VILLANOVA LAW REVIEW

VOLUME 61 2016 NUMBER 1

CONTENTS

ARTICLES

The text and citations conform to *The Bluebook: A Uniform System of Citation* (19th ed. 2010), copyright by the *Columbia, Harvard,* and *University of Pennsylvania Law Reviews,* and the *Yale Law Journal*; *Manual on Usage, Style & Editing* (12th ed. 2011), copyright by the *Texas Law Review*; and the *U.S. Government Printing Office Style Manual* (2008).

The *Villanova Law Review* seeks to publish materials representing meritorious viewpoints on subjects of interest to the legal profession. Publication does not indicate that the views expressed are adopted by the *Villanova Law Review.*

As a matter of policy, the *Villanova Law Review* encourages and promotes the use of gender-neutral writing.

The *Villanova Law Review* (ISSN-0042-6229) is published in February, March, June, August, and October by Villanova University Charles Widger School of Law at 299 North Spring Mill Road, Villanova, Pennsylvania 19085-1682. Periodicals postage paid at Villanova, Pennsylvania and at additional mailing offices. Annual Subscription $35.00 . . . . . Single Issue of Current Volume $8.00 . . . . . Special and Symposium Issues $10.00.

POSTMASTER: Send address changes to Villanova Law Review, Villanova University Charles Widger School of Law, 299 North Spring Mill Road, Villanova, Pennsylvania 19085-1682.

UNLESS NOTICE TO THE CONTRARY IS RECEIVED AT THE EDITORIAL OFFICE, IT IS ASSUMED THAT A RENEWAL OF THE SUBSCRIPTION IS DESIRED.

SUBSCRIPTIONS ARE BILLED ANNUALLY.

ONE MONTH'S NOTICE IS NECESSARY TO EFFECT A CHANGE OF ADDRESS.

Change of address notification, subscription requests,
article submissions, requests for expedited review,
and all other correspondence should be sent to:
Villanova Law Review
Villanova University Charles Widger School of Law
299 North Spring Mill Road
Villanova, Pennsylvania 19085-1682

Or via e-mail:
Review@law.villanova.edu
http://www.law.villanova.edu/academics/Journals/LawReview.aspx

Address requests for back issues to:
William S. Hein & Co., Inc.
2350 North Forest Road
Getzville, NY 14068

VILLANOVA UNIVERSITY
CHARLES WIDGER SCHOOL OF LAW
2015–2016

The Villanova University Charles Widger School of Law is a graduate professional school providing a three-year course of instruction in the law and its history and philosophy. The School of Law is fully accredited by the American Bar Association and is a member of the Association of American Law Schools.

A J.D. degree is granted on completion of a three-year full-time course of study. An LL.M. degree in Taxation is granted to those who complete the Graduate Tax Program. An LL.M. degree in International Study can be earned through joint programs offered with universities in London, Edinburgh, Leiden, the Netherlands, Singapore, and Hong Kong. The School of Law offers a joint J.D./M.P.A. program with Villanova University College of Liberal Arts and Sciences. In addition, the School of Law offers a joint J.D./M.B.A. program in collaboration with Villanova's top-ranked School of Business.

The School of Law is located on the Villanova campus on Philadelphia's suburban Main Line, fifteen miles west of Philadelphia, Pennsylvania. The building—brand new in 2009—is devoted exclusively to the study of law. The innovative building offers state-of-the-art classrooms and facilities focused on student needs. It has received LEED Gold certification by the U.S. Green Building Council and includes five state-of-the-art courtrooms, including the Martin G. McGuinn Ceremonial Courtroom. The building also features computer labs, dining facilities, student lounges, and a 500,000 volume law library.

LAW REVIEW

The *Villanova Law Review* is a national legal periodical publishing five issues per year by a board of student editors. The *Law Review* offers an opportunity for valuable experience in legal writing and research. In addition, the *Law Review* has created *Tolle Lege*, a dedicated online journal which showcases articles by members of the bench, bar, and academia on timely legal issues. Students are invited to join the *Law Review* on the basis of scholarship or performance in an open-writing competition.

ENVIRONMENTAL LAW JOURNAL

The *Villanova Environmental Law Journal* is a national legal publication dedicated to the scholarly review of environmentally related topics. The *Environmental Law Journal* is published two times per year in the winter and the summer by a staff of student editors and writers who have been invited to join the *Environmental Law Journal* based on their performance in an open-writing competition.

JEFFREY S. MOORAD SPORTS LAW JOURNAL

The *Jeffrey S. Moorad Sports Law Journal* is a national legal periodical published and edited two times per year by a staff of students chosen for membership based on their performance in an open-writing competition. The *Sports Law Journal* is committed to publishing scholarly articles on topics of importance in the sports law field.

JOURNAL OF CATHOLIC SOCIAL THOUGHT

The *Journal of Catholic Social Thought* is a national legal periodical published two times per year. It is a peer-reviewed interdisciplinary publication that brings together authors from the fields of law, theology, philosophy, and political theory. It is jointly published by Villanova University Charles Widger School of Law and the Office for Mission and Ministry at Villanova University.

EXPERIENTIAL LEARNING

The Villanova University Charles Widger School of Law provides students with a wealth of opportunities for practical, hands-on learning that goes beyond traditional legal theory and principles taught in the classroom. Students are encouraged to take part in a number of different programs: represent clients from the local community in one of our six legal Clinics; hone their litigation skills by participating in Moot Court, taking the Intensive or Advanced Trial Advocacy courses, or competing on the National Trial Team; put their legal writing and research skills to the test by joining the staff of one of our three published and two online journals; explore their interest in Sports Law by becoming involved with the Jeffrey S. Moorad Center for the Study of Sports Law; learn where the law meets business in Villanova's new Center for Law, Entrepreneurship, and Innovation; go beyond the law school's walls and complete an externship in a variety of legal settings from clerkships to small or large law firm employment; or explore the legal system of another nation through our Rome Summer Program or one of our semester abroad programs.

MOOT COURT

The Moot Court Board is the principal oral advocacy organization at Villanova University Charles Widger School of Law. The Board is chiefly responsible for the administration of the Theodore Reimel Moot Court Competition, an intra-school tournament in honor of the late Theodore L. Reimel, Judge for the Pennsylvania Court of Common Pleas from 1953 to 1973. The Board also competes in appellate advocacy competitions at law schools across the country. Finally, the board assists the legal writing faculty in the administration and judging of 1L oral arguments. Students are invited to join the board through a summer competition or by advancing to the Semi-Final round of the Reimel Moot Court Competition.

THE STUDENT BAR ASSOCIATION

The Student Bar Association of the School of Law is the official resource and liaison for the student body in matters involving the school administration, faculty, and students. Upon enrollment at the Villanova University Charles Widger School of Law, each student becomes a member of the organization. The SBA Board consists of nineteen positions that are elected by the student body and one appointed position. The Association conducts student government and extra-curricular activities at the School of Law and is affiliated with the Law Student Division of the American Bar Association.

THE ORDER OF THE COIF

The Order of the Coif, a national honor society with chapters in most leading law schools, selects members from among the School of Law's graduating students who are in the top ten percent of the class and who have contributed substantially to the Law School.

ADMISSION

A bachelor's degree, or equivalent, is required for admission. Applicants are required to take the Law School Admission Test. For further information, address the Admissions Office, Villanova University Charles Widger School of Law, 299 North Spring Mill Road, Villanova, Pennsylvania 19085-1682 or call 610-519-7010.

VILLANOVA UNIVERSITY
CHARLES WIDGER SCHOOL OF LAW
2015–2016

OFFICERS OF THE UNIVERSITY

THE REVEREND PETER M. DONOHUE, O.S.A., B.A., M.A., Ph.D., *President*
THE REVEREND KAIL C. ELLIS, O.S.A., Ph.D., *Vice President for Academic Affairs*

SENIOR ADMINISTRATION

JOHN Y. GOTANDA, B.B.A., J.D., *Arthur J. Kania Dean and Professor of Law*
TODD AAGAARD, B.A., J.D., M.S., *Vice Dean and Professor of Law*
APRIL M. BARTON, B.S., J.D., *Associate Dean for Academic Affairs, Director of the Reza Taleghani J.D./M.B.A. Programs, and Director of the J.D./M.P.A.*
MICHELLE MADDEN DEMPSEY, B.A., J.D., LL.M., D.Phil., *Associate Dean for Faculty Research and Professor of Law*
VICENÇ FELIÚ, B.A., M.S., J.D., LL.M., *Associate Dean for Library Services and Associate Professor of Law*
JAMES C. GARVEY, B.A., M.B.A., Ed.D., *Associate Dean for Development and Alumni Relations*
JENNIFER A. HENFEY, B.S., J.D., *Associate Dean for Finance and Administration, Executive Director of Career Strategy and Advancement*

