⁷ cites the iHub's success when recommending the establishment of additional co-working spaces, incubators, and accelerators by companies, government, and development organizations in Africa. The report argues that a comparative look at Silicon Valley is worthwhile since there, too, the proximity of institutions with similar interests, the concentration of expertise, and the pronounced culture of recommendations and networking are important factors for success in consolidating innovations.⁷⁸

Over the three years following iHub's launch, it incubated over 100 startups and was seeking further expansion at the Bishop Magua Centre, by then dubbed the "mecca for African ICT (information and communication technology) entrepreneurs."⁷⁹ Erik Hersman initially thought the space on the fourth floor of Bishop Magua was too big, but it was soon outgrown with a community of over

25, 2010), <http://whiteafrican.com/2010/01/25/ihub-nairobis-tech-innovation-hub-is-here/>; see also Geoffrey York, *The Entrepreneurs of Africa's Silicon Savannah*, GLOBE AND MAIL, Sept. 2, 2013, <http://www.theglobeandmail.com/report-on-business/international-business/african-and-mideast-business/the-entrepreneurs-of-silicon-savannah/article14073645/>; see also *We Are Moving!*, IHUB (Feb. 26, 2017) <https://ihub.co.ke/blogs/29373/we-are-movin> (announcing that in February 2017, after seven years at the Bishop Magua Centre, iHub relocated to Senteu Plaza, approximately two kilometers from its former location on Ngong Road).

⁷⁵ Dipolelo Moime, *Kenya, Africa's Silicon Valley, Epicentre of Innovation*, VENTURE CAPITAL FOR AFRICA (Apr. 25, 2016), <https://vc4a.com/blog/2016/04/25/kenya-africas-silicon-valley-epicentre-of-innovation/>.

⁷⁶ DROUILLARD ET AL., *supra* note 72.

⁷⁷ JULIA MANSKE, VODAFONE INST. FOR SOC'Y AND COMMC'NS, INNOVATIONS OUT OF AFRICA. THE EMERGENCE, CHALLENGES AND POTENTIAL OF THE KENYAN TECH ECOSYSTEM, VODAFONE INST. FOR SOC'Y AND COMMC'NS 29 (2014), http://www.vodafone-institut.de/wp-content/uploads/2015/09/VFI_InnovationsAfrica_EN.pdf.

⁷⁸ *Id.*

⁷⁹ Gemma Solés, *iHub a Mecca for African ICT Entrepreneurs*, URBAN AFRICA (Nov. 26, 2013), <http://www.urbanafrika.net/urban-voices/ihub-mecca-african-ict-entrepreneurs/>.

10,000 online and physical members.⁸⁰ According to iHub's own observations in 2013, their space at Bishop Magua had propelled them from being the space "where things happen, to the centre of where things happen, the centre of ideas on the future of this region and across Africa."⁸¹

During the period that iHub was located the Bishop Magua Centre, from 2010 to 2017, other hubs and a host of startups clustered in that complex and the surrounding area. In addition to iHub, three of the eleven tech hubs in Kenya listed in the World Bank Report are located at the Bishop Magua Centre: m:Lab East Africa, Nailab, and Akirachix.⁸² Each of these entities serves a distinct subset of the overall entrepreneurship community: m:Lab is primarily an incubator focusing on very early stage startups; Nailab is an accelerator focusing on early to medium-stage startups; and Akirachix focuses on increasing the number and role of women in the technology scene.

The iHub community is also responsible for the 2014 creation of a new makerspace at the Bishop Magua Centre called Gearbox, a Kenyan makerspace for design and rapid prototyping.⁸³ Savannah Fund is another entity located at the Bishop Magua Centre. It is an accelerator fund set up by several partners, including an iHub founder, focusing on finding and investing in East Africa's highest potential, pre-revenue startups.⁸⁴ As described in the section below, the Savannah Fund itself exhibits many characteristics of a company hub.⁸⁵

The Bishop Magua Centre is just one pole of the Ngong Road hub. Approximately four kilometers west of the Bishop Magua Centre is Piedmont Plaza—the base for Nairobi Garage, Nest and Sinapis. Outside Piedmont Plaza and the Bishop Magua Centre, the Ngong Road cluster is home to other key entities, including Growth Africa (formerly Growth Hub).

⁸⁰ Jimmy Gitonga, **iHub: A Growing Community*, iHUB (Mar. 2, 2013), <http://ihub.co.ke/blogs/12778>.

⁸¹ *Id.*

⁸² WORLD BANK GROUP, *supra* note 56, at 230.

⁸³ Jessica Colaço, *Gearbox - A Space for Tinkering and Making Things*, iHUB (Sept. 09, 2013), <http://ihub.co.ke/blogs/15711/>; *see also* Erik Hersman, *Launching Gearbox, A Kenyan Makerspace*, WHITE AFRICAN (Sept. 09, 2013), <http://whiteafrican.com/2013/09/09/launching-gearbox-a-kenyan-makerspace/>.

⁸⁴ Erik Hersman, *Launching the Savannah Fund in East Africa*, WHITE AFRICAN (June. 6, 2012), <http://www.whiteafrican.com-launching-the-savannah-fund-in-east-africa/>.

⁸⁵ *See infra* Part II.B.

1. Characteristics of a Cluster Hub

Based on our observations, we define a cluster hub as a relatively small geographical region containing a high density of hubs, as well as the infrastructure and organizations that support such hubs. The individual company hubs within a cluster hub are generally independent entities that interact with one another. Such interactions include sharing physical spaces, Internet access, human resources/capital, and other resources. Within the cluster hub, members and users of one hub may use another hub's facilities or even join an alternative hub. Individuals seen as "experts" (e.g., engineering or IT experts, entrepreneurship coaches, intellectual property experts, etc.) frequently split their time within one or among the various hubs.

Supporting entities, such as food and beverage vendors (particularly coffee shops), serve all of the hubs within a cluster and often function as ad hoc meeting spaces where members from the various hubs gather and interact.

The services offered and the scope of activities in a cluster hub can vary widely. A cluster hub is likely to have at least one incubator, accelerator, shared-working space, and investment entity. As the cluster hub develops, training and mentorship opportunities multiply. Specific experts, such as engineers and lawyers, may also be brought in when the cluster hub reaches a critical mass and the community requires such support on a regular basis.

The growth of a cluster hub can take several paths. In some cases, the founding member(s) of one hub may identify an unmet need and create a new hub to fill that void. Efficiency and proximity naturally lead the new hub to share facilities and resources with the original hub. The new entity may be an entirely separate entity (e.g., iHub's emergence from Ushahidi), or may remain owned and operated by the original hub (e.g., iHub Research, an entity that had been owned and operated by iHub). Further, one hub's success appears to attract others to create similar but separate (and sometimes complementary) spaces (e.g., Nailab's existence next to iHub). The founder of the new hub may be a former user of the original hub, a friend of the original hub's founder(s), or an unrelated person drawn by the hub's success. Through a combination of the varying types of growth, multiple hubs now occupy a substantial portion of the Bishop Magua Centre, a key point along the Ngong Road cluster.

B. Company Hubs

The concept of a hub as a company recognizes that most hubs, when viewed independently from their regional ecosystems, are similar to for-profit or non-profit

entities in structure and operation. Interestingly, critical language and analysis not present elsewhere tends to be used in commentary and literature on individual company hubs. When hubs are viewed as independent operating entities as opposed to regional agglomerations, new issues emerge, including long-term sustainability, scalability, and, in some cases, even profitability.

This lens typically reflects a strategic management approach, where individual hubs are structured and assessed like any other business. Financial monitoring and evaluation methods are essential in this context to quantify and value a hub's impact. In our view, this change in tone is a natural and expected progression, as these firms *per se* become increasingly important actors in Africa's economy.

While our classification of cluster hubs most easily aligns with established theories in strategic management and economic geography, most of the practice- and policy-oriented literature we reviewed defines hubs (implicitly) as particular companies. Based on the concept of a hub as a company, the literature then focuses on the scalability and sustainability of these ventures. This is particularly the case for hubs that offer support services to startups, such as co-working spaces, incubators, accelerators, and maker-spaces. Writings about such topics are varied both in terms of geographic focus and sectoral focus. Such facts further support the categorization of certain individual organizations as hubs and vice versa.

The application of conventional business principles and success metrics to company hubs has gained support among the organizations themselves. For example, in 2013, iHub Research (the research arm of iHub) released a comparative study of seven technology hubs in Africa: iHub, Hive Colab (Uganda), Activspaces (Cameroon), kLab (Rwanda), MEST (Ghana), Bongo Hive (Zambia), and Kinu (Tanzania).⁸⁶ The report aimed to document the various ICT hub models emerging across the continent in order to determine how they differ, and to identify the challenges they face. The study yielded lessons learned, recommendations, and strategies for success.

While recognizing that the appropriate hub model will differ based on the country context, the report outlined "critical common success factors for a strong ICT hub": government support (e.g., funding, market, infrastructure); science,

⁸⁶ Duncan Gathege & Hilda Moraa, IHUB RESEARCH, DRAFT REPORT ON COMPARATIVE STUDY ON INNOVATION HUBS ACROSS AFRICA, IHUB RESEARCH (2013), http://research.ihub.co.ke/uploads/2013/may/1367840837__923.pdf.

technology, and innovation; strategic partners (e.g., business partners, funders, mentors); a community of members (e.g., entrepreneurs); human capital (i.e., skills, education, experience); research and development; and monitoring and evaluation mechanisms.⁸⁷ Based on the high number of hub graduates and the emergence of numerous successful startups, the arguably optimistic report concluded, “the hype surrounding technological hubs can be justified.”⁸⁸ Despite this, the report notes that most African hubs are in their infancy and thus their long-term self-sustainability is not yet certain.⁸⁹ Furthermore, the report finds that the success of individual hub models also depends on external factors that impact a country’s potential for ICT growth: ICT GDP, government support for ICT, corruption levels, existing infrastructure, ICT budget allocation, investment in telecommunications, and prioritization of ICT initiatives.⁹⁰ Overall, the report suggests that hubs strive to achieve long-term sustainability by filling local gaps and resolving contextual needs rather than trying to replicate successful models achieved elsewhere.⁹¹

There are additional examples that show members within the hub community characterizing hubs as companies. Since November 2014, Akinyemi has penned two pieces on hub sustainability. Akinyemi’s first post argued that Africa’s hubs were fragile and had no clear path towards long-term sustainability. In March 2015, Akinyemi’s second post proposed eleven lessons learned regarding hub sustainability on the African continent. The advice was the product of a series of Google Hangouts on hub sustainability held in November and December 2014.⁹² Akinyemi explains that there was no single substitute for a typical technology hub, as “[a] hub is a workspace, Internet café, coffee shop, training centre, incubator, accelerator, event venue, maker space; [and] it’s usually many of these at the same time.”⁹³ While this presents opportunities, Akinyemi points out that it also poses challenges. Hubs fill the gaps in the enabling environments of Africa’s technology sectors by providing and serving as the necessary infrastructure to support local entrepreneurs.⁹⁴ However, problems arise when hubs overextend themselves and

⁸⁷ *Id.* at 31-33.

⁸⁸ *Id.* at 34.

⁸⁹ *Id.* at 26.

⁹⁰ *Id.* at 36.

⁹¹ *Id.* at 36-37.

⁹² Akinyemi, *supra* note 21.

⁹³ *Id.*

⁹⁴ *Id.*

engage in too many activities that do not generate revenue.⁹⁵

While noting that “the exact formula for a truly sustainable business model remains to be seen,” Akinyemi offers the continent’s technology hubs practical advice to achieve sustainability.⁹⁶ She suggests that hubs plan for sustainability from the outset, regardless of whether they receive donor funding. While hubs should not depend on government support, she recommends hubs align their priorities with government to ensure a cooperative relationship while still maintaining their independence.⁹⁷ To attract partnerships and investment, she advises hubs to publicly communicate their impact to key stakeholders using accessible metrics.⁹⁸ Finally, as there is no ‘one-size-fits-all’ model for sustainability, Akinyemi notes that the road to sustainability would differ from hub to hub.⁹⁹

Writings such as these firmly place the discourse about hubs into the business world. So it not surprising that in March 2016, the iHub announced that it had received new investors “in order to help it grow, to tighten up its service offerings and make them more profitable, and to help it figure out how not to just find startups but to grow the ones that are getting traction.”¹⁰⁰

A potential conflict exists, however, because at the same time some hubs are expected to exist solidly in the nonprofit space. In 2012, the Aspen Network of Development Entrepreneurs (ANDE) and Village Capital undertook a study on the role of social-impact focused accelerators, and followed their research with a report in 2013.¹⁰¹ The study examined the global accelerator landscape, but the majority of the fifty-two accelerator operations surveyed were based in Africa.¹⁰² The study provided a number of key findings with respect to the variables linked to the success and failure of accelerators as well as the sustainability of these ventures. First, the study found that a lower acceptance rate and rigorous selection

⁹⁵ *Id.*

⁹⁶ *Id.*

⁹⁷ *Id.*

⁹⁸ *Id.*

⁹⁹ *Id.*

¹⁰⁰ Hersman, *supra* note 25.

¹⁰¹ Ross Baird, Lily Bowles & Saurabh Lall, *Bridging the “Pioneer Gap”: The Role of Accelerators in Launching High-Impact Enterprises*, ASPEN NETWORK OF DEV. ENTREPRENEURS (2013), <https://www.aspeninstitute.org/publications/bridging-pioneer-gap-role-accelerators-launching-high-impact-enterprises/>.

¹⁰² *Id.* at 7.

process, as well as partnerships with in-country commercial investors, are factors in favor of an accelerator's success.¹⁰³ The study also found that while seventy-five percent of accelerators depended on philanthropy to survive and grants funded fifty-four percent of all accelerator budgets, such funding was not statistically related to the success of incubated startups.¹⁰⁴ Based on these findings, the study concluded that the business models of social-impact accelerators had not been proven to generate sustainable revenue streams.¹⁰⁵

The study further warned that accelerators may confront a “free rider” problem going forward: investors look to accelerators as “sourcing mechanisms,” but do not view it as their role to support accelerators—in fact, only twenty percent of investors help to fund the operations of accelerators.¹⁰⁶ This imbalance adds to the complexity of assessing and sustaining individual hubs: are they ordinary businesses themselves, or platforms for business that warrant different kinds of support?

Further critiques of hubs understood as companies are not difficult to locate. GSMA's Digital Entrepreneurship in Kenya 2014 survey of more than 230 startups across Kenya reveals that at least seventy percent of the country's startups are “not earning enough to maintain business and living expenses for a small team.”¹⁰⁷ Overall, these developments appear to have enlightened the African technology community and, in turn, spurred a flood of critical pieces on Africa's hub and startup ecosystem. In its summary, the GSMA Report notes that while hands-on support for entrepreneurs is available through hubs and accelerators, there is insufficient support to meet demand.¹⁰⁸ As a result, entrepreneurs appear to be very aware of the fact that they must increase their skills and balance out their teams but struggle to do so. Therefore, the report suggests that one-to-one mentorship across a broad variety of topics is desperately needed for start-ups, especially in marketing, technology access and skills, growth strategy, and business management.¹⁰⁹

The discourse of hubs as companies can creep into the goals, operations, and

¹⁰³ *Id.* at 22.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *Id.* at 23.

¹⁰⁷ DROUILLARD ET AL., *supra* note 72, at 4.

¹⁰⁸ *Id.* at 70.

¹⁰⁹ *Id.*

management of hubs, causing them to alter behaviors to better resemble those of traditional companies. In 2015, C4DLab, which was mentioned earlier in this Article,¹¹⁰ announced that it would run on a “lean model” until it reaches a break-even point.¹¹¹ The reworking of C4DLab’s business model reinforced the seriousness posed by the failure to confront the issue of hub sustainability. Startup accelerator 88mph’s recent departure from Kenya for Nigeria also raised concern. Nikolai Barwell, former Nairobi-based director of 88mph, explained the accelerator was exiting Kenya in favor of Nigeria, where “the tech ecosystem is more profit-focused and there is less fluff.”¹¹²

The now-common practice of applying the language of private companies to hubs is not without critics. Mostly such criticism is due to imposing corporate or non-profit oriented goals on entities that were originally intended merely to provide a community space where previously there was none. For example, prominent African technology entrepreneurs, including Mark Essien, have expressed concerns regarding the current trajectory.¹¹³

Essien, founder of successful Nigerian startup Hotels.ng, takes aim at the current approach taken by the majority of startup incubators in Africa. Essien’s critique centers on the fact that the reigning incubation model has yet to yield any “big successes,” as well as his belief that the fundamentals of entrepreneurship cannot be taught.¹¹⁴ Furthermore, the ANDE and Village Capital study has found that the difficulty in assessing accelerator performance arose partly from the fact that many accelerators are not collecting data, or tracking graduates of their programs.¹¹⁵ Such difficulty is echoed by a 2013 report by Dalberg, which notes that while the concept of incubators and accelerators is not new, additional research is needed to determine what truly drives growth and impact on entrepreneurs in Africa.¹¹⁶

¹¹⁰ See discussion *supra* Introduction.

¹¹¹ Jackson, *supra* note 38.

¹¹² Jorgic, *supra* note 39.

¹¹³ Essien, *supra* note 41.

¹¹⁴ *Id.*

¹¹⁵ BAIRD ET AL., *supra* note 101, at 1.

¹¹⁶ DALBERG GLOBAL DEV. ADVISORS, DIGITAL JOBS IN AFRICA: CATALYZING INCLUSIVE OPPORTUNITIES FOR YOUTH, DALBERG GLOBAL DEV. ADVISORS 13 (2013), http://www.dalberg.com/documents/Digital_Jobs_in_Africa.pdf.

Such criticisms point to a half-way approach at corporatizing hubs: hubs are expected to grow, become sustainable, and do many other things that businesses do, but rarely make the effort to monitor, evaluate, and communicate their activities. The identity crisis that company hubs face is unsustainable.

Applying the unforgiving language of the business world to hubs means that entities must prove their value or be assumed obsolete and subsequently abandoned, sold, or subject to takeover. USAID's Morgan McClain-McKinney questions the role and usefulness of incubators in the context of sub-Saharan Africa's emerging markets.¹¹⁷ McClain-McKinney sought to evaluate the success and utility of these ventures, but encountered numerous challenges in doing so. For McClain-McKinney, the challenge stemmed from the fact that the success of an incubator is generally measured on the basis of the number of program graduates or the number of startups that voluntarily exit the incubator, after receiving investments or other support.¹¹⁸ The problem with this metric is that a likelihood of failure persists following a startup's exit from an incubator. While McClain-McKinney notes that a better measure of success would be to track the number of graduates continuing to operate their startups three to five years post-graduation, she also recognizes that these figures were not available, as the vast majority of incubators in Africa have yet to be in existence for five years.¹¹⁹

1. Characteristics of a Company Hub

A company hub, as we define it, is characterized by its individuality. Unlike a cluster hub, a company hub is typically a separate legal entity. It may be for-profit or non-profit, and is able to enter into bilateral agreements with other entities. The company hub is also answerable to an identifiable entity, e.g., a group of shareholders, investors, or board members. Such stakeholders may or may not have in mind the best interests of the hub users. Company hubs may be structured with a variety of internal checks and balances, and may produce corporate-style documents such as strategic plans and earnings reports. In light of their independence, the services provided by a company hub are likely to be less extensive than those available at cluster hubs consisting of multiple loosely organized or unconnected companies. Company hubs can, however, overcome this shortcoming by partnering with other company hubs. Individual company hubs

¹¹⁷ McClain-McKinney, *supra* note 41.

¹¹⁸ *Id.*

¹¹⁹ *Id.*

may vary widely in their offerings to users and in other aspects, as shown in Tables 1 and 2.¹²⁰

Table 1: Company Hubs in Kenya Located Within Ngong Road Cluster

Name of Hub	Type of Hub	Main Offering of Hub
Nailab	Civil Society-led	Incubation/Accelerator
m:lab East Africa	Hybrid	Incubation
Nairobi Garage	Hybrid	Co-working space
Sinapis	Hybrid	Accelerator
Savannah Fund	Hybrid	Accelerator
Gearbox	Civil Society-led	Makerspace
Akirachix	Civil Society-led	Co-learning space
88mph (inactive)	Hybrid	Accelerator

Table 2: Company Hubs in Kenya Located Outside Ngong Road Cluster Hub

Name of Hub	Location of Hub	Type of Hub	Main Offering of Hub
Fab Lab Nairobi	University of Nairobi	Academic Institution-led	Makerspace
iLabAfrica/iBizAfrica	Strathmore University	Academic Institution-led	Incubation
Chandaria BIIC	Kenyatta University, Nairobi	Academic Institution-led	Incubation
Kenya CIC	Strathmore University, Nairobi	Academic Institution-led	Co-working space/Accelerator
C4D Lab	University of Nairobi	Academic Institution-led	Incubation
GrowthHub/Growth Africa	Kilimani area, Nairobi	Civil Society-led	Accelerator
Lake Hub	Kisumu	Civil Society-led	Incubation
Sote Hub	Voi Town	Civil Society-led	Co-working/Incubation/Accelerator
Swahili Box	Mombasa	Civil Society-led	Incubation
ARO Fab Lab	Kisumu	Civil Society-led	Makerspace

¹²⁰ WORLD BANK GROUP, *supra* note 56.

Name of Hub	Location of Hub	Type of Hub	Main Offering of Hub
KICTANET	Virtual (online only)	Hybrid	ICT reform catalyst (online platform)
iHub	Kilimani area, Nairobi	Civil Society-led	Co-working/Pre-Incubation
Konza Technopolis	Machakos Town	Government-led	Smart City
Matili Technology Hub (mtHub)	Bungoma Town	Government-led	Science and Technology Park
MakersHub	Mombasa	Civil Society-led	Makerspace
Metta	Riverside area, Nairobi	Hybrid	Co-working space
Nest VC	Riverside area, Nairobi	Hybrid	Accelerator
The Foundry Africa	Westlands area, Nairobi	Hybrid	Co-working space
SwahiliPot	Mombasa	Civil Society-led	Co-working space
Mombasa Tech	Virtual (online only)	Hybrid	Tech Community initiative (online platform)
Ubunifu Hub	Machakos Town	Civil Society-led	Co-working
PAWA254	Kilimani area, Nairobi	Civil Society-led	Co-working/Incubation
FabLab Kivuli	Nairobi	Civil Society-led	Makerspace
BitHub.Africa	Kilimani area, Nairobi	Hybrid	Co-working/Incubation

C. Country Hubs

The category of country hub recognizes that, in addition to small geographic areas emerging as cluster hubs, very large geographic areas can become known as centers of entrepreneurship activities. This is an especially important archetype in terms of government policy initiatives. While innovation policy is often made at the local level, it is also very often the focus of national-level attention. Therefore, discussions about countries as technology hubs are most closely aligned with theoretical concepts and literature on national systems of innovation.¹²¹

In Africa this is particularly, but not exclusively, evident in tech entrepreneurship. As discussed below, the country hub may evolve intentionally due to government policies (government-led country hubs) or unintentionally due to an organic clustering of like-minded entrepreneurs in a city, country, or region

¹²¹ NATIONAL SYSTEMS OF INNOVATION: TOWARDS A THEORY OF INNOVATION AND INTERACTIVE LEARNING (Bengt-Åke Lundvall ed., 1992).

(sector-led or “government follows” country hubs, or a combination thereof). Although the term “country hub” implies a political boundary, these hubs are not explicitly limited to entire countries—regions smaller and larger than a country may also qualify.

Kenya exemplifies the sector-led or “government follows” country hub. Since the launch of M-Pesa in 2007 and other technological milestones such as Ushahidi and iHub, Kenya has been recognized as Africa’s leading technology hub.¹²² The country has been hailed as the origin of technological ventures on the continent, leading to the “Silicon Savannah” moniker,¹²³ or our preferred term “Digital Savannah.” Going forward, Kenya is anticipated to maintain this lead ahead of other African countries.¹²⁴ Below, a variety of government actions are described that illustrate Kenya’s intentional and continued development as a country hub.

1. Vision 2030, a new Constitution, and the ICT Masterplan

The Kenyan government has set out a detailed action plan designed to further Kenya’s reputation as a country hub. These actions include the enactment of ICT friendly laws and policies, investment in critical ICT infrastructure, and the establishment of e-services.

Launched in 2008, Vision 2030 (the country’s development blueprint) is a foundation to Kenya’s development as a country hub. The blueprint recognizes the ICT sector’s potential to foster economic development and to improve quality of life. Vision 2030’s overall aim is to transform Kenya into an industrialized, middle-income country by 2030. Progress is to be achieved by way of five-year, medium-term plans. Currently, the blueprint is in its second medium-term plan.

Under the first medium-term plan, Kenya transitioned to a new constitutional dispensation that introduced a partially devolved government. Milestones of the first medium-term plan in the ICT sector included the laying of three undersea submarine fiber-optic cables, linking Kenya to the global broadband highway, and the completion of a 5500-kilometer terrestrial fiber-optic network.

¹²² Sophie Mongalvy, *Inside the African Tech Hub Rising in Nairobi*, BLOOMBERG, July 29, 2015, <https://www.bloomberg.com/news/articles/2015-07-29/inside-the-african-tech-hub-rising-in-nairobi/>.

¹²³ See Moime, *supra* note 75.

¹²⁴ Lily Kuo, *African Startups Are Defying the Global Tech Slowdown*, QUARTZ AFRICA (Jan. 12, 2016), <http://qz.com/592119/african-startups-are-defying-the-global-tech-slowdown/>.

These milestones were achieved in 2010, the same year that iHub was launched. High-speed Internet access enabled the development of Kenya's ICT industry. From 2009 to 2012, Internet subscriptions increased over 500%—from 1,579,387 to 8,506,748.¹²⁵ During this time, Kenya also established an open-data portal that provides public access to government data, and multinationals such as Google and IBM opened offices in Nairobi.¹²⁶

The second medium-term plan (2013-2017), entitled “Transforming Kenya: Pathway to Devolution, Socio-Economic Development, Equity and National Unity,” identifies ICT as one of the foundations for national transformation:

Kenya's vision of knowledge based economy aims at shifting the current industrial development path towards innovation where creation, adoption, adaptation and use of knowledge remain the key source of economic growth. ICT is a critical tool for expanding human skills and rests largely on a system of producing, distributing and utilizing information and knowledge that in turn plays a great role in driving productivity and economic prosperity.¹²⁷

The National ICT Masterplan guides the country's ICT transformation. Launched in 2014, the plan provides for four flagship projects: the upgrading of national ICT infrastructure, improvement of public service delivery through the use of ICT, development of the ICT industry, and the upgrading of ICT capacity. The Masterplan's aim is “to make Kenya an ICT hub and a globally competitive digital economy.”¹²⁸

The Masterplan notes certain ongoing activities that put Kenya on the map as an African ICT hub, including e-government, infrastructure projects, public-private partnerships (PPP's), and support of private initiatives.¹²⁹

The National Optic Fibre Backbone Infrastructure (NOFBI) project aims to increase connectivity and enable communication across Kenya's forty-seven counties. In phase one of the project, completed in 2009, NOFBI access points

¹²⁵ REPUBLIC OF KENYA, SECOND MEDIUM TERM PLAN, 2013–2017 (2013), www.vision2030.go.ke/lib.php?f=second-medium-term-plan-2013-2017.

¹²⁶ KENYA OPEN DATA PORTAL, <https://www.opendata.go.ke> (last visited Mar. 2, 2017).

¹²⁷ REPUBLIC OF KENYA, *supra* note 125, at 21.

¹²⁸ THE KENYA NATIONAL ICT MASTERPLAN 2014-2017, INFO. AND COMMC'N TECH. AUTH. 12 (2014), <http://icta.go.ke/national-ict-masterplan/>.

¹²⁹ *Id.* at 26-34.

were established in major towns in Kenya, covering a total distance of 4300 kilometers.¹³⁰ In phase two of the project, launched in 2014, the government aims to extend connectivity to all forty-seven counties, covering a total of 2100 kilometers.¹³¹ According to the 2015 Report of the Kenya National Bureau of Statistics, the implementation of the NOFBI project led to an exponential increase in the number of Internet users by twenty-three percent, from 21,300,000 in 2013 to 26,300,000 in 2014.¹³² The County Connectivity Project (CCP) utilizes the NOFBI to connect the national government to county governments and to interconnect county governments. It is aimed at enhancing the quality of e-government services, thereby improving service delivery to citizens.¹³³

The Masterplan also recognizes the importance of incubators and co-working spaces, referred to as “local ICT development groups.”¹³⁴ In turn, the government established incubation hubs at Kenyatta University (Chandaria Business Innovation and Incubation Centre) in 2011 and at the University of Nairobi (Computing for Development Lab) in 2013. These innovation centers were situated in public universities to afford citizens opportunities to innovate, thereby furthering national development.

From 2012 to 2015, the government partnered with Nailab to launch a KES 144,000,000 (equivalent to USD 1,400,000) incubation project.¹³⁵ The creation of the incubator was carried out as a Business Process Outsourcing project, which allowed the government’s commitment to supporting startups to be outsourced to a specialist hub. Over the three-year contract period, the program aimed to incubate thirty startups countrywide. Digital Villages, also known as ‘Pasha Centres’ (“pasha” is a Swahili word for “inform”), are another example of a PPP. Launched in 2009, Pasha Centres are essentially ICT hubs created to bring online services to marginalized communities.¹³⁶ Private entrepreneurs who secure loans from the ICT

¹³⁰*National Optic Fibre Backbone (NOFBI)*, INFO. AND COMMC’N TECH. AUTH., <http://www.icta.go.ke/national-optic-fibre-backbone-nofbi/#> (last visited Mar. 2, 2017).

¹³¹ *Id.*

¹³² Economic Survey 2015, KENYA NAT’L BUREAU OF STATISTICS, 19, 235 (2015), http://www.knbs.or.ke/index.php?option=com_phocadownload&view=category&id=107:economic-survey-publications&Itemid=1181.

¹³³ KENYA NATIONAL ICT MASTERPLAN 2014-2017, *supra* note 128.

¹³⁴ *Id.* at 30.

¹³⁵ *Nailab Business Incubation Project*, INFO. AND COMMC’N TECH. AUTH., <http://icta.go.ke/nailab/> (last visited Mar. 2, 2017).

¹³⁶ Peter Drury, *Kenya’s Pasha Centres: Development Ground for Digital Villages*, CISCO

Authority operate these centers.¹³⁷ The benefits of this initiative are two-fold: it not only provides jobs to youth but also ensures digital inclusion across Kenya. Statistics available via the Kenya Open Data portal reveal that approximately sixty-three Pasha Centers have been created since 2011.¹³⁸

Despite these advances, the development of Konza Technology City remains the kingpin of government efforts to advance Kenya to a “country hub.” Konza is a massive Vision 2030 flagship project with ambition “to be a sustainable, world class technology hub and a major economic driver for the nation, with a vibrant mix of businesses, workers, residents and urban amenities.”¹³⁹ The project was initiated in 2009 with the procurement of a 5000-acre parcel of land located sixty kilometers southeast of Nairobi. In phase one of the project, the authority made a call for investors to take up twenty-four parcels of land for development. It is estimated that the first phase of the project will be complete and ready for occupation in 2017. Konza is expected to host business processing, outsourcing, residential areas, a university focused on research and technology, hotels, shopping malls, schools, and hospitals.¹⁴⁰ The project, which will take twenty years to complete, is set to host the largest technology hub in East and Central Africa.

2. *Characteristics of a Country Hub*

A country hub, then, is a relatively large geographic area with several sub-regions that are similarly governed (typically this is a single country although, as we discuss below, characteristics of a country hub are discernable in multilateral regional political bodies). There are a number of critical factors for countries aiming to establish themselves as a hub: presence of an ICT regulatory framework, the existence and implementation of government policies, a highly-educated and skilled workforce, a business-friendly environment, incentives for private sector development, and a viable IT infrastructure, among others. Intentional efforts

(Jan. 2011), http://www.cisco.com/c/dam/en_us/about/ac79/docs/case/Kenya-Pasha-Centres_Engagement_Overview_IBSG.pdf.

¹³⁷ *Pasha Centres*, INFO. AND COMMC’N TECH. AUTH., <http://icta.go.ke/pasha-centres/> (last visited Mar. 2, 2017).

¹³⁸ *Kenya Open Data, Distribution of Pasha Centre*, KENYA OPEN DATA, <http://www.opendata.go.ke/datasets/distribution-of-pasha-centre/> (last visited Mar. 2, 2017); see also *Distribution of Pasha Centres in Kenya (2016 Update)*, OPENAFRICA, <http://africaopendata.org/dataset/distribution-of-pasha-centers-in-kenya-2016> (last visited Mar. 2, 2017).

¹³⁹ *The Vision*, KONZA TECHNO CITY, www.konzacity.go.ke/the-vision (last visited Mar. 2, 2017).

¹⁴⁰ *Id.*

toward formation of a country hub typically involve local and/or national governments addressing these factors through policies, procurement, and public relations.

A country can become a technology hub either led or followed by conscious government policy choices and decisions to encourage, facilitate and foster innovation while enhancing the uptake of ICT by locals. Kenya undertakes this process by building ICT capacity in its human resources, creating a foundation for Internet infrastructure, and creating an enabling environment for technology startups.¹⁴¹ The implementation and impact of projects under the named policies are addressing these areas in a bid to strengthen Kenya's position as Africa's leading technology hub.

Above, Kenya is discussed as one example of a country hub. Rwanda, Tanzania, and Uganda are also taking steps to assert themselves as country hubs, particularly in the area of technopreneurship.¹⁴² Country hubs may further evolve intentionally into multi-national regional hubs through joint collaboration and implementation of regional policies. There is evidence that the East African Community (EAC) is cooperating with a common goal of regional integration in the area of ICT, perhaps with the ultimate goal of becoming the equivalent of a regional country hub.¹⁴³ This cooperation, evidenced by a Protocol for Cooperation on ICT Networks and a Model ICT policy framework, is aimed at harmonizing ICT policies across all member states in order to increase penetration of ICT and to encourage innovation in ICT.¹⁴⁴

¹⁴¹ In 2015, a new Companies Act was enacted enabling individuals to form single member companies.

¹⁴² *Tanzania to Become ICT Hub in East and Central Africa*, TANZANIA INVEST (Aug. 5, 2015), <http://www.tanzaniainvest.com/telecoms/Tanzania-to-become-ict-hub-in-east-and-central-africa/>; Joseph Mayton, *Uganda: Government Plans to Set Up ICT Hub*, IT NEWS AFRICA (June 1, 2015), <http://www.itnewsafrika.com/2015/06/uganda-govt-plans-to-set-up-ict-hub/>; see also Jean Philbert Nsengimana, *Co-creating a Smart Rwanda, Smart Africa and Smart World*, REPUBLIC OF RWANDA- MINISTRY OF YOUTH & INFO. AND COMMC'N TECH., <http://www.myict.gov.rw/ict/smart-rwanda/smartwanda-concept/> (last visited Mar. 2, 2017).

¹⁴³ Maurice Okore, *East Africa to Expand One Network Area to Data*, NEW VISION (Apr. 3, 2015), http://www.newvision.co.ug/new_vision/news/1323586/east-africa-expand-network/; see also Frank Kanyesigye, *EAC Eyes Stronger ICT Agenda*, NEW TIMES (Sept. 4, 2013), <http://www.newtimes.co.rw/section/article/2013-09-04/68906/>.

¹⁴⁴ *EAC Ministers Approve Bill on One Stop Border Posts*, NEW TIMES (Feb. 26, 2012), <http://www.newtimes.co.rw/section/article/2012-02-26/88973/>; E. AFRICAN COMMC'NS ORG.,

Challenges to such regional integration remain, including a lack of convergence of the ICT regulatory frameworks¹⁴⁵ and different timelines for infrastructure development in the six EAC member states.¹⁴⁶ Despite high levels of political unity and shared development goals, diverse needs country-to-country will complicate establishment and maintenance of regional-level country hubs.

III DISCUSSION/ANALYSIS

The cluster hub and company hubs of Nairobi's vibrant technology scene developed organically and the Kenyan government has positioned itself to capitalize on that success by way of new complementary initiatives, aiming at developing Kenya into a country hub. While there are instances of collaborative interaction between the three tiers, there is also the potential for conflict between the country hub, Nairobi's company hubs, and the cluster hub of Ngong Road. Beyond collaboration and conflict, however, evidence suggests that the tiers also compete.

Table 3 contains a summary of the archetypes of hubs we have identified:

Table 3: Archetypes of African Technology Hubs

	Cluster Hub	Company Hub	Country Hub
Features	Small geographical region (e.g., a neighborhood, municipality, urban corridor) containing a high density of hubs, and supporting entities	Separate legal entity (for-profit or non-profit) able to enter agreements, accountable to investors/funders and stakeholders	Large geographic area with distinct political identity and/or several similarly governed sub-regions (e.g., a country, countries or region)

EAC MODEL ICT POLICY FRAMEWORK, E. AFRICAN COMMC'NS ORG. 4 (2015), http://www.eaco.int/docs/WGsReports/Draft_Model_ICT_Policy_KGJ_March_2015.pdf; *see also Report of the Committee on Communications, Trade and Investments on the On-Spot Assessment of Regional Cooperation in ICT*, E. AFRICAN LEGISLATIVE ASSEMBLY, REPORT OF THE COMMITTEE ON COMMUNICATIONS, TRADE AND INVESTMENTS ON THE ON-SPOT ASSESSMENT OF REGIONAL COOPERATION IN ICT (2013), <http://www.eala.org/documents/view/on-spot-assessment-of-regional-cooperation-in-ict>.

¹⁴⁵ Joseph Kariuki Nyaga et al., *Regulatory Convergence of ICT Sectors in the East African Community (EAC): Challenges for the Current Legislative and Regulatory Frameworks and Lessons from the European Union Experience*, 2013 IST-AFRICA PROCEEDINGS 1 (2013).

¹⁴⁶ Kanyesigye, *supra* note 143.

	Cluster Hub	Company Hub	Country Hub
Development	One company hub's success attracts others; member(s) of existing company hub create(s) a new hub in close proximity to original company hub; association/partnership (formal or informal) between separate company hubs	Unmet need identified and new entity created to satisfy void; driven by entrepreneurial individuals and/or investors	Government policymaking as leader (to strategically develop strengths in a particular field, e.g. ICTs) or follower (to identify and capitalize on emerging strengths)
Governance	Informal, community-led	Funders/investors, board, executives	Government officials
Factors for success	Hub-to-hub interaction and sharing of resources (e.g., physical spaces, Internet access, human resources, venture capital)	Sustainability, scalability, profitability	Regulatory framework; government policies; educated/skilled workforce; business environment; incentives for private sector development; IT
Kenyan examples	Ngong Road	iHub	Kenya

A. Hubs as Complementary and Collaborative

There are complementary relationships both between the cluster hubs and the company hubs, and between the company hubs/cluster hubs and the country hubs.

1. Company Hubs and Cluster Hubs

The settlement of many company hubs in one location forms a cluster hub. This co-location results in the geographical concentration of open-working spaces, incubators, accelerators, startups, and venture capitalists, which together derive the benefits of collective efficiency common in clusters. Having various actors in the cluster offering diverse services creates an opportunity for mutual benefits from the roles and responsibilities of the others.