FULL-TIME FACULTY

TODD AAGAARD, B.A., J.D., M.S., *Professor of Law*
HEATHER D. BAUM, B.A., J.D., *Professor of Law*
LESLIE M. BOOK, B.A., J.D., LL.M., *Professor of Law*
RICHARD BOOTH, A.B., J.D., *Professor of Law and Martin G. McGuinn Chair in Business Law*
PATRICK MCKINLEY BRENNAN, B.A., M.A., J.D., *Professor of Law and John F. Scarpa Chair in Catholic Legal Studies*
DORIS DEL TOSTO BROGAN, B.A., J.D., *Professor of Law*
JOHN J. CANNON, A.B., J.D., *Professor of Law*
MICHAEL CAMPBELL, A.B. J.D., *Professor of Law, Director of Clinical Programs, and Director of the Interdisciplinary Mental and Physical Health Clinic*
DAVID S. CAUDILL, B.A., J.D., Ph.D., *Professor of Law and Goldberg Family Chair in Law*
CANDACE MUELLER CENTENO, B.A., J.D., *Professor of Law*
STEVEN L. CHANENSON, B.A., M.S., J.D., *Professor of Law*
JOSEPH W. DELLAPENNA, B.B.A., J.D., LL.M., *Professor of Law*
MICHELLE MADDEN DEMPSEY, B.A., J.D., LL.M., D.Phil., *Associate Dean and Professor of Law*
JOHN F. DOBBYN, A.B., J.D., LL.M., *Professor of Law*
DIANE PENNEYS EDELMAN, A.B., J.D., *Professor of Law and Director of International Programs*
VICENÇ FELIÚ, B.A., M.S., J.D., LL.M., *Associate Dean for Library Services and Associate Professor of Law*

T. Keith Fogg, B.A., J.D., LL.M., *Professor of Law*
Abraham J. Gafni, B.A., B.H.L., J.D., *Professor of Law*
Ruth Gordon, B.A., J.D., LL.M., *Professor of Law*
John Y. Gotanda, B.B.A., J.D., *Dean and Professor of Law*
Ann Carey Juliano, B.A., J.D., *Professor of Law*
Catherine J. Lanctot, B.A., J.D., *Professor of Law*
James Edward Maule, B.S., J.D., LL.M., *Professor of Law*
Melanie McMenamin, B.S., B.A., J.D., *Assistant Professor of Law and Director of the Clinic for Law and Entrepreneurship*
Christine Mooney, B.A., B.S., J.D., *Professor of Law and Director of the Duane Morris LLP Legal Writing Program*
Michael P. Moreland, B.A., M.A., J.D., Ph.D., *Professor of Law*
Joy Sabino Mullane, B.A., J.D., LL.M., *Professor of Law*
Mitchell Nathanson, B.A., J.D., *Professor of Law*
Jennifer O'Hare, B.S., J.D., *Professor of Law*
Michele R. Pistone, B.S., J.D., LL.M., *Professor of Law*
Anne Bowen Poulin, B.A., J.D., LL.M., *Professor of Law*
Teressa E. Ravenell, B.A., J.D., *Professor of Law*
Michael Risch, A.B., J.D., *Professor of Law*
Mary Ann Robinson, B.A., J.D., *Associate Professor of Law*
Chaim N. Saiman, B.B.S., J.D., *Professor of Law*
Tuan Samahon, B.A., J.D., *Professor of Law*
Dveera Segal, B.A., M.A., J.D., *Professor of Law*
Louis J. Sirico, Jr., B.A., J.D., *Professor of Law and Director of Legal Writing*
Walter John Taggart, A.B., J.D., *Professor of Law*
Jessica Webb, B.A., J.D., *Assistant Professor of Law*
Ellen Wertheimer, B.A., J.D., *Professor of Law*

VISITING FACULTY

Andrew Brandt, B.A., J.D., *Executive Director of the Jeffrey S. Moorad Center for the Study of Sports Law and Visiting Professor*
Caitlin Barry, B.A., J.D., *Visiting Assistant Professor of Law*
David A. Breen, B.S., M.B.A., J.D., *Visiting Assistant Professor of Practice*

DISTINGUISHED PROFESSOR OF PRACTICE

J. Richard Harvey, Jr., B.S.B.A., MSc.

EMERITUS FACULTY

Gerald Abraham, A.B., J.D., *Professor Emeritus of Law*
Lewis Becker, B.S., LL.B., *Professor Emeritus of Law*
Donald N. Bersoff, B.S., M.A., Ph.D., J.D., *Professor Emeritus of Law*
Donald W. Dowd, A.B., J.D., *Professor Emeritus of Law*
John M. Hyson, B.A., M.A., LL.B., *Professor Emeritus of Law*
Don W. Llewellyn, B.A., J.D., LL.M., *Professor Emeritus of Law*
Howard R. Lurie, A.B., J.D., *Professor Emeritus of Law*
Michael Mulroney, B.S.C., J.D., *Professor Emeritus of Law*
John F. Murphy, B.A. LL.B., *Professor Emeritus of Law*
J. Willard O'Brien, B.S., J.D., *Dean and Professor Emeritus of Law*
Leonard Packel, B.S., J.D., *Professor Emeritus of Law*
Frederick P. Rothman, A.B., J.D., *Professor Emeritus of Law*

LEGAL RESEARCH INSTRUCTORS

Benjamin Carlson, B.A., J.D., M.L.S., *Emerging Technologies Librarian and Legal Research Instructor*
Lori Strickler Corso, B.A., B.S., J.D., M.L.I.S., *Electronic Services Librarian and Legal Research Instructor*
Robert Hegadorn, B.A., J.D., M.L.S., *Reference Librarian and Legal Research Instructor*
Matthew McGovern, B.S., J.D., M.S.L.I.S., *Director of Experiential Learning and Legal Research Instructor*
Amy Spare, B.A., J.D., M.L.I.S., *Faculty Services Librarian and Legal Research Instructor*

WRITING SPECIALIST

Steven Schultz, B.A.

ACADEMIC DIRECTORS

April M. Barton, B.S., J.D., *Associate Dean, Director of the Reza Taleghani J.D./M.B.A. Program and Director of the J.D./M.P.A. Program*
Matthew Carluzzo, A.B., J.D., *Director of Academic Success and Bar Support Programs*
Diane Penneys Edelman, A.B., J.D., *Professor of Law and Director of International Programs*
Rufus King, A.B., B.Phil., J.D., *Director of the Center for Law, Entrepreneurship and Innovation*
Matthew McGovern, B.S., J.D., M.S.L.I.S., *Director of Experiential Learning and Legal Research Instructor*
Christine Mooney, B.A., B.S., J.D., *Professor of Law and Director of Legal Writing Programs*
Linda Love Vines, B.S., J.D., LL.M., *Instructor and Interim Director of the Graduate Tax Program*

ADMINISTRATIVE DIRECTORS

Anthony Cancelli, *Director of Business Support Services*
Colleen Frascatore, *Executive Director of Development and Campaign*
Nicole Garafano, *Director of Events*
Jennifer A. Henfey, *Associate Dean and Executive Director of Career Strategy and Advancement*
Joshua Liss, *Director of Annual Giving and Donor Relations*
Joseph Mariani, *Director of Facilities*
Bayrex Martí, *Executive Director of Admissions and Financial Aid*
Brian McCabe, *Campus Minister for Outreach and Student Formation*
Susan K. Mesick, *Registrar*
Sarah N. Ponzio, *Director of Alumni Relations*
Nancy Whalen, *Director of Student Affairs*
Kristina Wilhelm-Nelson, *Associate Director of Financial Aid*

J.D. ADJUNCT FACULTY

Jon Abramczyk, B.A., B.S., J.D.
Stephen Anderer, B.A., J.D., Ph.D.
Aaron Avila, B.A.S., J.D.
David Bailey, B.S., J.D.
Joe Banner, B.A.
Ronald N. Brown III, B.A., J.D.
Mark Bullock, B.A., J.D.
April Byrd, B.A., J.D.
John Carle, B.A., J.D.
Peter Clauss, A.B., LL.B.
A. Roy DeCaro, B.A., J.D.
Donald DiCarlo, Jr., B.A., J.D., LL.M., M.A.
John F. Dobbyn, A.B., J.D.
Amy Donnella, B.A., J.D.
Bryan Franey, B.A., ISAT, J.D.
Douglas Gaston, B.A., J.D.
Richard J. Gering, M.A., Ph.D.
The Honorable Gary S. Glazer, B.A., J.D.
Richard Goldberg, B.A., LL.B.
Aleksander J. Goranin, B.A., J.D.
Christopher Harvey, B.S., J.D., LL.M.
Matthew A. Homyk, B.S.F.S., J.D.
James A. Hunter, Jr., A.B., LL.B.
Rosalind T. Kaplan, B.A., M.A., J.D.
Amy S. Kline, B.A., J.D.
Steven Koffs, B.S.M.E., M.S., J.D.
Dale Larrimore, B.A., J.D.
Jordan A. LaVine, B.A., J.D.
Dale Larrimore, B.A., J.D.
Sherry Lemonick, B.S., J.D.
Joseph W. Lesovitz, C.P.A., C.F.A., C.F.E., C.F.F.
Stephen L. Liedtka, B.S., M.S., Ph.D.
Christy Martin, B.A., J.D.
James McCartney, B.A., M.A., M.S., Ph.D.
Patricia McCausland, B.A., M.A., J.D.
Richard L. McMonigle, B.A., J.D.
Coleen Meehan, B.S., J.D.
Maranna Meehan, B.A., J.D.
Mark Momjian, B.A., J.D.
Michelle L. Morgan, B.A., J.D.
Cheryl Nicolson, B.A., J.D.
Wendy Norman, B.S., J.D.
Mary DeFusco Ochal, B.A., J.D.
Michael O'Mara, B.S., J.D.
The Honorable Sheldon K. Rennie, B.A., M.S., J.D.
Bernard M. Resnick, B.A., J.D.
The Honorable Eduardo Robreno, B.A., J.D.
Lee Rosengard, A.B., J.D.
Lisa Sabatino, B.A., J.D.
The Honorable Juan Sanchez, B.A., J.D.
David Santee, B.A., J.D.

THE HONORABLE TODD SEELIG, B.S., J.D.
ROBERT SIEGER, B.A., J.D.
JUDY GOLDSTEIN SMITH, B.A., J.D.
SHAUN SNITMAN, LL.B.
STEVEN D. SPENCER, B.S., J.D.
JONATHAN SPERGEL, A.B., J.D.
TOMIKA STEVENS PATTERSON, B.A., M.A., J.D.
STEVEN SUGARMAN, B.A., J.D.
MICHAEL TRACHTMAN, B.A., J.D.
COLLEEN VALLEN, B.S.
SHEILAH VANCE, B.A., J.D.A
AYANNA WILLIAMS, A.B., J.D.
R. SETH WILLIAMS, B.A., J.D.
ANDREW YOUMAN, B.A., J.D.
MARC J. ZUCKER, B.A., J.D.

GRADUATE TAX PROGRAM
FULL-TIME FACULTY

SCHOOL OF LAW

LESLIE M. BOOK, B.A., J.D., LL.M., *Professor of Law*
T. KEITH FOGG, B.A., J.D., LL.M., *Professor of Law*
JOY SABINO MULLANE, B.A., J.D., LL.M., *Associate Professor of Law*
LINDA LOVE VINES, B.S., J.D., LL.M., *Instructor and Interim Director of Graduate Tax Program*

SCHOOL OF BUSINESS

SHELLY C. RHOADES-CATANACH, B.S., Ph.D., *Associate Professor*

EMERITUS FACULTY

MICHAEL MULRONEY, B.S.C., J.D., *Professor Emeritus of Law*

DISTINGUISHED PROFESSOR OF PRACTICE

J. RICHARD HARVEY, JR., BSBA, MSC.

GRADUATE TAX ADJUNCT FACULTY

GREGORY T. ARMSTRONG, B.A., J.D., LL.M.
RYAN M. BORNSTEIN, B.S., J.D., LL.M.
TIMOTHY C. BOYLE, BSBA, M.B.A.
JOHN BRADY, B.S., J.D., LL.M.
DAVID A. BREEN, B.S., M.B.A., J.D.
RONALD S. COLLINS, B.A., J.D., LL.M.
THOMAS J. DONNELLY, B.S., J.D., LL.M.
JAMES C. FEE, JR., B.S., J.D., LL.M.
NEIL H. FEINSTEIN, B.S., J.D., LL.M.
JARED H. GORDON, B.A., J.D., LL.M.
PETER J. HARDY, B.A., J.D.
CHRISTOPHER M. HARVEY, B.S., J.D., LL.M.
TIMOTHY J. HOLMAN, B.A., J.D.
JOEL E. HOROWITZ, B.A., J.D., LL.M.
RICHARD S. HYMAN, B.S., M.T.
THOMAS J. IANNOZZI, B.A., J.D., LL.M.
JASON KURATNICK, B.S., J.D., LL.M.
JONATHAN LANDER, B.A., J.D., LL.M.
EDWARD A. LIVA, B.S., B.A., J.D., LL.M.
SCOTT MASAITIS, B.A., J.D., LL.M.
MICHAEL J. MCGOLDRICK, B.A., J.D., M.L.D., LL.M. (Taxation), LL.M. (Corporate)
ERICA R. MCREYNOLDS, B.A., J.D., LL.M.
DREW A. MORRIS, B.A., J.D., LL.M., M.B.A.
DANIEL C. MUNCE, B.A., J.D., LL.M.
THOMAS OSTRANDER, B.A., J.D., LL.M.
PHILLIP PILLAR, B.A., J.D., LL.M.
THOMAS M. RATH, B.B.A., J.D., LL.M.
JOSEPH E. RONAN, JR., B.A., J.D., LL.M.
DAVID SHAPIRO, B.A., J.D.
MICHAEL E. STEINSALTZ, B.S., J.D., LL.M.
CHRISTOPHER P. WOEHRLE, B.A., J.D., LL.M.
DAVID B. ZELIKOFF, B.A., J.D.

The Harry S. Cherken, Jr. Law Review Fund
Provides Financial Support For The Law Review

VILLANOVA LAW REVIEW

VOLUME 61 2016 NUMBER 1

Articles

THE COST OF RAISING A KILLER—PARENTAL LIABILITY FOR THE PARENTS OF ADULT MASS MURDERERS

SHAUNDRA K. LEWIS*

"People never really recover from the death of a child. How much more difficult, then, is it to deal with the deaths of other children whom your child has killed? . . . So much shame surrounds the issue of mental illness that many parents don't seek help for their kids or are afraid to talk about it publicly. . . . [T]hat needs to change."[1]

INTRODUCTION

ONE simply cannot ignore the significant number of mass shootings continuing to occur in places one would least expect gun violence to erupt.[2] In the last seven years, there has been a shooting rampage at a

* Associate Professor of Law at Thurgood Marshall School of Law at Texas Southern University; J.D. 1998, St. Thomas University School of Law, *summa cum laude*; B.A. 1995, Stetson University. This Article was supported by a 2013 summer research grant from Texas Southern University Thurgood Marshall School of Law. The author thanks Professor Nancy Levit at the University of Missouri-Kansas City School of Law for her invaluable feedback and encouragement on this project. This author also thanks Professor L. Darnell Weeden of the Thurgood Marshall School of Law Faculty, the Thurgood Marshall School of Law Faculty, Associate Dean Andrew Kim and the Concordia School of Law Faculty in Boise, Idaho, and the Southeastern Annual Legal Conference participants for their extremely helpful contributions. The author also greatly appreciates the research assistance provided by law students Jack Batiste and Daniel DeLuna. Finally, the author thanks Dean Dannye Holley and Associate Dean Asmara Tekle of Thurgood Marshall School of Law for their support.

1. Belinda Luscombe, *Elliot Rodger's Parents Break Silence on Shooting*, TIME (May 29, 2014), http://time.com/136764/elliot-rodger-parents/ [https://perma.cc/HK4X-ZP2V].

2. Shaundra K. Lewis, *Firearm Laws Redux–Legislative Proposals for Disarming the Mentally Ill Post-Heller and Newtown*, 3 MENTAL HEALTH L. & POL'Y J. 320, 323 (2014) (observing that "public mass shootings are now ubiquitous in America"). According to a recent FBI report, from 2000 to 2013, there were approximately 160 active mass shooter incidents across America, including "in urban and rural areas, and in 40 of 50 states and the District of Columbia." J. PETE BLAIR & KATHERINE W. SCHWEIT, TEX. STATE UNIV. & FBI, U.S. DEP'T OF JUSTICE, A STUDY OF ACTIVE SHOOTER INCIDENTS IN THE UNITED STATES BETWEEN 2000 AND 2013, at 8 (2014),

grocery store,[3] movie theater,[4] elementary school,[5] and even a church.[6] Far too often, the shooter displayed clear signs of mental illness prior to their murderous attack,[7] inevitably leading many to ponder whether the shooter's parents were partially to blame.[8] Were they too permissive? Were they not loving enough? Were they aware of their child's deranged state and had a responsibility to protect the public? Do they have a duty to supervise an adult, mentally ill child? At the very least, should parents warn the proper authorities about their potentially deadly kin?

The quintessential example raising these issues is the case of Adam Lanza's single mother, Nancy Lanza. Ms. Lanza's twenty-year-old son utilized her firearms to inexplicably gun down twenty school-children and six educators at Sandy Hook Elementary School in Newtown, Connecticut, in 2012.[9] At the time of the massacre, her son, who had a significant history of mental illness, resided with her in her large suburban home where, to the bewilderment of many, she stockpiled an arsenal of firearms—including a very powerful semi-automatic rifle.[10] Notwithstanding Adam's mental health issues, Ms. Lanza gave him unfettered access to her legally purchased the weaponry.[11] In fact, when the police searched the Lanzas'

---

available at https://www.fbi.gov/about-us/office-of-partner-engagement/active-shooter-incidents/a-study-of-active-shooter-incidents-in-the-u.s.-2000-2013 [https://perma.cc/R9DG-7AC3].

3. *See* Tung Yin, *Were Timothy McVeigh and the Unabomber the Only White Terrorists?: Race, Religion, and the Perception of Terrorism*, 4 ALA. C.R. & C.L. L. REV. 33, 50 (2013).

4. *See* Jana R. McCreary, *"Mentally Defective" Language in the Gun Control Act*, 45 CONN. L. REV. 813, 823 (2013).

5. *See* Lewis, *supra* note 2, at 322.

6. *See Manhunt on for South Carolina Church Gunman*, CBSNEWS (June 18, 2015, 10:21 AM), http://www.cbsnews.com/news/9-dead-in-shooting-in-historic-black-south-carolina-church/ [https://perma.cc/Q898-XYK3]; s*ee also* BLAIR & SCHWEIT, *supra* note 2, at 8 (averaging 16.4 active shooter incidents occurring from 2007–2013 in various places—including city streets, health care facilities, and houses of worship—which is more than double the number of such incidents from 2000–2006, when an average of only 6.4 such incidents occurred annually).

7. *See* Lewis, *supra* note 2, at 324, 325–38 (noting "the all too familiar storyline behind nearly every shooting spree, prior to the shootings, the gunman exhibited clear signs of mental illness[ ]").