For instance, the different hubs at the Bishop Magua Centre (i.e., iHub (formerly), Nailab, m:Lab, and AkiraChix) are made up of distinctively different models, offering different services to their startups but with the common goal of

cultivating early stage startups. These include an open working space where entrepreneurs meet (iHub), an incubator that provides training and mentorship to early stage mobile startups (m:Lab), and an accelerator that provides rapid fix-term mentorship, funding, and education to early-stage growth driven startups (Nailab). It is common to find startups that have been part of these different hubs at different stages of their growth.

The concentration of different services in one area results in the concentration of technology entrepreneurs in that area, which in turn facilitates the accumulation of knowledge and skills, thus generating more innovation.

Besides benefitting startups, company hubs attain sustainability by relying on the structure of a cluster hub, i.e., a geographical concentration of distinct hub models offering different services. Thus, company hubs do not stretch their budgets in order to meet all the needs of a given startup, one of the factors known to hamper the sustainability of hubs.

2. Company/Cluster Hubs and Country Hubs

The successes of company hubs and cluster hubs have the ability to elevate the status of a country hub on the international stage. Kenya's company and cluster hubs continue to position the country as the regional leader in technology. The activities of iHub, for instance, have highlighted Kenya's potential to be a center of innovation on the African continent.

The company and cluster hubs are also considered to be necessary factors in sustaining the entrepreneurship goals of a country hub. In this light, company and cluster hubs fulfill government mandates of training and creating a startup culture, factors that are key to the creation of a technology revolution in Kenya and other African countries. Through public/private partnerships, the government's commitment to supporting startups is outsourced to specialist company hubs. A key example is the ICT Authority's partnership with the accelerator Nailab. In the absence of these partnerships, company hubs fill the gap and rely on their own resources to train entrepreneurs.

Similarly, supports and policies to build country hubs assist in the sustainability, scalability, value, and impact of the cluster and company hubs. These policies relate to Internet infrastructure and the creation of an enabling environment necessary to attract foreign investment, to reduce the costs of doing business for company hubs and startups, and to eliminate barriers to entry for new entrants.

The promotion of country hubs continues to be an important part of innovation systems in a country, and consequently in the companies and clusters that exist there. For example, the availability of cheap, reliable Internet in Kenya is attributed to the laying of three undersea fiber-optic cables in 2010 and the ongoing National Optic Fibre Backbone Infrastructure Project.¹⁴⁷

The government has also changed many of its policies with a view to creating an enabling environment for startups and company hubs and attracting foreign investment. For example, a new Companies Act was enacted in 2015 enabling the formation of single-member companies and making it possible for foreign companies to be registered by the use of electronic documents.¹⁴⁸ It also did away with the mandatory requirement of a company seal and company secretary for private companies with a share capital of less than KES 5,000,000 (equivalent to USD 50,000). The Companies (General) Regulations also provide Model Articles of Association thus making it easier and faster for entrepreneurs to register their companies. Through these efforts, Kenya's Ease of Doing Business rank has improved from 129 in 2014 to 108 in 2015. The government has also put in place Huduma (Swahili for 'service') centers, which aim to transform public service delivery by providing a one-stop shop for government services including registration of business names, procurement of national identity cards and drivers' licenses, and filing of tax returns, among others. Through these and ongoing efforts, Kenya has become a leading destination for foreign direct investment in Africa.¹⁴⁹

B. Hubs in Conflict

The primary conflict among the three archetypes of hubs we have identified stems from a number of ICT-related laws and policies (including drafts and proposals) by the country hub, which appear to have a negative impact on the scalability and/or sustainability of both company hubs and cluster hubs. Three recent examples illustrate this conflict: the proposed law to regulate ICT

¹⁴⁷ Margaret Nyambura Ndung'u & Timothy Mwololo Waema, *Understanding What Is Happening in ICT in Kenya. A Supply- and Demand-Side Analysis of the ICT Sector*, RESEARCH ICT AFRICA 18 (2012), https://www.researchictafrica.net/publications/Evidence_for_ICT_Policy_Action/Policy_Paper_9_-_Understanding_what_is_happening_in_ICT_in_Kenya.pdf.

¹⁴⁸ The Companies Act No. 17 (2015), KENYA GAZETTE SUPPLEMENT NO. 158 §§ 32, 841.

¹⁴⁹ *Africa's Largest Economy, Kenya, Is One of the Leading Destinations for Foreign Direct Investment (FDI) in Africa*, INVEST IN GROUP (Apr. 2016), <http://investingroup.org/snapshot/268/kenya-fdi-kenya/>.

practitioners, the country hub's stance on Bitcoin and other crypto-currencies in Kenya, and the draft regulations on electronic transactions and cyber security. Each example will be discussed in turn.

In June 2016, the Information Communication Technology (ICT) Practitioners Bill, 2016 was published in the Kenya Gazette.¹⁵⁰ The Bill is set to be tabled in Parliament as a Private Member's Bill.¹⁵¹ The Preamble of this "Anti-Innovation" Bill¹⁵² states that it is an Act of Parliament to provide for the training, registration, licensing, practice, and standards of ICT practitioners.

According to iHub Executive Director Josiah Mugambi, the Bill is "hugely detrimental to the ICT industry" and has caused "much concern and angst among people in the ICT industry in Kenya."¹⁵³ As a result, iHub hosted a workshop session to familiarize its members with the contents of the Bill, particularly those relating to registration and licensing of ICT practitioners which have a direct impact on the iHub members who work as freelance software consultants.¹⁵⁴ In July 2016, iHub together with a number of ICT industry stakeholders drafted and presented a memorandum to Parliament¹⁵⁵ in which they described the Bill as "ill-advised and completely out of touch with the realities in the ICT industry today."¹⁵⁶ Among the issues raised by iHub and others in the memorandum are the lack of stakeholder participation in the drafting of the Bill and the lack of congruence between the Bill and various government policies on ICT, labor, youth, and employment. The memorandum by iHub and others concludes that innovation and learning/knowledge, technology transfer, and fostering technology entrepreneurship will suffer if the Bill is passed by Parliament in its current

¹⁵⁰ The Information Communication Technology Practitioners Bill No. 18 (2016), KENYA GAZETTE SUPPLEMENT NO. 84.

¹⁵¹ Sandra Chao-Blasto, *ICT Secretary Disowns Bill Seeking to Have Practitioners Licensed*, BUS. DAILY AFRICA (July 7, 2016), <http://www.businessdailyafrica.com/Corporate-News/ICT-secretary-disowns-Bill/-/539550/3284184/-/fmb8g1/-/index.html>.

¹⁵² Maina Waruru, *Opposition to Kenyan "Anti-Innovation" ICT Bill Grows*, INTELL. PROP. WATCH (July 13, 2016), <http://www.ip-watch.org/2016/07/13/opposition-to-kenyan-anti-innovation-ict-bill-grows/>.

¹⁵³ Josiah Mugambi, *An Update on the Proposed ICT Practitioners' Bill*, IHUB (Aug. 19, 2016), <http://ihub.co.ke/blogs/28245/>.

¹⁵⁴ *Id.*

¹⁵⁵ *Id.*

¹⁵⁶ A copy of the Stakeholders' Feedback on the ICT Practitioners Bill is available online: https://drive.google.com/file/d/0Bw6KfbaBAWJ_TU5jOWRBSnhOLUE/view/.

form.¹⁵⁷

The regulation of crypto-currencies such as Bitcoin and blockchain technology in Kenya is another example of conflict between the country hub and various company hubs. In December 2015, the Central Bank of Kenya issued a public notice cautioning the public on “virtual currencies such as Bitcoin.”¹⁵⁸ Further, the Central Bank declared that such currencies were not legal tender in Kenya and the public should “desist from transacting in Bitcoin and similar products.”¹⁵⁹

This move appears to be in direct conflict with the existence of company hubs such as BitHub.Africa, a commercial, Kenyan-based blockchain accelerator driving the adoption of blockchain technology and solutions across Africa.¹⁶⁰ According to the founder of BitHub.Africa, the Central Bank should take time to assess the potential of blockchain technology to reduce costs and enhance transparency across multiple sectors of the economy.¹⁶¹

The final example of conflict between the country hubs and the company hubs is the proposed draft of the Kenya Information and Communications Regulations on Electronic Transactions¹⁶² and Cyber-security,¹⁶³ prepared by the

¹⁵⁷ Mbugua Njihia, *3 Concerning Issues Regarding Kenya's Information Communication Technology Practitioners Bill*, IAFRIKAN (July 8, 2016), <http://www.iafrikan.com/2016/07/08/3-concerning-issues-regarding-kenyas-information-communication-technology-practitioners-bill-that-need-to-be-addressed/>; see also John Ngirachu, *Bloggers Raise Storm over Draft ICT Bill*, DAILY NATION (July 8, 2016), <http://www.nation.co.ke/news/Bloggers-raise-storm-over-draft-ICT-Bill/1056-3284286-rpw0tx/index.html>.

¹⁵⁸ *Public Notice: Caution to the Public on Virtual Currencies such as Bitcoin*, CENTR. BANK OF KENYA (Dec. 2015), https://www.centralbank.go.ke/images/docs/media/Public_Notice_on_virtual_currencies_such_as_Bitcoin.pdf/.

¹⁵⁹ *Id.*

¹⁶⁰ Diana Ngo, *BitHub Africa's Founder: 'Achieving the Dream of Global Inclusiveness' with Bitcoin*, BTC MANAGER (Jan. 23, 2016), <https://btcmanager.com/news/business/bithub-africas-founder-satoshis-innovation-brings-us-closer-to-achieving-the-dream-of-global-inclusiveness/>.

¹⁶¹ *Id.*

¹⁶² Victor Nzomo, *You Will Need a License to Sell Stuff on Whatsapp: Thoughts on the Electronic Transactions Regulations 2016 by Communications Authority*, CENTRE FOR INTELL. PROP. AND INFO. TECH. LAW (CIPIT) BLOG (Jan. 19, 2016), <http://blog.cipit.org/2016/01/19/you-will-need-a-government-license-to-sell-stuff-on-whatsapp-thoughts-on-the-electronic-transactions-regulations-2016-by-communications-authority/>.

¹⁶³ Josiah Mugambi, *Of Cyber Security and Public Wifi*, iHUB (July 2, 2015), <http://ihub.co.ke/blogs/24750>.

Communications Authority of Kenya. According to iHub Executive Director Josiah Mugambi, there are two main problems with the draft regulations: the requirement that owners of public Wi-Fi networks register their users and the requirement that all Kenyan companies host their websites locally.¹⁶⁴ The Wi-Fi registration requirement has elicited considerable criticism both locally¹⁶⁵ and abroad.¹⁶⁶ On the requirement to host websites by Kenyan companies locally, the view of iHub's Executive Director is that it ought to be a business decision for companies to make based on a determination of various factors, including cost, uptime, reliability, and security.¹⁶⁷

C. Hubs in Competition

Competition can be defined as interaction between the tiers, in which the growth, success, and/or failure of a particular tier depends on gaining a share of the limited market. One example of this potential for competition is the government's flagship project of Konza Technology City (i.e., the development of Kenya into a country hub).

Ultimately, Konza can take one (or some combination) of two paths: collaborative or competitive.¹⁶⁸ Under a collaborative or complementary approach,

¹⁶⁴ *Id.*

¹⁶⁵ *Id.*; see also Kenya's Communication Authority Goes All 'Big Brother' on Public WIFI Networks to Curb Cybercrime, MOSES KEMIBARO (July 1, 2015), <http://www.moseskemibaro.com/2015/07/01/kenyas-communications-authority-goes-all-big-brother-on-public-wifi-networks-to-curb-cybercrime/>; Victor Nzomo, *State Surveillance, Mixed Signals and Seven Years in Jail: Thoughts on Cybersecurity Regulations 2016 by Communications Authority* CENTRE FOR INTELL. PROP. AND INFO. TECH. LAW (CIPIT) BLOG (Jan. 18, 2016), <http://blog.cipit.org/2016/01/18/state-surveillance-mixed-signals-and-seven-years-in-jail-thoughts-on-cybersecurity-regulations-2016-by-communications-authority/>; Sidney Ochieng, *Deciphering the Kenya Information and Communications Act Amendment Drafts*, IAFRIKAN (Feb. 20, 2016), <http://www.iafrikan.com/2016/02/20/deciphering-the-kenya-information-and-communications-act-amendment/>.

¹⁶⁶ Jillian York, *Kenya to Require Public Wi-Fi Users to Register with Phone Number*, ELEC. FREEDOM FRONTIER (July 14, 2015), <https://www.eff.org/deeplinks/2015/07/kenya-require-wireless-users-register-phone-number>; see also Karl Bode, *Kenya's Ingenious Solution to Cybercrime: Register Every Wi-Fi User and Device with the Government* TECHDIRT (July 6, 2015), <https://www.techdirt.com/articles/20150701/13054431518/kenyas-ingenious-solution-to-cybercrime-register-every-wi-fi-user-device-with-government.shtml>.

¹⁶⁷ Mugambi, *supra* note 153.

¹⁶⁸ MICHAEL BLOWFIELD & LEO JOHNSON, *TURNAROUND CHALLENGE: BUSINESS AND THE CITY OF THE FUTURE* 214-15 (2013).

Konza could serve to accelerate synergy between Konza and the so-called “iHub community” (i.e., the Ngong Road Cluster Hub and Nairobi’s Company Hubs).¹⁶⁹ In this form, the country hub’s policies would not serve to displace the existing company hubs or cluster hubs of multiple companies. Rather, the country hub serves as a platform enabling members of the other tiers to operate more productively.¹⁷⁰

Alternatively, Konza also has the potential to undercut the organic ‘iHub community’ in at least three ways.¹⁷¹ First, there remains a limited supply of local talent in Kenya. As such, a competition for top talent is conceivable. In this scenario, Konza could compete with the other tiers by attracting talent to the country hub and away from existing hubs.

Second, the government’s focus on Konza could result in the prioritization of the country hub over the acceleration of the Ngong Road cluster hub and associated company hubs. Practically, this could take the form of government funds dedicated to improving infrastructure (e.g., roads, electricity, water, and sanitation) in Konza rather than within Nairobi.

Third, Konza’s development could disrupt the existing innovation ecosystem and encourage competition by (1) dividing the ‘iHub community’ between the cluster hub of Ngong Road and country hub of Konza and/or (2) isolating Konza-based developers and entrepreneurs from end-users in Nairobi and elsewhere in Kenya. The success of M-Pesa, among others, is commonly attributed to the ability of its innovators to liaise with end-users, become attuned to local needs, and update the product accordingly.¹⁷²

While the Konza development is still in progress, there is already evidence of tier-to-tier competition. The lack of local interest and investment in Konza is one example of this competitive sentiment.¹⁷³ Entrepreneurs and investors from the Ngong Road cluster hub and Nairobi’s individual company hubs have expressed misgivings regarding Konza’s likely impact, including the iHub’s Erik Hersman

¹⁶⁹ *Id.* at 216.

¹⁷⁰ *Id.*

¹⁷¹ *Id.* at 215-16.

¹⁷² *Id.* at 215.

¹⁷³ Tom Jackson, *Kenya’s Konza Tech City: A Step Too Far?*, VENTURES AFRICA (July 23, 2012), <http://venturesafrica.com/kenyas-konza-tech-city-a-step-too-far/>.

who noted that Konza's success was "a bit of a long shot."¹⁷⁴

CONCLUSION

This paper captures the current state of knowledge on African hubs, using the example of Kenyan-based entities, and proposes a framework for characterizing the continent's hubs and their practices. Reviews of the relevant literature and theory reveals gaps between academic scholarship and practice- and policy-oriented literature on this topic, which this paper addresses through the development of a categorization framework. Further, the growth and diversity of hubs across Africa makes the development of an analytical framework to facilitate future research especially prudent. Based on our hypothesis that Kenya is a microcosm of the continent, the framework advanced herein suggests a three-tiered system for categorizing African hubs as either a cluster, a company, or an entire country.

Our original framework is derived from and applied to Kenya's hubs, and leaves at least three clear areas for follow-up research, which we recommend.

The Kenya-specific approach in this article leaves, first, the opportunity for future works to consider the application of the framework to hubs across the continent. Indeed, some research is already underway, supported by the Open African Innovation Research network.¹⁷⁵ This framework will facilitate better and more informed research in countries like South Africa, Nigeria, Ghana, and Egypt.

Second, the government-led versus government-follow approach to technological innovation presents another avenue for further study. There is a gap in academic scholarship on the potential impact of government support and ICT friendly policy-making on hubs and their future development. As such, a comparative study of the workings, successes, and failures of organic technology communities (cluster hubs) and government-backed ventures (country hubs) could be undertaken. For example, we see much potential in a comparative analysis of the successes and failures of policies from countries within a particular region, such as Kenya, Rwanda, Uganda, and Tanzania. There is also significant potential in comparative analyses across geographic regions. Comparative analysis could be

¹⁷⁴ Jake Bright, *Billion Dollar Tech Cities Hope and Konza: Boon or Bane to Africa's Tech Movement?*, THIS IS AFRICA (July 3, 2014), <http://www.thisisafricaonline.com/News/Billion-dollar-tech-cities-Hope-and-Konza-Boon-or-bane-to-Africa-s-tech-movement?ct=true>.

¹⁷⁵ Elahi et al., *supra* note 11.

done in developing regions, such as between Africa and Latin America, South Asia, or South East Asia. Alternatively, it could be done between a developing region and a developed region, such as between Africa and Europe or North America.

Third, our framework will facilitate more nuanced empirical research focusing on individual company hubs. It will help researchers understand and distinguish existing literature that may use “hub” terminology, but is discussing a different one of the three distinct archetypes of hubs that we have identified.

By reviewing the relevant literature and synthesizing the state of current knowledge about high technology hubs, developing an original taxonomy to describe and assess technology hubs, and discussing implications of our research, we have laid the groundwork for other researchers to pursue further study in this area.

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 6

SPRING 2017

NUMBER 2

THROUGH THE LOOKING GLASS: PHOTOGRAPHY AND
THE IDEA/EXPRESSION DICHOTOMY

CARL MAZUREK*

Copyright law has always expressed an idea/expression dichotomy, where copyright protection extends not to an idea of a work but only to work's expression of that idea. Alas, this distinction walks a fine line with regard to non-textual and visual works. In particular, courts are prone to inconsistent outcomes and a violation of the fundamental precepts of copyright law because courts often succumb to shortcomings in grasping aesthetic theories of originality, realism, and ideas idiosyncratic to visual works. However, this dilemma may be solved within the existing framework of copyright law. This note argues that the solution should start by focusing less on visual works' subject matter, but rather elements of the work, such as the originality and realism of the expression that clarify the author's creativity. Moreover, the concept of an "idea" should be defined broadly as the residual locus of uncopyrightable elements in a work, rather than a cohesive concept that attempts to definitively pin down the "idea" behind that individual work. Taking this two-pronged solution would thus both recognize visual and photographic work's unique niche within copyright as well as align these forms of art with copyright's law's ultimate objective of authorship protection.

* J.D. Candidate, Class of 2017, New York University School of Law; Ph.D. Classics, University of Cambridge; M.Phil. Classics, University of Cambridge; B.A. Classics and History, McGill University. I am particularly grateful to Amy Adler for introducing me to the fascinating overlap between copyright law and the visual arts, and to the members of the JIPEL Notes Program for their helpful comments throughout the writing process.

INTRODUCTION.....	279
I. ANALYTICAL FRAMEWORKS.....	281
A. <i>What Are We Looking For? How Courts Determine Copyrightability</i>	281
B. <i>The Idea-Expression Dichotomy: Rationales and Application</i>	283
1. <i>Encouraging Authorship</i>	284
2. <i>The Originality Requirement</i>	286
3. <i>Baker-Selden and the Application of the Idea-Expression Dichotomy</i>	287
C. <i>Aesthetic Theory and Originality</i>	288
II. VISUAL WORKS AND THE PROBLEMS OF COPYRIGHT LAW	291
A. <i>Applying the Substantial Similarity Test</i>	292
1. <i>Defining Ideas</i>	294
2. <i>Expression as Idea: Substantial Similarity and the Merger Doctrine</i>	295
3. <i>Idea as Expression: The Transparency of Images</i>	297
B. <i>Beyond images</i>	299
III. REFRAMING COPYRIGHT IN PHOTOGRAPHY (AND BEYOND)	302
A. <i>Locating Originality in Photography and Realism</i>	303
1. <i>Administrability</i>	303
2. <i>Hindering Authorship: Pictorialism and the Goals of Copyright</i>	304
3. <i>Aligning Photography Copyright with Other Visual Images</i>	305
B. <i>Reimagining Ideas</i>	306
CONCLUSION	309

INTRODUCTION

The law has an uncomfortable relationship with visual images. In part, this is simply a matter of familiarity: lawyers and judges are generally trained to read and write much more thoroughly than to see or hear, and tend to feel much more at ease analyzing textual than non-textual works.¹ However, the divide in law between the textual and the visual seems to run deeper than that. According to the standard account, the former is the vehicle through which reason, the paramount legal virtue, finds expression, while the latter is associated instead with irrationality

¹ See, e.g., *Martin v. City of Indianapolis*, 192 F.3d 608, 610 (7th Cir. 1999) (opening an opinion centered on the destruction of a sculpture by stating that “We are not art critics, do not pretend to be and do not need to be to decide this case.”)

and emotion.² Although this is, of course, a gross oversimplification, difficulties regularly arise when courts and scholars attempt to venture outside of the realm of texts. The field of copyright is no exception. One of copyright's most fundamental concepts—the idea/expression dichotomy, which encapsulates the notion that copyright protection does not extend to the idea of a work, only to its expression of that idea—has proven most difficult to apply to non-textual works.

This note will focus on the especially problematic treatment of photography to highlight fundamental problems with copyright law's conception of "ideas" and expression in general, and with their application to visual works in particular. Part I will lay out the relevant analytical frameworks, beginning with the dominant method for analyzing copyrightability: the substantial similarity test.³ It will then provide an overview of the elements of works categorically excluded from copyright protection, and the reasons for doing so, followed by an explanation of aesthetic theories of originality and realism and the important ways in which these diverge from courts' analysis of images. Part II will explore two of the most important shortcomings in courts' analyses of visual images, namely a poor grasp of aesthetic theory and the failure to articulate a viable theory of ideas applicable to visual works. It will show how these deficiencies have led to the misapplication of the frameworks described in Section I with respect to photographic works, yielding analyses and outcomes that are both inconsistent and in violation of fundamental precepts of copyright law. Part III will argue that improvements to each of these problems can be found within the current framework of copyright law. Firstly, we should locate the originality of photographs and other "realistic" images in their expression, rather than their subject matter, bringing analysis of these works in line with that of other images and lending greater clarity to the question of how to

² For a discussion of the internalization of this division not only in legal thinking, but in western culture in general, see Costas Douzinas & Lynda Nead, *LAW AND THE IMAGE: THE AUTHORITY OF ART AND THE AESTHETICS OF LAW*, 1, 6-9 (Costas Douzinas & Lynda Nead eds., 1999) (tracing it back to Plato, who distinguished art and poetry from philosophy and reason); in the context of First Amendment law, see Amy Adler, *The Art of Censorship*, 103 W. VA. L. REV. 205, 210 (2000) ("[T]he First Amendment offers greater protection to speech that is verbal rather than visual. The preference for text over image surfaces in a variety of places in First Amendment thinking. It is, however, a peculiar preference: it is often assumed and rarely explained. I know of no scholarship that addresses it directly.") (footnotes omitted). For an example from case law of images as irrational, see *W. Va. State Bd. of Educ. v. Barnette*, 319 U.S. 624, 632 (1943) (describing images as "shortcuts from mind to the mind").

³ I acknowledge that the substantial similarity test, strictly speaking, is concerned with infringement rather than copyrightability as such. However, as will be discussed below, the first step of this test requires identifying the copyrightable aspects of the first work, and it is in this context that the great majority of courts' discussions of copyrightability arise.

assess the creativity of photography. Secondly, we should recognize “idea” as a copyright term of art designating a residual category of uncopyrightable elements, rather than a cohesive concept. The attempt affirmatively to identify the “idea” behind an individual work is therefore misguided, and should be supplanted by the application of other copyright principles. Doing so will help prevent the blurring of the line between protectable and unprotectable expression described in Section II.

I

ANALYTICAL FRAMEWORKS

A. *What Are We Looking For? How Courts Determine Copyrightability*

While there are a number of requirements in order for a work to obtain copyright protection, the crux of the issue may be (and often is) boiled down to the famous statement from Justice Holmes’s opinion in *Bleistein v. Donaldson Lithographing Co.*:⁴ “Personality always contains something unique. [...] [A] very modest grade of art has in it something irreducible, which is one man’s alone. That something he may copyright unless there is a restriction in the words of the act.”⁵ And although further limitations have since been placed on the copyrightability of works—most notably a creativity threshold⁶ and additions to the list of elements that are per se ineligible for protection⁷—the continued relevance of this formulation is attested by the frequency with which it is cited. As is often the case, the seemingly simple task of separating an individual’s copyrightable contributions from the elements of her work that enter or remain in the public domain is fraught with complication. As a result, a vast quantity of copyright litigation turns on attempts to parse works and sort their elements into these two categories.

The primary mechanism through which courts analyze works for the extent of their copyright protection is the so-called “substantial similarity” test, which compares an allegedly infringing work with the original for similarity sufficiently

⁴ 188 U.S. 239 (1903).

⁵ *Id.* at 250.

⁶ See *infra* Part I.B.2.

⁷ The current Act lists the following works or elements of works as categorically ineligible for copyright protection: ideas, procedures, processes, systems, methods of operation, concepts, principles, or discoveries. 17 U.S.C. §102(b)(1994).

substantial to warrant a finding of copyright infringement.⁸ The test takes two main forms, though these overlap significantly with each other.

The “ordinary observer test,” pioneered by the Second Circuit,⁹ first dissects the constituent elements of the relevant works to determine if there is sufficient similarity between the two to infer copying in fact.¹⁰ For this step, parties may introduce expert testimony to aid the trier of fact.¹¹ Once copying has been established, the second step of the inquiry turns to whether the appropriation was unlawful. The answer to this question relies not on expert testimony or dissection, which are deemed irrelevant, but on the response of the “ordinary observer” to the respective works.¹²

The second prominent substantial similarity test was developed in the Ninth Circuit, and similarly employs a two-part analysis. In the “extrinsic” part, courts analyze objective elements of the works, such as materials and subject matter.¹³ This inquiry looks for similarity of both ideas and expression, and like the first step of the ordinary observer test, may rely on expert testimony.¹⁴ In the second step, termed “intrinsic,” the question is turned over to the jury to determine whether, taking into account the work’s “total concept and feel,”¹⁵ improper infringement has occurred.¹⁶

In both cases, courts begin with what might be termed a deconstructionist analysis, in which they examine the individual elements of the respective works for objective similarities. It is at this stage that courts using either test generally

⁸ Although infringement is technically a distinct inquiry from copyrightability, the nature of litigation means that the latter question only tends to arise in the context of a question of infringement.

⁹ *Arnstein v. Porter*, 154 F.2d 464, 468 (2d Cir. 1946) (describing the test as a response to the “ordinary lay hearer”).

¹⁰ *Id.* at 468.

¹¹ *Id.*

¹² *Id.*

¹³ *Shaw v. Lindheim*, 919 F.2d 1353, 1356 (9th Cir. 1990). The case from which this test derives characterized the two parts as inquiries into, respectively, similarity of ideas and of expression, *Sid & Marty Krofft Television Prods., Inc. v. McDonald’s Corp.*, 562 F.2d 1157, 1164 (9th Cir. 1977), but later cases have abandoned this framework in favor of simply applying an objective, followed by a subjective, analysis. See Mark A. Lemley, *Our Bizarre System for Proving Copyright Infringement*, 57 J. COPYRIGHT SOC’Y 719, 723 (2010).

¹⁴ *Krofft*, 562 F.2d at 1164.

¹⁵ *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1110 (9th Cir. 1970).

¹⁶ *Krofft*, 562 F.2d at 1164.

identify and distinguish the protectable and unprotectable elements of the work.¹⁷ Courts then proceed to assess the works based on subjective reactions to the work as a whole to determine whether infringement has occurred.¹⁸

B. The Idea-Expression Dichotomy: Rationales and Application

The distinction between a work's protectable and unprotectable aspects is typically referred to as the "idea-expression dichotomy,"¹⁹ but, as noted above, ideas are only one of several categories that the Copyright Act excludes from protection.²⁰ The terminology thus tends to elide the various rationales that underlie the decision not to afford protection to certain elements; the distinction between a work's factual and expressive content, for example, is often discussed as an example of the idea-expression dichotomy,²¹ despite relying on different principles. Since both the idea-expression and the fact-expression dichotomies come into play in courts' analyses of visual images, it will be useful to review the reasons for excluding ideas and facts from the ambit of copyright.

¹⁷ See, e.g., *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 449 (S.D.N.Y. 2006); *Cavalier v. Random House, Inc.*, 297 F.3d 815, 822-23 (9th Cir. 2002); *Swirsky v. Carey*, 376 F.3d 841, 845 (9th Cir. 2004).

¹⁸ The two steps correspond roughly to determinations of what Nimmer terms "fragmented literal similarity" (where elements are taken verbatim from another work, without necessarily being combined in a way that resembles the original) and "comprehensive nonliteral similarity" (in which nothing is copied exactly, but the essence of the work is appropriated). 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT §13.03 [A][1] (2007).

¹⁹ I use the term "idea-expression dichotomy" here for the sake of convenience to refer to the distinctions that courts draw both between ideas and their expression, and between facts and their expression. This is due to the tendency that courts have to treat these two doctrines as one. See *infra* Part I.B.3.

²⁰ 17 U.S.C. §102(b)(1994).

²¹ E.g. *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 555 (1985) ("[C]opyright's idea/expression dichotomy strikes a definitional balance between the First Amendment and the Copyright Act by permitting free communication of facts while still protecting an author's expression.") (quotation marks omitted); cf. *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 350 (1991) ("This principle, known as the idea/expression or fact/expression dichotomy, applies to all works of authorship."). See further Alan L. Durham, *Speaking of the World: Fact, Opinion and the Originality Standard of Copyright*, 33 ARIZ. ST. L. J. 791, 801 (noting the tendency of courts to treat the idea-expression and fact-expression dichotomies as the same).

1. *Encouraging Authorship*²²

If the notion that ideas are unprotectable lies at the heart of the idea-expression dichotomy, it is because of the core rationale for American copyright protection, namely the promotion of progress,²³ which is generally assumed to be best accomplished through the proliferation of works of authorship.²⁴ Giving an individual copyright ownership of an idea, in the sense of the overarching concept behind a work,²⁵ would allow her to control or prevent the production of a wide range of other original works on the same subject, unduly hindering later authors.²⁶ This desire to keep the copyright monopoly from sweeping too broadly, thereby defeating its own purpose, comes up repeatedly in laws and doctrines designed to deny protection to the fundamental building blocks of works—in addition to the Copyright Act removing from its ambit concepts and principles,²⁷ similar reasoning lies behind the prohibition on copyrighting words or short phrases²⁸ and *scènes à faire*.²⁹ Indeed, the need to ensure that these unprotectable elements remain in the public domain is so strong that any protectable expression that cannot

²² The related concern for encouraging technical and scientific innovation (manifested, for example, in the unprotectability of procedures, processes, systems, or methods of operation) does not apply to courts' analysis of visual images, and therefore will not be discussed here.

²³ U.S. CONST. art. I, § 8, cl. 8 ("The Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.").

²⁴ See, e.g., 1 PAUL GOLDSTEIN, COPYRIGHT: PRINCIPLES, LAW AND PRACTICE § 2.2.1, at 63-64 (1989) ("The aim of copyright law is to direct investment toward the production of abundant information."). For a more thoroughly theorized exploration of this assumption, see Jeanne C. Fromer, *An Information Theory of Copyright*, 61 EMORY L. J. 71 (2014).

²⁵ For analysis of copyright's definition of an "idea," see *infra* Part III.B.

²⁶ NIMMER, *supra* note 18 at §13.03[A] ("To grant property status to a mere idea would permit withdrawing the idea from the stock of materials open to other authors, thereby narrowing the field of thought open for development and exploitation."). See also Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 997 (1997); Jessica Litman, *The Public Domain*, 39 EMORY L. J. 965, 1023 (1990). Some scholars have suggested that the uncopyrightability of ideas derives instead from the fact that the level of generality with which these are defined precludes their possessing the originality required for copyright protection. E.g., Michael Steven Green, *Copyrighting Facts*, 78 IND. L. J. 919, 941 (2003) ("[A]bstract material tends to be uncreative..."). I believe that this suggestion confuses correlation with causation. While it is undoubtedly true that an abstractly formulated idea is less likely to be original than an intricately described one, it is surely not impossible. The idea of wrapping a building in fabric, for example, is an extremely simple one, but few would characterize Christo Javacheff's work as unoriginal.

²⁷ 17 U.S.C. §102(b)(1994).

²⁸ 37 C.F.R. §202.1(a) (2004).

²⁹ For discussion of the *scènes à faire* doctrine, see *infra* note 52

be separated from them loses its protection.³⁰ The canonical expression of the idea-expression dichotomy comes from *Nichols v. Universal Pictures Corp.*,³¹ a case involving two plays about romance between a Jewish and an Irish individual, and the ensuing conflict between their families. In his opinion denying the plaintiff's claim of copyright infringement, Judge Hand remarked:

Upon any work, and especially upon a play, a great number of patterns of increasing generality will fit equally well, as more and more of the incident is left out. The last may perhaps be no more than the most general statement of what the play is about, and at times might consist only of its title; but there is a point in this series of abstractions where they are no longer protected, since otherwise the playwright could prevent the use of his "ideas," to which, apart from their expression, his property is never extended.³²

The difficulty of deciding where to draw the line between uncopyrightable abstraction and protectable expression (and indeed the impossibility of formulating a rule to provide greater predictability in doing so) has been remarked upon repeatedly since the first iteration of the so-called "abstraction test."³³

As a result of its admittedly and necessarily ad hoc nature, some scholars have suggested doing away altogether with the analytical separation of ideas and expression.³⁴ However, following on a 1970 Nimmer article,³⁵ courts have continued to apply it under the rationale of encouraging speech, often citing the idea-expression dichotomy as one of the two primary limitations on copyright (the other being the fair use doctrine) that render its constraints on speech compatible

³⁰ For discussion of the merger doctrine, see *infra* Part I.C.

³¹ *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2d Cir. 1930).

³² *Id.* at 121.

³³ *Id.* ("Nobody has ever been able to fix that boundary, and nobody ever can."). The term "abstraction test" is used here as a matter of convention; for criticism of Judge Hand's statement being characterized as a "test," see, for example, *Nash v. CBS, Inc.*, 899 F.2d 1537, 1540 (7th Cir. 1990) ("Sometimes called the 'abstractions test,' Hand's insight is not a 'test' at all. It is a clever way to pose the difficulties that require courts to avoid either extreme of the continuum of generality. It does little to help resolve a given case. . .").

³⁴ E.g., Robert Yale Libott, *Round the Prickly Pear: The Idea-Expression Fallacy in a Mass Communications World*, 14 UCLA L. REV. 735 (1967).

³⁵ Melville B. Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 UCLA L. REV. 1180, 1186-1204 (1970).

with the First Amendment.³⁶ It is debatable whether this protection remains adequate to vindicate free speech rights. Just as the idea-expression dichotomy and fair use doctrine were necessary to accommodate copyright and First Amendment rights that had both expanded significantly since their inception, so, some scholars have argued, has the continued expansion of these rights since Nimmer's article necessitated further protection of speech.³⁷ Nevertheless, the idea-expression dichotomy continues to be treated as a constitutionally necessary constraint on the copyright monopoly, and therefore retains a central place in copyright jurisprudence.

2. *The Originality Requirement*

Nearly as central to American copyright jurisprudence as is encouraging copious production of works of authorship is the requirement that these works be original (that is, independently created and displaying some degree of creativity),³⁸ termed by *Feist* the *sine qua non* of copyright.³⁹ This demand is derived from the Copyright Clause's specification that it protects works of *authorship*,⁴⁰ which was determined early on to mean works that contained some modicum of originality,⁴¹ however small.⁴² A line of twentieth century cases held that mere labor on the part of the creator was sufficient to assure copyright protection for his work, but this so-called "sweat of the brow" reasoning was emphatically rejected in *Feist*.⁴³ As applied to the "idea-expression" dichotomy, this accounts for the Copyright Act

³⁶ That is, the constraint on authors that would result from the ability to copyright ideas would unconstitutionally limit speech. *Golan v. Holder*, 609 F.3d 1076, 1091 n.9 (10th Cir. 2010); *Eldred v. Ashcroft*, 537 U.S. 186, 219-20 (2003). Cf. *Harper & Row, Publishers., Inc. v. Nation Enters.*, 471 U.S. 539, 556 (1985).

³⁷ E.g., Neil Weinstock Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1 (2001).

³⁸ *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 350 (1991) ("This principle, known as the idea/expression or fact/expression dichotomy, applies to all works of authorship.").

³⁹ *Id.* at 345.

⁴⁰ U.S. CONST. art. I, § 8, cl. 8 (emphasis added).

⁴¹ E.g., *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 58-60 (1884); *Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

⁴² The now-canonical formulation regarding the required degree of originality comes from *Feist*, 499 U.S. at 358-59 ("Originality requires only that the author make the selection or arrangement independently (*i.e.*, without copying that selection or arrangement from another work), and that it display some minimal level of creativity. Presumably, the vast majority of compilations will pass this test, but not all will. There remains a narrow category of works in which the creative spark is utterly lacking or so trivial as to be virtually nonexistent.").

⁴³ *Id.* at 352-56.

refusing protection for discoveries⁴⁴ and the related well-established tenet that facts are similarly uncopyrightable,⁴⁵ since neither is original to the discoverer, he can claim no ownership over it.⁴⁶ Only the author's expression of those facts, insofar as it is creative, may claim copyright protection.⁴⁷

3. Baker-Selden and the Application of the Idea-Expression Dichotomy

The application of the idea-expression dichotomy in American law is generally traced to *Baker v. Selden*,⁴⁸ in which the Supreme Court ruled that a bookkeeping system represented an idea that could not be copyrighted; only the author's individual description of that system could merit protection.⁴⁹ The Court added to this the corollary principle that became known as the merger doctrine, namely, that when an idea is capable of only one or a limited number of expressions, and unprotectable ideas and protectable expressions "merge" to such an extent that they are neither physically nor conceptually separable, the entirety becomes uncopyrightable.⁵⁰ In *Baker*, the charts through which the bookkeeping system was implemented were deemed uncopyrightable, since they represented the system's only possible expression, so granting a copyright to the charts would result in allowing the system itself to be copyrighted through the back door.⁵¹ Throughout the late nineteenth and twentieth centuries, the scope of the idea-expression dichotomy, and its application via the merger doctrine, expanded to apply to everything from jewelry,⁵² to computer programs,⁵³ to (increasingly) the visual arts.⁵⁴

⁴⁴ 17 U.S.C. §102(b)(1994).