8. *See, e.g.*, Garance Franke-Ruta, *What About the Parents?*, ATLANTIC (Jan. 10, 2010), http://www.theatlantic.com/politics/archive/2011/01/what-about-the-parents/69176/ [perma.cc/4JZZ-LAXZ]; *Is Nancy Lanza to Blame for Her Son's Actions*, DEBATE.ORG, http://www.debate.org/opinions/is-nancy-lanza-to-blame-for-hersons-actions [https://perma.cc/K2H4-BX24] (last visited Feb. 29, 2016) (citing online poll where 89% of respondents blame Nancy Lanza for Adam Lanza's shooting rampage at Sandy Hook Elementary School).

9. *See* Lewis, *supra* note 2, at 322.

10. STEPHEN J. SEDENSKY III, STATE OF CONN. DIV. OF CRIMINAL JUSTICE, REPORT OF THE STATE'S ATTORNEY FOR THE JUDICIAL DISTRICT OF DANBURY ON THE SHOOTINGS AT SANDY HOOK ELEMENTARY SCHOOL AND 36 YOGANANDA STREET, NEWTOWN, CONNECTICUT ON DECEMBER 14, 2012, at 28, 36 (2013).

11. *See id.* at 28 (observing Adam lived alone with his mother). A search of the Lanza home after the shooting rampage revealed a gun locker intact that had

home, they discovered a check Ms. Lanza had given her son for Christmas to purchase his own gun.[12] Moreover, Ms. Lanza frequently took her son to shooting ranges,[13] despite an unconfirmed report that she was contemplating having him involuntarily committed to a psychiatric institution around the same time of the shootings.[14] Unfortunately, Ms. Lanza paid the ultimate price for this lapse in parental judgment with her own life, as her son shot and killed her before embarking on his elementary school killing spree.[15] Nevertheless, this did not stop the Sandy Hook victims' loved ones from suing Nancy Lanza's estate for wrongful death,[16] raising the question of whether a parent can be civilly liable for a mass shooting committed by their adult offspring under general common law negligence principles.

This Article is the first to explore this issue in great detail, using Nancy Lanza and other parents of notorious shooters as examples. Specifically, this Article attempts to evaluate, clarify, and build upon parental negligence law. Part I discusses the evolution of parental liability. Part II addresses whether there can be parental liability for parents of adult mass shooters based upon a special relationship. Part III provides an overview of negligence in general and its complexities, as well as exploring whether a duty to protect or warn can be established in mass shooting cases. Part IV examines whether the parents in the real-life examples breached a duty to protect or warn. Part V analyzes whether the aforementioned parents' breach caused the shooting victims' injury or death. Part VI concludes that in some circumstances parents can and should be held liable for negligence that leads to their child committing a mass killing. It further submits that the mere possibility of parents being subjected to financial liability for their child's mass shooting will not only incentivize parents to take more aggressive measures to keep firearms out of their mentally ill

---

not been broken into—the inference being that Adam had access to his mother's arsenal and did not forcibly steal her weapons. *See* Bill Cummings, Ken Dixon & John Pirro, *Lanza's Secret Arsenal: Swords, Guns and 1,400 Rounds of Ammo*, HOUS. CHRON., Mar. 28, 2013, 9:05 PM, http://www.chron.com/news/nation-world/article/Lanza-s-secret-arsenal-swords-guns-and-1-400-4393243.php [https://perma.cc/6TNK-KBEF].

12. *See* SEDENSKY, *supra* note 10, at 28.

13. *See id.* at 31; *see also* Cummings et al., *supra* note 11; Matt Flegenheimer & Ravi Somaiya, *A Mother, a Gun Enthusiast and the First Victim*, N.Y. TIMES, Dec. 15, 2012, http://www.nytimes.com/2012/12/16/nyregion/friends-of-gunmans-mother-his-first-victim-recall-her-as-generous.html?_r=0 [https://perma.cc/U42J-5638].

14. *See* Jana Winter, *Exclusive: Fear of Being Committed May Have Caused Connecticut Gunman to Snap*, FOXNEWS (Dec. 18, 2012), http://www.foxnews.com/us/2012/12/18/fear-being-committed-may-have-caused-connecticut-madman-to-snap/ [https://perma.cc/M57Q-DLNH].

15. *See* Lewis, *supra* note 2, at 322.

16. Dave Altimari, *11th Sandy Hook Family Files Wrongful Death Claim Notice*, HARTFORD COURANT, Dec. 10, 2014, 11:10 AM, http://www.courant.com/news/connecticut/hc-sandy-hook-shooting-parents-may-sue-p-20141208-story.html [https://perma.cc/BF34-K4FU].

child's hands, but to obtain the mental health assistance their child so desperately needs—measures that in the end will make everyone safer. Finally, Part VI provides advice to parents for dealing with significantly mentally ill, adult children residing in their home.

I. A HISTORICAL PERSPECTIVE ON PARENTAL LIABILITY

As explained in more detail below, while in early American history parents generally were not personally liable for their children's torts, the law has evolved over time such that parents can now be held civilly responsible for their offspring's actions under certain circumstances, even for a criminal attack like a mass shooting.

A. *Early Common Law on Parental Liability*

Originally, under early common law, parents could not be found vicariously liable for their child's torts solely because of their parent-child relationship.[17] Children were considered separate legal entities responsible for their own torts, and any judgment had to be satisfied by them personally, which left most victims uncompensated since children are usually insolvent and cannot satisfy a judgment.[18] Parents were only responsible for their children's actions if the parents themselves were somehow independently negligent by, for example, entrusting a minor child with a dangerous weapon.[19]

B. *Parental Liability Statutes*

To both address the problem of victims going uncompensated for property damage or personal injuries caused by children and to deter juvenile delinquency, every state enacted some form of a parental liability stat-

17. *See, e.g.*, Alioto v. Marnell, 520 N.E.2d 1284, 1286 (Mass. 1988) (noting that one's parental status, without more, does not impose duty to supervise and control emancipated adult children); Guzy v. Gandel, 229 A.2d 809, 810 (N.J. Super. Ct. App. Div. 1967) (declaring "[t]he mere existence of the relationship of parent and child does not render a parent liable for his child's torts"); Childers v. A.S., 909 S.W.2d 282, 287 (Tex. Ct. App. 1995) (agreeing that "the mere fact of paternity or maternity does not make a parent liable to third parties for the torts of his or her minor child and that generally, minors are civilly liable for their own torts"); *see also* Elizabeth G. Porter, *Tort Liability in the Age of the Helicopter Parent*, 64 ALA. L. REV. 533, 555 (2013); Joan Morgridge, Comment, *When Does Parental Liability End?: Holding Parents Liable for the Acts of Their Adult Children*, 22 LOY. U. CHI. L.J. 335, 336 (1990).

18. *See* Lavin v. Jordon, 16 S.W.3d 362, 366 (Tenn. 2000) (citing W. PAGE KEETON ET AL., PROSSER & KEETON ON TORTS § 123 (5th ed. 1984)); *see also* Morgridge, *supra* note 17, at 336.

19. *See* B.C. Ricketts, *Validity & Construction of Statutes Making Parents Liable for Torts Committed by Their Minor Children*, 8 A.L.R.3d 612 § 1[a] (1966); *see also* Kaminski v. Town of Fairfield, 578 A.2d 1048, 1051 (Conn. 1990) (noting "[a]t common law, the torts of children do not impose vicarious liability upon parents qua parents, although parental liability may be created by statute, or by independently negligent behavior on the part of parents" (footnote omitted) (citation omitted)).

ute that permitted plaintiffs to sue parents for their child's actions, regardless of any wrongdoing on the parents' part.[20] The different variations of parental liability statutes include some permitting vicarious liability for property damage only, some for personal injury, others permitting both, and some covering only damage to school property.[21]

Clearly, these statutes were enacted in derogation of common law, because they premise liability on the parent-child relationship itself—presuming that if parents could themselves be sued, they will be encouraged to control their children and prevent juvenile delinquency.[22] Thus, this is the first time that the law tacitly recognized that parents have some obligation to supervise or control their youth so as to prevent them from harming others or others' property.

The problem with parental liability statutes, however, is that even though the statutes were purportedly created to compensate innocent victims by allowing them to go after the parents with deeper pockets than their children,[23] these statutes severely capped the amount of monetary recovery. In some jurisdictions, financial recovery is limited to as little as \$250–\$500.[24] Because of these low caps, scholars have recognized that parental liability statutes were not truly intended to compensate the victims, but to penalize the parents.[25] Nevertheless, the overwhelming majority of courts have found parental liability statutes constitutional because the statutes are reasonably related to the legitimate state interest of reducing juvenile delinquency.[26]

---

20. *See* Morgridge, *supra* note 17, at 337.

21. *See, e.g.*, FLA. STAT. § 741.24 (2015) (allowing a person to sue parents of child under eighteen who "maliciously or willfully destroys or steals property" of another); LA. CIV. CODE ANN. art. 225 (2016) ("Parents are responsible for damage occasioned by their child as provided by law."); N.J. STAT. ANN. § 18A:37-3 (West 2015) (making parents or guardians liable for property damage done to school property); OHIO REV. CODE ANN. § 3109.09(b) (West 2015) (creating cause of action and compensatory award of up to \$10,000 against parents or guardians for their minor child's willful damage to another's property).

22. *See* Morgridge, *supra* note 17, at 337; *see also* Distinctive Printing & Packaging Co. v. Cox, 443 N.W.2d 566, 572 (Neb. 1989) (noting that "[s]ome courts have held that where parental liability statutes limit liability, the statutory purpose must be only to deter juvenile delinquency, not to compensate victims").

23. *See* Morgridge, *supra* note 17, at 337.

24. *See* Ricketts, *supra* note 19.

25. *See id.*

26. *See* James Lockhart, *Cause of Action Under Parental Liability Statute for Tort Committed by Minor Child*, 19 CAUSES ACTION 1st 271 § 12 (1989 & Supp. 2015); Laura Pfeiffer, Note, *To Enhance or Not to Enhance: Civil Penalty Enhancement for Parents of Juvenile Hate Crime Offenders*, 41 VAL. U. L. REV. 1685, 1716 (2007) (noting parental liability statutes "resemble a form of strict liability because the parents are liable regardless of whether they acted reasonably"). At least one court has held that for parental liability statutes to be constitutional, compensation for injuries caused by minor children must be capped to avoid limitless liability for parents who may not be at fault for their child's conduct. *See* Corley v. Lewless, 182 S.E.2d 766, 770 (Ga. 1971) (holding Georgia statute permitting parents to be sued for property damage or personal injury caused by their minor child without any caps

C. *Sections 316 and 319 of the Restatement (Second) of Torts*

To address the concern that parental liability statutes failed to adequately compensate victims suffering personal injuries committed by a child, Section 316 of the Restatement (Second) of Torts was created to provide a cause of action for parental negligence.[27] Under common law negligence claims, there are no caps.[28] This Restatement has been adopted by many state courts and serves as the basis for negligent supervision and "negligent entrustment of a dangerous weapon to a child" lawsuits.[29] Section 316 provides:

> A parent is under a duty to exercise reasonable care so to control his minor child as to prevent it from intentionally harming others or from so conducting itself as to create an unreasonable risk of bodily harm to them, if the parent
>
> (a) knows or has reason to know that he has the ability to control his child, and
>
> (b) knows or should know of the necessity and opportunity for exercising such control.[30]

Subsection B's requirement that the parent "know of the necessity" to control the child apparently means that the parent must have some reason to believe their child is dangerous and poses a risk to others. Parental liability lawsuits based upon Section 316's theory of negligent supervision, often fail, however, because regardless of how disorderly or strange a child's behavior is at home, rarely would parents have the foresight that this behavior would manifest itself into the murdering of others.[31] This then poses an obstacle in trying to hold parents liable for their child's violent, criminal acts, including a mass killing.

Additionally, Section 316's application is restricted to parents of *minor* children, providing no recourse when a child who has reached the age of majority commits the violent attack.[32] This restriction makes it difficult

on amount of recovery violated Due Process Clause because, it would "authorize a recovery without liability, and would compel payment without fault" (internal quotation marks omitted)); *see also Distinctive Printing*, 443 N.W.2d at 572 (upholding constitutionality of parental liability statute that permitted full compensation for property damage intentionally committed by children residing with them but limiting compensatory damages for intentional personal injury to $1,000 per person, on grounds statute did not violate Equal Protection Clause because there was rational basis for classification: property damage happens more frequently and legislature might have wanted to incentivize parents to better control their children).

27. *See* Morgridge, *supra* note 17, at 338.

28. *See id.*

29. *See* Ralph D. Mawdsley, *Parent Liability for the Intentional Acts of Their Children Under Common Law Negligence Theories and Parental Responsibility Statutes*, 230 EDUC. L. REP. 469, 471–72 (2008).

30. RESTATEMENT (SECOND) OF TORTS § 316 (1965).

31. *See* Mawdsley, *supra* note 29, at 473.

32. *See, e.g.*, Carney v. Gambel, 751 So. 2d 653, 654 (Fla. Dist. Ct. App. 1999) (refusing to hold parents responsible for their "emancipated, adult child['s]" con-

for a plaintiff to invoke Section 316 to hold the parents of a mass shooter responsible for a shooting rampage because most mass shooters are adults.[33] To get around this obstacle, some plaintiffs began invoking Section 319 of the Restatement (Second) of Torts to establish a parent's duty to control their adult offspring. Section 319 provides:

> [o]ne who takes charge of a third person whom he knows or should know to be likely to cause bodily harm to others if not controlled is under a duty to exercise reasonable care to control the third person to prevent him from doing such harm.[34]

Plaintiffs have alleged that parents who permit their adult children to reside with them "take charge" over the child and thus have a duty to control them.[35]

Courts, however, have repeatedly rejected this as a legitimate basis for imposing a duty on a parent to control their adult offspring, absent a court order or other legal process expressly making the parents the adult-child's legal custodian.[36] For example, in *Kaminski v. Town of Fairfield*,[37] the Connecticut Supreme Court rejected the defendant police officer's counterclaim, premised upon the Restatement (Second) of Torts Section 319, that the parents of an adult male, who seriously wounded him with an axe, had a duty to control their adult, paranoid schizophrenic son.[38] In rejecting the defendant's counterclaim, the court observed that reported cases had only imposed a duty to control pursuant to Section 319 where professional custodians or institutions had formal, legal custody of the tortfeasor, such as in the case of an individual placed in the care of a hospital or under the supervision of a prison warden.[39] The court further pointed out that, at that time, no court had ever held that a parent who

---

duct, noting "where a special relationship has been found imposing liability on a parent for conduct of a child, the duty to exercise control is limited to a *minor* child").

33. *See* Mark B. Melter, *The Kids Are Alright; It's The Grown-Ups Who Scare Me: A Comparative Look at Mass Shootings in the United States and Australia*, 16 GONZ. J. INT'L L. 33, 39 (2012) (observing that "roughly 80% of all mass murderers are aged between 20 and 50 years old. . . . men make up 94% of all mass murderers, and 63% of offenders are white").

34. RESTATEMENT (SECOND) OF TORTS § 319.

35. *See, e.g.*, Kaminski v. Town of Fairfield, 578 A.2d 1048, 1051 (Conn. 1990).

36. *See id.*; *see also* Lott v. Goodkind, 867 So. 2d 407, 408 (Fla. Dist. Ct. App. 2003) (disagreeing that Section 319 can be utilized as basis to impose duty on parents to control adult children, reasoning "[l]iability of the type described in section 319 has typically been imposed on persons having someone committed to their legal custody, such as a jailer or superintendent of a residential institution which has the ability to control the actions of its residents"); Blevins v. Hartman, Nos. 12CA116, 12CA115, 2013 WL 3936039, at *8 (Ohio Ct. App. July 18, 2013).

37. 578 A.2d 1048 (Conn. 1990).

38. *See id.* at 1050–51.

39. *See id.* at 1051–52.

had made a home for an adult, mentally ill child had taken charge of that person within the meaning Section 319.[40]

Similarly, in *Knight v. Merhige*,[41] a Florida appellate court recently reaffirmed that Section 319 does not provide a basis for imposing a duty on parents to control an adult child, stressing that "the 'take charge' requirement of section 319 has generally been limited to 'the context of professional custodians with special competence to control the behavior of those in their charge,'" like jailers or superintendents of residential institutions.[42] In *Knight*, the adult son shot and killed four relatives at a Thanksgiving family gathering.[43] The court held that providing financial assistance to their son (who did not reside with them) and controlling whether he attended family gatherings was insufficient to place the adult son "within the functional equivalent of their 'legal custody.'"[44]

D. *A Special Relationship*

Thus, in the absence of any statute or Restatement provision explicitly permitting parental liability for an adult child, courts have been divided as to whether a parent can be held accountable for their emancipated offspring's miscreant acts. Part of this division is based upon the well-settled common law principle that no one has a duty to protect others from dangers posed by third parties—or a duty to control or warn a third party to prevent harm to others—unless they have a special relationship with either the victim or the perpetrator.[45] Since, as used in this Article, mass or rampage shootings refers to the *random* shootings of people, the parents of the shooter likely will not have a special relationship with the victims. Thus, the only way a parent can be held accountable for a mass shooting committed by their child is if the courts recognize a special relationship between the adult shooter and their parent—something most courts have been reluctant to conclude, because parents have no legal duty to control their emancipated offspring.[46]

For example, in *Carney v. Gambel*,[47] a Florida appellate court held that the parents of an adult child who assaulted the head of security in the country club community where the parents and adult child lived did not have a special relationship with their adult child and thus did not have a

40. *See id.* at 1052.

41. 133 So. 3d 1140 (Fla. Dist. Ct. App. 2014).

42. *See id.* at 1146–47 (quoting *Kaminski*, 578 A.2d at 1051).

43. *See id.* at 1143.

44. *See id.* at 1142, 1147.

45. *See id.* at 1145; *see also* RESTATEMENT (SECOND) OF TORTS § 315 (1965).

46. *See e.g.*, Hartsock v. Hartsock, 592 N.Y.S.2d 512, 513 (App. Div. 1993) ("No justification exists [ ] for imposing liability on parents for the conduct of their emancipated, adult child. Inasmuch as parents have no legal right to control their adult child's activities, they cannot be held liable for those activities." (citations omitted)).