⁴⁵ 37 C.F.R. §202.1(a) (2004).

⁴⁶ *E.g.*, *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 347-48 (1991); *Harper & Row, Publishers., Inc. v. Nation Enters.*, 471 U.S. 539, 556 (1985); Durham, *supra* note 21, at 802 ("The fact/expression dichotomy is closely tied to the concept of 'originality.'"). For an argument that looks to ground the (lack of) copyright protection for facts not in their lack of originality, but in a weighting of transaction costs, see Green, *supra* note 26, at 951-57. However, this theory has not received support in the case law.

⁴⁷ *Harper & Row*, 471 U.S. at 557-58.

⁴⁸ 101 U.S. 99 (1879).

⁴⁹ *Id.* at 103. §102(b) of the Copyright Act ostensibly represents the codification of this case, though the statute adds a number of categories to those discussed in *Baker*.

⁵⁰ NIMMER, *supra* note 18, at §13.03[B][3][a].

⁵¹ 101 U.S. at 104.

⁵² *Herbert Rosenthal Jewelry Corp. v. Kalpakian*, 446 F.2d 738, 742 (9th Cir. 1971) (finding that, because only a very limited number of expressions was possible of a pin in the shape of a bee, the idea and expression merged, rendering the design uncopyrightable).

Baker likewise set the stage for the lack of differentiation between the idea-expression and fact-expression dichotomies: because its discussion was concerned primarily with identifying the respective spheres of patent and copyright, it lumped together inventions (such as the system in question) that are too broad to be afforded copyright protection, with discoveries, which lack the requisite originality.⁵⁵ The result has been a blurring of lines, with the two dichotomies, if they are distinguished, often simply assumed to rely on the same rationale – a tendency succinctly encapsulated in the statement from *Harper & Row* that “no author may copyright facts or ideas. The copyright is limited to those aspects of the work—termed ‘expression’—that display the stamp of the author’s originality.”⁵⁶

C. Aesthetic Theory and Originality

Given the centrality of the originality requirement to a work’s copyrightability, and the denial of protection (via the merger doctrine) to any work representing the inevitable expression of an idea or fact, a number of scholars have begun to call attention to the facile manner in which the concept of originality has been applied to courts’ analyses of visual images.⁵⁷ For example, in one case, the court deemed a series of drawings of birds to deserve very little copyright protection due to the images’ realism: they simply reproduced a fact (the appearance of the birds) with minimal creative input from the artist, and so did not merit protection against anything other than identical copying.⁵⁸ However, as Rebecca Tushnet pointed out, these drawings, far from being mechanical reproductions of an objective fact, represented highly stylized works that were based not on nature, but on scenes constructed according to a decidedly unnatural idiom.⁵⁹ Her extensive discussion shows, through examples from a wide range of

⁵³ *Comput. Assocs. v. Altai*, 782 F.2d 693, 707-08 (2d Cir. 1992).

⁵⁴ For in-depth discussion of the application of the idea-expression dichotomy to the visual arts, see *infra* Part II.A.2. This period also saw the development of the closely-related doctrine of *scènes à faire*, which holds that elements fundamental to the expression of a particular idea may not be copyrighted, since this too would lead to an indirect means of copyrighting the idea. The rationales for *scènes à faire*, however, are mixed, since as a practical matter any element necessary to the expression of an idea is also unlikely to be original.

⁵⁵ 101 U.S. at 102-03.

⁵⁶ 471 U.S. at 547.

⁵⁷ E.g., Rebecca Tushnet, *Worth a Thousand Words: The Images of Copyright*, 125 HARV. L. REV. 683 (2012); Christina O. Spiesel et al., *Law in the Age of Images: The Challenge of Visual Literacy*, in CONTEMPORARY ISSUES IN THE SEMIOTICS OF LAW 231, 237 (Anne Wagner et al. eds., 2005).

⁵⁸ *Franklin Mint Corp. v. Nat’l Wildlife Art Exch., Inc.*, 575 F.2d 62, 65 (3d Cir. 1978).

⁵⁹ Tushnet, *supra* note 57, at 727-29.

disciplines, how even a concept as seemingly uncomplicated as “realism” turns out to be a highly culturally-specific construct,⁶⁰ a fact to which courts have tended to be blind.⁶¹ Similarly, photographs do not so much reproduce reality as they do interpret it. Insofar as all of the elements of a picture are in focus, for example, a photograph creates an image that no eye could see.⁶² Choice of color palette may appear to increase realism even when its hues do not correspond to the viewer’s experience.⁶³ Other decisions such as length of exposure, contrast, or the use of filters ensure that two photographs of the same subject, taken at the same place and time, may look completely different from one another. The photographs below, for example, both capture the same subject at the same time. However, the use of a long exposure in the first results in the water taking on a smooth, serene appearance, in contrast with the sharply-defined choppy waves captured by the shorter exposure in the second.

⁶⁰ *Id.* at 728 (noting, for example, at least six distinct “realisms” that developed over the course of the twentieth century, or the many ways in which scientific photographs—specifically aimed at a faithful, mechanical reproduction of reality—require the photographer to make a significant number of aesthetic choices); cf. Joel Snyder, *Picturing Vision*, 6 Critical Inquiry 499, 500 (1980) (noting “the facile use by many art historians and critics of the visual arts of such weighty but indeterminate expressions as ‘reality’ or ‘visual reality...’”). See generally Roman Jakobson, *On Realism in Art, in Language in Literature* 19, 21 (Krystyna Pomoroska & Stephen Rudy eds., 1987) (“The methods of projecting three-dimensional space onto a flat surface are established by convention; the use of color, the abstracting, the simplification, of the object depicted, and the choice of reproducible features are all based on convention. It is necessary to learn the conventional language of painting in order to ‘see’ a picture, just as it is impossible to understand what is said without knowing the language. This conventional, traditional aspect of painting to a great extent conditions the very act of our visual perception.”).

⁶¹ For other examples of courts according only thin copyright to “realistic” styles, see, e.g., *Satava v. Lowry*, 323 F.3d 805, 812 (9th Cir. 2003); *Yankee Candle Co. v. Bridgewater Candle Co.*, 259 F.3d 25, 36 (1st Cir. 2001) (finding that realistic representations of fruit merge with their subject matter); *Concrete Mach. Co. v. Classic Lawn Ornaments, Inc.*, 843 F.2d 600, 607 (1st Cir. 1988) (giving protection only against virtually identical copying to a realistic statue of a deer); *First Am. Artificial Flowers, Inc. v. Joseph Markovits, Inc.*, 342 F.Supp. 178, 186 (S.D.N.Y. 1972). But see *Kamar Int’l, Inc. v. Russ Berrie & Co.*, 657 F.2d 1059, 1061 (9th Cir. 1981) (rejecting the notion that realistic depictions of live animals are not copyrightable).

⁶² Snyder, *supra* note 60, at 501-02.

⁶³ *Id.* at 505.



Figure 1: Drew Zweller, *Camera Function Part 1*, DREW ZWELLER DIGITAL MEDIA (May 5, 2016), available at <http://drewzweller.weebly.com/blog/camera-function-part-1>.

This observation has important implications for how the concept of originality should be analyzed, and how the merger doctrine should be applied. With respect to the former, if a “realistic” style represents an aesthetic choice like any other, then denial or limitation of copyright protection on that ground seems particularly misguided. As for the merger doctrine, the above examples show how courts often overestimate the “inevitability” of the expression of a particular fact or idea, and therefore may apply the merger doctrine in cases to which it is inappropriate.⁶⁴

These problems are related to the so-called “transparency” of realistic images, which courts deem to embody their subject in an unmediated manner (as distinct from, for example, verbal descriptions, in which the subject is accepted as being filtered through the author’s voice).⁶⁵ In this view, the images, particularly

⁶⁴ It should be noted that this observation is not limited to situations in which the merger doctrine is explicitly invoked and copyright denied; it can also lead courts to according such a thin copyright that the effect is nearly the same as refusing protection. Thus, for example, the Third Circuit could truthfully claim that “this Court has never found an instance in which a completely aesthetic expression merged into an idea.” *Kay Berry, Inc. v. Taylor Gifts*, 421 F.3d 199, 209 (3d Cir. 2005). But in *Franklin Mint*, 575 F.2d at 65, the copyright on the paintings of birds was so thin, it extends only to a precise reproduction of them.

⁶⁵ E.g., Dominic McIver Lopes, *The Aesthetics of Photographic Transparency*, 112 MIND 433, 440 (2003); Kendall L. Walton, *Transparent Pictures: On the Nature of Photographic*

photographs, as the ultimate example of realism, do not require interpretation, but simply “are” their subjects.⁶⁶ This notion is reflected in the case law, beginning with the fountainhead of American photography jurisprudence, *Burrow-Giles*. In that case, the defendant copied a photograph that the plaintiff had taken of Oscar Wilde, arguing that a photograph, as a mechanical reproduction of reality, did not contain the level of originality required for copyright protection. In a move that remains characteristic of photography opinions, the Court did not seem to contest the defendant’s characterization of photography. Instead, it listed a number of elements that might demonstrate originality, it concentrated overwhelmingly on the selection and arrangement of the subject matter,⁶⁷ and not the fundamental relationship between the photograph as medium and its subject matter.⁶⁸ The continued influence of this mode of analysis can be seen in the focus on elements such as arrangement⁶⁹ and perspective/camera angle⁷⁰ as the relevant factors in determining a photograph’s originality.

II

VISUAL WORKS AND THE PROBLEMS OF COPYRIGHT LAW

The strands of doctrinal confusion discussed above—between ideas and facts, between copyright rationales, and between works and their subject matter—

Realism, 11 CRITICAL INQUIRY 246, 246 (1984). For discussion and criticism of this characterization, see, e.g., Tushnet, *supra* note 57 at 688-95.

⁶⁶ See Justin Hughes, *The Photographer’s Copyright – Photograph as Art, Photograph as Database*, 25 HARV. J. LAW & TECH. 339, 345 (2012); Tushnet, *supra* note 57, at 700-01 (noting the decision in *Harris v. Scott*, 550 U.S. 372, 372 (2007), which treated video footage of a car chase as giving them direct access to reality); cf. Walton, *supra* note 65, at 252 (“[W]e see, quite literally, our dead relatives themselves when we look at photographs of them.”).

⁶⁷ *Burrow-Giles Lithographic Co. v. Saroni*, 111 U.S. 53, 60 (1884) (“[Saroni] gave visible form by posing the said Oscar Wilde in front of the camera, selecting and arranging the costume, draperies, and other various accessories in said photograph, arranging the subject so as to present graceful outlines, arranging and disposing the light and shade, suggesting and evoking the desired expression. . .”).

⁶⁸ E.g., Hughes, *supra* note 66 at 356 (“Burrow-Giles maintains an extremely mechanical view of photography.”); Christine Haight Farley, *The Lingering Effects of Copyright’s Response to the Invention of Photography*, 65 U. PITT. L. REV. 385, 428 (2004) (“[T]he Court located authorship not in the act of capturing the image or in the post-photograph manipulation that many art photographers were doing, but in the preparation.”).

⁶⁹ E.g., *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 450 (S.D.N.Y. 2006); *Rogers v. Koons*, 960 F.2d 301, 307 (2d Cir. 1992).

⁷⁰ E.g., *Mannion*, 377 F. Supp. 2d at 452; *Rogers*, 960 F.2d at 307; *Kisch v. Ammirati & Puris Inc.*, 657 F. Supp. 380, 384 (S.D.N.Y. 1987).

converge in courts' analyses of visual works, and photographs in particular. This section will show how the substantial similarity test has been applied to visual works in such a way as to collapse the idea-expression dichotomy, leading courts to take one of two diametrically opposite, but equally questionable, approaches. Some courts, relying on a simplistic understanding of visual representation, have denied protection to arguably original expression by assimilating the expression into its underlying "idea" through the merger doctrine. Others, conversely, have used a misunderstanding of copyright's "ideas" doctrine to ascribe to uncopyrightable elements the quality of protectable expression, thereby extending copyright protection beyond its proper bounds. The result is a body of case law that is doctrinally unsound, and contributes to the inconsistency (and therefore unpredictability) of copyright decisions.

This section will further argue that many of these problems are not, as is generally assumed, merely the result of applying doctrine appropriate to textual analysis to a fundamentally different category of works. Rather, they expose weaknesses in core copyright concepts in a way that pertains equally to textual works, demonstrating the necessity of revising these doctrines.

A. Applying the Substantial Similarity Test

Copyright's analysis of visual works is built on a tension that derives from the law's fundamental uncertainty regarding the nature of images and how they ought to be analyzed. On the one hand, an image, like any other work, by definition consists of its various elements, some of which may not be copyrightable. But there is also a sense in which images, unlike other mediums, strike their audience all at once, protectable and unprotectable parts alike.⁷¹ Insofar as this overall impression represents an integral part of an image's appeal, it too must be taken into account in analyzing the work's copyrightable expression.

⁷¹ This is reflected in the frequent reference in opinions to the immediate impression that the images in question create on a viewer. *E.g.*, *TY, Inc. v. GMA Accessories, Inc.*, 132 F.3d 1167, 1169 (7th Cir. 1997) ("A glance at the first picture shows a striking similarity between the two. . ."); *Steinberg v. Columbia Pictures Indus., Inc.*, 663 F. Supp. 706, 712 (S.D.N.Y. 1987) (noting a striking similarity between two illustrations "even at first glance," but ascribing this initial impression to unprotectable elements such as the bird's-eye view of Manhattan or the band of blue to represent the sky); *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1110 (9th Cir. 1970) (noting that "the remarkable similarity . . . is apparent even to a casual observer," and supporting this statement with reference to a number of unprotectable elements shared by the two works).

The tension between the particular details of the work, and the overall impression that these elements, taken together as a whole, create—somehow different from the elements individually—is reflected in the two steps of both the Second and Ninth Circuits’ substantial similarity tests.⁷² In the deconstructionist first step, courts employ an approach whose application to images predates either test, and can be traced back to the first American photography copyright case. In *Burrow-Giles*, the court enumerated the individual elements of the photograph that it deemed to be original to the author, and therefore merit protection.⁷³ A century later, the Southern District of New York applied a similar method in *Steinberg v. Columbia Pictures*, in which a movie poster was alleged to have impermissibly copied an illustration from *The New Yorker*.⁷⁴ Noting the similarity in everything from the depiction of intersecting two-way streets and water towers, to the facades and other details of the buildings, the court held that the poster constituted an infringing image.⁷⁵ Another Southern District case two decades after *Steinberg* reinforced the continued importance of the deconstructionist mode of analysis, applying it once again to photography. *Mannion v. Coors* involved two photographs of a man in a white t-shirt wearing jewelry, set against a blue sky. En route to determining that the second photograph had infringed the first, the court went item by item through the stylistic and compositional similarities of two photographs, noting choices such as angle, lighting, and selection and arrangement of jewelry on the subject.⁷⁶

In the second step of both substantial similarity tests, courts compare the works as a whole, an exercise which is supposed to be mindful of, but not limit itself to, the copyrightable elements identified in step one.⁷⁷ This holistic consideration has sometimes been analogized to the so-called “compilation copyright,”⁷⁸ explained most famously in *Feist*, whereby even a work composed entirely of uncopyrightable elements may be eligible for copyright protection of

⁷² See *supra* Part I.A.

⁷³ 111 U.S. at 60.

⁷⁴ 663 F. Supp. at 708.

⁷⁵ *Id.* at 712-13.

⁷⁶ *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 462-63 (S.D.N.Y. 2006).

⁷⁷ Cf. *Mannion*, 377 F. Supp. 2d at 462 (“Key elements . . . may not be copyrightable in and of themselves, but their existence and arrangement in this photograph indisputably contribute to its originality.”).

⁷⁸ *E.g.*, *Boisson v. Banian, Ltd.*, 273 F.3d 262, 271, 272 (2d Cir. 2001); *Matthews v. Freedman*, 157 F.3d 25, 28 (1st Cir. 1998).

the whole if these elements are arranged in a sufficiently original manner.⁷⁹ As applied to images, this may involve comparing the cumulative effect created by elements such as compositional layout,⁸⁰ artistic style,⁸¹ typeface,⁸² and even the sentiments conveyed by the respective works.⁸³

The exact nature of the logical leap between dissection of the work into protectable and unprotectable elements, and consideration of its “look and feel” has never been fully explained,⁸⁴ and numerous scholars have noted the problems that it raises. Most notably, the dual requirements of taking into account the protectability of the original work’s elements, but not limiting the comparison to these, are at best in tension with each other, and at worst constitute mutually contradictory demands.⁸⁵ Moreover, in wrestling with the paradox of a copyrightable whole composed entirely of unprotectable parts, courts have sometimes read this copyrightability back into the elements themselves, leading to suggestions that an author may control copying of aspects such as artistic style.⁸⁶ At the center of this interplay and confusion between the specific and the general, the protectable and the unprotectable, is the definition of an “idea.”

1. *Defining Ideas*

The copyright concept of an idea, as distinct from its expression by an author, appears in two guises. First, the term may be used to describe what I will term “micro ideas” – individual components of a work that are too general to

⁷⁹ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 350 (1991) (“Facts, whether alone or as part of a compilation, are not original and therefore may not be copyrighted. A factual compilation is eligible for copyright if it features an original selection or arrangement of facts...”).

⁸⁰ *E.g.*, *Steinberg v. Columbia Pictures Indus., Inc.*, 663 F. Supp. 706, 712 (S.D.N.Y. 1987); *Boisson*, 273 F.3d at 269-70.

⁸¹ *E.g.*, *Steinberg*; 663 F. Supp. at 712; *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1110 (9th Cir. 1970).

⁸² *E.g.*, *Steinberg*, 663 F. Supp. at 712; *Roth Greeting Cards*, 429 F.2d at 1110; *cf. Boisson*, 273 F.3d at 271 (“[W]e hesitate to say that letter shapes are unprotectible in this context...”).

⁸³ *Roth Greeting Cards*, 429 F.2d at 1110.

⁸⁴ Tushnet, *supra* note 57, at 718 (“There is a sort of magic by which unprotectable parts together become protected.”); *cf. Pamela Samuelson, Essay: A Fresh Look at Tests for Nonliteral Copyright Infringement*, 107 NW. U.L. REV. 1821, 1830-34 (2013).

⁸⁵ NIMMER, *supra* note 18, at §13.03[A][1][c] (“[T]he touchstone of ‘total concept and feel’ threatens to subvert the very essence of copyright, namely the protection of original expression.”); Tushnet, *supra* note 57, at 719.

⁸⁶ *E.g.*, *Steinberg*; 663 F. Supp. at 712; *Roth Greeting Cards*, 429 F.2d at 1110.

receive copyright protection, such as a stock character in a play.⁸⁷ More importantly, though, courts will inquire as to a work's "macro idea," or the "idea" of the work as a whole. This form of "idea" is based on the famous abstraction test laid out in *Nichols*, which states that a work may be described in increasingly broad terms, until at some point what is being described is no more than the "idea" of the work, which may not be copyrighted.⁸⁸ In other words, the idea of a work is, as *Mannion* noted in the context of photographs, simply a description of that work in general terms.⁸⁹ This definition is problematic for a number of reasons, beginning with the ways in which it interacts with other aspects of courts' analyses of images.

2. *Expression as Idea: Substantial Similarity and the Merger Doctrine*

As outlined above,⁹⁰ the first step in the substantial similarity test for visual works involves a close analysis of the various elements of the original work to determine which are copyrightable, a process that encourages (or even necessitates) a highly deconstructionist approach to the analysis of visual images. Combined with the tendency to define the work's "idea" as a combination of its various elements, this has resulted in courts defining ideas with increasing specificity. Thus, in several cases, we find the court slipping imperceptibly between a detailed analysis of a work's elements and a definition of the work's idea that simply amalgamates those elements.⁹¹ This slippage, in turn, helps to

⁸⁷ *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2d Cir. 1930).

⁸⁸ *Nichols*, 45 F.2d at 121 ("Upon any work, and especially upon a play, a great number of patterns of increasing generality will fit equally well, as more and more of the incident is left out. The last may perhaps be no more than the most general statement of what the play is about, and at times might consist only of its title; but there is a point in this series of abstractions where they are no longer protected, since otherwise the playwright could prevent the use of his 'ideas,' to which, apart from their expression, his property is never extended.").

⁸⁹ *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 460 (S.D.N.Y. 2006) ("Other copyright cases that have referred to the 'idea' of a photograph also used 'idea' to mean a general description of the subject or subject matter.") (citing *SHL Imaging, Inc. v. Artisan House, Inc.*, 117 F. Supp. 2d 301, 314 (S.D.N.Y. 2000); *Andersson v. Sony Corp.*, 1997 WL 226310, at *3 (S.D.N.Y. May 2, 1997); *Gentieu v. Tony Stone Images/Chicago, Inc.*, 255 F. Supp. 2d 838, 849 (N.D. Ill. 2003)).

⁹⁰ See *supra* Parts I.A; II.A.

⁹¹ Perhaps the most succinct encapsulation of this tendency comes from *Kaplan v. Stock Market Photo Agency, Inc.*, which shifts back and forth within a single sentence between referring to the same elements as subject matter and idea. 133 F. Supp. 2d 317, 323 (S.D.N.Y. 2001) ("The subject matter of both photographs is a businessperson contemplating a leap from a

explain the increasing application of the merger doctrine to visual works, since the more specific an idea's description, the more "inevitable" the expression of that idea becomes.⁹² To take as an example arguably the most influential work of art of the twentieth century, the idea behind Marcel Duchamp's *Fountain* could be described as "a challenge to the definition of art through the elevation of a mundane object to the status of artwork," and most would agree that this concept is capable of a wide range of expressions. However, if a court, after analyzing the work, found that the idea behind Duchamp's fountain was "laying a urinal on its back, inscribing 'R. Mutt 1917' on its lower left-hand side, and calling it a fountain," it might well be said that the idea is capable of only a limited number of expressions,⁹³ and that the two therefore merge. If this hypothetical seems exaggerated, it is not significantly different from courts' practice of defining works' ideas by reference to their numerous constituent elements, which were merely listed and analyzed.⁹⁴ Indeed, it is instructive to contrast the difference in the level of detail recent cases have deemed to be encompassed by the work's "idea" with the extremely broad description applied to the plaintiff's work in the seminal *Nichols* case ("a comedy based upon conflicts between Irish and Jews, into which the marriage of their children enters. . .").⁹⁵ Viewed in this light, the increasing willingness to find that expression merges with idea is unsurprising.

tall building onto the city street below. As the photograph's central idea, rather than Kaplan's expression of the idea, this subject matter is unprotectable in and of itself."); *see also, e.g., Roth Greeting Cards*, 429 F.2d at 1110-11.

⁹² Cf. Jarrod M. Mohler, *Toward a Better Understanding of Substantial Similarity in Copyright Infringement Cases*, 68 U. CIN. L. REV. 971, 991 (2000) ("The court broke down the works into such minute segments that it could not have possibly found originality..."); Jane C. Ginsburg, *Four Reasons and a Paradox: The Manifest Superiority of Copyright over Sui Generis Protection of Computer Software*, 94 COLUM. L. REV. 2559, 2561 (1994) ("There is a danger, but not only with respect to computer programs, that courts, in seeking to distinguish the public domain 'idea' from the protected 'expression,' will so 'dissect' the work as to classify all its elements as unprotectable."). The phenomenon was alluded to in *Krofft*, though without acknowledgement of the corresponding problems that arise from it. 562 F.2d at 1168 n. 10. ("If, in describing how a work is expressed, the description differs little from a simple description of what the work is, then ideas and expression coincide.").

⁹³ The range of which might, incidentally, be thought to be conveyed by the various urinal shapes that Duchamp employed for the several iterations of that sculpture.

⁹⁴ *See, e.g., Satava v. Lowry*, 323 F.3d 805, 812 (9th Cir. 2003) (describing the idea of the sculpture in question as "lifelike glass-in-glass sculptures of single jellyfish with vertical tentacles."); *Matthews*, 157 F.3d at 27 (finding that the idea for a t-shirt's design was a phrase indicating that the t-shirt's purchase represents someone's love of the donee, combined with the use of childish lettering and emblems reminiscent of a location).

⁹⁵ *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 122 (2d Cir. 1930).

The problem is not novel, and indeed the impossibility of pinpointing where to draw the line between idea and expression has been noted ever since the abstractions test was first introduced.⁹⁶ However, courts have often taken this difficulty as a license to declare the correct standard by fiat, so that even when opinions recognize the potential for the concept to be described at different levels of generality, they often give little reason for choosing to apply one over the other.⁹⁷

3. *Idea as Expression: The Transparency of Images*

The approach outlined above describing works' ideas has resulted in a particular paradox in the case of photographs and other "realistic" visual images. As noted, courts have tended to view such works as transparent reproductions of their "factual" subject matter,⁹⁸ according them the thinner copyright that applies to fact-based (and therefore, implicitly, less original) works.⁹⁹ Such a position creates difficulties, since photographs are generally considered copyrightable,¹⁰⁰ even though merely reproductive works lack the creativity necessary for copyright protection.¹⁰¹ The requisite originality must therefore be sought not in the photographic medium, but in the photograph's subject – that is, its "idea." The

⁹⁶ *Id.* at 121 ("[T]here is a point in this series of abstractions where they are no longer protected, since otherwise the playwright could prevent the use of his 'ideas,' to which, apart from their expression, his property is never extended. Nobody has ever been able to fix that boundary, and nobody ever can.").

⁹⁷ See, e.g., *Kaplan v. Stock Market Photo Agency, Inc.*, 133 F. Supp. 2d 317, 324 (S.D.N.Y. 2001) (stating without any further explanation that "the concept of Kaplan's photograph 'may' be expressed more generally [...] however, this idea is clearly not the most accurate characterization of the concept embraced by the photographs at issue in this case.").

⁹⁸ Cf. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 59 (1884) (acknowledging, without deciding, that the characterization of a photograph as mere reproduction of its subject matter "may be true in regard to the ordinary production of a photograph, and that in such case a copyright is no protection.").

⁹⁹ *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 350 (1991). ("This Court has long recognized that the fact/expression dichotomy limits severely the scope of protection in fact-based works."). For courts' dim view of the originality represented by photographs, see, e.g., *Steinberg v. Columbia Pictures Indus., Inc.*, 663 F. Supp. 706, 712 (S.D.N.Y. 1987) ("[E]ven a photograph may be copyrighted.").

¹⁰⁰ Eva E. Subotnik, *Originality Proxies: Toward a Theory of Copyright and Creativity*, 76 BROOKLYN L. REV. 1487, 1489 (2011) (footnote omitted). ("[P]hotographs—especially those of human subjects—have long been deemed, on the highest authority, a worthy subject for copyright protection.").

¹⁰¹ *Meshrwerks v. Toyota Motor Sales U.S.A., Inc.*, 528 F.3d 1258, 1264 (10th Cir. 2008).

law's answer to this problem has been to adopt a "pictorialist" approach to analyzing photographs,¹⁰² in which courts focus especially on the author's hand in creating the scene depicted, and the originality involved in this creation.¹⁰³

It is here that the failure to distinguish between ideas and facts becomes important. By locating the originality of photographs in their ideas, and then declaring these copyrightable, these decisions are implicitly based on the assumption that what prevents ideas from receiving copyright protection is their lack of originality. Thus, in *Burrow-Giles*, the Court found that the photograph displayed the requisite originality, but did so largely on the basis of factors (composition, subject's pose, etc.) that describe *what* is depicted, not *how* it is depicted,¹⁰⁴ and that therefore belong quintessentially to the realm of ideas. This line of analysis was taken to its logical extreme in *Mannion*, in which the court came to the conclusion that in the case of photographs, the distinction between idea and expression collapses, but that both may be protected due to their originality.¹⁰⁵ As a result, the photographer in that case was able to copyright not only his photograph, but the scene itself, and prevent others from recreating and photographing a similar scene.¹⁰⁶

The trouble with this reasoning is that, as discussed above, the originality of an idea has no bearing on its copyrightability. The resulting problem of courts' granting protection to ideas is the reverse of the over-application of the merger doctrine: both problems spring from the difficulty that courts have in applying the notion of an "idea" to a visual work, but pull in different directions. Whereas the merger doctrine defines ideas so specifically that they subsume the work's expressive elements, the *Mannion*-style analysis treats them so much like expression that they end up being regarded as such. The former restricts copyright by removing protection even from seemingly original expression, while the latter greatly expands copyright by removing ideas from the public domain, thereby effecting the overbroad sweep of copyright that copyright law sought to prevent by denying protection to ideas.

¹⁰² For discussion of the term "pictorialism" as applied to photographs, see Farley, *supra* note 68, at 421; Teresa M. Bruce, *In the Language of Pictures: How Copyright Law Fails to Adequately Account for Photography*, 115 W. VA. L. REV. 93, 95-96 (2012).

¹⁰³ E.g., Farley *supra* note 68, at 425-29; Hughes, *supra* note 66, at 389-96. Cf. Subotnik, *supra* note 100, at 1517-23 (discussing courts' use of a "proxy of narrative" – that is, the author's narrative of her creative process as a substitute for finding originality in the image itself).

¹⁰⁴ *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 60 (1884).

¹⁰⁵ *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 454-61 (S.D.N.Y. 2006).

¹⁰⁶ *Id.*

The problem of copyrighting ideas is not limited to courts granting protection to “macro ideas.” One effect of the “look and feel” test finding copyrightability in collections of otherwise unprotectable elements has been for courts to sometimes read the copyrightability of the whole back into the individual elements. In *Steinberg*, for example, the court, in determining the copyright of the original illustration, implied that the plaintiff’s artistic style was copyrightable, committing the same error of protecting an idea based on its originality.¹⁰⁷ *Roth*, finding that the plaintiff’s cards were copyrightable, indicated that elements such as arrangement of words and characters’ emotions were protectable.¹⁰⁸ As with “macro ideas,” such reasoning fails to recognize that, whatever the originality of the elements in question (relatively high in *Steinberg*, extremely low in *Roth*), it should not overcome their unprotectability, which is based on different grounds.

B. Beyond images

Copyright originally developed as a system of protection for written works. This fact is regularly invoked by both courts and scholars to explain the difficulty encountered in applying copyright principles to non-textual works:¹⁰⁹ it is almost a truism that textual and non-textual works differ in certain fundamental respects,¹¹⁰ so it is inevitable that doctrines and modes of analysis that were tailored to the former will sometimes be a poor fit for the latter.¹¹¹ For example, both scholars and courts have regularly held up the ability to define ideas and distinguish them from

¹⁰⁷ *Steinberg v. Columbia Pictures Indus., Inc.*, 663 F. Supp. 706, 712 (S.D.N.Y. 1987) (“[O]ne can see the striking stylistic relationship between the posters, and since style is one ingredient of ‘expression,’ this relationship is significant.”).

¹⁰⁸ *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1110 (9th Cir. 1970).

¹⁰⁹ E.g., Tushnet, *supra* note 57, at 684 (“Copyright is literal. It starts with the written word as its model, then tries to fit everything else into the literary mode.”); Michael Murray, *Copyright, Originality, and the End of the Scenes a Faire and Merger Doctrines for Visual Works*, 58 BAYLOR L. REV. 779, 792-95 (2006); *Mannion*, 377 F. Supp. 2d at 458; *Warner Bros. v. Am. Broad. Co., Inc.*, 720 F.2d 231, 241 (2d Cir. 1983) (describing the tension between two doctrines “result[ing] from their formulation in the context of literary works and their subsequent application to graphic and three-dimensional works.”).

¹¹⁰ In general on the differences between images and words, see especially Christina Spiesel, *Reflections on Reading: Words and Pictures and Law*, in *LAW, MIND AND BRAIN*, 391 (Michael Freeman & Oliver R. Goodenough eds., 2009).

¹¹¹ See *Warner Bros.*, 720 F.2d at 241.

their expression as an analysis to which texts are much more amenable than other forms of expression.¹¹²

Given the broad consensus about the distinctiveness of textual works, it is surprising to note the disagreement about exactly what features distinguish texts from visual works.¹¹³ Beyond the widely-noted observation that books have a built-in order in which they are read,¹¹⁴ some have argued that text is a more inherently restrictive medium than visual artwork, and therefore more prone to merging ideas and expression.¹¹⁵ Others have claimed that it is visual artists who are more constrained by their medium.¹¹⁶ Some have seen the divide as being primarily between words, which owe their appeal to rationality, and other mediums that are less rooted in reason.¹¹⁷

Part of the explanation for the lack of consensus may be that the supposedly fundamental gap between textual and visual works is less of a chasm than is often assumed.¹¹⁸ Take the example of courts finding that certain visual representations merge ideas and expression, and are therefore uncopyrightable. As noted in Part I.C, this logic relies on a naive belief in the inevitability of certain forms of

¹¹² E.g., *Mannion*, 377 F. Supp. 2d at 458; Murray, *supra* note 109, at 791.

¹¹³ Cf. Tushnet, *supra* note 57, at 703-04 (“Images are different, courts agree. They just can’t agree what that difference is. . .”).

¹¹⁴ Even this seemingly obvious point is both less universally true than it might first seem (reference works, for example are read in whatever order the user wishes) and not unique to textual works (movies and music are two other prominent examples of categories of works that dictate the order in which they are consumed). However, it is true that this is a notable difference between textual and visual works. See Christina Spiesel, *More Than a Thousand Words in Response to Rebecca Tushnet*, 125 HARV. L. REV. F. 40, 41 (Feb. 22, 2012) (“Pictures are different from words. They are perceptually immediate, they can be vivid, and under some circumstances they can be confused with reality itself. [...] [T]heir elements are displayed in space; we read the relationships between the parts within a framing edge as they are visually bound together and related through many possible qualities (not grammatical order).”).

¹¹⁵ E.g., Murray, *supra* note 109, at 851-53.

¹¹⁶ E.g., Jay Dratler, Jr., *Distilling the Witches’ Brew of Fair Use in Copyright Law*, 43 U. MIAMI L. REV. 233, 306 (1988) (“Because it is a rare fact or idea that cannot be expressed verbally in more than a few ways, the merger doctrine may be more applicable to nonverbal works such as pictures, in which the content and expression may be inextricably intertwined.”) (footnotes omitted).

¹¹⁷ E.g., NEAL FEIGENSON & CHRISTINA SPIESEL, *LAW ON DISPLAY: THE DIGITAL TRANSFORMATION OF LEGAL PERSUASION AND JUDGMENT* xi (2009).

¹¹⁸ For a reflection on the similarities between textual and visual works in the copyright context, see Zahr Kassim Said, *Only Part of the Picture: A Response to Professor Tushnet’s Worth A Thousand Words*, 16 STAN. TECH. L. REV. 349 (2013) (focusing on the necessity in both cases of applying a sophisticated interpretative approach to expressive works).

expressions. However, those who rightly argue that these cases demonstrate a lack of imagination regarding the range of possibilities for visual expression often fail to recognize a related temptation to assume that, insofar as writers are bound by the vocabulary and structure of language, the written expression of some ideas will be “inevitable.”¹¹⁹ As Christina Spiesel has pointed out, while treating written expression as a relatively straightforward area of copyright law provides a “convenient foil” against which to compare the problematic treatment of pictures, it can have the effect of understating the flexibility and fluidity of language.¹²⁰ The meaning of a word may change depending on its context, or an author may convey her meaning through metaphor or other indirect language, or borrow words from different languages, or even invent entirely new words.

As with images, the distortion created by treating textual works as the settled core of copyright law similarly disguises fundamental problems with the definition of “ideas” in copyright law. Although both scholars and courts have long recognized the difficulty of applying the concept to visual works, this difficulty is usually contrasted with the suitability of the doctrine to the analysis of textual works.¹²¹ In that context, “idea” is generally treated as synonymous with “plot,”¹²² undoubtedly due at least in part to the enduring influence of Justice Hand’s opinion in *Nichols*, which equated the two concepts in the context of a play.¹²³ But many textual works do not have a plot, and, like visual works, resist easy application of the idea-expression dichotomy. Poetry is perhaps the best example of a literary genre in which identifying the “idea” behind a work will frequently be difficult or impossible.¹²⁴ Yet we could equally imagine the challenge of mapping the idea-expression framework onto unconventional literary works. How would a court go about deciding what “idea” lies behind *Finnegan’s Wake*? Or, for that matter, even as ubiquitous a work as the Bible? In the case of the poem or the Joycean novel, the overlap with the problems of visual image copyright is succinctly illustrated by the nearly perfect applicability to them of a judge’s statement regarding a picture:

¹¹⁹ See, e.g., Murray, *supra* note 109, at 850-51 (“[I]f you want to express the idea of darkness, you have to use the word ‘dark’ or one of a limited number of synonyms of that word.”).

¹²⁰ Spiesel, *supra* note 114, at 44.

¹²¹ E.g., *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 458 (S.D.N.Y. 2006); Murray, *supra* note 109, at 850-52.

¹²² Cf. *Warner Bros. v. Am. Broad. Co., Inc.*, 720 F.2d 231, 240 (2d Cir. 1983) (noting the difficulty of applying the notion of an “idea” to a concept other than plot).

¹²³ *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121-22 (1930).

¹²⁴ I have borrowed the apt analogy to poetry from Rebecca Tushnet, *supra* note 57, at 715.

“An artist's idea, among other things, is to depict a particular subject in a particular way.”¹²⁵

None of this is to deny that important differences exist between textual and visual works. For one, judges are generally much more comfortable analyzing texts than they are other kinds of works,¹²⁶ a fact that is sometimes admitted in opinions.¹²⁷ Moreover, several scholars have written about the different ways in which we both consume and process texts as opposed to images.¹²⁸ It is, rather, to say that several problems that are most acutely illustrated in the courts' struggles to fit visual images into the copyright framework are not limited to such cases; they are symptomatic of broader instabilities and inconsistencies that underlie fundamental copyright doctrines. Here too, though, judges confronted with texts could learn from scholarship on the legal analysis of visual works: just as there has been a growing call cautioning against judicial overconfidence in interpreting even seemingly straightforward pictures,¹²⁹ so might it be salutary to remember the pitfalls that texts present, even (or, perhaps, especially) for those as familiar with them as lawyers.¹³⁰

III

REFRAMING COPYRIGHT IN PHOTOGRAPHY (AND BEYOND)

In light of the myriad problems that have accompanied copyright law's attempt to accommodate visual works, it is unsurprising that suggestions have proliferated regarding ways to fix this area of law, some of them calling for a radical overhaul of entire areas of copyright doctrine.¹³¹ While there is substantial merit to many of these, the aims of this section are more limited, namely to

¹²⁵ *Mannion*, 377 F. Supp. 2d at 458.

¹²⁶ Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1172 (2007).

¹²⁷ *E.g.*, *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251 (1903) (“It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest and most obvious limits.”).

¹²⁸ *E.g.*, FEIGENSON & SPIESEL, *supra* note 117, at 7-9; Jennifer L. Mnookin, *The Image of Truth: Photographic Evidence and the Power of Analogy*, 10 YALE J.L. & HUMAN. 1, 1-2 (1998).