47. 751 So. 2d 653 (Fla. Dist. Ct. App. 1999).

duty to control his conduct.[48] In so holding, the court reasoned that in order to have a special relationship with their adult offspring, the parents had to have the right or ability to control his conduct, and parents have no legal right to control an emancipated, adult child.[49] The court further emphasized that parental liability had been limited to cases involving a *minor* child.[50] However, the court left open the possibility for parental liability to incur for an adult child where the adult child is insane or mentally deficient, citing to *Thorne v. Ramirez*.[51]

Similarly, in *Grover v. Stechel*,[52] a New Mexico appellate court concluded that no special relationship existed between a mother and her adult son who had stabbed the plaintiff. The court reasoned that the mother lived across the country from her adult son when the assault happened, and it was too much of a logical leap to find that, by merely assisting her adult son financially, she had the ability to control his conduct.[53] The court further explained that without the right or ability to control another's conduct, no duty to control could arise.[54]

In both of these cases, although the courts did not specifically reference Section 316 of the Restatement (Second) of Torts, they apparently incorporated Section 316's ability-to-control consideration in their analyses to determine whether there was a special relationship between the parents and their adult offspring. In these cases, the courts found the parents lacked the ability to control their emancipated offspring because either (1) the child had obtained the age of majority *and* had no mental illness, as in the *Carney* case; or (2) the parents did not have the opportunity to control the adult because the adult child did not reside with them, as in the *Grover* case. As demonstrated by the next few case illustrations, however, where these conditions do not exist, parental liability for adult children could attach.

For instance, contrary to the *Carney* and *Grover* cases, in *Silberstein v. Cordie*,[55] a Minnesota court held that the defendant-parents had a special relationship with their schizophrenic, twenty-seven-year-old son who resided with them.[56] In that case, the parents knew their son was delusional and nonverbal, had hallucinations about the devil and beasts, and had threatened to harm the victim in the past while suffering from delusions.[57]

48. *See id.* at 654.
49. *See id.*
50. *See id.*
51. *See id.*; *see also* Thorne v. Ramirez, 346 So. 2d 121, 122 (Fla. Dist. Ct. App. 1977).
52. 45 P.3d 80 (N.M. Ct. App. 2002).
53. *See id.* at 84.
54. *See id.*
55. 474 N.W.2d 850 (Minn. Ct. App.), *rev. granted in part, cause remanded by* 477 N.W.2d 713 (Minn. 1991).
56. *See id.* at 856–57.
57. *See id.* at 852.

In fact, the mother had her son involuntarily committed to a mental healthcare facility prior to the incident in question.[58] Four months after the son was released from involuntary commitment, the son moved in with his parents, and the parents noticed he had stopped taking his medication.[59] Nevertheless, the parents traveled out of town, leaving their 12-gauge shotgun readily accessible to him.[60] The son used the gun to kill his friend whom he believed was in collusion with the devil.[61] In recognizing the parents had a special relationship with their adult son, the court stressed that months before the murder, the mother assumed responsibility for her son's day-to-day care, and the son's mental condition "was like that of a child . . . ."[62]

Similarly, in *Frederic v. Willoughby*,[63] an Ohio court determined that a special relationship existed between parents and their mentally disturbed adult offspring who resided in the same household.[64] There, the adult son sexually assaulted the plaintiff while her four-year-old daughter lied next to her in the bed.[65] Prior to this incident, the father had struck his son in the head after the son had broken into the plaintiff's house and masturbated with a pair of her underwear.[66] The court explained that a "special relation" exists between a parent and an adult child when the parent either "has accepted such responsibility in a legally recognized way, such as a guardianship, or *in a situation where an adult child's dependence and the parent's overt acceptance of responsibility for the adult child establish a de facto guardianship*."[67]

In determining whether there was a de facto guardianship, the court considered whether the perpetrator posed a risk to others; whether the parents were aware of the perpetrator's mental condition and the risk he posed to others; and whether the parents accepted responsibility for their son-perpetrator's actions and took charge of him.[68] In finding that the parents were aware of their son's mental condition and posed a risk to others, the court highlighted that the parents knew their son was "not right," had violent criminal convictions including for a sexual battery on a blind woman and domestic violence, and was drawing Social Security benefits for his mental illness since his illness prevented him from working.[69] Additionally, the parents were aware their son posed a risk to others be-

58. *See id.*
59. *See id.*
60. *See id.* at 856.
61. *See id.* at 852–53.
62. *See id.*
63. No. 2007-P-0084, 2008 WL 2582593, at *1 (Ohio Ct. App. June 27, 2008).
64. *See id.*
65. *See id.*
66. *See id.* at *1–2.
67. *See id.* at *5 (emphasis added) (internal quotation marks omitted).
68. *See id.*
69. *See id.* at *5–6.

cause they adopted his son (their grandson) and raised him in their residence upstairs to protect the grandson from their son, who lived downstairs in the basement.[70]

The court held there was sufficient evidence for a jury to determine whether the parents assumed responsibility for the perpetrator and took charge of him.[71] That included evidence that the parents bailed the child out of jail; attended his court hearings; paid for his legal expenses and vasectomy; confiscated the perpetrator's keys to prevent him from leaving the house late at night; drove him to see a counselor; monitored him closely by restricting him from leaving their property; beat the perpetrator for masturbating in the neighbor's house; and, after the first masturbation incident, told the plaintiff-victim to contact them if she had any additional problems with their son.[72] The court noted that all of this evidence not only was sufficient for a jury to conclude that the defendant-parents owed a duty to the plaintiff neighbor, but that they breached the duty, and the breach was the proximate cause of the neighbor's rape.[73]

Finally, in *Smith v. Freund*,[74] a California appellate court found that a special relationship existed between a nineteen-year-old mentally disturbed man and his parents, but ultimately decided the parents owed no special relationship-based duty to the victims because their son's criminal act was unforeseeable.[75] In *Smith*, the plaintiffs alleged that the parents negligently supervised their nineteen-year-old son, who shot and killed their immediate family members.[76] At the time of the shootings, the adult son lived with his parents, and they were aware he had been diagnosed with a major depressive disorder with psychotic features, Asperger's Syndrome, and attention deficit disorder.[77] In the past, the son had flown into violent rages and attacked his parents.[78] The plaintiffs argued the parents had a special relationship with their son, notwithstanding that he was an adult, because he "was autistic and lived with his parents in the same way as a minor child," and thus they had the ability to control and monitor him.[79] The appellate court agreed, reiterating that a defendant has a duty to take affirmative action for the protection of another if a special relationship exists between the defendant and the person whose conduct needs to be controlled.[80] Most importantly, the court held that a "parent and child is one such special relationship," and "[a]nother special

70. *See id.* at *1.
71. *See id.* at *6.
72. *See id.* at *6–8.
73. *See id.* at *8.
74. 121 Cal. Rptr. 3d 427 (Ct. App. 2011).
75. *See id.* at 429.
76. *See id.*
77. *See id.* at 429–30.
78. *See id.* at 430.
79. *See id.* at 432 n.3 (internal quotation marks omitted).
80. *See id.* at 432.

relationship ensues when a party takes charge of a third person whom he or she knows or should know to be likely to cause bodily harm to others if not controlled,"[81] apparently relying on Sections 316 and 319 of the Restatement (Second) of Torts.

A synthesis of the above-referenced cases demonstrates that courts are willing to find a special relationship between a parent and their adult child based upon a de facto guardianship where three factors coalesce: (1) the parents knew, or should have known, their adult child had *significant* mental health issues and posed a risk to others; (2) the child resided with his parents; and (3) the parents acted as if they believed they had the ability to control the child.[82] This proposed test is consistent with Section 316 of the Restatement (Second) of Torts. Knowing that the child has significant mental health issues addresses Section 316's parental liability criterion requiring the parent to know of the need to control the adult child. Residing with the adult child is at least prima facie evidence that the parent had the ability and opportunity to control the child.

II. ESTABLISHING A SPECIAL RELATIONSHIP BETWEEN PARENTS AND ADULT CHILDREN IN MASS SHOOTING CASES

Having clarified the standard for determining whether a special relationship exists between parents and an adult child, it is necessary to see if it is a workable test that does not lead to absurd results. As demonstrated by the real-life illustrations that follow, the test is workable.

A. *Application of Special Relationship Test to Nancy Lanza*

Utilizing the test in Nancy Lanza's case proves that the special relationship test above is useful; it yields a finding that a special relationship existed between Nancy and her son—a result that is not only reasonable, but instinctively feels correct.

The first requirement of the test, that the parent knows or has reason to know their child has significant mental health issues and poses a security risk, is satisfied. Nancy knew Adam had a long history of mental health problems. At age six, Adam was diagnosed with sensory integration disorder, a condition that makes one overly sensitive to stimuli in the environment and that made him dislike being touched and around many people.[83] By age thirteen, a psychiatrist diagnosed Adam with Asperger's

81. *See id.* (internal quotation marks omitted).

82. *See* Silberstein v. Cordie, 474 N.W.2d 850, 855–56 (Minn. Ct. App.), *rev. granted in part, cause remanded by* 477 N.W.2d 713 (Minn. 1991); *see also* Todd v. Dow, 23 Cal. Rptr. 2d 490, 493–94 (Ct. App. 1993) (holding there was no special relationship between parents and their adult son absent any facts showing they had ability to control him); *cf.* Thorne v. Ramirez, 346 So. 2d 121, 122 (Fla. Dist. Ct. App. 1977) (suggesting that, had plaintiff alleged defendant-parents' child was "dependent, insane or mentally deficient," plaintiff's complaint against parents for child's intentional tort may have alleged sustainable cause of action).

83. *See* Lewis, *supra* note 2, at 334–35.