¹²⁹ *E.g.*, Alfred C. Yen, *Copyright Opinions and Aesthetic Theory*, 71 S. CAL. L. REV. 247 (1998).

¹³⁰ *Cf.* Tushnet, *supra* note 57, at 702 (“Those rules [of textual interpretation] might be, in fact, indeterminate and manipulable, but they feel predictable and rational.”).

¹³¹ *E.g.* Tushnet, *supra* note 57 (arguing that copyright in visual works should protect only against literal infringement); Amy Adler, *Why We Should Abolish Copyright Protection for Visual Art* (on file with the author) (suggesting the abolishment of copyright protection with respect to fine art).

propose improvements that work within the existing structure of copyright law to achieve greater coherence and consistency. I have argued many of the issues that have accompanied the analysis of realistic visual works stem from the combined effects of a poor understanding of aesthetic theory and a lack of a coherent theory of ideas. This part will therefore address each of these problems and suggest improvements in copyright's approach to them. The first section will argue for a rethinking of the basis of copyright in realistic visual works, resulting in a standard that is both more administrable than the current approach, and more consistent with the analysis of other visual works. The second section will address the problem of how ideas are thought of in copyright law, why the current conception has resulted in incoherent case law, and how it can be ameliorated.

A. Locating Originality in Photography and Realism

As discussed above,¹³² the pictorialist lens through which courts have analyzed photographs has contributed to the dissolution of the idea/expression dichotomy and resulted in the extension of copyright protection to unprotectable elements. However, this is only one of the flaws with such an approach, which is also difficult to administer consistently, undermines core copyright purposes, and fails to account for an increasingly large proportion of photographic works. Courts should therefore cease to employ it and instead bring the analysis of photographs into line with the treatment of other visual images.

1. Administrability

The mode of analysis that courts have applied to photographs relies above all on the photographer's pre-shutter actions to determine originality, and therefore protected expression.¹³³ This approach at least maintains a veneer of coherence when applied to scenes that the author has played a significant part in constructing, such as those in question in *Burrow-Giles* and *Mannion*, but can become strained to the point of absurdity in the context of other types of photographs. Thus, for example, in assessing the copyrightability of Abraham Zapruder's film of the Kennedy assassination, the court emphasized the "creative" choices involved in the selection of camera, film, lens, location, and time.¹³⁴

¹³² See *supra* Part II.A.3.

¹³³ *Id.*

¹³⁴ *Time Inc. v. Bernard Geis Assocs.*, 293 F. Supp. 130, 143 (S.D.N.Y. 1968); cf. Farley, *supra* note 68 at 449 (noting that frequently "the court cannot conceive of an author acting through a machine, but only in advance of it.").

Far from ensuring that the work in question represents a series of creative decisions by the photographer, the primary effect of such a line of inquiry is to incentivize authors to shape their testimony to the court's tastes, and punish those who are unwilling or unable to do so. A recent series of cases, in which courts based fair use on testimony of authorial intent, is instructive: without any significant change in working methods, artists have learned to adapt their description of the creative process to the courts' requirements,¹³⁵ suggesting testimony that is at best without value, and at worst unreliable.

2. *Hindering Authorship: Pictorialism and the Goals of Copyright*

The evidentiary problems inherent in relying on a narrative of creation bear on the fundamental question of copyright, namely identification of the expression original to the author. Since every work will contain elements that are not original to the author, by granting copyright to the creation of a scene courts risk granting ownership over aspects of a work for which the author is not responsible. The photograph at issue in *Burrow-Giles* illustrates the point well. Among the elements of the photograph that the Court describes as springing "entirely from [Sarony's] original mental conception" are Wilde's pose and costume.¹³⁶ But neither of these was original to Sarony in any significant way: the pose was one for which Wilde was already known,¹³⁷ as was his style of clothing. Nevertheless, under subsequent copyright jurisprudence that prevents later authors from recreating and photographing the same scene,¹³⁸ Sarony may well have been able to claim ownership over these elements and prevent others from photographing them.¹³⁹

The dangers of granting such a monopoly over elements for which the author is not responsible are closely related to the dangers of according copyright protection to ideas: both risk unduly constraining the future creation of works,¹⁴⁰ and therefore hindering the proliferation of works that copyright law seeks to encourage.

¹³⁵ Amy Adler, *Fair Use and the Future of Art*, 91 N.Y.U. L. REV. 559, 582-83 (2016).

¹³⁶ *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 54-55 (1884).

¹³⁷ Farley, *supra* note 68 at 433.

¹³⁸ E.g., *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 463 (S.D.N.Y. 2006); *Gross v. Seligman*, 212 F. 930, 931 (2d Cir. 1914).

¹³⁹ Note that this is only a hypothetical intended to illustrate problems with current copyright jurisprudence; the actual *Burrow-Giles* case involved a reprint of the original photograph rather than a recreation of it, and therefore did not consider the latter scenario.

¹⁴⁰ See *supra* Part II.A.3.

3. Aligning Photography Copyright with Other Visual Images

Copyright no longer protects only posed photographs like the Oscar Wilde portrait from *Burrow-Giles* or the Kevin Garnett photograph from *Mannion*, but extends to “[a]lmost any photograph.”¹⁴¹ However, as Christine Farley has pointed out, the logic for protecting photographs has remained essentially unchanged.¹⁴² Instead, courts have strained to shoehorn works like documentary photographs¹⁴³ into the *Burrow-Giles* framework, focusing on pre-shutter scene-setting and attributing unlikely aesthetic considerations to the photographers.¹⁴⁴ Confronted with the disconnect between the case law’s protection for photographs and its reasons for doing so, opinions have had to resort to a legal fiction. In light of the increasingly large proportion of photographs that conform much more closely to a documentary paradigm than a pictorialist one, it is untenable to continue applying a framework that fails to account for these.

I have argued that the need to analyze photographs through a pictorialist lens grew out of an untheorized acceptance of photography as a transparent medium, but that such a view ignores the creative choices involved in the rendering of a scene via a photograph.¹⁴⁵ However, if photography is a medium that interprets its subject in a manner akin to painting or the other visual arts, then there is no reason to accord it different treatment. The subject matter would therefore remain an element of the analysis of photographs, since substantial similarity requires at minimum that two works depict the same thing. However, Photographer A would not be able to prevent later authors from depicting the same scene—even if it was one that he had created—provided they did so in a manner that was not substantially similar.

A famous example from the world of painting may illustrate the point. When Pablo Picasso painted *Le déjeuner sur l’herbe (d’après Edouard Manet)*, he copied the scene from Manet’s *Le Déjeuner sur l’herbe* but depicted it in such a different manner that it is inconceivable a court would find the works substantially similar.

¹⁴¹ *Mannion*, 377 F. Supp. 2d. at 450.

¹⁴² Farley, *supra* note 68, at 438-51.

¹⁴³ I use the term “documentary photographs” to refer not to professional works in a documentary style, but to photographs whose primary value is evidentiary rather than aesthetic.

¹⁴⁴ Farley, *supra* note 68, at 447-48.

¹⁴⁵ See *supra* Part I.C.

That Manet had created the scene from his imagination would be of no importance. In an analogous manner, we can imagine a photograph of Oscar Wilde in the same pose, wearing the same clothes, from the same angle as Sarony's photograph, but that looks nothing like it – the background could be in focus, and Wilde himself little more than an unidentifiable fuzzy shape; or a long exposure could cause the picture to depict several blurred outlines of Wilde. How much of difference from the original would be required remains open to question, and would be for a jury to determine.¹⁴⁶ But importantly, it would accomplish the same end as a pictorialist approach by protecting original expression, while leaving room for the age-old practice of artists providing new takes on existing works. Moreover, it would better account for the vast majority of photographs that do not depict elaborately staged scenes, and free the courts from attempting to manipulate an inapplicable framework.

This leaves the problem of how to deal with works such as documentary footage that are produced with little if any aesthetic consideration. Here too, though, the distinction between replication and recreation of photographs may be of help: insofar as such works should be protected—a proposition concerning which this note is agnostic, but that describes the current approach of the law—one possibility would be to adopt a presumption of copyrightability for photographs, which would protect against exact copying of pictures, but not against their recreation. This would allow those who had captured something truly unique such as the Kennedy assassination to capitalize on their work, without entailing the problems associated with overbreadth of protection.

B. Reimagining Ideas

So far this section has discussed solutions to problems limited (or at least, especially relevant) to visual works. However, the broader theme underlying this note—the problematic separation of ideas from expression—is one that I have argued is thrown into sharpest relief in analyses of visual images, but certainly is not limited to such works. At the heart of the confusion as to what constitutes an idea is a widespread failure to acknowledge the fundamental difference between the colloquial definition of “ideas,” and “ideas” as they are understood for the purposes of copyright law.

¹⁴⁶ It is possible, for example, that the result of *Mannion* would be the same under this analytic framework, since elements such as shading and focus may have simply reinforced the substantial similarity between the works.

The ordinary meaning of the word “idea” has a germinal connotation as the first step of an endeavor, which may manifest itself in a number of ways, including creative expression. Courts frequently use the term in this way, alluding to a causal and logical chain that originates with the author’s idea, and yields the end result of the work that is the object of analysis.¹⁴⁷ A number of difficulties attend the practical application of this concept, beginning with the question of whether it is relevant, even theoretically, to many works of authorship. As a number of scholars have pointed out,¹⁴⁸ creators often are not driven by a single idea that they proceed to execute. Rather, works may represent the combination of several ideas, none of which predominates over the others. Alternatively, authors may simply create gratuitously without any conscious or articulable motivation. To seek a single idea behind a work as the centerpiece of copyright analysis therefore risks in many cases being an illogical quest.

Perhaps more importantly for the purposes of copyright analysis, however, is the fact that this definition does not align with the “idea” analysis that courts employ, in which the “macro idea” of the work is simply a description of the work at a level too broad to receive protection.¹⁴⁹ Essentially, the “idea” that courts arrive at through the *Nichols* abstraction test consists of what is left once the work’s protectable elements have been filtered out. However, no reason is given for thinking that the result of such an analytical process will be a reasonable proxy for the work’s “idea” in the colloquial sense, even though the two are frequently treated interchangeably. Certainly, in some cases the definitions will coincide, and the failure of courts to distinguish between them may reflect the fact that the case that introduced the abstractions test involved two such works: the general plots of

¹⁴⁷ See, e.g., *Kaplan v. Stock Market Photo Agency, Inc.*, 133 F. Supp. 2d 317, 323 (S.D.N.Y. 2001) (describing the photograph’s idea as the attempt to convey a state of mind); *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1110 (9th Cir. 1970) (describing greeting cards as protectable due to their “embodiment of humor, praise, regret or some other message in a pictorial and literary arrangement.”).

¹⁴⁸ E.g. Adler, *supra* note 135, at 584-87.

¹⁴⁹ Judge Kaplan made this very point in *Mannion*, 377 F. Supp. 2d at 460, while discussing a previous case (“‘a couple with eight small puppies seated on a bench’ [identified in *Rogers v. Koons*, 960 F.2d 301, 304 (2d Cir. 1992) as the idea of the plaintiff’s work] is not necessarily the idea of *Puppies*, which just as easily could be ‘people with dogs on their laps,’ ‘the bliss of owning puppies,’ or even a sheepishly ironic thought such as ‘Ha ha! This might look cute now, but boy are these puppies going to be a lot of work!’ Rather, ‘a couple with eight small puppies seated on a bench’ is nothing more or less than . . . a description of the subject at a level of generality sufficient to avoid implicating copyright protection for an original photograph.”).

the plays at issue in *Nichols* were ideas both in the copyright sense of elements too general to merit protection, and in the general usage sense of initial conceptions from which the works sprung.¹⁵⁰ However, in other works the two definitions of “idea” will diverge widely. The plot of the Samuel Beckett play *Waiting for Godot*, for example, may be broadly described as two men waiting for an individual who never shows up, but this hardly captures the “idea” that Beckett was looking to convey and that prompted him to write the work.

A similar observation may be made of visual images, some of which seek merely, or at least primarily, to depict their subjects, and whose “ideas” in the colloquial sense may therefore correspond roughly to a general description of them.¹⁵¹ Conversely, few would describe the idea of Edvard Munch’s *The Scream* as “an individual with his mouth open,” or even “an individual screaming,” although both of these provide a broad description of the work.

The common definition of a work’s “idea,” then, is different from that yielded by the analysis of copyright cases, though courts treat them as equivalent. Once we understand this, we can begin to make sense of courts’ struggle to pin down the “ideas” behind works: since the “idea” of a work as determined through the abstraction analysis corresponds simply to whatever aspects cannot be copyrighted, it is fundamentally a residual category, defined by what it is not. As a result, it has no necessary internal coherence, and is therefore not amenable to being unified under the label of the work’s overarching “idea.” Attempting to locate and define such a singular idea thus risks being an impossible task.

If locating a work’s overall “idea” is a difficult or impossible task, it begs the question of whether there are any benefits to the exercise. The primary role in copyright analysis of identifying and describing a work’s idea is to determine whether the doctrines of merger or *scènes à faire* apply. I have argued in Part I.C that the notion of inevitability of expression is generally greatly overstated, and may even be inapplicable, in the case of non-utilitarian visual works. As a result, the merger doctrine for such works is of questionable value even aside from the problem of identifying their “ideas.”¹⁵² The application of *scènes à faire* to visual works is similarly of limited usefulness. To the extent that elements encompassed

¹⁵⁰ That is, they were plot-driven works whose primary purpose seems to have been to convey stories.

¹⁵¹ Works with a documentary purpose, such as a court sketch, would fall into this category.

¹⁵² See, e.g., Murray, *supra* note 109 (arguing that merger and *scènes à faire* doctrines are inapplicable to visual works, due to the infinite number of ways to convey something visually). I differ from him primarily in that I do not believe merger or *scènes à faire* to be any better applicable to textual works.

by this doctrine are necessary to convey an idea, they are likely to be too unoriginal or unspecific to meet the requirements of copyright anyway. For example, a character with a hook for a hand would not be protectable regardless of whether the hook were considered a *scène à faire* with respect to pirates, since it is neither original nor sufficiently detailed to rise to the level of copyrightability.

The identification of “macro ideas” is therefore difficult or impossible, is of little value for copyrightability analyses, and leads to inconsistent results of the kind described in Part II.B. As a result, courts should cease the attempt. This would not eliminate the idea/expression dichotomy from copyright law, but simply reframe how we identify it: rather than attempting affirmatively to identify ideas, courts would identify all of the aspects of a work that are protectable, without worrying how to characterize those that are not. As I have argued in this section, this is essentially what courts already do through the abstractions test, with the identification of macro ideas simply adding a layer of confusion to the analysis.¹⁵³

The adjustment that I propose here would by no means solve all of the problems inherent in attempting to distinguish between protectable and unprotectable expression. Most notably, it does nothing to aid the line-drawing problem that Justice Hand famously identified in *Nichols*.¹⁵⁴ However, by eliminating the inquiry into a work’s overarching idea, courts can cut out one perennial source of confusion, and the effects that flow from it. In doing so, they can refocus the copyrightability analysis on its proper object, namely the tangible elements of the works in question.

CONCLUSION

Among the recurring challenges of copyright law is the adaptation of existing law to technological developments that allow new means of creating, storing, or displaying works of authorship. While photography has posed enduring problems since its invention, the tools of existing copyright doctrine can be refined to significantly ameliorate some of the more pronounced areas of confusion. But these improvements require coming to grips with the unique characteristics of photographs, and a willingness to recognize and rectify the ways in which they have been largely ignored to this point in copyright jurisprudence.

¹⁵³ Cf. *Mannion*, 377 F. Supp. 2d at 461 (“The idea/expression distinction in photography, and probably the other visual arts, thus achieves nothing beyond what other, clearer copyright principles already accomplish.”).

¹⁵⁴ 45 F.2d at 121.

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 6

SPRING 2017

NUMBER 2

TOMATO, TAMATIE?
REVISING THE DOCTRINE OF FOREIGN EQUIVALENTS
IN AMERICAN TRADEMARK LAW

THOMAS MERANTE*

The growing specter of globalization impacts industries from communication to transportation, resulting in an unparalleled proliferation of cultural diffusion unmatched throughout history. Naturally, this cultural diffusion has familiarized American consumers with foreign brands and foreign languages despite the obvious English dominance domestically, resulting in a trademark quagmire. Under the current American doctrine of foreign equivalents, trademark examiners and courts translate non-English words into English to determine whether they meet the general United States Patent & Trademark Office (USPTO) registration requirements. However, by treating English and non-English words alike, the pool of source-identifying marks is unnecessarily restricted. This note argues that a clear rule-like form that relaxes restrictions of registering descriptive foreign language marks through offering 'descriptive' foreign-language a presumption of eligibility for protection would mitigate inconsistent application of the doctrine. Such a rule would also limit costs on consumers and producers that are caused by restricting the range of available marks and inhibiting creative and communicative branding.

* J.D. Candidate, Class of 2017, New York University School of Law; B.A. Philosophy and International Political Economy, Fordham University. I am grateful to Professor Christopher Sprigman for his oversight, the JIPEL Notes Program participants for their many rounds of feedback, and Kathleen for her unfailing support. Errors or omissions are my own.

INTRODUCTION.....	312
I. TRADEMARK FOUNDATIONS AND THE ROLE OF THE DOCTRINE OF FOREIGN EQUIVALENTS	313
A. <i>Locating the Origins and Sources of Trademark Law</i>	314
1. <i>From Consumer Protection to Brand Protection: Understanding the Goals and Justifications for Contemporary Trademark Law</i>	315
i. Consumer Protection as the Historical Basis for Trademark Protection	315
ii. Protecting the Valuable Goodwill and Communicative Branding Abilities of Mark Holders	316
2. <i>Optional, but Essential: The Value of Trademark Registration</i>	320
i. Distinctiveness as the Principal Requirement.....	320
ii. Use and the Geographic Scope of Protection	321
B. <i>Explaining the Doctrine of Foreign Equivalents</i>	322
1. <i>The Doctrine's Elements</i>	322
2. <i>Justifications for the Doctrine: Domestic Competition and International Comity</i>	324
II. ISSUES WITH THE CURRENT STATE OF THE DOCTRINE	325
A. <i>Inconsistent Application Based on Disparate Judicial Intuitions</i>	325
B. <i>Mismanaging the Relationship Between Information Gains Consumer Confusion</i>	328
1. <i>Information vs. Confusion: Determining the Proper Relationship</i>	329
2. <i>The American Marketplace and Overprotection of Consumers</i>	330
C. <i>Inhibition of Imagination and Communicative Branding</i>	332
III. REEVALUATING THE DOCTRINE AND PROPOSING SOLUTIONS	337
A. <i>A Rule is Preferable to the Current Guideline</i>	338
1. <i>Benefits of Adopting a Foreign Equivalents Rule</i>	339
2. <i>Features of the Proposed Rule: Standardizing Translation Methods and Adopting a Confusion Threshold</i>	339
B. <i>Permitting Registration of Foreign Descriptive Terms</i>	342
1. <i>Informational and Branding Advantages from Protecting Descriptive Foreign-Language Marks</i>	342
2. <i>Precedent for Protecting Descriptive Foreign-Language Marks</i>	344
C. <i>Applying the Proposed Rule: "DELICIOSA" Versus "NÓSTIMA"</i>	346
D. <i>Addressing Competition Concerns through Descriptive Fair Use and Reexamination</i>	347
CONCLUSION	349

INTRODUCTION

International trade has grown tremendously over the last thirty years due to significant decreases in communication and transportation costs,¹ and with such growth the impact of effective brand and trademark protection has increased correspondingly. Trademarks move across national borders as they never have before, reaching new populations of diverse and multilingual consumers.² Although English remains the dominant language in the United States, American consumers are growing more familiar with foreign brands and foreign languages.³ Accordingly, American trademark law, which addresses foreign language trademarks under the doctrine of foreign equivalents, must critically assess its foundational objectives and its means for achieving them in today's complex marketplace.

Under the American doctrine of foreign equivalents, trademark examiners and courts translate non-English words into English to determine whether they meet the general United States Patent & Trademark Office (USPTO) registration requirements. The doctrine aims to provide a guideline for the registration of foreign words in a country where the vast majority of consumers exclusively speak English, but twenty-one percent of the population speaks a language other than English in the home.⁴ However, as commentators have noted, the precarious "guideline" nature of the doctrine has occasioned unequal application of the law and disparate results.⁵ When the doctrine *is* applied, it treats foreign words exactly the same as English words in assessing their distinctiveness and their likelihood of causing consumer confusion. By treating English and non-English words alike, however, those applying the doctrine often excessively restrict the pool of source-identifying marks and waste creative branding opportunities.

This note will argue that the doctrine of foreign equivalents requires a clearer, rule-like form and a reformed substance. A rule-like form, as opposed to the doctrine's current form as a guideline, would mitigate inconsistent application of the doctrine. Moreover, relaxing the restrictions on registering certain descriptive

¹ See generally WORLD TRADE ORGANIZATION, WORLD TRADE REPORT (2015), available at https://www.wto.org/english/res_e/booksp_e/world_trade_report15_e.pdf.

² See CAMILLE RYAN, U.S. CENSUS BUREAU, LANGUAGE USE IN THE UNITED STATES: 2011, at 1 (2013).

³ *Id.* at 2.

⁴ *Id.*

⁵ See generally Elizabeth J. Rest, *Lost in Translation: A Critical Examination of Conflicting Decisions Applying the Doctrine of Foreign Equivalents*, 96 TRADEMARK REP. 1211 (2006).

foreign-language marks would allow consumers to recognize informational gains and mark holders to establish communicative branding strategies.

Part I of this note will examine American trademark law's general purposes and the scope of the trademark holder's rights. This section will then proceed to place the doctrine's treatment of foreign-language marks within trademark law's broader framework. Part II explains how and why examiners and courts apply the doctrine both inconsistently and in a manner that fails to assess properly what information consumers are losing or gaining through foreign-language marks. This part concludes that the doctrine imposes costs on consumers and producers by restricting the range of available marks, thereby inhibiting creative and communicative branding.

Part III outlines how Congress or the courts should adapt the doctrine to increase the consistency and predictability of its application as well as the substantive benefits for the marketplace. Specifically, this part advocates that 'descriptive' foreign-language marks gain a presumption of eligibility for protection.⁶ This section justifies the presumption by assessing the advantages to both consumers and producers and showing how many marks with descriptive features or elements already gain registration in the United States. Recent foreign precedent from both the European Court of Justice and the Australian High Court enhance the viability of the proposal.

I

TRADEMARK FOUNDATIONS AND THE ROLE OF THE DOCTRINE OF FOREIGN EQUIVALENTS

This section aims to provide the reader with an understanding of how the doctrine of foreign equivalents functions within American trademark law. After identifying the bases for trademark protection and describing the Lanham Act's registration requirements, this section will offer the rationales and general vision of the doctrine of foreign equivalents using case law and the USPTO's Trademark Manual of Examining Procedure (TMPE).

⁶ Descriptive marks refer to one or more features of the products to which they are attached, and typically are ineligible for trademark protection unless they have acquired a "secondary meaning" to consumers. *See infra* Part I.A.ii.a.

A. Locating the Origins and Sources of Trademark Law

Unlike patents and copyrights, trademarks do not have a specific constitutional basis for protection in the United States.⁷ Rather, trademarks derive their status as protectable intellectual property from common law and the Lanham Act.⁸ In its earliest trademark case, the United States Supreme Court explicitly distinguished trademarks from patents and copyrights: “the ordinary trade-mark has no necessary relation to invention or discovery The trade-mark may be, and generally is, the adoption of something already in existence as the distinctive symbol of the party using it.”⁹ From their birth, trademarks held a different status than other intellectual property. Rather than protecting creation of an artistic or scientific work, trademarks were “useful and valuable aid[s] or instrument[s] of commerce.”¹⁰

The Lanham Act, passed in 1946, marked a new era for trademarks while continuing to emphasize the commerce facilitation aspect of protection found in the common law. The Lanham Act defines a trademark as

any word, name, symbol, or device, or any combination thereof used by a person, or which a person has a bona fide intention to use in commerce and applies to register on the principal register . . . to identify and distinguish his or her goods, including a unique product, from those manufactured or sold by others and to indicate the source of the goods, even if that source is unknown.¹¹

Several elements of this definition require definition and explanation to understand the doctrine of foreign equivalents: namely, “use,” “register,” and “indicate the source.” “Use” and “register” relate to the creation and perfection of trademark rights, while source identification pertains to the essence, or purpose, of trademark protection. We begin with the latter.

⁷ The Constitution grants Congress the power “[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” U.S. CONST. art. I, § 8, cl. 8. In the seminal *Trade-Mark Cases*, the U.S. Supreme Court held that trademarks do not fall under this clause’s umbrella. *In re Trade-Mark Cases*, 100 U.S. 82, 93–94 (1879).

⁸ 15 U.S.C. § 1051 et seq. (2013).

⁹ *Trade-Mark Cases*, 100 U.S. at 94.

¹⁰ *Id.* at 95.

¹¹ 15 U.S.C. § 1127.

1. From Consumer Protection to Brand Protection: Understanding the Goals and Justifications for Contemporary Trademark Law

To identify the basis or bases for trademark protection, it is essential to begin with the historical emphasis on consumer protection through source identification. However, the twenty-first century observer must also consider the significance of goodwill and ‘branding’ to fully understand the foundation for trademark protection in the contemporary marketplace.

i. Consumer Protection as the Historical Basis for Trademark Protection

The most prevalent trademark protection rationale is the desire to guard against consumer confusion by assisting consumers in identifying a product’s source.¹² Recall that the Lanham Act’s definition of a trademark refers to a mark’s ability to “identify” and “distinguish” one firm’s product or service from those of others.¹³ In this sense, trademarks function as tools (source identifiers) that reduce information costs in a competitive marketplace. Thus, the source identification function is consumer-based and enhances competition.¹⁴ The consumer-protection philosophy dominates American trademark law and is, for instance, the impetus behind the trademark infringement cause of action.¹⁵ Note that although it is mark holders, and not consumers, that bring infringement actions, the legal end of such lawsuits is to abate harms to consumers.¹⁶ This requirement that mark holders frame the harm in an infringement suit as one to their consumers underscores the traditional understanding of trademarks as pro-competitive, consumer-protecting tools rather than proprietary assets of mark holders.¹⁷

¹² *See id.*

¹³ *Id.*

¹⁴ *See* Smith v. Chanel, Inc., 402 F.2d 562, 566 (9th Cir. 1968) (“Without some such method of product identification, informed consumer choice, and hence meaningful competition in quality, could not exist.”); William M. Landes & Richard A. Posner, *The Economics of Trademark Law*, 78 TRADEMARK REP. 267, 271 (1988).

¹⁵ *See* 15 U.S.C. § 1114(1)(a).

¹⁶ Vincent Chiappetta, *Trademarks: More Than Meets the Eye*, 2003 U. ILL. J.L. TECH. & POL’Y 35, 43 (2003).

¹⁷ The conception of trademarks as “tools” rather than “property rights in gross” distinguishes trademarks from other forms of intellectual property. *See, e.g.,* Deven R. Desai, *From Trademarks to Brands*, 64 FLA. L. REV. 981, 1011–12 (2012).

ii. Protecting the Valuable Goodwill and Communicative Branding Abilities of Mark Holders

In today's marketplace, trademarks also serve two significant functions beyond consumer protection through source identification.¹⁸ Although consumer protection remains the dominant foundation for trademark law, these alternative justifications highlight the economic significance of branding for mark holders, as well as the ability of marks to communicate to consumers the values or characteristics of an individual product or the firm that produces the product.

First, trademarks embody the mark holder's goodwill and contribute to a brand's value in a manner that results in trademarks being perceived more as rights in gross than merely pro-competitive tools.¹⁹ To understand this position, one must distinguish and understand the relationship between trademarks and brands. Though the law does not define the term "brand," a fair definition is an exclusive identity that a firm creates through marketing its products or services and garnering goodwill – for example, a reputation for quality, value, or even prestige.²⁰ Trademarks encapsulate, and present to the marketplace, the goodwill that a brand generates. This goodwill, though intangible, is a highly valuable asset.²¹ Indeed, if goodwill were readily quantifiable, it would represent a large proportion of many firms' overall value – in some instances, eighty percent or more.²² Therefore, one alternative justification for trademark protection holds that consumers purchase products or services in part or in whole because of their trademarks, and that

¹⁸ See, e.g., Andrew Griffiths, *Quality in European Trade Mark Law*, 11 NW. J. TECH. & INTELL. PROP. 621, 627 (2013) ("Trade marks. . . provide the means whereby undertakings can cultivate an image and other associations for their products in advertising and other promotional activity. This role goes beyond providing reassurance about the quality of the marked products, to a more active role of conferring intangible quality onto the products or even forming part of their overall quality.").

¹⁹ Lisa H. Johnston, *Drifting Toward Trademark Rights in Gross*, 85 TRADEMARK REP. 19, 22 (1995) ("Despite the well-established principle that trademarks exist only in connection with the good will of a business, trademark and unfair competition law has drifted toward a recognition of trademark rights in gross.").

²⁰ Andrew Griffiths, *Brands, Firms, and Competition*, in BRANDS, COMPETITION LAW AND IP 241 (Deven R. Desai et al. eds., 2015).

²¹ *Id.* at 246.

²² Irina D. Manta, *Privatizing Trademarks*, 51 ARIZ. L. REV. 381, 389–90 (2009) ("In 1988, for example, Kraft, which owns trademarks such as Kraft cheese, Miracle Whip, and Breyers ice cream, was purchased for a total of \$12.9 billion, a sum worth four times Kraft's tangible assets. Similarly, the impressive growth of Starbucks's revenues from \$975 million to over \$2.6 billion in only four years has been attributed to the company's strong brand name Trademarks are therefore obviously extremely valuable assets in their own right.").

trademarks encapsulate highly significant value for firms. Indeed, some contend that even more than symbolizing the goodwill, trademarks *become* part of a particular product's goodwill.²³ For this reason, trademarks deserve and require protection for purposes beyond their consumer protection function. As one scholar argues, "Trademark law continues to face business realities that see the brand as valuable in and of itself and as a way to engage in a range of business activities."²⁴ Contemporary trademarks' branding functions highlight the increasingly prevalent conception of trademarks as evolving from consumer protection tools into property rights in gross.

Second, trademarks communicate characteristics about the products and brands to which they are attached.²⁵ Ironically, the communicative function was an original purpose of trademarks before the source identification function superseded it.²⁶ Trademarks' communication to consumers may be direct or indirect. Generally, trademarks that directly communicate information about their accompanying products cannot gain registration without an additional showing of a source-identifying function.²⁷ Nevertheless, businesses seek to adorn their products with terms as descriptive as possible because there is value in the "richness of evocation that inferably descriptive expressions allow."²⁸ Other trademarks communicate aspects of their accompanying products indirectly. "JAGUAR," for example, does

²³ Frank I. Schechter, *The Rational Basis of Trademark Protection*, 40 HARV. L. REV. 813, 818 (1927) ("To describe a trademark merely as a symbol of good will, without recognizing in it an agency for the actual creation and perpetuation of good will, ignores the most potent aspect of the nature of a trademark and that phase most in need of protection.").

²⁴ Desai, *supra* note 17, at 1019. Note, however, that not everyone sees this a desirable shift. *E.g.*, Johnston, *supra* note 19, at 53 ("According trademark rights in gross would stifle competition — the very premise upon which capitalism is based — and create monopolistic enterprises.").

²⁵ Sonia K. Katyal, *Cosmopolitanism and the Transnational Trademark*, in *THE LUXURY ECONOMY AND INTELLECTUAL PROPERTY: CRITICAL REFLECTIONS* 309, 316 (Haochen Sun et al. eds. 2015) ("Trademarks, today, are less about identifications of origin; instead, their earlier function has been surpassed by their growing role as 'messengers' that convey a broad range of information to the public about the product, the company, the people behind the company, and the attributes of each."); Chiappetta, *supra* note 16, at 45 ("At its most basic level, a brand carries the assurance of product-specific 'experiential' characteristic consistency — qualities such as freshness, energy efficiency, low maintenance, or enjoyable taste.").

²⁶ Chiappetta, *supra* note 16, at 44.

²⁷ These marks are termed "descriptive." *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4 (2d Cir. 1976); see *infra* Part I.A.2.i.

²⁸ Alan Durant, *How Can I Tell the Trade Mark on a Piece of Gingerbread from All the Other Marks on It? Naming and Meaning in Verbal Trademark Signs*, in *TRADE MARKS AND BRANDS: AN INTERDISCIPLINARY CRITIQUE* 107, 129 (Bently et al. eds. 2008).

not directly describe anything about the luxury automobile to which it is attached; however, it certainly conveys to consumers that the car is fast and powerful.²⁹

Whether and how trademark law should recognize these alternative trademark functions is controversial.³⁰ American trademark law has not explicitly acknowledged a proprietary interest in trademark rights (unlike some European trademark law)³¹ and at times has strongly resisted fuller recognition of goodwill as a property right in gross. The naked licensing doctrine, for example, holds that a mark holder abandons (i.e., forfeits its right to) its registered, protected mark if the mark holder licenses the mark without maintaining quality control over the licensee's products.³² Nevertheless, trademark law has demonstrated openness to the expansion of trademark rights through the birth of the dilution cause of action codified in U.S.C. § 1125(c) and the evolution of trademark licensing.

Dilution, for instance, is fueled by mark holder protection rather than consumer protection.³³ Dilution takes two forms: blurring and tarnishment.³⁴ Blurring occurs when a defendant's mark, though not confusing as to source, still interrupts the nexus between the plaintiff's mark and the plaintiff's product.³⁵ The

²⁹ ADRIAN ROOM, NTC'S DICTIONARY OF TRADE NAME ORIGINS 98 (1990). In many cases, such marks are deemed "suggestive" and worthy of protection. *Abercrombie*, 537 F.2d, at 10–11.

³⁰ Compare Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1713–14 (1999) (arguing against expansion and propertization of trademark law as impinging on free speech) with Schechter, *supra* note 23, at 813, 823–26 (arguing for broader scope of trademark causes of action). See also Alex Kozinsky, *Trademarks Unplugged*, 68 N.Y.U. L. REV. 960 (1993) (weighing the moral, utilitarian, and free expression implications of expanded trademark rights).

³¹ See, e.g., Loi n° 64-1360 du 31 décembre 1964 sur les marques de fabrique, de commerce ou de service (Fr.) ("La propriété de la marque s'acquiert par le premier dépôt, valablement effectué conformément aux dispositions de la présente loi et des décrets pris pour son application, qui déterminent les modalités et conditions du dit dépôt, ainsi que les actes ou paiements de taxes qui en perpétuent l'existence.").

³² See, e.g., *Barcamerica Int'l USA Trust v. Tyfield Imps., Inc.*, 289 F.3d 589, 592 (9th Cir. 2002).

³³ See, e.g., *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 429 (2003) ("Unlike traditional infringement law, the prohibitions against trademark dilution are not the product of common-law development, and are not motivated by an interest in protecting consumers.").

³⁴ 15 U.S.C. § 1125(c).

³⁵ For example, the Ninth Circuit held that 'eVisa,' a "multilingual education and information business," blurred Visa International's mark for credit cards. Visa International already had gained an association between its products and services and the "VISA" mark, and eVisa's mark would cause harm to Visa International because it would disrupt the association between product and mark in consumers' minds. *Visa Int'l Serv. Ass'n v. JSL Corp.*, 610 F.3d 1088, 1090–91 (9th Cir. 2010).

dilution by tarnishment cause of action further suggests a proprietary status of trademarks in the United States, protecting mark holders' goodwill from undesirable associations in the marketplace.³⁶ Tarnishment occurs when a defendant's product somehow reduces or deteriorates the reputation of another's mark.³⁷

Like the dilution cause of action, the evolution of trademark licensing supports the notion that trademarks contain a proprietary element in the contemporary marketplace. Common law and statutory law initially prohibited licensing because many viewed licensing as violating a mark's source identification function.³⁸ That is, the licensed mark failed to function as a source identifier if the producer of the trademarked good was not the trademark owner.³⁹ However, the Lanham Act changed the landscape of trademark licensing in the United States.⁴⁰ Although the law has maintained elements of the source-identification function through requirements of quality control and use "as [not] to deceive the public," licensing's primary advantage—and arguably the reason for its inception—is the capitalization of goodwill and the intrinsic value of trademarks.⁴¹

Together, dilution and licensing indicate that American trademark law has taken a step beyond merely consumer protection via source identification. Dilution focuses on harms to mark holders, while licensing opens a market in the goodwill of certain brands. Ultimately, this note recommends that trademark law should continue to recognize its consumer-protection function, but afford greater recognition to these alternative bases for trademark protection.⁴² Part II will establish the reasons for such an expansion.

³⁶ 15 U.S.C. § 1125(c)(2)(C).

³⁷ *E.g.*, *V Secret Catalogue, Inc. v. Moseley*, 558 F. Supp. 2d 734, 750 (W.D. Ky. 2008) *aff'd*, 605 F.3d 382 (6th Cir. 2010) (finding that an adult video and novelty store operating under the name "Victor's Little Secret" was likely to tarnish the "VICTORIA'S SECRET" mark).

³⁸ Neil Wilkof, *Trademark Licensing: The Once and Future Narrative*, 104 TRADEMARK REP. 895, 898 (2014) (citing *Bowden Wire v. Bowden Brake*, [1914] R.P.C. 385 (Eng.)).

³⁹ *Id.*

⁴⁰ 15 U.S.C. § 1055; Irene Calboli, *The Sunset of "Quality Control" in Trademark Licensing*, 57 AM. U. L. REV. 341, 354 (2007) (suggesting that changes in manufacturing and distribution in the early twentieth century spawned the opportunity for legitimate licensing practices).

⁴¹ *See* Wilkof, *supra* note 38, at 915.

⁴² *But see* Sandra L. Rierson, *The Myth and Reality of Dilution*, 11 DUKE L. & TECH. REV. 212, 214 (2012) ("[T]he federal dilution statute turns competition on its head and serves to entrench and further concentrate economic power in the hands of dominant corporate firms at the expense of consumers and competitors alike. Dilution law should be repealed or, at the very least, reformed.").

2. *Optional, but Essential: The Value of Trademark Registration*

The Lanham Act establishes the means by which an entity gains rights in a mark. Firms and individuals may gain common law protection for their marks through “use in commerce.”⁴³ Registration on the USPTO’s Federal Register, though not necessary to gain rights in a mark, provides additional benefits to those interested in expanding their brands: a legal presumption to use the mark nationwide, a presumption that the mark is distinctive, public notice of the mark holder’s claim of ownership, the ability to use the ‘®’ federal registration symbol, and a basis for obtaining protection in foreign countries under the Madrid Agreement.⁴⁴ Registration requires that the mark be “distinctive” and used in commerce.⁴⁵

i. Distinctiveness as the Principal Requirement

In keeping with the original source identification function of trademarks, a mark must distinguish one good’s source from the sources of other goods in the market. That is, it must be “distinctive.”⁴⁶ A seminal 1976 Second Circuit case, *Abercrombie & Fitch Co. v. Hunting World, Inc.*,⁴⁷ generated trademark law’s metric of distinctiveness for word marks.⁴⁸ According to *Abercrombie* there are five categories of distinctiveness: fanciful, arbitrary, suggestive, descriptive, and generic.⁴⁹ Fanciful, arbitrary, and suggestive marks are the strongest on the spectrum.⁵⁰ These three categories of marks are “inherently distinctive” because consumers will readily identify marks such as “Kodak” (fanciful), “Apple Computers” (arbitrary), and “Greyhound” (suggestive) as indicative of the products’ sources.⁵¹ Descriptive marks describe a characteristic of the product on which the mark is placed, such as “Bran-Nut” for cereal containing bran and walnuts. Because they lack inherent distinctiveness, descriptive marks may be registered only with a showing of “secondary meaning.”⁵² That is, they may gain protection only if the

⁴³ 15 U.S.C. § 1127.

⁴⁴ U.S. PATENT & TRADEMARK OFFICE, PROTECTING YOUR TRADEMARK: ENHANCING YOUR RIGHTS THROUGH FEDERAL REGISTRATION (2014), available at <http://www.uspto.gov/sites/default/files/trademarks/basics/BasicFacts.pdf>.

⁴⁵ *Id.*

⁴⁶ See 15 U.S.C. § 1052(f).

⁴⁷ 537 F.2d 4 (2d Cir. 1976).

⁴⁸ *Id.*

⁴⁹ *Id.* at 9.

⁵⁰ *Id.* Fanciful marks are “words invented solely for their use as trademarks.” *Id.* at 11 n.12. Arbitrary marks involve the uncommon use of a word or phrase. *Id.*

⁵¹ See *id.*

⁵² 15 U.S.C. § 1052(f); *Abercrombie*, 537 F.2d at 9.

consuming public attaches source-identifying significance to them.⁵³ Finally, generic marks, such as “chocolate,” describe a genus of goods, and can never be protected because protection would preclude competitors from using the word on their products.⁵⁴ Granting one firm or individual the exclusive use of a generic word would undoubtedly inhibit competition: competitors would be unable to use words that are necessary to describe their products, consequently limiting vital consumer information and causing consumer confusion.⁵⁵

Understanding the relevant consumer market is vital to determining whether a mark is distinctive, for one must appreciate the general vocabulary of the marketplace to understand a word’s significance to that marketplace.⁵⁶ Differences across linguistic populations mean that a mark that is inherently distinctive in one territory is not necessarily so in another territory.⁵⁷ For example, the mark “GREENGROCER” can be generic in the United Kingdom but inherently distinctive in the United States.⁵⁸ The significance of this point is that the relevant consumers for marks gaining protection in the USPTO must be consumers in the American market for a particular product.⁵⁹ Trademark examiners and judges must assess distinctiveness with the American purchasing public in mind.⁶⁰

ii. Use and the Geographic Scope of Protection

Since the passage of the Trademark Law Revision Act in 1988 and its subsequent revisions, satisfying the second registration requirement, “use in commerce,” is rarely an issue.⁶¹ A registrant must demonstrate actual use of the mark

⁵³ An example of a descriptive term that has acquired secondary meaning is “Best Buy.”

⁵⁴ *Abercrombie*, 537 F.2d at 9.

⁵⁵ *Id.* (“[N]o matter how much money and effort the user of a generic term has poured into promoting the sale of its merchandise and what success it has achieved in securing public identification, it cannot deprive competing manufacturers of the product of the right to call an article by its name.”).

⁵⁶ See Jake Linford, *A Linguistic Justification for Protecting “Generic” Trademarks*, 17 YALE J. L. & TECH. 110, 163 (2015).

⁵⁷ *Abercrombie*, 537 F.2d at 9.

⁵⁸ *Carcione v. Greengrocer, Inc.*, No. S-78-561, 1979 WL 25110, at *1 (E.D. Cal. Oct. 12, 1979).

⁵⁹ Rest, *supra* note 5, at 1244.

⁶⁰ *Id.*

⁶¹ 15 U.S.C. §§ 1051–1128 (1988). However, modern computer technology has raised questions about defining “use in commerce.” *E.g.*, *Rescuecom Corp. v. Google Inc.*, 562 F.3d 123 (2d Cir. 2009) (finding Google’s suggestion of competitors’ trademarks as search terms for advertisers constituted “use in commerce” under the Lanham Act).

in commerce, or offer a bona fide intent to use the mark in commerce.⁶² The use requirement also has significant implications even *after* a mark gains registration because even if a mark holder gains nationwide priority with registration in the PTO, he must show actual use when seeking an injunction to suppress another's use of the mark in a particular location.⁶³ Thus, the mark holder's radius of use defines the geographic scope of the trademark rights even after the time of registration.⁶⁴ Ultimately, distinctiveness is the primary registration concern for registrants of foreign-language marks, but one should be aware that the use in commerce requirement always remains.

B. Explaining the Doctrine of Foreign Equivalents

Consider now a hypothetical. A young entrepreneur decides to bottle his family's tomato sauce and sell it under the mark "FRUTTO DELLA TERRA." If the entrepreneur submits a registration application, the question arises: how ought an examiner or a court of review determine whether "FRUTTO DELLA TERRA" is eligible for protection? Alternatively, if the mark gains registration, what happens if a competitor using the mark "FRUIT OF THE EARTH" for tomato sauce brings a trademark infringement suit under Section 43(a) of the Lanham Act?

1. The Doctrine's Elements

Using the doctrine of foreign equivalents, examiners and courts translate non-English words into English to determine their place on the *Abercrombie* spectrum (i.e., whether they are fanciful, arbitrary, suggestive, descriptive, or generic) and consequently, whether they are eligible for registration. Likewise, in infringement cases, courts translate the word to see if it is likely to cause confusion with an existing English-language mark (or vice versa, if a foreign-language mark is registered and another firm seeks to register an English-language mark).

Courts and the TMEP have made clear that the doctrine is not a rule, but rather a "guideline" to help determine whether a mark is registrable.⁶⁵ The fact that the doctrine is a guideline, rather than a rule, means there is no strict procedure for its application. Rather, trademark examiners and judges have discretion to choose

⁶² 15 U.S.C. § 1051(b) (1988).

⁶³ *E.g.*, *Dawn Donut Co. v. Hart's Food Stores, Inc.*, 267 F.2d 358, 364 (2d Cir. 1959) (holding that the senior user of a mark could not enjoin the junior user from using "Dawn" for donuts until it could show that it was expanding into the junior user's market, even though the senior user had nationwide rights).

⁶⁴ *Id.*

⁶⁵ TMEP §§ 1209.03(g), 1210.10, 1211.01(a)(vii) (Jan. 2017).

whether to employ the doctrine at all, and then have fairly wide latitude to determine whether the foreign-language mark is distinctive.

In applying this guideline, examiners and courts are tasked with considering how American consumers understand non-English trademarks in the marketplace. The doctrine is meant to apply only to words from “common” foreign languages.⁶⁶ “Common” languages include Spanish, Italian, Russian, and Yiddish.⁶⁷ By contrast, the doctrine does not apply to dead or obscure languages such as Latin.⁶⁸ If the doctrine acts as a proxy for determining whether consumers would find the foreign-language mark inherently distinctive or not, then the rationale for the common language requirement seems apparent: the more consumers that speak the language, the greater the probability that the consuming public will understand the meaning of the foreign-language mark in its original form. The Federal Circuit has stated that the doctrine “should be applied only when it is likely that the ordinary American purchaser would stop and translate the foreign word into its English equivalent.”⁶⁹ Thus, examiners and courts should translate foreign-word marks as if they were consumers purchasing the product to which the mark is affixed.

If, upon translation, the TTAB or court finds that the mark is generic or merely descriptive in the foreign language, then the mark is ineligible for registration. In other words, if the word is from a “common” language, the “ordinary American consumer” would translate the word, and if the word would not be inherently distinctive of source in the foreign language, then it cannot enter the Federal Register.⁷⁰ Likewise, if the TTAB or court finds that the ordinary American consumer is likely to confuse the foreign-language mark with a registered English-language mark, the foreign-language mark is ineligible for protection.⁷¹

Return now to the “FRUTTO DELLA TERRA” tomato sauce example. Were an examiner to receive an application for this trademark, she would have to determine whether to apply the doctrine of foreign equivalents. Upon realizing that

⁶⁶ *Palm Bay Imports, Inc. v. Veuve Clicquot Ponsardin Maison Fondée En 1772*, 396 F.3d 1369, 1377 (Fed. Cir. 2005); TMEP § 1209.03(g) (Jan. 2017).

⁶⁷ 2 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 11:34 (4th ed. 2017). Others include French, German, Chinese, Japanese, Polish, Hungarian, and Serbian.

⁶⁸ TMEP § 1209.03(g) (Jan. 2017).

⁶⁹ *Id.*; *Palm Bay Imports*, 396 F.3d at 1377.

⁷⁰ Note, however, that the “ordinary consumer” has assumed a multitude of meanings over time, making this part of the definition quite murky. *Rest, supra* note 5, at 1235. *See infra*, Part II.A.

⁷¹ *E.g., In re Ithaca Indus.*, 230 U.S.P.Q. 702 (T.T.A.B. 1986) (refusing registration of “LUPO” mark due to confusing similarity with a previously registered “WOLF” mark.).

the mark does not employ English words, she would next determine whether the doctrine should apply to Italian words. Based on the number of consumers in the market that understand Italian and English, is the language common or obscure?⁷² If it is a common language, then the examiner must determine who the ordinary consumer is, and whether he would be likely to “stop and translate” the mark into English.⁷³ Finally, the examiner would have to determine whether the English translation is distinctive according to the *Abercrombie* spectrum, and whether it is confusingly similar to any other registered marks. In other words, the examiner must decide whether “FRUTTO DELLA TERRA” is confusingly similar in sight, sound, or meaning to another registered mark, e.g., “EARTH’S FRUIT.”⁷⁴ One can see that the doctrine, while perhaps not overly complex, is flexible and affords examiners a fair amount of discretion in determining whether a mark is eligible for protection.

2. *Justifications for the Doctrine: Domestic Competition and International Comity*

The stated rationales for the doctrine are twofold.⁷⁵ First is the promotion of domestic competition in a diverse contemporary American marketplace. The Second Circuit explained that barring non-inherently-distinctive foreign-language marks from registration

[R]ests on the assumption that there are (or someday will be) customers in the United States who speak that foreign language. . . . [C]ommerce in the United States utilizes innumerable foreign languages. No merchant may obtain the exclusive right over a trademark designation if that exclusivity would prevent competitors from designating a product as what it is in the foreign language their customers know best.⁷⁶

The court’s rationale, unsurprisingly, is consumer-oriented. It emphasizes the source identification function in its concern that non-English speakers who are in the United States, either permanently or temporarily, will be harmed by generic, descriptive, or otherwise confusing foreign-language trademark uses.⁷⁷ The second rationale for the doctrine rests on a notion of international comity; namely, that there

⁷² MCCARTHY, *supra* note 67, at § 11:34.

⁷³ TMEP § 1209.03(g) (Oct. 2015). However, note that the relevant purchasing public is one of the inconsistencies noted in Part II, *infra*.

⁷⁴ For the factors used to determine whether there is a likelihood of confusion, see *In re E. I. DuPont de Nemours & Co.*, 476 F.2d 1357, 1361 (C.C.P.A. 1973).

⁷⁵ Sujata Chaudhri, *Trademark Doctrine of Foreign Equivalents*, NYIPLA BULLETIN, Jan./Feb. 2007, at 12.

⁷⁶ *Otokoyama Co. v. Wine of Japan Imp.*, 175 F.3d 266, 270–71 (2d Cir. 1999).

⁷⁷ *See id.*

ought to be parity in the international treatment of foreign-language marks. Because the United States has previously sought to bar foreign registration of English-language generic trademarks, “to permit registration here of terms in a foreign language which are generic for products sold in a foreign country would be inconsistent with the rationale supporting these international protests.”⁷⁸ Like the first rationale, international comity sounds in consumer protection, but on a larger geographic scale.

Thus, consumer protection lies at the heart of the two principal justifications for the doctrine of foreign equivalents. The next part will show the limits of relying on consumer protection as the foundation for the doctrine, especially when the doctrine fails to address the proper range of consumers.

II

ISSUES WITH THE CURRENT STATE OF THE DOCTRINE

Although the doctrine of foreign equivalents has applied to cases for many years, legitimate questions of its continued usefulness and application have grown over the past decade.⁷⁹ This section reveals the inconsistencies between various applications of the doctrine of foreign equivalents and demonstrates how such inconsistencies impose costs on both producers and consumers. First, this section will describe how the “guideline” nature of the doctrine provides insufficient guidance for examiners and courts, which thereby produces uncertainty for prospective mark registrants. Next, this section will demonstrate how application of the current doctrine often overestimates the potential for consumer confusion, consequently hindering information gains to American consumers by rejecting marks that should be accepted. Finally, this section will suggest that the doctrine of foreign equivalents leaves little room for imaginative branding, thereby harming producers and reducing the range of valuable source identifiers for consumers.

A. Inconsistent Application Based on Disparate Judicial Intuitions

Because the doctrine of foreign equivalents is a guideline rather than a rule, examiners’ and judges’ discretion and disparate intuitions about the doctrine’s features result in some unpredictability in the doctrine’s application. Uncertainty, in turn, places costs on prospective mark registrants, who must guess as to how their marks will be received and how much to invest in a brand or mark that ultimately may be rejected. This is by no means to suggest that judges or examiners are poor at

⁷⁸ *In re Le Sorbet, Inc.*, 1985 T.T.A.B. LEXIS 27, at *9 (T.T.A.B. 1985).

⁷⁹ See, e.g., Serge Krimnus, *The Doctrine of Foreign Equivalents at Death’s Door*, 12 N.C. J. L. & TECH. 159 (2010).

assessing trademark distinctiveness; trademark cases often rely on intuition about how the consuming public will receive a mark.⁸⁰ Rather, in the context of foreign equivalents, there are remaining ambiguities in the doctrine that could be fixed with a more consistent application procedure and definition of terms.

One of the issues arising from decision makers' discretion relates to the basic mechanics of translation. The USPTO permits consultation of a wide variety of online resources (X-search, foreign language dictionaries, "free online translation tools") or the Office's Translations Branch.⁸¹ While the purpose presumably is to provide examiners with an array of resources to gain the meaning of a specific foreign term, an unfixed pool of sources makes it unnecessarily difficult for registrants to predict how a mark will be received. Translation mechanics are complicated even further in cases involving characters outside of the Latin alphabet. Take, for example, the application for the mark, "DARK HORSE."⁸² The examiner rejected the applicant's "DARK HORSE" mark for spirits on the grounds that it was confusingly similar to "KUROUMA," which was already on the Federal Register for spirits in both the Latin and Japanese alphabets.⁸³ The TTAB reversed the examiner's decision, holding that the doctrine of foreign equivalents did not apply because there was no single and literal translation of "KUROUMA."⁸⁴ In fact, there are twelve possible interpretations depending on how one reads the Japanese characters.⁸⁵ Moreover, even once one chooses a reliable source for translation, there is no guarantee of direct equivalence between the foreign and English words.⁸⁶ Indeed, the Fifth Circuit explicitly acknowledged common ambiguities of translation and attempted to clarify its translation procedure: "The act of translation, of course, can itself be an imprecise task, as foreign words sometimes have no exact equivalent in English; therefore, courts may rely on the 'primary and common translation' in

⁸⁰ See Michael Grynberg, *The Judicial Role in Trademark Law*, 52 B.C. L. REV. 1283, 1302–18 (2011) (discussing judicial discretion in trademark cases, particularly in assessing likelihood of confusion using multifactor tests).

⁸¹ *Examination Guide 1-08*, USPTO (Apr. 23, 2008), http://www.uspto.gov/trademarks/resources/exam/examguide1_08.jsp#ftn2; TMEP § 809.01 (Oct. 2015). For example, in *In re Tokutake Indus. Co.*, 87 U.S.P.Q.2d 1697 (T.T.A.B. 2008), the T.T.A.B. permitted as evidence a translation provided by freedict.com. While such a site may suffice for the casual user, its lack of reputation for accuracy renders it unsuitable as a basis for judicial decision-making.

⁸² *In re Dark Horse Distillery, LLC*, No. 85104448, 2012 WL 4832274, at *2 (T.T.A.B. Sept. 30, 2012) (not precedential).

⁸³ *Id.* "KUROUMA" is registered for "Japanese distilled spirits of barley soju."

⁸⁴ *Id.* at *4–5.

⁸⁵ *Id.* at *4.

⁸⁶ *Cf. Enrique Bernat F., S.A. v. Guadalajara, Inc.*, 210 F.3d 439 (5th Cir. 2000) (debating the translation of the Spanish word "CHUPA" as applied to lollipops).

determining English equivalency.”⁸⁷ Of course, requiring examiners and judges to select the “primary” meaning of a word raises its own issues, especially when they are working with foreign words that are likely unfamiliar to them. Expert testimony, or even a clearer hierarchy of translation references, would ease the registration process for foreign-language marks, but has not yet gained widespread use.

Beyond the mechanics of translation, an important question of interpretation arises in regards to the doctrine’s substance. When do we know that consumers are likely (or unlikely) to “stop and translate” a particular foreign word into English? Is the implication that an examiner or court should not translate a Spanish-language mark marketed to a Portuguese-speaking population, because it is unlikely that native Portuguese speakers would translate the mark into English? As one commentator has noted, it is difficult to ascertain a meaningful likelihood of translation based on pure intuition.⁸⁸ Surveys, though “expensive and time-consuming,” would provide a more concrete determination of the likelihood of confusion.⁸⁹ The TMEP aims to clarify what it means for consumers to be likely to “stop and translate” a mark in Section 1207.01(b)(vi)(B), but merely compounds the analysis by referring to whether any given translation is “literal and direct” (in which case the ordinary consumer would be more likely to stop and translate) or ambiguous.⁹⁰ If the point of the “stop and translate” component of the doctrine is merely to determine whether the mark contains a “literal and direct” translation, it is unclear why it is necessary at all: it is merely a proxy for whether consumers will be confused. The wide discretion granted to examiners and judges consequently leads to broad conclusions lacking a basis in survey evidence or expert testimony. For example, in *Palm Bay Imports*,⁹¹ the Federal Circuit concluded, without any support or explanation, “that it is improbable that the average American purchaser would stop and translate ‘VEUVE’ into ‘widow.’”⁹² The TTAB similarly found that French-speaking American consumers would not stop and translate a mark that is “grammatically incorrect” because it wouldn’t have any meaningful translation to them, even though they would understand what the individual words in the phrase meant.⁹³

⁸⁷ *Id.* at 443.

⁸⁸ Rest, *supra* note 5, at 1233.

⁸⁹ *Id.*

⁹⁰ TMEP § 1207.01(b)(vi)(B) (Oct. 2015).

⁹¹ *Palm Bay Imports, Inc. v. Veuve Clicquot Ponsardin Maison Fondée En 1772*, 396 F.3d 1369 (Fed. Cir. 2005).

⁹² *Id.* at 1377.

⁹³ *In re Helen Trimarchi & Michael Merr*, No. 77222086, 2009 WL 1692509, at *6 (T.T.A.B. May 14, 2009).

Prospective registrants consequently experience uncertainty because of imprecise language in the guideline and unpredictable application of the doctrine. Explicit rules and procedures would offer a relatively easy solution and minimize registrants' uncertainty when approaching the USPTO.⁹⁴

B. Mismanaging the Relationship Between Information Gains and Consumer Confusion

The doctrine of foreign equivalents mismanages the balance between information gain and loss when a new mark is introduced into the marketplace. As it stands, the doctrine may force or influence the rejection of a mark that only a minor fraction of the population could interpret, and with an English equivalent that an even smaller fraction would be confused by.⁹⁵ Barring registration of marks with such low potential for confusion represents an overall informational loss to the market.⁹⁶

To understand the informational loss, one must begin with the concept of consumer protection through source identification.⁹⁷ A properly exercised trademark assists the consumer in making a purchasing decision by providing her with information as to the source or quality of the product to which it is attached.⁹⁸ Thus, the introduction of a new trademark adds to the pool of information in a given market. However, if a mark causes confusion among consumers, either because it reduces competition by removing words from the public domain or by mimicking a competitor's mark too closely, then that mark reduces the flow of information in the market. Imagine, for instance, that trademark 'X' is introduced into the market for perfume. For some portion of the marketplace, 'y' percent, the mark will distinguish the perfume's source from that of other perfumes on the shelf; these consumers gain from the ability of the mark to distinguish source. To the rest of consumers, 'z' percent,⁹⁹ the mark is so confusing that it disrupts their purchasing decision, or they buy the product marked 'X' thinking it is another product. The consumers

⁹⁴ See discussion *infra* Part III.A.

⁹⁵ Krimnus, *supra* note 79, at 159–60.

⁹⁶ *Id.* at 200 (“The doctrine may ensure that a negligible portion of the public is not confused or otherwise negatively affected by a mark, but in doing so, it harms the public at large by eliminating a valuable opportunity to distinguish goods in the marketplace through a registered trademark.”).

⁹⁷ See discussion *supra* Part I.A.

⁹⁸ See generally Landes & Posner, *supra* note 14.

⁹⁹ This assumes that the sum of 'y' and 'z' is one hundred. In other words, every consumer in the marketplace either gains or loses from the introduction of the mark, and there is no overlap between the two groups.

comprising ‘z’ percent of the population lose from the introduction of mark ‘X’ into the marketplace. Thus, every mark provides a net informational gain or net informational loss to the marketplace: when ‘y’ exceeds ‘z’, there is a net gain; when ‘z’ exceeds ‘y’, there is a net loss.

1. Information vs. Confusion: Determining the Proper Relationship

Because trademarks ought to add information to the marketplace for the purposes of limiting consumer confusion, only those marks that offer a net informational gain should earn protection. Nevertheless, trademark law typically requires, and should require, more than a simple showing of a net informational gain to offer protection to a mark. It is not desirable for the law to protect a mark that will assist fifty-five percent of the population but harm the other forty-five percent, because forty-five percent is a substantial portion of the marketplace. In the ideal marketplace, zero consumer confusion would exist because that would imply perfect information.¹⁰⁰ However, because zero confusion is extremely unlikely in any circumstance,¹⁰¹ the law must tolerate a certain threshold of consumer confusion or risk being unworkable.¹⁰² In other words, *zero* confusion is not the aim.¹⁰³

What, then, is the desirable relationship between information and confusion? For a successful trademark infringement claim, plaintiffs often must show a minimum of eleven percent confusion.¹⁰⁴ Note that this implies that the law has determined that a mark that potentially provides informational benefits to eighty-nine percent of the population still hurts the market overall. This eleven percent confusion threshold has led one commentator to argue, “In [situations where consumers face minor confusion], the likelihood of meaningful disruption to markets is low, so society will probably gain overall by tolerating minor confusion in exchange for the benefits that arise from consumers’ exposure to confusion.”¹⁰⁵ Let us assume, however, that the eighty-nine to eleven split is the proper tipping point for whether a foreign-language mark provides sufficient net benefit to the marketplace to warrant protection. Under this standard, one will find that examiners

¹⁰⁰ See William McGeeveran & Mark P. McKenna, *Confusion Isn’t Everything*, 89 NOTRE DAME L. REV. 253, 274 (2013).

¹⁰¹ See generally Alfred C. Yen, *The Constructive Role of Confusion in Trademark*, 93 N.C. L. REV. 77, 79 (2014) (arguing that a certain level of confusion in fact benefits consumers in the market because it facilitates more sophisticated purchasing decisions).

¹⁰² McGeeveran & McKenna, *supra* note 100, at 256.

¹⁰³ See *id.*

¹⁰⁴ 4 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 23:2 (4th ed. 2017).

¹⁰⁵ Yen, *supra* note 101, at 86–87.

and the TTAB have exercised extreme overprotection of consumers, to a point where the marketplace is harmed by the rejection of beneficial source-identifying marks.

2. *The American Marketplace and Overprotection of Consumers*

Part of the reason for the overprotective information-confusion assessments relates to inconsistencies in how those applying the doctrine perceive the relevant national marketplace. There is uncertainty regarding whether the doctrine primarily serves the current American purchasing public, a hypothetical, more linguistically-diverse American purchasing public of the future, or the international purchasing public.¹⁰⁶ The typical consequence is that trademarks found to be non-inherently-distinctive or confusingly similar to English-language marks for infinitesimal percentages of the population are not protected. For example, in *In re Savisa*, a South African corporation appealed to the TTAB to reverse a rejection of its application for the “SONOP” mark for alcoholic beverages.¹⁰⁷ The original examiner had refused registration on the grounds that “SONOP” was confusingly similar to “SUNRISE,” already registered for the same class of goods.¹⁰⁸

Applying the doctrine of foreign equivalents, the examiner found that although the “SONOP” mark applicant provided reliable evidence that less than .01 of one percent of the American population spoke Afrikaans in the home, Afrikaans met the common language requirement because other consumers may have learned Afrikaans from family or trips abroad, despite not speaking it in the home.¹⁰⁹ However, consider that even if one takes an extremely generous estimation of what that number constitutes—e.g., nine times the number of people that actually speak Afrikaans at home—still far less than one percent of the American consuming public would be familiar with the language.¹¹⁰ Moreover, the TTAB went one step further by taking into account the number of people *worldwide* that speak Afrikaans.¹¹¹ This information is irrelevant at best; at worst, it unduly influenced the board in determining whether the mark warrants protection by shifting focus away from the

¹⁰⁶ Recall the international comity rationale for the doctrine explained in Part I.B.2, *supra*.

¹⁰⁷ *In re Savisa* (Pty) Ltd., No. 78154196, 2005 WL 548058, at *1 (T.T.A.B. Feb. 24, 2005) (not precedential).

¹⁰⁸ *Id.* at *2.

¹⁰⁹ *Id.* at *3.

¹¹⁰ *In re Savisa* is not the only case involving application of the doctrine in circumstances where an infinitesimal percentage of the population might be confused. The Board similarly refused registration of “OIVA” for cutlery on the grounds that it translates to “excellent” in Finnish and therefore was not distinctive. *In re Marimekko Oyj*, No. 85320832, 2012 WL 3224736, at *1-2 (T.T.A.B. July 10, 2012). Less than .02 of one percent of the American population speaks Finnish. CAMILLE RYAN, U.S. CENSUS BUREAU, LANGUAGE USE IN THE UNITED STATES: 2011, at 3 (2013).

¹¹¹ *In re Savisa* (Pty) Ltd., 2005 WL 548058, at *8.

mark's source identifying potential in the actual marketplace in which "SONOP" competes to the mark's potential for causing confusion in an amorphous and undefined market.

There are also occasionally instances where marks in modestly-spoken languages fail to gain registration on the first attempt due to an overestimation of the potential for confusion, but are saved by the TTAB on appeal. In March 2014, for instance, the TTAB in *In re Dunville Peat*¹¹² reversed a refusal of registration for the mark "ÓGRA" for beauty care products.¹¹³ The mark translates to "youth" in Gaelic and was therefore originally found to be descriptive and not inherently distinctive for beauty products.¹¹⁴ In reversing the refusal to register, the TTAB suggested that there was insufficient evidence to show that Gaelic was a common language warranting translation to English.¹¹⁵ This decision properly respects the informational assessment that ought to occur during the registration process of foreign-language marks because it understands that the potential informational gains to consumers in the beauty care market vastly outweigh the possibility of confusion.

Yet even in *In re Christopher A. Fahey*, a case similar to *In re Dunville Peat* where the TTAB properly weighed the informational gains and losses, the Board revealed a temptation to stretch the relevant consumer population too far.¹¹⁶ Like *In re Savisa*, a "sunrise" mark was at issue. "PUKANA LA," the Hawaiian translation of "sunrise," was already registered as a trademark for guitar picks when a California resident attempted to register "SUNRISE" for the same class of goods.¹¹⁷ Although the examiner initially rejected the "SUNRISE" application for registration, the TTAB reversed and permitted registration because the number of Hawaiian-language speakers in the United States, approximately 25,000, did not constitute "an appreciable number of individuals sufficient to sustain a finding of a likelihood of confusion."¹¹⁸ Indeed, according to the evidence on the record, and the Census Bureau's estimate of the current United States population, only .008 of one percent

¹¹² *In re Dunville Peat & Herbal Products Ltd.*, No. 79111854, 2014 WL 1390503 (T.T.A.B. Mar. 24, 2014).

¹¹³ *Id.* at *9–10.

¹¹⁴ Although the decision was reversed and the mark was registered after appeal, such determinations impose economic and temporal costs on mark registrants. See discussion *supra* Part II.A.

¹¹⁵ *In re Dunville Peat & Herbal Products Ltd.*, 2014 WL 1390503, at *9–10.

¹¹⁶ *In re Christopher A. Fahey*, DBA Gravity Guitar Picks, No. 86250337 (T.T.A.B. Apr. 13, 2015) (not precedential).

¹¹⁷ *Id.* at *1–2.

¹¹⁸ *Id.* at *3.

of the population speaks Hawaiian.¹¹⁹ Assuming generously that a full one-half of the Hawaiian-speaking population would be confused by the mark, one will conclude that .004 of one percent of the American consuming population would be harmed by the registration of the mark. Therefore, keeping in mind the eleven percent confusion threshold, this decision properly weighed the potential information gain to the marketplace against the extremely minor possibility of confusion.

However, one troubling aspect of the TTAB's opinion in *In re Christopher Fahey* is a statement made by the board after remarking on how few American consumers understand Hawaiian: "Moreover, there is substantially no population of Hawaiian speakers elsewhere around the globe."¹²⁰ Thus, like in *In re Savisa*, the TTAB in *In re Christopher Fahey* considered the distinctiveness of marks to foreign populations or to hypothetical populations beyond the current American market to which the product is sold.¹²¹ Cases like these fail to consider that even if some consumers are confused by descriptive foreign-language marks, ninety-nine percent of the market may use those marks as source identifiers. Since zero confusion is an unworkable target, the law must balance the interests of potentially confused consumers with those who benefit from the product differentiation signals provided by unique marks such as "SONOP."

Ultimately, the doctrine of foreign equivalents as it stands may force or influence the rejection of a mark that could at worst confuse a minute fraction of the American marketplace. By rejecting such marks, the doctrine overestimates the likelihood of confusion to the marketplace and constricts potential information gains for consumers. This leads us to the third problem with the doctrine's current state; namely, constrictions on mark holders' ability to present information to the marketplace through creative trademarks.

C. Inhibition of Imaginative and Communicative Branding

The third problem with the doctrine's current state relates to the alternative justifications for trademark protection discussed in Part I, particularly the promotion

¹¹⁹ *Id.*; *State & County QuickFacts*, U.S. CENSUS BUREAU, <https://www.census.gov/quickfacts/table/PST045216/00> (last visited Mar. 1, 2017). Interestingly, the number of Hawaiian speakers approximates the number of Gaelic speakers in the United States at .008 of one percent of the population as found in the "ÓGRA" case. U.S. CENSUS BUREAU, DETAILED LANGUAGE SPOKEN AT HOME AND ABILITY TO SPEAK ENGLISH FOR THE POPULATION FIVE YEARS AND OLDER BY STATES: 2006–2008 (2010).

¹²⁰ *In re Christopher A. Fahey, DBA Gravity Guitar Picks*, No. 86250337, at *3.

¹²¹ *In re Savisa (Pty) Ltd.*, No. 78154196, 2005 WL 548058, at *3 (T.T.A.B. Feb. 24, 2005) (not precedential); Rest, *supra* note 5, at 1242.

of communicative branding.¹²² The doctrine leaves little to no room for imaginative branding because it focuses exclusively on the source identification function of trademarks and neglects the benefits of trademark protection for alternative reasons. Its one-track focus on consumer protection, while desirable to an extent, detracts from the ability of firms to employ creative branding strategies that convey product characteristics to consumers. In other words, the doctrine does little to protect firms' branding desires and investments or consumers' ability to distinguish goods based on a mark's communicative elements.

To be fair, this shortcoming is not unique to the doctrine of foreign equivalents; consumer protection continues to largely dominate the focus of trademark law.¹²³ Yet foreign-language marks are a particularly significant type of trademark in today's globalized culture, and language can provide consumers with insights into the characteristics of a product. Marketing literature has examined extensively phenomena such as the "country-of-origin effect," which highlights the effectiveness of drawing comparisons between the trademarked products and the qualities, characteristics, or emotions associated with a particular foreign country.¹²⁴ Branding has an emotional appeal to consumers, and by stimulating particular emotions or memories, a brand can communicate values that attract consumers.¹²⁵ Foreign languages, like certain geographical names, evoke emotions or associations that resonate with consumers.¹²⁶

The concept of "sound symbolism" further helps to explain how trademarks convey product characteristics to consumers.¹²⁷ Consumers consciously or subconsciously associate certain sounds with desirable qualities.¹²⁸ The letter "S" connotes quickness and ease of use for mops, such as with "SWIFFER."¹²⁹

¹²² See discussion *supra* Part I.A.1.ii.

¹²³ See discussion *supra* Part I.A.1.i.

¹²⁴ See generally Keith Dinnie, *Country-of-Origin 1965-2004: A Literature Review*, 3 JOURNAL OF CUSTOMER BEHAVIOUR 165 (2004).

¹²⁵ Griffiths, *supra* note 20, at 255–58.

¹²⁶ Cf. Alan L. Durham, *Trademarks and the Landscape of Imagination*, 79 TEMP. L. REV. 1181, 1183 (2006) ("Marketers today emphasize the importance of emotion or fantasy in selling even ordinary goods, and geographic names are a potent source of ready-made associations.").

¹²⁷ See Jake Linford, *Are Trademarks Ever Fanciful?*, 105 GEO L.J. (forthcoming 2017), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2732582&download=yes (arguing that "fanciful" marks on the *Abercrombie* spectrum are strategically selected by firms because of their sound symbolism, and that consequently courts should resist granting protection for such marks to prevent anticompetitive effects).

¹²⁸ *Id.* (manuscript at 5).

¹²⁹ *Id.* (manuscript at 31).

“FROSH” sounds “smoother, richer, and creamier” than “FRISH” for ice cream.¹³⁰ Likewise, by drawing from easily identifiable linguistic patterns from certain languages, the phonetic composition of trademarks such as “OIKOS” or “HÄAGEN-DAZS” can convey significant qualities of their respective products, even if American consumers are unable to translate those words to gain a precise meaning. Tapping into these common linguistic patterns can facilitate the exchange of information in the marketplace through trademarks.

Of course, one might be hesitant to give firms and their marketing departments such power. The question arises whether creative and communicative branding provides a sufficient advantage to warrant increased protection for firms’ trademarks. One skeptical commentator remarks:

In considering how far branding operates in the public interest and justifies legal protection, it is necessary to consider both what it achieves for consumers and how it influences the behaviour of firms. Having exclusive control over a source of emotional appeal to consumers can give a firm significant market power and increase its bargaining power within the streams and value chains in which it operates.¹³¹

However, despite such potential objection, strong branding provides consumers with a variety of economic and social benefits, ranging from lower transaction costs in the marketplace to expression of social values.¹³² Moreover, any notion that communicative branding only confuses or misdirects consumers assumes that consumers are largely unable to interpret information on packaging. It again fails to consider that such branding may indeed help consumers learn about the general features of a product rapidly. Consider the earlier “FRUTTO DELLA TERRA” example for tomato sauce. The name may evoke an idea of natural and fresh ingredients that may signal to the consumer such qualities in the product. Moreover, the Italian language may connote to the consumer, consciously or

¹³⁰ *Id.* (manuscript at 32).

¹³¹ Griffiths, *supra* note 20, at 241. *See also* Linford, *supra* note 127 (manuscript at 49) (“Component sounds can convey meaning about product features, and picking the right sounds can increase product desirability. Failing to account for sound symbolism may therefore lead to a level of protection for fanciful marks that imposes unacceptable costs on competitors.”).

¹³² Jerre B. Swann, *An Interdisciplinary Approach to Brand Strength*, 96 TRADEMARK REP. 943, 949–54 (2006).

subconsciously, that the product will enhance their next Sunday spaghetti dinner with rich flavor.¹³³

Both the firm and the consumer benefit from this communication. Consider, for example, the Greek yogurt trademarked “OIKOS,” a Greek word meaning “house.” Produced by Dannon, the American subsidiary of French company Danone, OIKOS entered the Greek yogurt market in 2010.¹³⁴ Why did the American subsidiary of a French company choose a Greek trademark for its yogurt? Greek-style yogurt has surged in popularity among American yogurt consumers, rising from one percent of the yogurt market in 2007 to over fifty percent of the market in 2015.¹³⁵ Consumers need not understand Greek—indeed, they need not be able to translate the word “oikos”—to appreciate that it is a Greek word. What matters is that the general sight, sound, and recognition of “oikos” as a Greek word differentiates “OIKOS” yogurt from other yogurts in the dairy aisle, including Dannon’s other yogurt varieties. “OIKOS” designates that the product is Greek-style yogurt because it has a Greek-style name.¹³⁶ Although it is impossible to state conclusively that the word mark *alone* conveys the Greek yogurt characteristics of the product without survey evidence, Dannon’s decision to use a Greek word remains rather insightful because it parallels the behavior of other firms.¹³⁷ Marks

¹³³ See ROGER SHUY, *LINGUISTIC BATTLES IN TRADEMARK DISPUTES* 36 (2002).

¹³⁴ *Our Heritage*, DANNON, <http://www.dannon.com/our-history/> (last visited Jan. 16, 2016).

¹³⁵ Carey Polis, *Greek Yogurt Food Invasion Continues As Product Is Added To Cream Cheese, Hummus And More*, HUFFINGTON POST (Mar. 5, 2013, 9:31 AM), http://www.huffingtonpost.com/2013/03/05/greek-yogurt-food-product_n_2807818.html; Hal Conick, *US Greek Yogurt Market to Reach \$4bn by 2019: Technavio*, DAIRYREPORTER.COM (Nov. 19, 2015, 9:22 AM), <http://www.dairyreporter.com/Trends/Greek-Yogurt-Revolution/US-Greek-yogurt-market-to-reach-4bn-by-2019-Technavio>.

¹³⁶ While one might argue that such inventive branding might *misinform* consumers by giving the false impression of the geographic source of the products, such concerns are unwarranted. The OIKOS products clearly state that the product is not made in Greece, and the Dannon website reinforces that point. *Frequently Asked Questions*, DANNON, <http://www.oikosyogurt.com/what-is-greek-yogurt/faq.aspx> (last visited Jan. 16, 2016) (“Though Oikos Greek yogurt is not produced in Greece, it is made according to the traditional Greek manner of production, which removes some of the whey through straining, leaving more of the milk solids.”).