Syndrome, a high-functioning form of autism resulting in unusual social awkwardness.[84] While neither one of these conditions have been linked to violence,[85] Adam displayed other disturbing behavior over the years that would be disconcerting to most parents. For example, in the fifth grade, Adam wrote a document entitled the "Big Book of Granny," wherein the main character had a gun in her cane and shot people, including children.[86] In the seventh grade, Adam's writing assignments contained descriptions of battles and war far more than others his age, and "[t]he level of violence in the writing was disturbing."[87] Additionally, a month before the shootings in November of 2012, Adam's mother confided in a friend that she was worried about Adam because he had not left the house in three months, even refusing to go to a hotel with her when Hurricane Sandy caused a power outage in their home.[88]

Those who knew Adam described him as a "shut-in who rarely left home and played military-style video games."[89] Additionally, Adam would only communicate with his mother through email, even though they resided in the same household.[90] Adam also had been estranged from his father and only brother—one of many indicators he had some mental health issues.[91] Medical doctors found that Adam "lacked empathy and had very rigid thought processes."[92] For instance, shortly before Adam killed his mother and the children at Sandy Hook, he told his mother that he would not care if she had passed away.[93] The most compelling evidence that Nancy Lanza knew, or had reason to know, that her son was mentally ill and dangerous was the fact that she had completed paperwork to have Adam involuntarily committed to a mental health treatment facil-

84. *See id.*; *see also* Flegenheimer & Somaiya, *supra* note 13.

85. *See* Adam Clark Estes, *Revelations About Adam Lanza's Mental Health Still Don't Explain the Violence*, WIRE (Feb. 19, 2013, 11:26 PM), http://www.thewire.com/national/2013/02/revelations-about-adam-lanzas-mental-health-still-dont-explain-violence/62317/ [https://perma.cc/G65D-6ZC2].

86. *See* SEDENSKY, *supra* note 10, at 33 (internal quotation marks omitted).

87. *See id.* at 34.

88. *See id.* at 28; John Christoffersen, *New Report Provides the Most Disturbing Picture Yet of Nancy Lanza's Relationship with Her Son*, BUS. INSIDER (Dec. 28, 2013, 11:34 AM), http://www.businessinsider.com/paramedic-at-sandy-hook-was-warned-this-will-be-the-worst-day-of-your-life-2013-12 [https://perma.cc/H3KR-EC4W].

89. *See* Susan Candiotti, Greg Botelho & Tom Watkins, *Newtown Shooting Details Revealed in Newly Released Documents*, CNN (Mar. 29, 2013, 9:53 AM), http://www.cnn.com/2013/03/28/us/Connecticut-shooting-documents/ [perma.cc.G26Y-J9XB].

90. *See* SEDENSKY, *supra* note 10, at 28; Christoffersen, *supra* note 88.

91. *See* SEDENSKY, *supra* note 10, at 3, 29–30 (noting Adam's brother, who was four years older, had not had contact with Adam since 2010, despite attempts to reach him, and Adam's father had not seen him since 2010, despite the father's invitations for Adam to join him for various activities).

92. *See id.* at 34.

93. *See id.* at 30.

ity.[94] Since an involuntary civil commitment in Connecticut requires a finding that a person is mentally ill *and* poses a danger to themselves or others,[95] one may conclude that Nancy had to believe Adam would harm himself or someone else. Indeed, a week before the Sandy Hook tragedy, Nancy told an acquaintance, "I'm worried I'm losing him," referencing his illness.[96]

The second and third requirements for finding a special relationship between a parent and their emancipated offspring—that the adult child resides with his parent and the parent believes he or she has the ability to control the child—are also satisfied in this case. Adam lived at home with his mother at the time of his shooting spree.[97] In fact, Adam's mother indicated that she did not work because of Adam's mental condition and apparently devoted her life to caring for him.[98] Like the mother in *Silberstein*, Nancy assumed responsibility for his day-to-day care by washing his clothes daily, cooking for him, shopping for him, and taking care of him as if he were a minor child, even though he was an adult.[99] Nancy's care of her son is distinguishable from a situation where a parent cares for an adult child out of the goodness of their heart. Nancy provided for all of Adam's needs because he was incapable of caring for himself due to his mental issues. For example, she purchased a car for Adam that was registered in her name only.[100] One can infer from the fact that she owned the car that she could have taken it away from Adam at any time by, for example, taking his keys away or disabling the vehicle.

Additionally, because Adam did not work, he was wholly dependent on Nancy for food and his daily needs just like a minor child.[101] Adam's mother had the ability to and, in fact, did control Adam's environment. Persons hired to work on the Lanzas' home reported that Adam's mother never let them in the house, instructed them not to ring the doorbell, and

---

94. *See* Winter, *supra* note 14.

95. CONN. GEN. STAT. ANN. § 17a-502 (West 2015) (permitting psychiatric commitment under an "emergency certificate" if physician determines person has "psychiatric disabilities" and poses danger to himself or others).

96. *See* Matthew Lysiak, Kerry Wills & Stephen Rex Brown, *Exclusive: Nancy Lanza Feared Son, Adam, Was 'Getting Worse'; Told Friend 'He Was Burning Himself with a Lighter' and That She Was 'Losing Him,'* N.Y. DAILY NEWS (Dec. 17, 2012, 9:17 AM), http://www.nydailynews.com/news/national/nancy-lanza-feared-son-adam-worse-article-1.1221505 [https://perma.cc/5D5E-5GMC] (internal quotation marks omitted).

97. *See* SEDENSKY, *supra* note 10, at 28 (noting that only Nancy and her son had resided in their family home for extended time prior to shootings on December 14, 2012).

98. *See id.* at 30.

99. *See id.* at 30, 35.

100. *See id.* at 23.

101. *See* Alaine Griffin & Josh Kovner, *New Report on Lanza: Parental Denial, Breakdowns, Missed Opportunities*, HARTFORD COURANT (Nov. 21, 2014), http://www.courant.com/news/connecticut/hc-newtown-adam-lanza-child-advocate-report-20141121-story.html [https://perma.cc/WZ9B-PJTS].

advised them to make prior arrangements for using power equipment because Adam had problems with loud noises.[102] In fact, a number of people who knew the mother for years had neither been inside her home nor met Adam.[103]

In sum, the overwhelming evidence demonstrates that this was not a case of a parent simply providing for their child financially; rather, Adam's mother had assumed responsibility and charge over Adam. Accordingly, a court could easily conclude a relationship between a parent(s) or child like the Lanzas' would meet the special relationship test formulated in Part I(D) for parental liability for adult children.

B. *Application of Special Relationship Test to Jared Loughner's Parents*

Likewise, the aforementioned special relationship test could lead a court to reasonably conclude that Jared Loughner and his parents, Randy and Amy Loughner, had a "special relationship" for purposes of establishing the first element of negligence—a duty.[104] Jared is the twenty-two-year-old who shot United States Congresswoman Gabrielle Giffords in the head and then fired at seventeen others who were waiting to meet Giffords outside a grocery store in Tucson, Arizona on January 8, 2011.[105]

Prior to his shooting spree, Jared's parents knew he had significant mental health issues and posed a risk of violence to the public. In the months leading up to the shootings, Jared was clearly unraveling. While attending Pima Community College, he would make inappropriate outbursts during class, including remarks about "blowing up babies" and asking the professor if he believed in mind control.[106] He would repeatedly utter nonsensical statements such as, "How can you deny math instead of accepting it."[107] Jared also was contacted five times by the police for disruptive behavior on campus.[108] Jared posted a disturbing video on YouTube showing him walking around the campus at night with a video camera threatening to torture students and calling his college a "genocide

102. *See* SEDENSKY, *supra* note 10, at 31.

103. *See id.*

104. *See* Claire Martin & Masada Siegel, *Loughner's Parents: Did They Know About His Mental Illness?*, DAILY BEAST (Jan. 11, 2011, 9:31 AM), http://www.thedailybeast.com/articles/2011/01/11/jared-loughners-parents-randy-and-amy-did-they-know-about-his-mental-illness.html [https://perma.cc/M4J7-44UK] (discussing Jared's parents).

105. *See* Marc Lacey & David M. Herszenhorn, *In Attack's Wake, Political Repercussions*, N.Y. TIMES, Jan. 8, 2011, http://www.nytimes.com/2011/01/09/us/politics/09giffords.html?pagewanted=all&_r=0 [https://perma.cc/AC2W-RYAY] (reporting at least seventeen persons were shot and six murdered by twenty-two-year-old Jared Loughner). Miraculously, Congresswoman Giffords survived the shooting.

106. *See* Lewis, *supra* note 2, at 332 (internal quotation marks omitted).

107. *See id.* (internal quotation marks omitted).