¹³⁷ Interestingly, however, Dannon’s Vice President of Marketing noted that the company was careful not to overemphasize Greece in marketing because a Greek origin suggests to some consumers the idea of a product being “too old, so it’s not modern.” E.J. Schultz, *Dannon Goes Greek, Takes on (Former) Little Guy*, ADVERTISINGAGE (Sept. 26, 2011) <http://adage.com/article/cmo-interviews/dannon-s-greek-yogurt-oikos-sights-set-1-chobani/230009/>.

such as “HÄAGEN-DAZS”¹³⁸ and “FRUSEN GLÄDJÉ,”¹³⁹ both used for ice cream, are designed by firms to sound foreign, and consequently to attract consumers based on the associations between certain products, sounds, and places of origin.¹⁴⁰

The doctrine of foreign equivalents restricts such opportunities for creative branding by shifting marks that are inherently distinctive to the vast majority of consumers to the “descriptive” *Abercrombie* category, rendering them ineligible for protection without a showing of acquired distinctiveness. For example, Hormel was unable to gain registration for the mark “SAPORITO” for sausages because the USPTO found the word to mean “tasty” in Italian, and classified it as descriptive without secondary meaning.¹⁴¹ Likewise, Tokutake, a Japanese shoe firm, was unable to register “AYUMI” for footwear because the TTAB found that its translation, “walking,” was merely descriptive.¹⁴² In yet another case, the TTAB rejected an application for “MARCHE NOIR” for jewelry on the grounds that it was confusingly similar to the mark “BLACK MARKET” in the same class of goods.¹⁴³ The Board stated that although the marks are “decidedly different in sound and appearance, [they] have the same connotations.”¹⁴⁴

The law currently fails to recognize the costs of such barriers to registration and the benefits of creative and communicative branding to both firms and consumers. As a result, firms are often prevented from using marks with strong sound symbolism to distinguish the source or qualities of their products,¹⁴⁵ and consumers lose the opportunity to make more knowledgeable purchasing decisions based on trademarks that communicate information about elements of the product and assurances of quality.¹⁴⁶

¹³⁸ *Our Story*, HÄAGEN-DAZS, <http://www.haagendazs.us/Learn/History/> (last visited Feb. 13, 2017) (“Mr. Mattus, supported by his wife Rose, decided to form a new company dedicated to his ice cream vision. He called his new brand Häagen-Dazs, to convey an aura of the old-world traditions and craftsmanship, values which he held close to his heart.”).

¹³⁹ Alix M. Freedman, *Rolls-Royces of Ice Cream*, N.Y. TIMES, Feb. 22, 1981, <http://www.nytimes.com/1981/02/22/business/rolls-royces-of-ice-cream.html>.

¹⁴⁰ See Linford, *supra* note 127 (manuscript at 29).

¹⁴¹ *In re Geo. A. Hormel & Co.*, 227 U.S.P.Q. 813 (T.T.A.B. 1985).

¹⁴² *In re Tokutake Indus.*, 87 U.S.P.Q.2d 1697 (T.T.A.B. 2008).

¹⁴³ *In re Thomas*, 79 U.S.P.Q.2d 1021, 1028 (T.T.A.B. 2006).

¹⁴⁴ *Id.* at 1025.

¹⁴⁵ See discussion *supra* Part I.A.1.

¹⁴⁶ Griffiths, *supra* note 20, at 240. *But see* Ralph S. Brown, Jr., *Advertising and the Public Interest: Legal Protection of Trade Symbols*, 57 YALE L.J. 1165, 1186 (1948) (“Since the user of the symbol probably guarantees by it nothing more than his hope that the buyer will come back for more, the term smacks strongly of the ad-man’s desire to create the illusion of a guarantee

III

REEVALUATING THE DOCTRINE AND PROPOSING SOLUTIONS

This section argues that the doctrine of foreign equivalents has merit and should continue to guide the registration of foreign-language marks. However, Congress or the USPTO ought to refine the doctrine so that it may provide optimal results for both producers and consumers. First, this section will argue that the doctrine should apply as a rule rather than as a guideline. Currently, the doctrine is applied only when it is likely that the “ordinary American purchaser would ‘stop and translate the foreign word into its English equivalent.’”¹⁴⁷ Examiners and courts should apply the doctrine in *all* cases involving a non-English word absent from a trusted and recognized English-language dictionary; words with foreign origins that have entered the American lexicon such as “sombrero” and “carte blanche” should be treated as English words.¹⁴⁸ Second, Congress or the USPTO must reconsider the doctrine’s ability to assess informational gains or losses to the marketplace by taking as the defined demographic not only those bilingual consumers who might stop and translate a non-English mark, but the entire American market for the product.¹⁴⁹ As it stands, the doctrine’s application leads to instances in which words from uncommon foreign languages are barred from registration, despite the fact that not many consumers would take the mark as non-distinctive of source. While its stated purpose—the protection of immigrant and tourist populations from confusion in the marketplace—is laudable, its consequences are entirely inconsistent with the typical registration standards and impose high costs on firms and other consumers. This note proposes that as a compromise, the doctrine allow greater flexibility for non-English descriptive marks to be registered in the United States, while still barring generic non-English words.¹⁵⁰ There exists both foreign and domestic precedent for such an approach, one that would better respect the balance of interests involved in the registration process while also simplifying the registration and review process. Third, and relatedly, although the justifications and purposes of trademark law are primarily consumer-based, trademark law ought to acknowledge the potential benefit to trademark holders stemming from a mark’s ability to signal

without in fact making more than the minimum warranty of merchantable quality.”). *See also* discussion *supra* Part II.B.

¹⁴⁷ *E.g.*, *Palm Bay Imports, Inc. v. Veuve Clicquot Ponsardin Maison Fondée En 1772*, 396 F.3d 1369, 1377 (Fed. Cir. 2005).

¹⁴⁸ *See, e.g.*, MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY (11th ed. 2014); OXFORD ENGLISH DICTIONARY, <http://www.oed.com/>. The TMEP already offers this step. TMEP § 809.01(b)(i) (Jan. 2017).

¹⁴⁹ *Rest, supra* note 5, at 1244.

¹⁵⁰ This refers to marks that would otherwise be “descriptive” on the *Abercrombie* spectrum.

particular qualities of their goods via foreign-language words. Such imaginative product differentiation would be an additional benefit from allowing registration of some descriptive non-English terms.

A. A Rule is Preferable to the Current Guideline

One might question why it is necessary to keep the doctrine of foreign equivalents at all. Why translate a word to English when it is possible to assess its distinctiveness independently, that is, as the word appears on the packaging? The importance of the doctrine relates to the international comity rationale,¹⁵¹ but extends further than that. Some application of the doctrine is necessary to prevent truly harmful consumer confusion.

While it is possible to justify the protection of descriptive terms in some cases, there is a legitimate public interest in preventing the registration of generic terms in *all* cases. This is because removing a generic word from the public commons to the sole possession of a trademark holder unduly inhibits speech and poses a high likelihood of confusion in the marketplace.¹⁵² The only way to prevent registration of generic foreign-language terms is to translate the words to a language understood by trademark examiners and judges before assessing their distinctiveness. Moreover, in infringement claims where the plaintiff's mark is both conceptually strong (i.e., arbitrary or fanciful) and famous, there is compelling reason to prevent registration of marks that will confuse bilingual speakers.

"MANZANA," the Spanish equivalent of "APPLE," for computers and technology products, would be one such example. Consumers who understand both the Spanish word "manzana" and the English word "apple" might be confused as to the source of the products, especially because "APPLE" is a conceptually strong, arbitrary mark.¹⁵³ Employing the mark "MANZANA" for computers would be a cheeky attempt to free-ride on the goodwill encapsulated by "APPLE." Of course, the overall ceiling of potential confusion in such cases may still be extremely low; that is where substantive rethinking of the doctrine is necessary.¹⁵⁴

Ultimately, a rule is better suited to focus trademark examiners and courts on the core issues the doctrine of foreign equivalents aims to address. Trademark law

¹⁵¹ See discussion *supra* Part I.B.2.

¹⁵² Jared Stipelman, *A Failure to Communicate: How Linguistics Can Inform Trademark Law*, 42 AIPLA Q.J. 69, 99 (2014) (offering the example of an immigrant consumer purchasing a branded product based on a generic trademark on the false assumption that it is cheaper).

¹⁵³ It is conceptually strong per the *Abercrombie* spectrum because "APPLE" is arbitrary for the computers and technology classes of goods.

¹⁵⁴ See discussion *infra* Part III.B.

must rein in application of the doctrine of foreign equivalents and use it to invalidate mark registrations only when necessary in order to strike the proper balance between confusion and information.¹⁵⁵ Therefore, a rule is preferable to the current guideline, which allows for over-application of the doctrine in some instances.

1. Benefits of Adopting a Foreign Equivalents Rule

Treating the doctrine of foreign equivalents as a mandatory rule, rather than a guideline, would eliminate misinterpretation and the disparate results by making the doctrine easier to apply, and therefore more predictable for firms or individuals registering marks. A mandatory rule must apply to every scenario in which stated conditions are met.¹⁵⁶ A guideline, on the other hand, invites interpretation by examiners and judges. While this may offer versatility for considering unique features of individual cases, the costs of uncertainty are high.¹⁵⁷ An alternative to making the doctrine a strict rule would be to keep the doctrine a guideline while creating more consistent definitions and standards.¹⁵⁸ However, this still leaves a significant, unpredictable decision to the examiner or judge: whether the doctrine should apply in the first place, based on whether consumers are likely to stop and translate a mark. By confining the universe of easily recognizable English-language words to those found in selected dictionaries and translating *all* marks absent from them, registrants can determine whether their mark will be translated before the mark is assessed for distinctiveness or likelihood of confusion. Registrants can then predict whether their marks will gain protection based on the clarified positions explained next.

2. Features of the Proposed Rule: Standardizing Translation Methods and Adopting a Confusion Threshold

To treat the doctrine as a rule, it is necessary to solidify the terms of the doctrine and how to apply them. First, if the doctrine is to apply, then it is desirable, for the sake of predictability and consistency, that a defined list of references assist in the determination of whether a word is part of the English lexicon, and subsequently, in the translation of non-English words to English. While one might

¹⁵⁵ *But see* MCCARTHY, *supra* note 67, at § 11:35 (“A rigid, unthinking application of the “doctrine” of foreign equivalents can result in a finding quite out of phase with the reality of customer perception.”).

¹⁵⁶ *Rule*, BLACK’S LAW DICTIONARY (10th ed. 2014).

¹⁵⁷ Stipelman, *supra* note 152, at 86 (“The DFE is thus meant to vindicate ambivalent policies, governed by a vague standard, and is applied by judges who, by and large, are not privy to special insights regarding what constitutes translation. The problem is emblematic of issues in trademark law writ large.”).

¹⁵⁸ Rest, *supra* note 5, at 1223.

argue that granting examiners and judges a broad ability to consult myriad sources encourages exhaustive research for word meaning, the reality is that such wide latitude in choosing translation references results in dependence on often unreliable sources and overly-restrictive registration decisions. For example, the Second Circuit held on remand that the Southern District of New York should consider the popular meaning of the mark “OTOKOYAMA” for sake in Japan.¹⁵⁹ However, considering popular meaning opens possibilities for introduction of just about any evidence to prove use of the word without reliable, verifiable support from a dictionary or translation service.¹⁶⁰ Examiners and judges should consult dictionaries and translation services that are reputable for their accuracy.¹⁶¹ Even slang words and formerly trademarked words that have entered the American lexicon by ‘genericide’ are now present in dictionaries such as The Oxford English Dictionary and The Merriam-Webster Dictionary.¹⁶² As such, trademark examiners should consult only trusted dictionaries and translation services to determine whether a word is part of the American lexicon or inherently distinctive. The TMEP can specify exactly which references an examiner may cite.¹⁶³ Creating a specific list of translation references will guide registrants and assist them to predict more accurately whether their marks may gain registration.

Next, the law would benefit from a clear statement of the acceptable threshold of potential confusion to enable examiners and courts to make more consistent decisions on which trademark applications are eligible for protection. The Board has explained correctly that to require infringement-level confusion of ten percent or more effectively “would write the doctrine out of existence.”¹⁶⁴ Nevertheless, it is crucial for the integrity of the rule to establish a minimum percentage of the national

¹⁵⁹ *Otokoyama Co. v. Wine of Japan Imp., Inc.*, 175 F.3d 266, 272 (2d Cir. 1999).

¹⁶⁰ While one may argue that this places too much emphasis on dictionaries, dictionaries remain the most reliable reference for what constitutes a word in a given language. *Cf.* Landes & Posner, *supra* note 14, at 294 (suggesting that dictionaries are “accurate inventor[ies] of words in general use by the relevant publics” to determine whether a word has become generic).

¹⁶¹ For translation functions, services such as “Gengo” offer a far more reliable and professional translations than free online translators, and at extremely low prices – often less than ten cents per word.

¹⁶² Examples include “aspirin,” “cellophane,” “twerk,” and “selfie.” OXFORD ENGLISH DICTIONARY, <http://www.oed.com/>.

¹⁶³ Although some courts have looked beyond reputable dictionaries such as Merriam-Webster’s and the Oxford English Dictionary to sources such as “Urban Dictionary,” this practice has obvious shortcomings. Leslie Kaufman, *For the Word on the Street, Courts Call Up an Online Witness*, N.Y. TIMES, May 21, 2013, at A1.

¹⁶⁴ See *In re Thomas*, 79 U.S.P.Q.2d 1021, 1024 (T.T.A.B. 2006). Indeed, at least one commentator believes that for this very reason the doctrine should cease to exist. Krimnus, *supra* note 79, at 161.

population that speaks the foreign language to warrant translation. Determining the minimum population percentage could induce debate, but even an arbitrary figure would be better than no figure at all for the doctrine's consistency and registrants' ability to predict outcomes.

The point is best illustrated with numbers. Spanish is the most popular language after English in the United States, and the current Spanish-speaking population comprises approximately thirteen percent of the American population.¹⁶⁵ Chinese is the second-most popular non-English language after Spanish;¹⁶⁶ yet Chinese speakers comprise only about one percent of the population.¹⁶⁷ This means that other "common" languages are recognizable by less than one percent of the total population, and by extension, generic or descriptive foreign-language words in those languages are capable of causing a maximum of one percent confusion for products with a national market.¹⁶⁸ Instead of granting each individual examiner or reviewer the ability to determine whether a language is sufficiently common to warrant translation, a census-based threshold should be stated to provide trademark reviewers with a fixed guidance for what constitutes a "common" language. This note suggests that even a threshold that is by all accounts arbitrary—for example, 0.2 percent—would provide all parties with a set standard by which to predict the outcome of a foreign-word mark registration.

Another alternative would be to eliminate the "common language" requirement completely and simply translate all non-English words to English. There appears to be no material difference between a language recognized by 0.19 percent of the population as opposed to one recognized by 0.2 percent of the population that warrants differential treatment. Therefore, the advantage of this position is that there would be no claims of discrimination over minor differences in population percentages. However, note that the same problem arises whether we set a bright-line threshold for the number of speakers required to constitute a "common language" or we eliminate the "common language" requirement altogether. In either case the law would be refusing protection to foreign-language marks that are intelligible to, and consequently pose a maximum likelihood of confusion for, percentages of the American population that fall well below what is normally

¹⁶⁵ CAMILLE RYAN, U.S. CENSUS BUREAU, LANGUAGE USE IN THE UNITED STATES: 2011, at 3 (2013).

¹⁶⁶ *Id.* Note that the Census statistics group Mandarin and Cantonese into the same category, "Chinese."

¹⁶⁷ *Id.*

¹⁶⁸ This assumes, of course, that the relevant market for a particular product spans the entire country.

required for infringement claims.¹⁶⁹ This is where substantive changes to the doctrine of foreign equivalents are necessary.

B. Permitting Registration of Foreign Descriptive Terms

Beyond refining the doctrine of foreign equivalents by tightening its application, this section advocates for a significant substantive change to the doctrine: permitting the registration of foreign descriptive terms. If, as Part II of this note suggests, trademark law should focus on the present-day American marketplace and not on a future or hypothetical marketplace, then one improvement would grant a presumption of registrability for descriptive foreign-language marks. The only exception would be for cognates from common modern languages, which would lose the presumption but be eligible for registration with a showing of secondary meaning. Currently, the doctrine of foreign equivalents so overprotects consumers as to reject potential information gains, and inhibits mark holders' abilities to communicate features of their products through their marks. Extending protection to descriptive foreign-language marks would recalibrate the information-confusion balance and open avenues for effective, communicative branding.

1. Informational and Branding Advantages from Protecting Descriptive Foreign-Language Marks

The American doctrine would benefit from protection of descriptive foreign-language marks for two reasons. First, extending protection to descriptive foreign words would encourage registration of marks that provide high net informational gains to the marketplace, and would not bar from protection many marks that are inherently distinctive to a vast majority of the population.¹⁷⁰ Because in many instances non-English marks are understood by minute percentages of the American marketplace, even descriptive foreign-language marks are likely to cause negligible consumer confusion, and consequently offer information to consumers in the form of strong source identifiers.¹⁷¹ The only descriptive foreign-language marks that would lose the presumption of registrability are cognates from common languages.¹⁷² Such words so closely resemble their English counterparts that they ought to be included under the 'misspellings' rule, which the TMEP states as, "[a] slight misspelling of a word will not turn a descriptive or generic word into a non-

¹⁶⁹ See MCCARTHY, *supra* note 104, at § 23:2 (showing that a minimum of eleven percent confusion is typically necessary to win an infringement claim).

¹⁷⁰ See discussion *supra* Part II.B.

¹⁷¹ See discussion *supra* Part II.B.

¹⁷² Recall that languages would be deemed common if they meet an established threshold. See discussion *supra* Part III.A.2.

descriptive mark.”¹⁷³ The rationale behind this rule is that consumers will treat the marks as the correctly-spelled word, and when that happens, there is a potential for information loss or confusion. One may think of common cognates in the same way as misspelled words. In effect, the doctrine would treat cognates as misspelled English words for the purposes of a distinctiveness analysis because they are phonetic equivalents.¹⁷⁴ If the descriptive mark is clearly a cognate and loses the presumption of registrability, then the mark registrant must show secondary meaning as with any other descriptive mark in order to gain protection.

Second, opening the opportunity for registration of descriptive foreign language marks would promote imaginative branding efforts and recognize the current status of trademarks as communicative tools. As one scholar noted, “[t]he descriptive or near-descriptive sign has accordingly a kind of condensed power: it triggers mental effects that, if suitably guided, significantly enrich its resonance and make it more effective as a means of communication.”¹⁷⁵ Firms could use foreign-language marks to communicate characteristics of their products via linguistic sights and sounds, and consumers would perceive such marks as both indicators of source and quality communicators. Current highly-visible examples of firms that employ foreign-language marks as descriptors include the restaurant chains “PRET A MANGER” (“ready to eat” in French)¹⁷⁶ and “AU BON PAIN” (roughly translated to “place of good bread” in French),¹⁷⁷ as well as clothing retailer “SUPERDRY”¹⁷⁸ (in Japanese characters: 極度乾燥(しなさい)). The French names convey an aura of fine quality, while the Japanese characters and “SUPERDRY” name communicate high-performance clothing bearing “Japanese-inspired graphics.”¹⁷⁹

¹⁷³ TMEP § 1209.03(j) (Jan. 2017).

¹⁷⁴ See Stipelman, *supra* note 152, at 100–08 (explaining how an understanding of phonetic dissimilarities ought to permit the registration of descriptive foreign-language trademarks).

¹⁷⁵ Durant, *supra* note 28, at 131.

¹⁷⁶ Kim Bhasin, *Here’s What Pret A Manger’s President Thinks Every Time You Pronounce The Name Wrong*, BUSINESS INSIDER (Mar. 8, 2013, 4:38 PM), <http://www.businessinsider.com/how-to-pronounce-pret-a-manger-2013-3>.

¹⁷⁷ Leeann Cannon, *Au Bon Pain: Bakery-Café Weighs in on Diet Fads, Offers More Healthful Fare to Concerned Customers*, 39 NATION’S RESTAURANT NEWS, Jan. 31, 2005, at 28, available at EBSCO.

¹⁷⁸ James Hall, *How Superdry Became 2010’s £1bn Fashion Success Story*, THE TELEGRAPH (Dec. 27, 2010, 5:45 AM), <http://www.telegraph.co.uk/finance/newsbysector/retailandconsumer/8221598/How-Superdry-became-2010s-1bn-fashion-success-story.html>.

¹⁷⁹ *About Us*, SUPERDRY, <http://www.superdry.com/about-us> (last visited Apr. 6, 2016).

2. Precedent for Protecting Descriptive Foreign-Language Marks

Precedent exists for extending trademark protection to descriptive foreign-language marks. Two major foreign courts have demonstrated direct openness to offering such protection, and one could reasonably extend principles in existing American case law to cover descriptive foreign-language marks. While foreign cases have no binding effect on any American court, they remain theoretically instructive and offer an attractive alternative to the current application of the doctrine of foreign equivalents.¹⁸⁰ First, the European Court of Justice has supported protection for descriptive foreign-language marks, going so far as to uphold protection for a *generic* foreign-language mark. In *Matratzen Concord AG v. Hukla Gernany*,¹⁸¹ Hukla, a German company, held the registration for the “MATRATZEN” mark in Spain for bedding products. “MATRATZEN” translates to “mattresses” in German.¹⁸² When a competitor attempted to gain a Community Trade Mark for “MATRATZEN CONCORD,” Hukla opposed.¹⁸³ Seeking to turn the tables, Matratzen Concord attempted to cancel Hukla’s “MATRATZEN” mark in Spain.¹⁸⁴ The Court of First Instance rejected Matratzen Concord’s effort,¹⁸⁵ and the European Court of Justice affirmed the lower court, holding,

Article 3(1)(b) and (c) of the Directive does not preclude the registration in a Member State, as a national trade mark, of a term borrowed from the language of another Member State in which it is devoid of distinctive character or descriptive of the goods or services in respect of which registration is sought, unless the relevant parties in the Member State in which registration is sought are capable of identifying the meaning of the term.¹⁸⁶

Because the average Spanish consumer would not take “MATRATZEN” to be merely descriptive of the products to which the mark was attached, it could gain trademark protection. This holding was confirmed by *Bimbo SA v. OHIM* in 2012, a case involving the registration of “DOUGHNUT” in Spain.¹⁸⁷ Although there was

¹⁸⁰ See Roger P. Alford, *Federal Courts, International Tribunals, and the Continuum of Deference*, 43 VA. J. INT’L L. 675, 677 (2003).

¹⁸¹ Case C-421/04, *Matratzen Concord AG v Hukla Germany SA*, 2006 E.C.R. I-2306.

¹⁸² *Id.*

¹⁸³ *Id.*

¹⁸⁴ *Id.*

¹⁸⁵ *Id.*

¹⁸⁶ *Id.*

¹⁸⁷ Case C-591/12, *Bimbo SA v. OHIM*, 2014 E.C.R. 305.

a proposal to overturn the *Matratzen* decision,¹⁸⁸ recent amendments to the Community Trade Mark Regulations in December 2015 opted not to do so.¹⁸⁹

In similar fashion, the Australian High Court recently accepted the registration and protection of descriptive foreign-language marks. The High Court stated that although a word mark should be “substantially different from any word in ordinary and common use...[it] need not be wholly meaningless and it is not a disqualification that it may be traced to a foreign source or that it may contain a covert and skilful allusion to the character or quality of the goods.”¹⁹⁰ In *Cantarella Bros. Pty Ltd. v. Modena Trading Pty. Ltd.*,¹⁹¹ Cantarella had registered the marks “ORO” and “CINQUE STELLE” for coffee products. The marks translate to “gold” and “five-star” in Italian, and approximately 1.5 percent of the Australian population speaks Italian.¹⁹² In analyzing the mark, the Court sought to ascertain the “ordinary significance” of the mark to Australian consumers and whether it was “inherently adapted to distinguish” the source of the goods – that is, whether it was distinctive. The High Court concluded that “ORO” and “CINQUE STELLE” are inherently adapted to distinguish the source of Cantarella’s coffee because they do not “convey a meaning or idea sufficiently tangible to anyone in Australia concerned with coffee goods as to be words having a direct reference to the character or quality of the goods.”¹⁹³

Beyond these foreign cases, there is a viable justification within the current American registration precedent to support registration of foreign descriptive terms. Recall that descriptive marks are not per se unprotectable. However, they typically

¹⁸⁸ *Commission Proposal for a Regulation of the European Parliament and of the Council Amending Council Regulation No. 207/2009 on the Community Trade Mark*, at 10, COM (2013) 161 final (Mar. 27, 2013); see also James Nurton, *EU Trade Mark Proposals Sound Wrong, In Any Language*, MANAGING INTELLECTUAL PROPERTY BLOG (June 25, 2013), <http://www.managingip.com/Blog/3223159/EU-trade-mark-proposals-sound-wrong-in-any-language.html>.

¹⁸⁹ Council Regulation 2015/2424 amending Council Regulation (EC) No. 207/2009 on the Community Trade Mark and Commission Regulation (EC) No. 2868/95 Implementing Council Regulation (EC) No. 40/94 on the Community Trade Mark, and Repealing Commission Regulation (EC) No. 2869/95 on the Fees Payable to the Office for Harmonization in the Internal Market (Trade Marks and Designs), 2015 O.J. (L 341) 58.

¹⁹⁰ *Cantarella Bros Pty Ltd. v Modena Trading Pty. Ltd.* [2014] HCA 48 (Austl.) (internal quotations omitted).

¹⁹¹ *Id.*

¹⁹² Australian Bureau of Statistics, *2011 Census QuickStats*, http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/0 (last updated Mar. 28, 2013).

¹⁹³ *Cantarella Bros* [2014] HCA 48.

require a showing of secondary meaning.¹⁹⁴ Despite the acquired distinctiveness restriction on descriptive marks, American trademark law has already on several occasions permitted the registration of multilingual terms or phrases that would certainly seem descriptive, and perhaps even generic, without such a showing. For instance, the USPTO granted the marks “LA YOGURT” and “LE CASE” for yogurt and briefcases, respectively.¹⁹⁵ Contrast these trademarks with “LA LINGERIE,” “LE SORBET,” and “LE CROISSANT SHOP,” to each of which the USPTO denied registration.¹⁹⁶ The primary difference between the first and second sets of marks are the correctness of the two-letter articles in each group. In French, the correct article for “yogurt” is “le,” not “la.”¹⁹⁷ Although the courts justified the registrations of “LA YOGURT” and “LE CASE” on the grounds that they presented unique commercial impressions, such findings are highly questionable for two reasons. First and most fundamentally, a mark is descriptive if it describes any feature of a product. Consumers will still glean from these marks the products to which they are attached: certainly, “LA YOGURT” is in some sense descriptive of the product to which it is attached. Indeed, these marks contain the *generic*, genus terms used to describe their respective products. Second, slight misspellings ought not to change the analysis per the TMEP.¹⁹⁸ Thus, in each of these instances the USPTO implicitly acknowledges that there is indeed something unique about marks employing a non-English term that warrants a distinct analysis, even when there is a significant chance that consumers will take the mark as descriptive or generic. Protecting wholly descriptive foreign-language marks does not require much of a step from such current TTAB precedent.

C. Applying the Proposed Rule: “DELICIOSA” Versus “NÓSTIMA”

To envisage how this note’s proposal for a revised treatment of descriptive foreign-language marks would operate, consider two examples. First, imagine a brand of salsa marketed nationally under the trademark “DELICIOSA.” Assuming that this word is absent from the established list of English language dictionaries, under this note’s recommended analysis the examiner should assess the distinctiveness of the word based on the entire national market. Since many people in the United States speak Spanish, it will most certainly cross the threshold for

¹⁹⁴ *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, 7, 9 (2d Cir. 1976).

¹⁹⁵ *In re Johanna Farms, Inc.*, 222 U.S.P.Q. 607 (T.T.A.B. 1984); *In re Universal Package Corp.*, 222 U.S.P.Q. 344 (T.T.A.B. 1984).

¹⁹⁶ Note, however, that even this part of the doctrine is sloppy, as the Board recently refused registration for “Uchi Sweets Café.” *In re Kabushiki Kaisha Lawson*, No. 79138995, 2015 WL 5579957 (T.T.A.B. Aug. 25, 2015).

¹⁹⁷ *In re Johanna Farms, Inc.*, 222 U.S.P.Q. at 607.

¹⁹⁸ TMEP § 1209.03(j) (Jan. 2017).

qualifying as a common language. This means that the examiner should translate the word to English using an accepted, established reference or translation service. “DELICIOSA” will translate to “delicious” in English, which is descriptive of salsa. This means that “DELICIOSA” would initially gain a presumption of registrability under this note’s framework; however, since “DELICIOSA” for salsa clearly is a cognate of “delicious,” it should lose the presumption. Now apply the same analysis to “NÓSTIMA,” which translates to “delicious” in Greek, as it applies to pita bread. The examiner here would have to determine whether Greek is a common language in the United States. If she does, and finds that “NÓSTIMA” translates to “delicious,” the mark will gain a presumption of registrability since the two words are not cognates.

D. Addressing Competition Concerns through Descriptive Fair Use and Reexamination

It is important at this stage to address an important critique that likely will arise in response to this note’s proposal. One might argue that the registration of descriptive foreign-language marks at the present time would, in the future, undercut the consumer protection purposes of trademark law. For example, one writer has critiqued the European Court of Justice’s *Matratzen* decision using the following hypothetical: “Under the current policy for Community trademarks, “Ha-Lush-Ka” could be registered in the European Union and retain trademark rights despite its generic nature after enlargement of the Union to include Hungary. Such an occurrence would seem to present problems for the free movement of goods in the future.”¹⁹⁹ Beyond the argument that the trademark law should focus on the *current* American market,²⁰⁰ which this section previously has addressed, this critique is inadequate for two reasons.

First, the descriptive fair use doctrine could apply to foreign-language marks.²⁰¹ Descriptive fair use functions as a defense to trademark infringement

¹⁹⁹ Eric E. Bowman, *Trademark Distinctiveness in A Multilingual Context: Harmonization of the Treatment of Marks in the European Union and the United States*, 4 SAN DIEGO INT’L L.J. 513, 526 (2003).

²⁰⁰ If the TTAB’s concern is that marks need to be distinctive for future American markets, then ‘genericide’ shows how to address the problem. If, for example, the Gaelic-speaking population of the United States increases to a point at which it comprises an appreciable portion of the purchasing public, then non-inherently-distinctive Gaelic marks could forfeit their registration just like English marks that have lost their protection to ‘genericide.’ Cf. Linford, *supra* note 56, at 110.

²⁰¹ Descriptive fair use is codified in 15 U.S.C. § 1115(b)(4). “[T]he right to use the registered mark . . . shall be subject to the following defenses or defects . . . That the use of the name, term, or device charged to be an infringement is a use, otherwise than as a mark, of the party’s individual

claims, even claims against marks that have achieved incontestable status. It permits Person X to use Person Y's trademark to describe Person X's goods, so long as Person X's use is not descriptive of Person Y's product, and is in good faith.²⁰² This is important because it means that if one trademarks a descriptive term from a foreign language, and a foreign competitor enters that market, the foreign competitor would not be subject to a claim of trademark infringement for the use of the descriptor. For example, imagine that under this note's framework, Hormel reapplies for the "SAPORITO" mark to apply to sausages.²⁰³ The examiner translates the mark from Italian to English, finding it means "tasty." "Tasty" undoubtedly is descriptive of sausage, yet under the proposed solution, foreign descriptive terms that are not cognates of English words may gain protection. Therefore, Hormel would gain rights in the "SAPORITO" mark for sausages. Soon after, an Italian firm begins to export its sausages to New Jersey under the mark "SALSICCIA DI MERANTE." On the side of the packaging, the Italian firm writes, "Il nostro prodotto è dolce, piccante e saporito."²⁰⁴ Under the descriptive fair use doctrine, since the Italian firm is not using the word "saporito" as a trademark, but as a descriptor, presumably in good faith, the Italian firm would not be liable for trademark infringement. Hormel would have the advantage of using the "SAPORITO" mark, but other firms are not prohibited from using the same descriptive term in good faith. Consumers would not be confused because they would understand one term, when seen with the Hormel house mark, as an indicator of source, and the other as descriptive.²⁰⁵

The critique is inadequate for a second reason. If a foreign-language mark wrongfully gains registration (that is, if it is generic, yet passes examination in the USPTO) it may, like any other mark in the Register, face cancellation and lose its registration. In fact, the TTAB applied the doctrine of foreign equivalents in 2009 for this very purpose.²⁰⁶ It found the registered mark "AZUCAR MORENA" to be

name in his own business, or of the individual name of anyone in privity with such party, or of a term or device which is descriptive of and used fairly and in good faith only to describe the goods or services of such party, or their geographic origin."

²⁰² *E.g.*, *KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc.*, 543 U.S. 111, 118 (2004).

²⁰³ Hormel has since claimed common law rights in the mark. *See Saporito Prosciutto, Boneless, 2-1/2 PC*, HORMEL, <http://www.hormelfoodservice.com/products/saporitotm-prosciutto-ham-bone-out-2half-pc/> (last visited Apr. 3, 2016).

²⁰⁴ This translates to "Our product is sweet, spicy, and tasty."

²⁰⁵ This assumes that the Hormel house mark is prominent and recognized by consumers.

²⁰⁶ *Marquez Bros. Int'l, Inc. v. Zucrum Foods, LLC*, Cancellation No. 9204826, 2009 WL 4956033, at *7 (T.T.A.B. Dec. 11, 2009).

generic for brown sugar, and consequently cancelled the mark.²⁰⁷ This is an instance where the doctrine worked extremely well: Spanish is a common language, and many American consumers would understand “AZUCAR MORENA” to designate the genus of brown sugar. Therefore, it is likely that many consumers would translate “AZUCAR MORENA” to English and be confused as to whether it was a source designator or not. Moreover, it is not sensible to grant one producer of brown sugar a linguistic monopoly over the Spanish translation of its genus of goods.

CONCLUSION

As the world grows more connected and markets expand, products in the American marketplace increasingly bear foreign trademarks or domestic marks that seek to communicate ‘foreign’ characteristics. Despite this economic connectivity, trademark law must continue to operate on a localized level until the entire world comes to recognize the same terms as generic, descriptive, or distinctive. Excessive trademark protection holds the potential to inhibit competition, information, and even free speech rights; however, treating trademarks with an enhanced proprietary status does not necessarily require such an ill-fated result. Indeed, for the sake of intellectual honesty and practical necessity, the law ought to formally recognize the purpose and value of trademarks beyond their source-identification functions. The doctrine of foreign equivalents in particular is one area of the trademark law that calls for a more nuanced understanding of contemporary trademarks.

Foreign-language marks have immense branding potential, but this potential will never be realized if prospective mark registrants are wary of investing in a mark that may be rejected or face excessive word choice restrictions at the PTO. Thus, revising the doctrine of foreign equivalents in both form and substance would benefit consumers and firms alike. Consumers benefit from the source communicating information the newly registrable marks would offer and would gain new tools by which to distinguish products on the market. Firms such as Pret A Manger, Au Bon Pain, and Häagen-Dazs, among others, have shown that there is interest in employing foreign language terms that communicate their products’ qualities to consumers in creative ways. The law ought not inhibit these information exchanges as it does through its current doctrine.

²⁰⁷ Indeed, with the exception of a lack of a diacritical mark over the “U” in “AZUCAR,” this is a direct translation of brown sugar. The Board stated that this minor difference did not create a distinct commercial impression, and therefore was insignificant. *Id.* at *7 n.3.

Implementation of this note's proposal would require a focused effort by Congress or the USPTO to revisit the advantages of foreign-language marks in the American marketplace.

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 6

SPRING 2017

NUMBER 2

FITTING MARRAKESH INTO A CONSEQUENTIALIST
COPYRIGHT FRAMEWORK

NEIL YAP*

The Marrakesh Treaty to Facilitate Access to Published Works by Visually Impaired Persons and Persons with Print Disabilities entered into force on September 30, 2016. The treaty aims to alleviate what has been described as the “book famine,” and has been lauded as a significant achievement in advancing the rights of and promoting equal opportunity for the visually disabled. Contracting states are required to implement copyright limitations and exceptions to facilitate access to copyrighted material for the global print-disabled community. This note will argue that, notwithstanding the treaty’s strong rights-based underpinnings, the treaty aligns comfortably with U.S. consequentialist copyright justifications. This note will also demonstrate the limitations of other copyright justificatory theories while discussing their incompatibility with the treaty’s philosophy.

* LLM candidate, New York University School of Law, 2017. LLB (First Class), University College London. The author would like to thank Professors Christopher Sprigman and Barton Beebe for their comments and for conducting the “Theories of IP” seminar for which this note was submitted. The author is grateful for the encouragement and feedback from fellow JIPEL Notes Editors: Julian Pymonto, Gia Wakil, Ryan Jin and Vincent Honrubia. All errors and views expressed are the author’s own.

INTRODUCTION.....	352
I. COPYRIGHT JUSTIFICATIONS AND MARRAKESH IMPLEMENTATION	355
A. <i>Philosophical Justifications of Copyright in the U.S.</i>	355
1. <i>The Dominance of Consequentialism</i>	355
2. <i>Consequentialism in Fair Use</i>	356
3. <i>Labor, Personality and Rawlsian Justifications</i>	357
B. <i>From Marrakesh to America</i>	360
1. <i>History and Development of the Marrakesh Treaty</i>	360
2. <i>The Marrakesh Treaty's Provisions</i>	362
3. <i>Implementation of Marrakesh in the U.S.</i>	364
II. CONSEQUENCES FOR CONSEQUENTIALISM	365
A. <i>Is the Marrakesh Treaty a Copyright Game-Changer?</i>	365
1. <i>Paradigm Shift?</i>	365
2. <i>Copyright and Human Rights</i>	366
B. <i>Consequences for U.S. Copyright Law</i>	367
1. <i>Minimal Consequences for Consequentialism</i>	367
2. <i>Empowerment, Output, Progress</i>	368
III. CONSEQUENTIALISM IS MARRAKESH AT ITS BEST.....	370
A. <i>Consequentialism and Rights</i>	370
B. <i>The Utility of Rights Rhetoric</i>	371
1. <i>Sharpening the Fair Use Doctrine</i>	372
2. <i>Weighing Rights with Rights</i>	373
3. <i>Constitutional Compatibility</i>	374
C. <i>Distributive Justice</i>	376
D. <i>Natural Rights</i>	376
1. <i>Distinguishing Human Rights from Natural Rights</i>	376
2. <i>Problems with Labor Theory</i>	377
3. <i>Means or End?</i>	379
E. <i>Personality Theories</i>	379
1. <i>Users Missing</i>	379
2. <i>Author-Centrism</i>	381
CONCLUSION	381

INTRODUCTION

The Marrakesh Treaty to Facilitate Access to Published Works by Visually Impaired Persons and Persons with Print Disabilities¹ has been lauded as a significant achievement in advancing the rights of, and promoting equal opportunity for, the visually disabled.² The treaty aims to alleviate what has

¹ Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled, June 27, 2013, 52 I.L.M. 1312 (2013).

² See, e.g., Catherine Saez, *Vibrant Lauding of "Historic" Marrakesh Treaty For The*

been described as the “book famine”³ by requiring contracting states to implement copyright limitations and exceptions to facilitate access to copyrighted material for the global print-disabled community.⁴

Although the Marrakesh Treaty’s potential impact on the visually disabled is significant, one should not confuse this with having a significant impact on the U.S. copyright regime, were it to be implemented. Commentators have described the treaty as a paradigm shift in the international approach to copyright law, as it is the first instrument that harmonizes a minimum standard for copyright limitations and exceptions, focusing on users’ rights instead of authors’ rights as prior instruments have done.⁵ Nonetheless, the U.S. copyright system already substantially accounts for the importance of user access to works in the larger scheme of copyright’s consequentialist aim of promoting progress.⁶

This note argues that, in the United States, implementing the Marrakesh Treaty’s provisions will hardly alter the status quo of the copyright paradigm, as the treaty can be construed to be considerably consequentialist. In fact, the prominence of users’ rights in the treaty reinforces its compatibility with a consequentialist utilitarian model over other alternative theories, such as labor and personality justifications.⁷ The utilitarian model accounts for users’ rights better than alternative theories, which place more emphasis on the rights of the author or publisher.⁸

This note will not focus on analyzing the merits of the Marrakesh Treaty.⁹ Instead, this note seeks to demonstrate the compatibility of the treaty’s

Blind at WIPO, INTELLECTUAL PROPERTY WATCH (Oct. 6, 2016), <http://www.ip-watch.org/2016/10/06/vibrant-lauding-of-historic-marrakesh-treaty-for-the-blind-at-wipo/>.

³ *The Marrakesh Treaty in Action*, WORLD INTELLECTUAL PROPERTY ORGANIZATION, <http://www.wipo.int/pressroom/en/stories/marrakesh-treaty.html> (last visited Dec. 6, 2016).

⁴ See Marrakesh Treaty, *supra* note 1.

⁵ See, e.g., Danielle Conway, *The Miracle at Marrakesh: Doing Justice for the Blind and Visually Impaired While Changing the Culture of Norm Setting at WIPO*, in DIVERSITY IN INTELLECTUAL PROPERTY 35, 36 (Irene Calboli & Srividhya Ragavan eds., 2015); Paul Harpur & Nicolas Suzor, *Copyright Protections and Disability Rights: Turning the Page to a New International Paradigm*, 36 UNSW L. J. 745, 746 (2013).

⁶ See U.S. CONST. art. I, § 8, cl. 8.

⁷ See generally Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287 (1988) (laying out theories justifying intellectual property law in the U.S., such as those based on Lockean labor theory and Hegelian personality theory).

⁸ See *infra* Part III.

⁹ As the Marrakesh Treaty is already in force, this note will not examine whether a treaty was the optimal way of bringing about the aims of Marrakesh. For an analysis of whether an international treaty was the best option, see Margot E. Kaminski & Dr. Schlomit Yanisky-Ravid, *The Marrakesh Treaty for Visually Impaired Persons: Why a Treaty was Preferable to Soft Law*, 75 U. PITT. L. REV. 255 (2014).

conception of copyright justifications with the current U.S. copyright framework on a philosophical level.¹⁰ Compatibility between the respective theoretical justifications of the Marrakesh Treaty and U.S. copyright law is pertinent because it provides a convincing case for its ratification beyond the more general moral appeal of equality for the disabled.

Part I begins by reiterating the conventional view that copyright in the U.S. is founded on primarily utilitarian consequentialist justifications, zooming in on the consequentialist rationales underpinning the doctrine of fair use in copyright. Alternative philosophical justifications for copyright in the U.S. will be briefly introduced. Subsequently, the legislative history of the Marrakesh Treaty will be recounted to analyze the underlying justifications of the treaty. Part I will then explain the legislative amendments required to implement the Marrakesh Treaty in the U.S.

Part II critically examines the impact of the Marrakesh Treaty on the U.S. copyright framework. Part II assesses the arguments that the treaty constitutes a paradigm shift in copyright law, and goes on to demonstrate that copyright and human rights are not strangers to one another. Despite its strong rights-based underpinnings, the Marrakesh Treaty's provisions still fit comfortably into the U.S. consequentialist copyright framework. First, the Marrakesh Treaty itself contains consequentialist provisions and has utilitarian aims overall. Second, even the human rights references in the treaty can be construed to be aligned with consequentialism.

Part III will argue that implementing the Marrakesh Treaty, notwithstanding its human rights emphasis, is best understood as being harmonious with a consequentialist copyright framework. The tension between rights and consequentialism will be explored but it will be shown that, in the context of the treaty, rights nomenclature does more to assist than hinder the existing utilitarian framework. Part III will finish by arguing that justifications for the Marrakesh Treaty's provisions are less suitably aligned with distributive justice, natural rights, or personality theories of copyright. In doing so, it will highlight some limitations of these theories, as well as the strengths of the consequentialist interpretation.

¹⁰ The focus will be on arguing the Marrakesh Treaty's compatibility with U.S. copyright law. For an analysis of the treaty's desirability in the context of the international regime, see, for example, Aaron Scheinwald, *Who Could Possibly be Against a Treaty for the Blind?*, 22 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 445 (2012).

I

COPYRIGHT JUSTIFICATIONS AND MARRAKESH IMPLEMENTATION

A. *Philosophical Justifications of Copyright in the U.S.*1. *The Dominance of Consequentialism*

The dominant position is that copyright in the U.S. is founded on utilitarian consequentialist justifications.¹¹ Support for this position comes directly from the U.S. Constitution. Article I, Section 8, Clause 8 gives Congress the power to “promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”¹² The copyright system achieves “progress” by recognizing rights in works to incentivize creation.

In *Feist Publications, Inc., v. Rural Telephone Service Co.*,¹³ the Supreme Court held that to achieve copyright’s primary objective of promoting progress, “copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work.”¹⁴ To effectuate the consequentialist outcome of progress, both the proprietary interests of authors and accessibility interests of users must be balanced carefully.¹⁵

The consequentialist model in America is often approached in economic terms,¹⁶ in part owing to the rise of the law and economics movement.¹⁷ Landes and Posner’s landmark work, *The Economic Structure of Intellectual Property Law*, prescribes the “efficient level of protection . . . at which the social benefits from further protection just equal the social costs.”¹⁸ They propose that a

¹¹ See, e.g., Jeanne C. Fromer, *Expressive Incentives in Intellectual Property*, 98 VA. L. REV. 1745, 1750-52 (2012); Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1107-10 (1990); Pamela Samuelson, *Economic and Constitutional Influences on Copyright Law in the United States*, 23 EUR. INTELL. PROP. REV. 409, 422 (2001); Christopher Sprigman, *Copyright and the Rule of Reason*, 7 J. ON TELECOMM. & HIGH TECH. L. 317, 317-19 (2009).

¹² U.S. CONST. art. I, § 8, cl. 8.

¹³ 499 U.S. 340 (1991).

¹⁴ *Id.* at 349-50.

¹⁵ See *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 479 (1984).

¹⁶ See Samuelson, *supra* note 11, at 410-11.

¹⁷ See Richard A. Posner, *Intellectual Property: The Law and Economics Approach*, 19 J. ECON. PERSP. 57, 57 (2005).

¹⁸ WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 66 (2003). Note, however, that Richard Posner sought to distinguish between utilitarianism and the economic theory of wealth maximization. See Richard A. Posner, *Utilitarianism, Economics, and Legal Theory*, 8 J. LEGAL STUD. 103, 103 (1979).

“fundamental task of copyright law viewed economically . . . is to strike the optimal balance between . . . encouraging the creation of new works by reducing copying and its effect in discouraging the creation of new works by raising the cost of creating them”.¹⁹ They pay particular attention to the dynamic costs and benefits of copyright,²⁰ determining the ideal balance that will ensure that the “public domain is nourished.”²¹ Evidently, their consequentialist approach to copyright is largely in accord with the Progress Clause.

2. *Consequentialism in Fair Use*

Consequentialist justifications do not only appear in the courts’ observations about the copyright system in general. They also ground specific doctrines in the copyright framework. The most relevant for the purposes of this note is that of fair use.

The fair use doctrine is codified in section 107 of the Copyright Act of 1976 as a four-factor test.²² Economic considerations such as those identified by Landes and Posner are pervasive. In *Harper & Row, Publishers, Inc. v. Nation Enterprises*,²³ the Supreme Court noted that the fourth factor, the use’s impact on the work’s market, is the “single most important element of fair use.”²⁴ Jeanne Fromer proposes that reference to market harms and benefits in determining fair use helps to protect works and provide incentivization for the “overall benefit of society.”²⁵

The principle of transformativeness in determining fair use is also grounded in consequentialism. This concept used to guide fair use decisions stems from Judge Leval’s article, *Towards a Fair Use Standard*,²⁶ on which the Supreme Court heavily relied in *Campbell v. Acuff-Rose Music, Inc.*²⁷ Judge Leval’s and the Court’s reasoning is based on consequentialism. Judge Leval stresses the need to “focus on the utilitarian, public-enriching objectives of copyright” in addressing the fair use doctrine.²⁸ Transformativeness, he proposes, helps to determine whether the new use constitutes the “very type of activity that the fair use doctrine intends to protect for the enrichment of

¹⁹ LANDES & POSNER, *supra* note 18, at 69.

²⁰ *Id.* at 70.

²¹ *Id.* at 69.

²² See 17 U.S.C. §107 (2012).

²³ 471 U.S. 539 (1985).

²⁴ *Id.* at 566. But see Barton Beebe, *An Empirical Study on U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549, 617-20 (2008).

²⁵ Jeanne C. Fromer, *Market Effects Bearing on Fair Use*, 90 WASH. L. REV. 615, 649 (2015).

²⁶ Leval, *supra* note 11.

²⁷ 510 U.S. 569 (1994).

²⁸ Leval, *supra* note 11, at 1135.

society.”²⁹ In *Campbell*, the Supreme Court noted that “the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works.”³⁰ Fair use decisions such as *Campbell* and an earlier case, *Sony of America v. Universal City Studios*,³¹ have also been interpreted by commentators in economic terms, on the bases of market efficiency and avoiding market failure.³²

However, with regards to granting access to copyright works for the visually disabled, the role of consequentialism is not entirely clear. This right of access is found both in the §121 limitation of the Copyright Act and within the fair use doctrine itself.³³ The Supreme Court in *Sony* noted in passing that “[m]aking a copy of a copyrighted work for the convenience of a blind person is . . . an example of fair use.”³⁴ However, in *Authors Guild, Inc. v. Hathitrust*,³⁵ the Second Circuit noted that “providing access to the print disabled is not transformative.”³⁶ They did hold that the fourth factor weighed heavily in favor of finding fair use because “the present-day market for books accessible to the handicapped is so insignificant.”³⁷ But they ultimately noted that making accessible copies for the print disabled is a “special instance” of fair use, derived from Congress’s “commitment to ameliorating the hardships faced by the blind and the print disabled.”³⁸ To support this proposition, the Second Circuit cited the Chafee Amendment³⁹ and Congress’s declaration in the Americans with Disabilities Act, particularly their goal to “assure equality of opportunity, full participation, independent living, and economic self-sufficiency for such individuals.”⁴⁰ The Second Circuit decision seems to hint that there is possibly more at play than market efficiency.

3. Labor, Personality and Rawlsian Justifications

Despite the clear indication of the Constitution and authority of the courts recognizing consequentialism as the dominant justification for copyright

²⁹ *Id.* at 1111.

³⁰ *See Campbell*, 510 U.S. at 579.

³¹ 464 U.S. 417 (1984).

³² Samuelson, *supra* note 11, at 412-13.

³³ 17 U.S.C. §121(a) (2012) (“[I]t is not an infringement of copyright for an authorized entity to reproduce or to distribute copies or phonorecords of a previously published, nondramatic literary work if such copies or phonorecords are reproduced or distributed in specialized formats exclusively for use by blind or other persons with disabilities.”).

³⁴ *Sony*, 464 U.S. at 455 n.40.

³⁵ 755 F.3d 87 (2d Cir. 2014).

³⁶ *Id.* at 101.

³⁷ *Id.* at 103.

³⁸ *Id.* at 102.

³⁹ 17 U.S.C. §121 (2012).

⁴⁰ *Authors Guild*, 755 F.3d at 102; 42 U.S.C. §12101 (1990).

protection, alternative justificatory theories exist in copyright discourse.

The first category consists of natural rights theories, the most prominent being Lockean labor theory.⁴¹ Notwithstanding the Supreme Court's position that the "primary objective of copyright is not to reward the labor of authors,"⁴² some commentators believe that labor theory justifies intellectual property rights,⁴³ or should at least play a bigger role in doing so in the U.S.⁴⁴ Locke's labor theory can be briefly summarized as follows. God gave the world to mankind "for the support and comfort of their being."⁴⁵ Man thus has property in his personhood and therefore in his labor.⁴⁶ When man mixes his labor with things in the commons, he acquires property rights in them.⁴⁷ This acquisition is provided that one does not take more than one can properly use (waste proviso), and that enough is left for the rest of humanity to enjoy the inherited world (enough and as good proviso).⁴⁸ Locke's theory has been adapted to explain desert in intellectual property, with the commons being ideas, and limitations such as the idea-expression dichotomy serving a similar function to the provisos.⁴⁹

Personality theories have also gained prominence due to the rise of moral rights in the U.S. copyright regime.⁵⁰ Moral rights are often attributed to continental philosophy,⁵¹ a fair characterization due to the fact that moral rights have been introduced into U.S. copyright law to comply with their obligations under the Berne Convention.⁵² These theories are often based on the philosophy of Immanuel Kant⁵³ and G.W.F. Hegel.⁵⁴ For Kant, the wrongness in unauthorized publishing stems from the lack of agency and consent by authors to speak in their name.⁵⁵ According to Hegel, property rights stem from an

⁴¹ See JOHN LOCKE, SECOND TREATISE OF GOVERNMENT 116 (1689).

⁴² See *Feist Publ'ns, Inc., v. Rural Telephone Serv. Co.*, 499 U.S. 340, 349 (1991).

⁴³ See, e.g., William Fisher, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1661, 1688 (1988); Hughes, *supra* note 7, at 287-366.

⁴⁴ See, e.g., Wendy J. Gordon, *A Property Right in Self-Expression: Equality and Individualism in the Natural Law of Intellectual Property*, 102 YALE L.J. 1533, 1535 (1993).

⁴⁵ LOCKE, *supra* note 41, ch.5 §44.

⁴⁶ *Id.* ch.5 §27.

⁴⁷ *Id.*

⁴⁸ *Id.* ch.5 §33.

⁴⁹ See Hughes, *supra* note 7, at 298-329; Gordon, *supra* note 44, at 1581-82.

⁵⁰ See Lior Zemer, *Moral Rights: Limited Edition*, 91 B.U. L. REV. 1519, 1523-27 (2011).

⁵¹ See Hughes, *supra* note 7, at 330; Amy M. Adler, *Against Moral Rights*, 97 CAL. L. REV. 263, 264 (2009).

⁵² See Zemer, *supra* note 50, at 1523-27; Adler, *supra* note 51, at 266-86.

⁵³ See generally IMMANUEL KANT, ON THE INJUSTICE OF REPRINTING BOOKS (1785).

⁵⁴ See generally G.W.F. HEGEL, PHILOSOPHY OF RIGHT (1820).

⁵⁵ Kant, *supra* note 53, at 30-33.

individual's personality and are acquired through the exercise of their free will.⁵⁶ Autonomy-based theories are also proposed by contemporary commentators. Abraham Drassinower theorizes that copyright is primarily about authorship and the right of authors to have control over their speech.⁵⁷ This view has some degree of resonance in the United States. Many see First Amendment concerns as being within the ambit of what copyright should, but does not adequately, address.⁵⁸

Academics have also explored the role of copyright law in upholding distributive justice,⁵⁹ in the spirit of John Rawls' seminal work, *A Theory of Justice*.⁶⁰ Generally speaking, Rawls believed that society should be arranged to ensure that the least-advantaged in society benefitted the greatest, and that there should be equality of opportunity.⁶¹ Justin Hughes and Robert Merges have explored the extent to which copyright law conforms with the Rawlsian standard of justice⁶² and whether copyright could be used as a tool for distributive justice.⁶³ However, they do not claim that Rawlsian justifications ground U.S. copyright law.⁶⁴ Furthermore, evidence of the correlation between distributive justice and U.S. copyright law is scant and selective.⁶⁵ However, Rawls proposed his version of justice as a better alternative to utilitarianism.⁶⁶ Accepting that the copyright regime is Rawlsian would implicitly mean rejecting much of the utilitarian foundations grounding existing copyright doctrine.⁶⁷

⁵⁶ Hegel, *supra* note 54.

⁵⁷ See ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYING? (2015).

⁵⁸ See Melville B. Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 UCLA L. REV. 1180, 1181-82 (1970).

⁵⁹ See Margaret Chon, *Copyright's Other Functions*, 15 CHI. KENT J. INTELL. PROP. 101 (2016); Justin Hughes & Robert P. Merges, *Copyright and Distributive Justice*, 92 NOTRE DAME L. REV. 513 (2016).

⁶⁰ JOHN RAWLS, A THEORY OF JUSTICE (1971).

⁶¹ See *id.*

⁶² Hughes & Merges, *supra* note 59, at 526-28.

⁶³ *Id.* at 573.

⁶⁴ *Id.* at 575-6.

⁶⁵ See *id.* at 552-53 (noting the importance of copyright to the success of many high-earning African Americans).

⁶⁶ See, e.g., RAWLS, *supra* note 60, at 52, 91, 181. See also David Lyons, *Rawls Versus Utilitarianism*, 69 J. PHIL. 535 (1972).

⁶⁷ See Lyons, *supra* note 66 (discussing the rivalry between utilitarianism and Rawls' theory of distributive justice).

B. From Marrakesh to America

1. History and Development of the Marrakesh Treaty

The Marrakesh Treaty's "main goal is to create a set of mandatory limitations and exceptions for the benefit of the blind, visually impaired and otherwise print disabled."⁶⁸ The origins of this instrument can be traced back to as early as 1981, when WIPO and UNESCO started a Working Group to examine access to copyrighted works for the visually and auditory handicapped.⁶⁹ More studies subsequently followed, including those by Sam Ricketson⁷⁰ and Judith Sullivan,⁷¹ which piqued the international community's interest in this issue.⁷² Coinciding with these developments, the United Nations adopted the Convention on the Rights of Persons with Disabilities.⁷³ It remained clear that market forces were insufficient to create accessible works for the disabled.⁷⁴

In 2008, the World Blind Union (WBU) and Knowledge Ecology International arranged for an expert group to propose a treaty to address access to copyrighted material for those with reading disabilities.⁷⁵ The proposal was presented to WIPO by Brazil, Ecuador, and Paraguay.⁷⁶ Their approach was to present the right to read as a fundamental human right, pushing for obligations to uphold user rights and equality in access to information.⁷⁷ They emphasized measures necessary for the publication and distribution of works in accessible formats, and stressed the need for international harmonization of copyright

⁶⁸ See WORLD INTELLECTUAL PROPERTY ORGANIZATION, *supra* note 3.

⁶⁹ Conway, *supra* note 5, at 41.

⁷⁰ Sam Ricketson, WIPO STANDING COMMITTEE ON COPYRIGHT AND RELATED RIGHTS, STUDY ON LIMITATIONS AND EXCEPTIONS OF COPYRIGHT AND RELATED RIGHTS IN THE DIGITAL ENVIRONMENT (2003), http://www.wipo.int/edocs/mdocs/copyright/en/sccr_9/sccr_9_7.pdf.

⁷¹ Judith Sullivan, WIPO STANDING COMMITTEE ON COPYRIGHT AND RELATED RIGHTS, STUDY ON COPYRIGHT LIMITATIONS AND EXCEPTIONS FOR THE VISUALLY IMPAIRED (2007), http://www.wipo.int/edocs/mdocs/copyright/en/sccr_15/sccr_15_7.pdf.

⁷² Conway, *supra* note 5, at 41-42.

⁷³ United Nations Convention on the Rights of Persons with Disabilities, Dec. 13, 2006, 2515 U.N.T.S. 3.

⁷⁴ LIONEL BENTLEY & BRAD SHERMAN, INTELLECTUAL PROPERTY LAW 250 (4th ed. 2014).

⁷⁵ Conway, *supra* note 5, at 42.

⁷⁶ *Standing Comm. on Copyright and Related Rights, Proposal by Brazil, Ecuador and Paraguay, Relating to Limitations And Exceptions: Treaty Proposed by the World Blind Union (WBU)*, WIPO Doc. SCCR/18/5 (May 25, 2009), http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=122732.

⁷⁷ Conway, *supra* note 5, at 43.

limitations and exceptions.⁷⁸

Proposals by more WIPO member states followed. In 2010, the African Group proposed a draft treaty which intended to have a broader group of beneficiaries, including, for example, those with intellectual disabilities and even educational institutions and libraries.⁷⁹ The U.S. and European Union also simultaneously submitted proposals which required that the copyright exceptions complied with the Berne Convention's Three-Step-Test.⁸⁰ The U.S. Draft Consensus Instrument, "[r]ecognizing the public interest in maintaining a balance between the interests of authors and users, particularly the needs of those persons with print disabilities or impairment of their vision," took a soft law approach that focused on the import and export of accessible formats.⁸¹ The European Union similarly favored a non-binding approach in its Joint Recommendation.⁸² The European Union suggested that member states provide for such copyright exceptions in their national regimes, which would only apply when market solutions are inadequate.⁸³

Disagreements between developing countries and interest groups stalled negotiations.⁸⁴ Those acting on behalf of authors and publishers feared that the treaty could become "a Trojan horse for a future weakening of copyright protection through international treaties."⁸⁵ For instance, Allan Adler, counsel to the Association of American Publishers, noted that, up until the Marrakesh Treaty, international treaties had only been establishing the minimal rights of copyright owners.⁸⁶ The concern was that the treaty would establish a

⁷⁸ *Id.* at 44.

⁷⁹ *Standing Comm. on Copyright and Related Rights, Draft WIPO Treaty on Exceptions and Limitations for the Persons with Disabilities, Educational and Research Institutions, Libraries and Archives (Proposal by the African Group)*, WIPO Doc. SCCR/22/12 (June 3, 2011), http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=169397. See Conway, *supra* note 5, at 44.

⁸⁰ Lior Zemer & Aviv Gaon, *Copyright, Disability and Social Inclusion: The Marrakesh Treaty and the Role of Non-Signatories*, 10 J. INTELL. PROP. L. & PRAC. 836, 840 (2015).

⁸¹ *Standing Comm. on Copyright and Related Rights, Draft Consensus Instrument (Proposal by the Delegation of the United States of America)*, WIPO Doc. SCCR/20/10 (June 10, 2010), http://www.wipo.int/edocs/mdocs/copyright/en/sccr_20/sccr_20_10.pdf. See Conway, *supra* note 5, at 45.

⁸² *Standing Comm. on Copyright and Related Rights, Draft Joint Recommendation Concerning the Improved Access to Works Protected by Copyright for Persons with a Print Disability (Proposal by the Delegation of the European Union)* WIPO Doc. SCCR/20/12 (June 17, 2010), http://www.wipo.int/edocs/mdocs/copyright/en/sccr_20/sccr_20_12.pdf.

⁸³ Conway, *supra* note 5, at 46.

⁸⁴ Zemer & Gaon, *supra* note 80, at 840.

⁸⁵ Marketa Trimble, *The Multiplicity of Copyright Laws on the Internet*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 339, 354 (2015).

⁸⁶ KEIWashDC, Alan [sic] Adler on WIPO Negotiations on Copyright Exceptions,

“precedent of developing a series of treaties that specifically focus on trying to set forth minimal limitations and exceptions to the right of copyright owners.”⁸⁷ Adler supported a solution for the visually disabled, as it was an area where “there was no other way to accomplish the needs” that copyright exceptions and limitations aim to fulfill.⁸⁸ However, he cautioned that this treaty may become the “nose of the camel” for the international community to push for more treaties on exceptions and limitations for other users and beneficiaries.⁸⁹ Limitations and exceptions to serve the needs of educational institutions that should be adopted at the domestic level have not happened as many countries “don’t have adequate copyright laws to begin with.”⁹⁰ The problem was the infeasibility of “adopt[ing] appropriate limitations and exceptions on rights when [there is no] clear establishment of rights.”⁹¹ It took almost four years to reach a compromise.⁹²

The Marrakesh Treaty was finally adopted in June 2013, requiring twenty ratifications to become binding.⁹³ It entered into force on September 30, 2016, three months after Canada became the twentieth nation to accede to it.⁹⁴ The treaty currently has twenty-six contracting parties.⁹⁵ The U.S., despite being a signatory since October 2013, has yet to ratify the treaty as of February 2017.⁹⁶

2. *The Marrakesh Treaty’s Provisions*

The final product reflects elements raised by the different parties to the negotiation. The WBU got its wish of a binding treaty that obliged states to take active measures to facilitate the publication and distribution of works accessible to the print disabled.⁹⁷ The active measures required of states are mainly to be found in Articles 4, 5, 6 and 10. Article 4(1)(a) requires Contracting Parties to “provide in their national copyright laws for a limitation or exception to the

YOUTUBE (July 18, 2012), <https://www.youtube.com/watch?v=dxVcmOwBAsY> [hereinafter *Allan Adler Interview*].

⁸⁷ *Id.*

⁸⁸ Adler was also involved in drafting the Chafee amendment. *See id.*

⁸⁹ *See id.*

⁹⁰ *Id.*

⁹¹ *Id.*

⁹² Zemer & Gaon, *supra* note 80, at 840.

⁹³ Marrakesh Treaty, *supra* note 1, art. 18.

⁹⁴ *See* Press Release, WIPO, WIPO Director General Hails a Success for Visually Impaired People and International Community as Marrakesh Treaty Enters Into Force (Sept. 20, 2016), http://www.wipo.int/pressroom/en/articles/2016/article_0009.html [hereinafter WIPO Press Release].

⁹⁵ *See* WIPO-Administered Treaties, WIPO, http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&treaty_id=843 (last visited Feb. 22, 2016).

⁹⁶ *Id.*

⁹⁷ Marrakesh Treaty, *supra* note 1, art. 10.

right of reproduction, right of distribution, and the right of making available to the public . . . to facilitate the availability of works in accessible format copies for beneficiary persons.”⁹⁸ Article 4(2) provides an example of how that requirement may be fulfilled, by providing an exception or limitation that permits “authorized entities” to make or obtain an accessible format copy and supply them to beneficiaries “without the authorization of the copyright holder.”⁹⁹ Article 2 provides definitions of “works,” “accessible format copy,” and “authorized entity” for the purposes of the treaty.¹⁰⁰

Articles 5 and 6 contain provisions on the import and export of accessible formats, as put forward in the U.S. draft. Article 5 obliges states to allow these accessible format copies to be “distributed or made available by an authorized entity to a beneficiary person or an authorized entity in another Contracting Party.”¹⁰¹ Article 6 mirrors that provision with respect to importing accessible format copies without the authorization of the right holder.¹⁰²

As strategized by the WBU, the treaty explicitly mentions fundamental human rights in its preamble, but only those of “non-discrimination, equal opportunity . . . and full and effective partition and inclusion in society.”¹⁰³ The right to education, as proposed by the WBU and pushed further by the African Group, is referred to in the preamble but not in relation to human rights obligations.¹⁰⁴ And it is specifically the right to education of “persons with visual impairments or with other print disabilities.”¹⁰⁵ The beneficiary group is thus strictly limited to what was originally proposed by the WBU, and not the broader scope the African Group had wanted. Indeed, the full title of the treaty itself is the most obvious evidence for this—the treaty is clearly for the benefit of “Visually Impaired Persons and Persons with Print Disabilities.”¹⁰⁶ Furthermore, Article 3 clearly and exhaustively defines the list of beneficiary persons under the treaty.¹⁰⁷

The rights of authors remain close to the status quo. As the U.S. and E.U. had proposed, the Berne Three-Step-Test was incorporated into the Marrakesh Treaty via Article 11(a), referring to Article 9(2) of the Berne Convention, that a reproduction under the Marrakesh Treaty may be permitted only if it “does not

⁹⁸ *Id.* art. 4(1)(a).

⁹⁹ *Id.* art. 4(2). Note this is very similar to the §121 limitation in the U.S. Copyright Act.

¹⁰⁰ *Id.* art. 2.

¹⁰¹ *Id.* art. 5(1).

¹⁰² *Id.* art. 6.

¹⁰³ *Id.* pmbl.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *Id.*

¹⁰⁷ *Id.* art. 3.

conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.”¹⁰⁸ This entire phrase is repeated verbatim at the end of subparagraphs 11(b), (c) and (d) in relation to Article 13 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) and Articles 10(1) and 10(2) of the WIPO Copyright Treaty.¹⁰⁹ The message could not be clearer: the interests of authors in their works still remain dominant. Article 11 also reiterates compliance to the obligations under Berne, TRIPS, and the WIPO Copyright Treaty when adopting measures to implement the Marrakesh Treaty.¹¹⁰ Subordination to other treaties is set out clearly under Article 1.¹¹¹

The Marrakesh Treaty recalls authors not only in substance but also in spirit. Authors’ rights, as found in the Berne Convention and more generally, are acknowledged in the preamble.¹¹² Nevertheless, the Marrakesh Treaty is predominantly focused on its beneficiaries. Most of the preamble spells out the plight of the visually impaired, especially in developing countries, and how the copyright system and technology have failed to grant them adequate access to information and societal inclusion.¹¹³ However, the treaty remains faithful to the fundamental goals of copyright as “an incentive and reward for literary and artistic creations and of enhancing opportunities for everyone,” by “maintain[ing] a balance between the effective protection of the rights of authors and the larger public interest.”¹¹⁴

3. *Implementation of Marrakesh in the U.S.*

In February 2016, President Obama sent the Marrakesh Treaty to Congress for ratification.¹¹⁵ The Department of Commerce provided the President with the draft legislation, the “Marrakesh Treaty Implementation Act of 2016.” In Secretary of State John Kerry’s Letter of Submittal to the White House, he described the provisions of the Marrakesh Treaty as “compatible with existing U.S. law,” requiring only “[n]arrow statutory changes” for

¹⁰⁸ *Id.* art. 11(a).

¹⁰⁹ *Id.* art. 11(b), (c), (d).

¹¹⁰ *Id.* art. 11.

¹¹¹ *Id.* art. 1.

¹¹² *Id.* pmbl.

¹¹³ *See id.*

¹¹⁴ *Id.*

¹¹⁵ Message from the President of the United States Transmitting the Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled, S. TREATY DOC. NO. 114-6 (2016) [hereinafter Message from the President], <https://www.gpo.gov/fdsys/pkg/CDOC-114tdoc6/pdf/CDOC-114tdoc6.pdf>.

implementation.¹¹⁶

Since the Chafee Amendment introduced the §121 limitation two decades ago, the U.S. already has a limitation to copyright permitting authorized entities to reproduce accessible copies of works for the benefit of the visually disabled. This is very similar to what is required under Article 4 of the Marrakesh Treaty. As such, implementation would only require tweaking of §121 and integrating the cross-border exchange provisions of the treaty into the limitation.

The textual amendments to the Copyright Act required to implement the Marrakesh Treaty are as follows. Firstly, the scope of works covered by the §121 exception have to be broadened to align with Article 2 of the treaty, which follows the meaning of “works” under Article 2(1) of the Berne Convention.¹¹⁷ The draft legislation removes the word “nondramatic” from §121(a) so as to include dramatic scripts, and will include musical works in textual or notation form such as sheet music.¹¹⁸ The second minor adjustment to §121 is to match the definition of beneficiaries to Article 3 of the treaty.¹¹⁹

Finally, to implement Article 5 of the treaty on the export of accessible copies, §121 must be amended to “specify that accessible-format copies may be distributed to Marrakesh Treaty parties and to eligible persons abroad who are citizens and domiciliaries of the United States.”¹²⁰

II

CONSEQUENCES FOR CONSEQUENTIALISM

A. *Is the Marrakesh Treaty a Copyright Game-Changer?*

1. *Paradigm Shift?*

Commentators have described the Marrakesh Treaty as representing “an important change in how lawmakers balance the demands of copyright owners against the interests of people with disabilities in particular.”¹²¹ Some regard the Marrakesh Treaty as “usher[ing] in a new way of thinking about the global IPR

¹¹⁶ Letter of Submittal from the Secretary of State to the President of the United States, *reprinted in* Message from the President, *supra* note 115, at V, VI.

¹¹⁷ Marrakesh Treaty, *supra* note 1, art. 2.

¹¹⁸ Marrakesh Treaty Implementation Act of 2016, S. 6, 114th Cong. (2016).

¹¹⁹ *Id.*

¹²⁰ DEPARTMENT OF COMMERCE, THE “MARRAKESH TREATY IMPLEMENTATION ACT OF 2016” STATEMENT OF PURPOSE AND NEED AND SECTIONAL ANALYSIS DRAFT 4 (Feb. 7, 2016), https://ogc.commerce.gov/sites/ogc.commerce.gov/files/media/files/2016/sopan_-_marrakesh_treaty_only_sopan_final_02-11-16_clean.pdf.

¹²¹ Harpur & Suzor, *supra* note 5, at 746.

regime,”¹²² and being “the first time WIPO has addressed the rights of users in the copyright regime.”¹²³ Zemer and Goan call it “an historic landmark” and “the first time a treaty has been exclusively directed to the creation of a minimum standard for copyright exceptions.”¹²⁴

However, other commentators point out that fifty-seven countries, including many developed countries, already have specific provisions on making copyright works more accessible for the visually impaired.¹²⁵ Thus, it seems that, with regard to implementing exceptions and limitations, the Marrakesh Treaty can be viewed as a paradigm shift only for countries who do not already have such measures in place. The newly-created international obligations, however, particularly the provisions on cross-border distribution of accessible format copies, represent a significant change for all the WIPO nations.

2. Copyright and Human Rights

Another prominent aspect about the Marrakesh Treaty is that, unlike previous intellectual property treaties, it expressly considers fundamental human rights.¹²⁶ However, the relationship between human rights and intellectual property is not novel.¹²⁷ Intellectual property rights are alluded to in several human rights instruments.¹²⁸ Article 27(2) of the Universal Declaration of Human Rights refers to an author’s right to “moral and material interests” in his “scientific, literary or artistic production,”¹²⁹ essentially what we see in copyright and patents. Article 27(1) of the Universal Declaration refers to the right to “participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits,”¹³⁰ which mirrors the importance of the public domain and the interests of the public and users to

¹²² Conway, *supra* note 5, at 57.

¹²³ *Id.*

¹²⁴ Zemer & Goan, *supra* note 80, at 838.

¹²⁵ Kaminski & Yanisky-Ravid, *supra* note 9, at 269.

¹²⁶ See, e.g., Conway, *supra* note 5, at 57.

¹²⁷ See, e.g., Saleh Al-Sharieh, *Toward a Human Rights Method for Measuring International Copyright Law’s Compliance with International Human Rights Law*, 32 UTRECHT J. INT’L & EUR. L. 5 (2016); Jingyi Li & Niloufer Selvadurai, *Reconciling the Enforcement of Copyright with the Upholding of Human Rights: A Consideration of the Marrakesh Treaty to Facilitate Access to Published Works for the Blind, Visually Impaired and Print Disabled*, 36 EUR. INTELL. PROP. REV. 653 (2014).

¹²⁸ See, e.g., International Covenant on Economic, Social and Cultural Rights, G.A. Res. 2200 (XXI), art. 15, U.N. Doc. A/RES/2200(XXI) (Dec. 16, 1966) [hereinafter ICESCR]; Universal Declaration of Human Rights, G.A. Res. 217 (III) A, art. 27(2), U.N. Doc. A/RES/217(III) (Dec. 10, 1948) [hereinafter UDHR].

¹²⁹ UDHR, *supra* note 128, art. 27(2).

¹³⁰ *Id.* art. 27(1).

fairly access works. Article 27(1) strikingly resembles the U.S. Constitution's Progress Clause, which is perhaps no coincidence given Eleanor Roosevelt's role in drafting the Universal Declaration. Furthermore, the United Nations High Commissioner for Human Rights remarked that the "balance between public and private interests found under" human rights instruments "is one familiar to intellectual property law."¹³¹

Human rights instruments also echo the consequentialist structure of the American copyright system, specifically how intellectual property is used as an instrument in achieving a defined end—cultural and scientific progress. For example, the "ends" are reflected in Article 15(1) of the International Covenant on Economic, Social and Cultural Rights, which recognizes the "right of everyone" to "take part in cultural life" and "enjoy the benefits of scientific progress and its applications."¹³² The consequentialist "means" are seen in Article 15(2), which provides that "[t]he steps to be taken by the States Parties to the present Covenant to achieve the full realization of this right shall include those necessary for the conservation, the development and the diffusion of science and culture."¹³³

These examples serve to demonstrate that intellectual property is not a foreign concept when it comes to human rights or vice versa. Intellectual property and human rights are not the same thing, but they are surely more fraternal than adversarial. Likewise, the human rights references in the Marrakesh Treaty, which also cites the Universal Declaration, should not necessarily be seen as a deviation from consequentialism.

B. Consequences for U.S. Copyright Law

1. Minimal Consequences for Consequentialism

When the Marrakesh Treaty is holistically examined, it becomes apparent that there are strong consequentialist underpinnings. The treaty's emphasis on the importance of "copyright protection as an incentive and reward for literary and artistic creations and of enhancing opportunities for everyone . . . to participate in the cultural life of the community, to enjoy the arts and to share scientific progress and its benefits,"¹³⁴ mirrors the Progress Clause. One

¹³¹ High Comm'r for Human Rights, *The Impact of the Agreement on Trade-Related Aspects of Intellectual Property Rights on Human Rights*, ¶11, U.N. Doc. E/CN.4/Sub.2/2001/13 (June 27, 2001).

¹³² ICESCR, *supra* note 128, art. 15(1).

¹³³ *Id.* art. 15(2); David Vaver, *Copyright Defenses as User Rights*, 60 J. COPYRIGHT SOC'Y. U.S. 661, 671 (2013) (observing that IP is a lesser right than the rights to participate in culture and enjoy the benefits of scientific progress).

¹³⁴ Marrakesh Treaty, *supra* note 1, pmb1.¶ 3.

commentator has even suggested that the Marrakesh Treaty is an example of the American fair use doctrine being implemented at an international level.¹³⁵ The role reversal is no coincidence: the U.S. was one of the first nations to push for this treaty.¹³⁶ The affinity between the Marrakesh Treaty and consequentialist American copyright law both proves and explains the consequentialist principles of the treaty. However, two added aspects that appear on the return journey from Marrakesh are worth noting: the requirement of facilitating cross-border exchange of accessible copies,¹³⁷ and the reference to human rights.¹³⁸

2. *Empowerment, Output, Progress*

Inclusion and accessibility are not just about respecting rights. Equal participation is enriching to society as a whole. The WIPO Director General remarked that the Marrakesh Treaty is not just about literacy, but about what one can do with literacy, that is, to “become a fully empowered economic agent.”¹³⁹ Acknowledging equal opportunity is thus aimed at increasing overall societal output. This is in line with consequentialist aims of maximizing artistic and scientific contributions. Note that the human right stressed in the Marrakesh Treaty is equal opportunity, not the right to education as such.¹⁴⁰ It is in paying attention to substantive equality that we rectify the disparity in access to information affecting the print disabled. Their disabilities have hindered their enjoyment of works. But the market has also played a role. Copyright law can resolve the latter.¹⁴¹

The philosophy behind pushing for an international treaty is in line with consequentialist economic thinking such as that of Landes and Posner. The Marrakesh Treaty has been described as a solution to “market failure.”¹⁴² The

¹³⁵ Conway, *supra* note 5, at 57.

¹³⁶ See Shae Fitzpatrick, *Setting Its Sights on the Marrakesh Treaty: The U.S. Role in Alleviating the Book Famine for Persons with Print Disabilities*, 37 B.C. INT’L & COMP. L. REV. 139, 146-47 (2014) (describing U.S. involvement in the early negotiation stages of the Marrakesh Treaty).

¹³⁷ The cross-border provisions were actually proposed by the U.S. in the negotiations. See *supra* text accompanying note 79.

¹³⁸ See Marrakesh Treaty, *supra* note 1, pmbl.

¹³⁹ WIPO, *WIPO Director General Lauds the Entry into Force of the Marrakesh Treaty*, YOUTUBE (Sep. 22, 2016), <https://www.youtube.com/watch?v=ZNjPYx-U4h4>.

¹⁴⁰ Education is mentioned in the preamble but not as part of human rights instruments. The “right to education” is only mentioned in the subsequent clause that is not in reference to human rights instruments. Education is also subsequently referred to in the preamble as a type of public interest that has to be weighed against the rights of authors. See Marrakesh Treaty, *supra* note 1, pmbl. ¶¶2, 9.

¹⁴¹ Landes & Posner, *supra* note 18, at 56 (describing copyright as a tool to correct distortions in the market).

¹⁴² See Kaminski & Yanisky-Ravid, *supra* note 9, at 269, 272; Fitzpatrick, *supra* note

preamble of the treaty also stresses the “importance of appropriate exceptions and limitations to make works accessible ... particularly when the market is unable to provide such access.”¹⁴³ A commentator notes that, “[f]ew commercial entities seek to create accessible works because copyright holders or licensees do not perceive the accessible format market as economically viable.”¹⁴⁴ Thus, the treaty would be an example of tapering copyright law to address the social inefficiencies described by Landes and Posner.¹⁴⁵ In implementing the Marrakesh Treaty provisions, the social benefits of lessening protection (improved accessibility and cultural participation for the visually impaired) outweigh the social costs (the negligible deterrence of rights holders, who are reluctant to enter that market to begin with).¹⁴⁶

How does the cross-border element fit into consequentialist justifications? One line of argument proposes that opening up to international market creates economies of scale for accessible formats.¹⁴⁷ Additionally, legalizing the importation and exportation of these copies reduces “costly, duplicative efforts.”¹⁴⁸ This would reduce what Lander and Posner refer to as “socially wasteful expenditures on creating and producing such works,”¹⁴⁹ unnecessary costs that have minimal or no bearing on incentivizing creativity.

The import and export provisions raise another issue, namely, whether distributing accessible copies to non-U.S. beneficiaries falls in line with the Progress Clause. Is American copyright law meant to promote progress globally or just domestically?¹⁵⁰ The answer, however, has no bearing on whether the Marrakesh Treaty itself is consequentialist. At most, the consequentialist ends of the U.S. copyright system and the justifications of the Marrakesh Treaty would differ in scope, not philosophy. Nonetheless, it is best to agree with Justin Hughes’ view in his statement to the WIPO General Assembly on behalf of the U.S. on the importance of the “copyright system and the incentives it provides for the creation and dissemination of works *for all people*.”¹⁵¹

136, at 143, 157-58.

¹⁴³ Marrakesh Treaty, *supra* note 1, pmbl. ¶8.

¹⁴⁴ Fitzpatrick, *supra* note 136, at 156.

¹⁴⁵ See Landes & Posner, *supra* note 18, at 66.

¹⁴⁶ See *id.*

¹⁴⁷ Fitzpatrick, *supra* note 136, at 158-59.

¹⁴⁸ *Id.* at 167; see also WIPO Press Release, *supra* note 94 (“[S]haring of works in accessible formats should increase the overall number of works available because it will eliminate duplication and increase efficiency.”).

¹⁴⁹ Landes & Posner, *supra* note 18, at 56.

¹⁵⁰ This is assuming this is a dichotomy at all, given our globalized market and the transnational collaborative nature of scientific and creative industries.

¹⁵¹ Justin Hughes, *US Statement at the WIPO General Assembly*, U.S. MISSION GENEVA

Although only one part of its justifications, the human rights element has been a dominant source of attention for the Marrakesh Treaty.¹⁵² As such, the relationship between human rights and consequentialism will be analyzed in the following section.

III

CONSEQUENTIALISM IS MARRAKESH AT ITS BEST

A. *Consequentialism and Rights*

The language of rights can sound the alarm for consequentialists. Deontologists, such as Kant, consider rights to be ends in themselves, never a means to an end.¹⁵³ However, the view that rights are not absolute is trite.¹⁵⁴ There have been consequentialists who incorporate rights as secondary principles instrumental to achieving desired consequences.¹⁵⁵ Rights are treated by utilitarians as “solutions to problems of institutional design.”¹⁵⁶ John Stuart Mill famously incorporated rights into his model of utilitarianism.¹⁵⁷ He argued that although man’s “independence, is of right, absolute,” we are answerable for our actions to others in society.¹⁵⁸ Even in Mill’s utilitarian system, there are “certain social utilities which are vastly more important, and therefore more absolute and imperative, than any others are as a class.”¹⁵⁹ Samuel Freeman argues, however, that the question that “bedevils traditional consequentialist views” is how these rights ought to be “*equally* or fairly distributed.”¹⁶⁰ In relation to copyright, that question would be, for instance, how should the rights of copyright owners and users be distributed? The Marrakesh Treaty does attempt to address that question.

SWITZ., (Dec. 17, 2012), <http://geneva.usmission.gov/2012/12/17/wipo/> (emphasis added).

¹⁵² See, e.g., Zemer & Gaon, *supra* note 80, at 837; Li & Selvadurai, *supra* note 127.

¹⁵³ See generally IMMANUEL KANT, *GROUNDWORK FOR THE METAPHYSICS OF MORALS* (1785).

¹⁵⁴ See e.g., John Finnis, *Absolute Rights: Some Problems Illustrated*, 61 AM. J. JURIS. 195 (2016).

¹⁵⁵ See Samuel Freeman, *Problems with Some Consequentialist Arguments for Basic Rights*, in *THE PHILOSOPHY OF HUMAN RIGHTS: CONTEMPORARY CONTROVERSIES* 107, 107 (Gerhard Ernts & Jan-Christoph Heilinger eds., 2012) (discussing how several consequentialists incorporate rights into their framework of morality).

¹⁵⁶ Allan Gibbard, *Utilitarianism and Human Rights*, 1 SOC. PHIL. & POL’Y. 92, 94 (1984).

¹⁵⁷ John Stuart Mill, *Utilitarianism*, in *UTILITARIANISM AND ON LIBERTY* 181 (Mary Warnock ed., 2008).

¹⁵⁸ John Stuart Mill, *On Liberty*, in *UTILITARIANISM AND ON LIBERTY* 88, 95 (Mary Warnock ed., 2008).

¹⁵⁹ Mill, *supra* note 157, at 235.

¹⁶⁰ Freeman, *supra* note 155, at 112.

There have been consequentialists who propose a stronger incorporation of rights into their model of justice.¹⁶¹ Amartya Sen rejected both traditional utilitarianism and constraint-based deontology, instead proposing a mathematical approach to a consequentialist system.¹⁶² Sen's model does not persuade Freeman, who remarks that to avoid circularity, "the broad consequentialist would seem to have to concede that respecting these principles of right is intrinsically good and hence the right thing to do for its own sake."¹⁶³

Others have taken a more pragmatic approach. For instance, Allan Gibbard contends that "[p]eople in general, utilitarians and non-utilitarians, can be strongly moved by a principle with a coherent rationale."¹⁶⁴ He explains that insofar as non-utilitarian rationales, such as rights, are used in a utilitarian argument for moral conviction, that conviction itself is of great utility.¹⁶⁵ Specifically regarding human rights, William Talbott proposes that "good consequences are not simply a fortunate by-product of the protection of human rights; they are, ultimately, the ground of their moral importance."¹⁶⁶

These arguments show that tensions between rights and consequentialism exist, but those tensions can be reconciled. The concepts are not mutually exclusive. The above scholarly debate attempts to tackle the arduous task of creating a general theory of morality. For American copyright law, the task is easier (if only slightly) in that we already have a consequentialist premise to work on and a more clearly defined end. We need not work on the absolute conception of rights as did Freeman and Sen do because American copyright law has explicitly rejected an absolute approach to rights.¹⁶⁷ The Marrakesh Treaty requires authors' rights to be balanced against those of other stakeholders.¹⁶⁸ We could settle for John Stuart Mill's approach to make sense of rights' role in copyright consequentialism.

B. The Utility of Rights Rhetoric

If the Marrakesh Treaty is primarily consequentialist, why does it focus on rights? Why not? Take the word "copyright." More than half of the word is "right," yet it remains predominantly utilitarian in the U.S. This statement illustrates the futility of suspicion based on semantics. Why not then, just do

¹⁶¹ See, e.g., WILLIAM J. TALBOTT, WHICH RIGHTS SHOULD BE UNIVERSAL? (2005).

¹⁶² See Amartya Sen, *Rights and Agency*, 11 PHIL. & PUB. AFF. 3 (1982).

¹⁶³ Freeman, *supra* note 155, at 124.

¹⁶⁴ Gibbard, *supra* note 156, at 100.

¹⁶⁵ *Id.*

¹⁶⁶ William J. Talbott, *Consequentialism and Human Rights*, 8 PHIL. COMPASS 1030, 1039 (2013).

¹⁶⁷ See, e.g., *Cariou v. Prince*, 714 F.3d 694, 705 (2d Cir. 2013).

¹⁶⁸ Marrakesh Treaty, *supra* note 1, pmbl. ¶9.

away with rights rhetoric altogether? An obvious answer is that rights terminology is already embedded in copyright law, used with respect to authors. An obvious example is found in 17 U.S.C. §106 on “exclusive rights” in works,¹⁶⁹ and §106A on “rights of certain authors.”¹⁷⁰ The Berne Convention, to which the United States is party, also clearly obligates States to respect authors’ rights.¹⁷¹ In fact, out of the troika of main IP rights, *copyright* is the only one with “right” in its name.¹⁷² The current copyright regime itself is thus evidence that a consequentialist system which acknowledges rights of individuals is clearly conceivable.

Beyond that, this discussion will show that the rights rhetoric, particularly with reference to those other than copyright holders, is compatible with and assistive to the current consequentialist regime.

1. *Sharpening the Fair Use Doctrine*

The utilitarian doctrine of transformativeness serves a primary role in determining fair use.¹⁷³ Nonetheless, “transformative use is not absolutely necessary for a finding of fair use.”¹⁷⁴ The market-based rationale behind the Marrakesh Treaty can aid in explaining the consequentialist justifications behind the Copyright Act’s §121 limitation and §107 non-transformative fair use provision, as they relate to accessible copies for the visually impaired, where the courts have not comprehensively done so.

In *Authors Guild, Inc. v. HathiTrust*, the Second Circuit addressed market factors in its analysis, finding that the fourth factor favored a finding of fair use because “the present day market for books accessible to the handicapped is so insignificant.”¹⁷⁵ However, there are shortcomings in their explanation. First, under that rationale, fair use defenses should be open to other minority stakeholder groups. Second, it implies that once that stakeholder group is enlarged to, for example, the *global* community of visually disabled, the fourth factor will start to tip against finding fair use. Third, they do not provide a nexus

¹⁶⁹ 17 U.S.C. § 106 (2012).

¹⁷⁰ *Id.* § 106A (2012).

¹⁷¹ See Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, revised at Paris July 24, 1971, 1161 U.N.T.S 31, 31 [hereinafter Berne Convention].

¹⁷² Solely based on etymology, trademarks seem to indicate that their function is to serve an indicative function of trade origin, in line with the conventional understanding of trademark law. Patent, originating from the Latin term that means “to open/to spread” seems to be in line with its incentivization for inventors to disclose inventions in exchange for their monopolies. Copyright thus seems to be about the right to make copies/the right to copy.

¹⁷³ See Lloyd L. Weinreb, *Fair’s Fair: A Comment on the Fair Use Doctrine*, 103 HARV. L. REV. 1137, 1141-42 (1990) (critiquing the utilitarian approach to fair use).

¹⁷⁴ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).

¹⁷⁵ *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 103 (2d Cir. 2014).

between their market-based justifications and their “special instance” commitment to aiding the disabled.¹⁷⁶

The Marrakesh Treaty’s economic rationale and rights sensitivities can be used to address these gaps. The treaty combats market inefficiencies that apply to this particular group of treaty beneficiaries. It requires adjustments that will reduce publishers’ costs to cater to the visually disabled by increasing efficiency through means like incentivizing other players to fill gaps in the market.¹⁷⁷ Due to the agent-sensitivity of copyright rights, the target beneficiaries are succinctly defined so authorized entities may more effectively fill the demand gap by creating and distributing accessible format copies made under national copyright exceptions.¹⁷⁸ When authorized entities take on the distributive roles that the rights holders have failed to assume, the publishers’ costs are reduced for each beneficiary receiving an accessible copy. That way, as the number of beneficiaries increases to a global scale, the marginal costs to publishers do not increase. Hence, the fourth fair use factor on market effects will still favor a finding of fair use.

2. *Weighing Rights with Rights*

David Vaver points out that the language of user rights posits authors and users as equals.¹⁷⁹ Vaver’s observation happens to fit appropriately within the Marrakesh Treaty’s call for equality. It is also faithful to the consequentialist goal of scientific and cultural progress as it treats both authors and users as equally important parts of this goal. More emphasis on user rights helps to alleviate the criticisms against consequentialism for lacking a just distribution of rights.¹⁸⁰

Vaver discusses the Canadian Supreme Court decision of *CCH Canadian, Ltd. v. Law Society of Upper Canada*,¹⁸¹ which demonstrates Canada’s shift to a user-rights approach to copyright defenses.¹⁸² He acknowledges that while fair use has been accepted as affirming user rights in American courts,¹⁸³ it remains a “minority view in U.S. law.”¹⁸⁴ Thus he accepts

¹⁷⁶ See *supra* text accompanying note 36.

¹⁷⁷ See Fitzpatrick, *supra* note 136, at 166.

¹⁷⁸ See *id.* at 143, 158 (discussing how the Marrakesh Treaty’s cross-border provisions could stimulate an accessible-format copies market to meet the demand of those with print disabilities).

¹⁷⁹ See Vaver, *supra* note 133, at 669.

¹⁸⁰ See Freeman, *supra* note 155, at 109-18 (arguing that consequentialist theories do not adequately take distribution of rights into account).

¹⁸¹ [2004] 1 S.C.R. 339 (Can.).

¹⁸² Vaver, *supra* note 133, at 667.

¹⁸³ See *e.g.*, *Bateman v. Mnemonics, Inc.*, 79 F.3d 1532, 1542 n.22 (11th Cir. 1996).

that the conventional view still focuses on fair use as a defense to infringement of copyright owners' works rather than a positive exercise of rights by users. Nonetheless, he observes that "[t]hese decisions represent 'a move away from an earlier author-centric view' to one that emphasizes user rights as an important tool to balance 'protection and access' sensitively, so as to further the public interest in making culture widely available."¹⁸⁵ This demonstrates how reinforcing rights on the users' side can contribute to a consequentialist copyright framework. Moving away from author-centricity is not a betrayal of the utilitarian copyright system, but perfectly consistent with it. This is exactly what the Marrakesh Treaty is trying to do. Striving for equality is consistent with the goal of optimal societal progress. Rights language makes this clearer.

Drassinower notes the Canadian *CCH* position on user rights, but remains critical of the "hegemony of instrumentalist thinking in the United States."¹⁸⁶ He acknowledges that the American approach does strike a balance, but is one that is "entirely devoted to the public interest."¹⁸⁷ He is critical of this, and suggests it would be better to focus on the dignity of authors, and user rights as embedded in any description of the dignity of authorship.¹⁸⁸ Drassinower is right that one must not neglect the human dignity of authors and users. However, adopting a consequentialist approach does not necessarily forsake their dignity from not regarding their rights "as ends in themselves." Not treating authors' or users' rights as absolute ends by subjecting them to a public interest requirement is not the same as not treating human dignity as an absolute end. Drassinower wrongly equates rights being ends in themselves with *persons* being ends in themselves. He seems to overlook the proposition that the consequentialist goal of progress, after all, has the purpose of benefitting humanity. The Marrakesh Treaty is an example of how accepting equality as being in the public interest does not mean rejecting the dignity of individuals.

3. *Constitutional Compatibility*

Users' rights may also be useful in addressing the incongruence between copyright law and other rights prevalent in the American legal landscape, such as, most prominently, First Amendment rights.¹⁸⁹ As Pamela Samuelson notes, "one would think that just as speakers have First Amendment rights, they

¹⁸⁴ Vaver, *supra* note 133, at 668.

¹⁸⁵ *Id.* at 669 (citing *SOCAN v. Bell Canada*, [2012] 2 S.C.R. 326 ¶ 9-11 (Can.)).

¹⁸⁶ Abraham Drassinower, *Taking User Rights Seriously*, in *IN THE PUBLIC INTEREST: THE FUTURE OF CANADIAN COPYRIGHT LAW* 462, 479 (Michael Geist ed. 2005).

¹⁸⁷ *Id.* at 478.

¹⁸⁸ *Id.* at 479.

¹⁸⁹ See, e.g., Rebecca Tushnet, *Copy this Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 *YALE L.J.* 535 (2004).

should have fair use rights.”¹⁹⁰ Melville Nimmer has discussed the importance of reconciling copyright and the First Amendment.¹⁹¹ However, Rebecca Tushnet maintains that modern copyright law is “incompatible with the First Amendment,” and does not believe that a solution to resolve the conflict between owners’ rights and free speech rights is possible.¹⁹² Yet if we were to try to tackle this formidable task, rights talk introduces consistency that would make reconciliation slightly less difficult.

The Supreme Court’s decision in *Harper & Row*¹⁹³ held that First Amendment interests were already embodied in the idea-expression distinction and fair use defense, rejecting an expansion of the fair use doctrine to create a new public interest exception based on First Amendment rights.¹⁹⁴ Tushnet warns that this causes litigants to use the fair use defense as a “back-door” to invoke the First Amendment, leading to worrisome outcomes.¹⁹⁵ One such outcome is that it “obscures the speech-enhancing role” of other copyright limitations, “allowing those limits to be dissolved in favor of copyright owners.”¹⁹⁶ Another is that it “distorts the communal, reciprocal nature of copyright’s theory of free speech.”¹⁹⁷ However, as the Marrakesh Treaty has shown, adding weight to user rights can help to explain fair use doctrine in a manner that is sensitive to users’ interests while remaining faithful to the communal interests of consequentialist copyright reasoning.¹⁹⁸ Yes, the treaty addresses the right of equal opportunity and social inclusion, not free speech. But the utility of its approach applies *mutatis mutandis*. If we wish to fruitfully address the First Amendment in the copyright realm, focusing on user rights in their relation to utilitarian copyright aims could be helpful in ironing out their disparities.

Ultimately, further incorporation of rights rhetoric unavoidably allows stronger claims for alternative copyright justifications. However, it will be shown that these justifications are incongruent with the objectives of the Marrakesh Treaty, and have inherent shortcomings.

¹⁹⁰ Pamela Samuelson, *Possible Futures of Fair Use*, 90 WASH. L. REV. 815, 857 (2015).

¹⁹¹ See Nimmer, *supra* note 58.

¹⁹² Tushnet, *supra* note 189, at 537-38.

¹⁹³ *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539 (1985).

¹⁹⁴ *Id.* at 560.

¹⁹⁵ Tushnet, *supra* note 189, at 590.

¹⁹⁶ *Id.*

¹⁹⁷ *Id.*

¹⁹⁸ See *supra* Part III.B.1.

C. Distributive Justice

The Marrakesh Treaty clearly has strong Rawlsian elements¹⁹⁹ because it is grounded on the principle of equal opportunity, and it clearly addresses social and economic equalities to the benefit of the least advantaged.²⁰⁰ The treaty notes the challenges “prejudicial to the complete development” of its beneficiaries.²⁰¹ It also pays particular attention to the print disabled who “live in developing and least-developed countries.”²⁰² This is clearly in the spirit of Rawls’ difference principle.²⁰³

Admittedly, the Marrakesh Treaty can be interpreted as having *both* utilitarianism and Rawlsian justice rationales, despite the rivalry of the theories. However, it just happens to work out on *this particular* permutation of interest groups. Still, siding with Mill over Rawls would be preferable. Accepting a Rawlsian justification to support users’ rights would certainly rain on the slippery slope which alarmed early opponents of the treaty.²⁰⁴ Accepting that the visually disabled are a possible class of the open subset of the “least advantaged” may allow others to start claiming that copyright should be limited in favor of other “less fortunate” stakeholders.²⁰⁵ Saying that this would necessarily result in market failure might be an exaggeration. But accepting the Marrakesh Treaty as Rawlsian would indeed destine it to be the “camel’s nose.”²⁰⁶ The Marrakesh Treaty was decidedly agreed on to benefit a specific community—the print-disabled.²⁰⁷ It would not be appropriate to interpret it as allowing anything more than that.

D. Natural Rights

1. Distinguishing Human Rights from Natural Rights

As there is no conclusive evidence of the Marrakesh Treaty taking an absolutist approach to rights,²⁰⁸ we must assess natural rights theory’s compatibility with the treaty as a legal *instrument* of human rights law. To dispel confusion, it must be clarified that natural law and human rights law are

¹⁹⁹ See *supra* Part I.A.3.

²⁰⁰ Marrakesh Treaty, *supra* note 1, pmbl. ¶¶ 1, 4.

²⁰¹ *Id.*, pmbl. ¶ 2.

²⁰² *Id.*, pmbl. ¶ 5.

²⁰³ See *supra* text accompanying note 59.

²⁰⁴ See Allan Adler Interview, *supra* note 86.

²⁰⁵ See *supra* text accompanying note 79.

²⁰⁶ See Allan Adler Interview, *supra* note 86.

²⁰⁷ See *supra* text accompanying note 107.

²⁰⁸ See Marrakesh Treaty, *supra* note 1, pmbl. ¶ 9 (noting “the need to maintain a balance between effective protection of the rights of authors and the larger public interest.”).

not the same.²⁰⁹ Natural law is concerned with “specifying the first and most general principle of morality . . . that one should choose and act in ways that are compatible with a will towards integral human fulfillment.”²¹⁰ International human rights law is about “obligations which States are bound to respect . . . to help ensure that international human rights standards are indeed respected, implemented, and enforced at the local level.”²¹¹

For this note’s purposes, the key difference is this—establishing a primordial standard of morality is not essential for the latter. It merely requires compliance with international law obligations, and ideally, normative acceptance of a particular international standard of rights.²¹² The justifications for those standards of rights need not be based on natural law parameters.²¹³ To implement the Marrakesh Treaty is to accept the standards set out in human rights instruments and the legal obligations required to uphold those standards. Implementation does not, however, require endorsing any particular interpretation of the moral bases on which those rights are founded.

The main natural rights theory that arises in justifying copyright is Lockean Labor theory, which functions on a primordial standard of morality.²¹⁴ Accepting its coherence in explaining (intellectual) property rights requires accepting the moral basis those rights arise out of, which raises doubts.

2. *Problems with Labor Theory*

The demise of natural law justifications in human rights law,²¹⁵ a field with it intuitively seems congruent, is telling. One reason for its demise was the growing opposition to absolutism over time.²¹⁶ The position of American law regarding the relationship between natural rights and copyright is very clear. In *Cariou v. Prince*,²¹⁷ the Second Circuit reiterated that copyright “is not an inevitable, divine, or natural right that confers on authors the absolute

²⁰⁹ *But cf.* JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS (2d ed. 2011) (using the terms human rights and natural rights interchangeably).

²¹⁰ Robert P. George, *Natural Law*, 31 HARV. J. L. & PUB. POL’Y. 171, 172 (2008).

²¹¹ *International Human Rights Law*, UNITED NATIONS HUMAN RIGHTS OFFICE OF THE HIGH COMMISSIONER, <http://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx> (last visited Nov. 27, 2016).

²¹² See generally Makau Mutua, *Standard Setting in Human Rights: Critique and Prognosis*, 29 HUM. RTS. Q 547 (2007).

²¹³ See PHILIP ALSTON & RYAN GOODMAN, INTERNATIONAL HUMAN RIGHTS LAW 494 (2013) (describing how human rights law is no longer necessarily viewed as tied to natural law).

²¹⁴ See *supra* text accompanying note 43.

²¹⁵ See, e.g., ALSTON & GOODMAN, *supra* note 213, at 490-94.

²¹⁶ *Id.* at 494.

²¹⁷ 714 F.3d 694 (2d. Cir. 2013).

ownership of their creations.”²¹⁸ Natural law justifications of Lockean labor theory are grounded precisely on such questionable absolutist premises.²¹⁹ Rights arise from annexing labor to objects.²²⁰ Therefore, Lockean labor is premised on self-ownership sanctioned by a divine creator.²²¹ The enough and as good proviso arises from a duty to comply with the wishes of that creator for man to equally enjoy the inherited world.²²² For agnostics and secularists,²²³ removing the creator from the equation causes the Lockean Jenga tower to collapse. Mark Lemley notes that “even the die-hard natural law theorists have mostly abandoned that way of thinking,”²²⁴ turning to some notion of “I made it and so I own it.”²²⁵ Without the God endorsement, one has to grasp at alternatives such as exalting the laboring itself. This approach however, is quickly vitiated when A. John Simmons asks “why we should be inclined to take *making itself* to be morally significant.”²²⁶

The absolutism of the premise clearly plagues the logic of the theory. But assuming we accept the premise, is it a useful justification for explaining fair use and the Marrakesh Treaty’s motives? A Lockean theory of copyright may still account for users. Wendy Gordon, one of the strongest proponents of Lockean justifications in copyright, argues that fair use can be justified based on the public’s right to the commons.²²⁷ In the context of copyright, the commons would include expired or abandoned works in the public domain and abstract ideas that are not protectable *ab initio*.²²⁸ As the proviso would prohibit ownership of abstract ideas,²²⁹ Gordon rationalizes fair use findings in noncommercial uses, such as scholarly or technical work, on the basis that they would be using the plaintiff’s work in its capacity as facts, which are part of the commons.²³⁰ She also explains fair use findings in parody cases based on the “needed access . . . to criticize” the work.²³¹ Both examples seem to be premised

²¹⁸ *Id.* at 705.

²¹⁹ *See supra* text accompanying note 43.

²²⁰ *See supra* text accompanying note 45.

²²¹ *See supra* text accompanying note 43.

²²² Locke, *supra* note 41, ch.5 §33.

²²³ *See* Gabe Bullard, *The World's Newest Major Religion: No Religion*, NATIONAL GEOGRAPHIC, Apr. 22, 2016 (on the growing secularism in the United States and across the globe).

²²⁴ Mark Lemley, *Faith-Based Intellectual Property*, 62 UCLA LAW. REV. 1328, 1338 (2015).

²²⁵ *Id.*

²²⁶ A. John Simmons, *Makers Rights*, 2 J. ETHICS 197, 217 (1998).

²²⁷ *See* Gordon, *supra* note 44.

²²⁸ *Id.* at 1559.

²²⁹ *Id.* at 1581182

²³⁰ *Id.* at 1593.

²³¹ *Id.* at 1605.

on the fair user handling the work with “different purposes” from the owner.²³² However, in the context of reproducing accessible copies for the print disabled, the proviso does not seem compatible. In *Hathitrust*, the Second Circuit clearly noted that when works are recast into accessible formats for the print disabled, “the underlying purpose of the [defendant’s] use is the same as the author’s original purpose.”²³³ Authorized entities, as defined by the Marrakesh Treaty, would thus be in violation of the proviso, handling these works in their capacity as the labor of rights holders rather than the commons.

3. *Means or End?*

Deontic theories also suffer from the identity crisis that Freeman accuses rights-sensitive consequentialist theories of having.²³⁴ Locke’s theory of natural rights is built on a deontic premise, but it seems to function on several consequentialist criteria.²³⁵ It has been shown how consequentialists purport to coherently incorporate rights as secondary rules.²³⁶ However, for deontic theories, in the case of Locke at least, the reverse cannot be said. Consequentialism can accommodate for a plurality of rules, but deontic theories struggle to accept the cohesiveness of various consequences.

Using the Marrakesh Treaty to substantiate the above point, a consequentialist approach allows that further rights be granted to a particular stakeholder group because the cost-savings to the owners and the increased benefits to the visually disabled, together with the dynamic benefits of cultural nourishment of a portion of society neglected by the market, justify it as the right move. However, a Lockean approach to the treaty struggles to explain why “enough and as good” warrants granting further access to the print disabled but not to other less fortunate individuals in developing countries whom a divine creator would deem equally deserving.

E. *Personality Theories*

1. *Users Missing*

A clear shortcoming of personality justifications is that they inadequately account for users. Justin Hughes refers to Margaret Radin’s personhood theory of property,²³⁷ identifying the “enough and as good” equivalent in personality

²³² *Id.* at 1604.

²³³ *Authors Guild, Inc. v. Hathitrust*, 755 F.3d 87, 101 (2d Cir. 2014).

²³⁴ See *supra* text accompanying note 164.

²³⁵ See e.g., Hughes, *supra* note 7, at 301-08 (discussing Lockean labor theory in instrumentalist terms); Gordon, *supra* note 44, at 1544 (mentioning utilitarian strands of Locke’s theory of property).

²³⁶ See *supra* Part III.A.

²³⁷ Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957 (1982).

theory.²³⁸ The proposition is that since property rights perform a self-actualizing function,²³⁹ these rights should not be given to person X if doing so would deny person Y's self-actualization.²⁴⁰ Appropriation of property is conditional upon "whether it has deleterious effects on others."²⁴¹

This conditional limitation does sound similar to the Marrakesh Treaty's goal of limiting authorial rights in order to increase access for the disabled with the aim of allowing the latter to realize their full potential.²⁴² However, the deleterious effects proviso is unhelpfully broad and seems to neglect incentivization considerations. In the copyright context, what is deleterious enough to warrant restriction? On this interpretation, authors' rights could be, ironically, left overly limited. Adhering to it could lead to the overly-extensive right to education feared by parties in the treaty's negotiations.²⁴³ Restricting access to educational materials could be argued to be deleterious, restricting copyright.

The consequentialist view does better in identifying the harm that should be avoided in reference to the system as a whole, namely, harm that stifles overall net progress. The consequentialist approach is also more balanced than the personhood approach, as it adequately considers both authors' and users' interest by using a consequentialist standard as a reference point.

Furthermore, even Hegel acknowledges the instrumentalism of intellectual property law. He admits that protecting intellectual property rights is the primary "means of advancing the science and arts."²⁴⁴ Thus, for Hegel, to the extent that he considers users (limiting personality-based rights to make the "resource" available for others),²⁴⁵ it is underscored by a consequentialist end.

Furthermore, the Marrakesh Treaty's main objectives provide the antithesis to Kant's autonomy explanation and agency theory.²⁴⁶ The treaty requires bypassing the very premise Kant's explanation of author's rights is based on: the permission of the author.²⁴⁷ This discrepancy can be attributed to the fact that unlike the Marrakesh Treaty, the rights of users are outside the

²³⁸ Hughes, *supra* note 7, at 336.

²³⁹ *See Id.*; HEGEL, *supra* note 54, at 62; Radin, *supra* note 237.

²⁴⁰ Hughes, *supra* note 7, at 336.

²⁴¹ *Id.*

²⁴² *See* Marrakesh Treaty, *supra* note 1, pmb. ¶¶ 2-4.

²⁴³ *See* Allan Adler Interview, *supra* note 86.

²⁴⁴ HEGEL, *supra* note 54, at 81.

²⁴⁵ *Id.* at 80.

²⁴⁶ *See supra* text accompanying note 52.

²⁴⁷ *See* Marrakesh Treaty, *supra* note 1, arts. 4-6 (on not requiring authorization by the rights holder to make, import or export accessible format copies).

periphery of Kant's agency justification.

2. Author-Centrism

Drassinower provides a more modern theory of autonomy for copyright that is not explicitly based on Hegel or Kant's theories.²⁴⁸ His focus on authors' and users' rights provides helpful insight into how to strike a balance in a consequentialist system. But as seen, he remains skeptical of the American consequentialist approach to copyright.²⁴⁹ His theory rightfully considers the free speech interests of authors,²⁵⁰ and it accounts for users much more substantially than Hegel or Kant.²⁵¹ However, it appears slightly over-individualistic, for it neglects the contributory aspects of works to societal progress. The tendency to exalt the status of authors may unintentionally (or intentionally) tilt the scale too far out of users' reach. One commentator notes that even in human rights instruments, there is no reflection of a hierarchy between authors' rights and users' rights.²⁵²

Therefore, the author-centric approach of Drassinower is aligned with the spirit of the Marrakesh Treaty insofar as it supports empowering the neglected disabled to become authors in their own right.²⁵³ Where Drassinower and the Marrakesh Treaty lack harmony is in their placement of normative force. Drassinower's priority for copyright is protecting the interests of individuals in their *capacity* as authors.²⁵⁴ The normative force of the Marrakesh Treaty lies in unencumbering societal participation so as to achieve the outcome of optimal progress.²⁵⁵ Paving the way for the disabled to maximize their authorial capabilities is the means of getting there.²⁵⁶

CONCLUSION

This note does not claim that consequentialism should be the best way to justify copyright. It simply acknowledges that consequentialism is the theory that is supported in the United States by its Constitution, statutes, courts, and many academics, undoubtedly much more so than the alternative theories.²⁵⁷

²⁴⁸ See DRASSINOWER, *supra* note 57.

²⁴⁹ Drassinower, *supra* note 186, at 479.

²⁵⁰ See generally DRASSINOWER, *supra* note 57 (on copyright as protecting author's speech).

²⁵¹ See Drassinower, *supra* note 186 (emphasizing the importance of user rights).

²⁵² Al-Sharieh, *supra* note 127, at 13.

²⁵³ Drassinower, *supra* note 186, at 469.

²⁵⁴ See generally DRASSINOWER, *supra* note 57 (focusing on the importance of authorship in copyright).

²⁵⁵ See *supra* text accompanying note 138.

²⁵⁶ See *supra* text accompanying note 133.

²⁵⁷ It would be worth exploring whether utilitarianism is really the class of

What this note does argue is that the consequentialist thinking that underpins U.S. copyright law has a strong presence in the Marrakesh Treaty. This note also asserts that the alternative theories have clear weaknesses in themselves, in justifying and explaining copyright, and specifically in justifying the Marrakesh Treaty. Admittedly, their weaknesses have not been exhaustively covered in this note, nor have their roles in foreign copyright systems been addressed.

I do not doubt that analyzing the Marrakesh Treaty against a consequentialist framework may have swayed me to seek out the utilitarian aspects of the treaty. But that does not mean they were never there. In fact, I may have missed out on much of the treaty's consequentialist underpinnings if I had assessed its philosophical justifications in isolation, perhaps myopically focusing on its prominent human rights references, neglecting the careful balance it attempts to strike between different stakeholders. A question that remains open is whether the Marrakesh Treaty has shifted that balance toward users in general, for it has certainly shifted the balance in favor of its specific beneficiaries. However, this note is focused on *why* the balance is struck, and not *how* it has shifted. The note demonstrates that the Marrakesh Treaty is consequentialist, but what could be further explored is whether its implementation would move U.S. copyright law toward a brand of consequentialism that determines that progress is better fueled by broadening users' rights.

I recognize that some may find a consequentialist approach to the rights of the visually disabled too clinical. Perhaps outside the intellectual property realm, a deontic approach would sound kinder. But I believe that in order to uphold the durability and cohesiveness of the copyright system, the Marrakesh Treaty's evidently consequentialist purpose (in the United States at least) cannot be flouted. Accepting the Marrakesh Treaty's compatibility as such recognizes that the reasons the visually disabled should be granted more access to works can be primarily found within copyright law itself, and not just taken to be a benevolent act of charity. This approach upholds both the integrity of the copyright system and the disabled beneficiaries of the treaty.

Undoubtedly, one could look at the Marrakesh Treaty as advocating inclusion and equality for their own sake. Accepting its consequentialist connotations, however, allows us to see that assisting the visually disabled is for the benefit of society as a whole. Not only does it make society more inclusive and respectful of the rights of their fellow human beings, but it will hopefully, in time, reward its members with the future contributions of many individuals who have thus far been denied the means to flourish.

consequentialism that the American intellectual property system is, or should be, based on.



JIPEL's 2017 "Careers in IP Symposium" brought together a broad cross-section of attorneys specializing in intellectual property law. Speakers included both established partners and young associates with experience at large and boutique law firms, as well as in-house counsel for a major pharmaceutical company and the New York Times. The panelists' collective expertise ranged from concentrations in cybersecurity and data privacy to patent law to fashion law.

Anne Hassett, the executive director of NYU Law's Engelberg Center on Innovation Law & Policy, moderated the evening's first panel, which focused on IP attorneys with litigation practices. The panelists emphasized that many roads lead to IP litigation. Whereas Skadden partner Lauren Aguiar ('95) began her legal career without a definite interest in IP, Wilkie Farr associate Alexandra Awai joked that being an IP attorney had been in her plans since early childhood. The panel also discussed some recurring themes of the litigation practice area. Paul, Weiss partner Jennifer Wu ('04) talked about the challenges of presenting complex and highly technical factual scenarios to lay juries, and Teva Pharmaceutical's Colman Ragan emphasized a similar challenge in presenting legal issues within the context of a business case to non-lawyer managers and directors. Finally, Gibson Dunn's George Stamatopoulos ('12) recounted his experience pursuing a JD as a Greek international student with a background in mechanical engineering.



The second panel of the evening focused on attorneys with transactional practices, and NYU Professor of Law Christopher Sprigman facilitated discussion. Lori Lesser, the Head of Simpson Thacher's Intellectual Property Transactions Practice, advised students to be receptive to unplanned opportunities. Indeed, the panelists all agreed that elements of luck and "being in the right place at the right time" played a bigger part in their career trajectories than they had anticipated as law students. Panelists also talked about their experiences developing niche practices. Douglas Hand ('97) described the transition from working as an M&A associate at a big firm to starting his own boutique, Hand Baldachin & Amburgey LLP, which represents clients in the fashion industry. Conversely, White & Case's Jeremy Apple discussed how his public sector experience in cybersecurity and privacy led him to a position in big law. In addition, Preeta Reddy talked about life as a junior IP associate at Wilkie Farr, and Richard Samson reflected on his work reviewing content for the New York Times as part of its legal team.

After a question-and-answer period, the panelists and attendees enjoyed the opportunity to talk further over cocktails and refreshments. All in all, it was a lively and informative night for students excited by the prospect of working in IP.



© NYU Photo Bureau: Slezak