108. *See id.*

school."[109] Pima Community College then notified Jared's parents of what he had done and that he was suspended from school, warning that Jared had mental health issues and may be dangerous and that Jared could not return to school until he had a mental health clearance.[110]

Randy and Amy Loughner also personally observed their son talking to himself and displaying other strange behavior like refusing to communicate with them and making worrisome noises.[111] Randy found some of Jared's journals that were written in script that was "indecipherable."[112] Jared's parents indicated by their actions that they believed Jared was dangerous; for example, they confiscated the only gun they believed he had and disabled his car every night to keep him home.[113] Additionally, both of Jared's parents pleaded with him to get help, but Jared refused, and his parents never had Jared psychologically evaluated.[114] In light of these facts, it is clear that the Loughners knew their son had significant mental health issues and was potentially dangerous; thus, the first criterion of the special relationship test is met. The second requirement that the adult child live with his parents is clearly satisfied because Jared resided with his parents.[115]

The third criterion—that the parents believed they had the ability to control Jared—is a closer call. On one hand, Amy and Randy acted as if they believed they had the ability to control their son by routinely disabling his car and taking away his shotgun after he started acting erratically.[116] The night before Jared's shooting spree, however, his father

109. *See id.*; Michael Martinez & Chelsea J. Carter, *New Details: Loughner's Parents Took Gun, Disabled Car to Keep Him Home*, CNN (Mar. 28, 2013, 11:45 AM), http://www.cnn.com/2013/03/27/justice/arizona-loughner-details/ [https://perma.cc/3JM9-HCJN]; Tim Steller, *What Loughner's Parents Knew: They Watched His Decline but Didn't Arrange for Treatment*, ARIZ. DAILY STAR, Mar. 28, 2013, 12:00 AM, http://tucson.com/news/local/crime/what-loughner-s-parents-knew-they-watched-his-decline-but/article_fb334f4e-4d20-5d80-baac-b467e8f56cb9.html [https://perma.cc/3GDV-CMQS] (internal quotation marks omitted).

110. *See* Franke-Ruta, *supra* note 8; Courtney Hutchison & ABC News Med. Unit, *Violence and Mental Illness: Is Loughner a Case for Involuntary Commitment?*, ABC NEWS (Jan. 12, 2011), http://abcnews.go.com/Health/MindMoodNews/jared-loughner-involuntarily-committed/story?id=12593355 [https://perma.cc/5FW9-NHK6]; Martinez & Carter, *supra* note 109; Dennis Wagner, *Records Detail Shooter's Agitation Before Ariz. Rampage*, USA TODAY (Mar. 27, 2013, 10:47 PM), http://www.usatoday.com/story/news/nation/2013/03/27/gabby-giffords-shooting-records/2024589/ [https://perma.cc/XBK7-WWPR].

111. *See* Wagner, *supra* note 110.

112. *See Jared Lee Loughner Grew Delusional in Months Before Tucson Rampage, Police Reports Show*, CBSNEWS (Mar. 27, 2013, 6:29 PM) [hereinafter Loughner CBS Article], http://www.cbsnews.com/news/jared-lee-loughner-grew-delusional-in-months-before-tucson-rampage-police-reports-show/ [https://perma.cc/L9GF-AA65].

113. *See* Martinez & Carter, *supra* note 109.

114. *See id.*; Hutchison & ABC News Med. Unit, *supra* note 110; Wagner, *supra* note 110.

115. *See* Martinez & Carter, *supra* note 109.

116. *See id.*

forgot to disable the car.[117] Additionally, on the day of the shooting, Jared tried to leave the house with a black bag, and when his parents asked what was in the bag, Jared bolted out the door.[118] Jared's father chased him down the street on foot but, unfortunately, could not catch him.[119] The father then jumped in his car to follow Jared and Jared ran off into the desert.[120] Apparently, the bag contained a gun Jared had legally purchased on November 30, 2010, unbeknown to his parents.[121] Jared then caught a cab to the grocery store where he shot Congresswoman Giffords and seventeen others.[122]

While some could argue that Jared's parents' inability to stop him from leaving the house on the day of the murders shows that they did not have the ability to control him, the test requires only that the parents *believe* they have the ability to control their adult offspring; it does not require the ability to control in actuality. On the other hand, Jared's parents stated that during this time period they could not have a rational conversation with Jared, who had stopped speaking with them. They also reported to investigators "they had lost control of their son long before the shootings."[123] At times, Jared would leave the Loughners' home and check-in at a hotel, and they would go looking for him and bring him home. Thus, a court could come out either way on whether a special relationship existed in this case.

C. *Applying Special Relationship Test to Elliot Rodgers's Parents*

In contrast to the Lanza and Loughner cases, the special relationship test articulated in Part I(D) of this Article would lead a court to conclude that no special relationship existed between Elliot Rodgers and his parents. Elliot is the twenty-two-year-old community college student who, in May of 2014, stabbed his three roommates to death and then randomly shot and killed three other college students from his BMW in Santa Barbara, California.[124] A month before Elliot's heinous acts, his mother saw

117. David Trifunov, *Jared Loughner Evidence Reveals Parents Hid Guns Before Tucson Shooting (VIDEO)*, GLOBAL POST (Mar. 27, 2013, 6:01 PM), http://www.globalpost.com/dispatch/news/regions/americas/united-states/130327/jared-loughner-evidence-gabby-giffords-tucson-shooting [https://perma.cc/5WKS-9L3P].

118. *See* Martinez & Carter, *supra* note 109.

119. *See* Wagner, *supra* note 110.

120. *See id.*

121. *See* Steller, *supra* note 109.

122. *See id.*; *see also* Martinez & Carter, *supra* note 109.

123. *See* Wagner, *supra* note 110.

124. *See Elliot Rodger's Family Tried to Intervene Before Deadly Rampage*, FOX NEWS (May 27, 2014) [hereinafter Fox News Elliot Rodger's Article], http://www.foxnews.com/us/2014/05/27/elliot-rodgers-family-tried-to-intervene-before-deadly-rampage/ [https://perma.cc/9P7M-YRCV]; Christopher Weber & Alicia Chang, *Elliot Rodger's Family Tried to Intervene Before Rampage*, HUFFINGTON POST (May 26, 2014, 8:57 AM), http://www.huffingtonpost.com/2014/05/26/elliot-rodgers-family_n_5392211.html [https://perma.cc/AVS8-8QMZ] (updated July 26, 2014).

some disturbing videos he had posted on YouTube and alerted one of her son's mental health counselors.[125] The mental health counselor notified the police, who went to Elliot's doorstep to check on him, but ultimately took no action because Elliot presented himself as a well-mannered young man who posed no risk for violence.[126] As Elliot divulged in his 140-page manifesto, which was found after the killings, if the police searched his room, they would have uncovered a cache of firearms and his murderous plot.[127] In fact, hours before Elliot committed his shooting spree, he emailed his manifesto about committing a deadly rampage "to a couple of dozen people, including his parents" and one of his therapists.[128] Upon receiving this manifesto, Elliot's mother promptly called the authorities and raced to Santa Barbara with her husband to stop Elliot but, unfortunately, arrived too late.[129]

While under these facts, Elliot's parents knew he had significant mental health issues and had reason to believe he may harm someone due to his Internet postings, the second and third criterion of the special relationship test are not met. Elliot did not live with his parents, and there was no evidence that they acted in a manner that would cause one to believe they had the ability to control him. In fact, the evidence points to the contrary: when they learned of his inflammatory internet postings, they did not try and handle the matter alone but contacted the authorities. Moreover, they did not assume responsibility for Elliot's day-to-day care, as he was taking care of himself and living independently in his own apartment hours away.[130] Therefore, no special relationship was established, and thus no duty to warn or protect others arose in this case nor could arise in others like it.

In sum, the special relationship test formulated in the Part I(D) is workable. Its application will only yield findings of a special relationship between a parent and their adult child when the parent has assumed de facto guardianship of their adult offspring by taking full responsibility for that child, allowing the child to reside with them, and trying to control their behavior. Establishing a special relationship, however, does not end the duty inquiry. It is only a prerequisite for finding a special relationship-based duty. Should a court recognize a special relationship in the context of a mass shooting case, the next inquiry is whether the majority of policy

---

125. *See* Fox News Elliot Rodger's Article, *supra* note 124.

126. *See id.*

127. *See id.*

128. *See* Daily Mail Reporters, *Virgin Killer's Parents Read His Hate-Filled Manifesto Then Called the Police and Rushed to Stop Him When They Heard of Murder Spree on Their Car Radio,* DAILYMAIL.COM (May 25, 2014, 7:44 PM), http://www.dailymail.co.uk/news/article-2639177/Parents-shooter-read-manifesto-driving-stop-son-heard-massacre-radio-revealed-investigators-search-moms-house.html [https://perma.cc/3PRK-BXFH] (updated May 2, 2014, 10:47 PM).

129. *See id.*

130. *See id.*

considerations weigh in favor of recognizing a duty, which is addressed in Part III.

III. OVERVIEW OF PARENTAL NEGLIGENCE AND ESTABLISHING A DUTY TO PROTECT OR WARN

A. *The Negligence Cause of Action and Its Pitfalls*

To establish parental negligence, the plaintiff must prove that: (1) the parents owed a duty to the plaintiff; (2) the parents breached that duty; and (3) the parents' breach caused the plaintiff's injury.[131] While this formula seems rather straightforward to apply, it is fraught with difficulties, particularly the first and third elements.[132]

The first element of parental negligence—duty—can be problematic; it is essentially a policy and moral question concerning how far parental liability should reasonably extend, which varies from court to court and is constantly evolving over time, hampering the ability to predict, with any certainty, how a court will resolve a future similar case.[133] This fluctuation, however, is not necessarily a negative when it comes to negligence in the context of mass murder, which is a relatively new phenomenon. Additionally, the duty analysis in the third-party criminal act context is complicated because it is partially dependent upon whether the third party's criminal act was foreseeable, and in the majority of mass shooting cases, the perpetrators had no prior history of shooting violence, making their sudden criminal attack arguably capricious.[134] Further adding to the con-

131. *See* Lisa Lockwood, Comment, *Where Are the Parents? Parental Criminal Responsibility for the Acts of Children*, 30 GOLDEN GATE U. L. REV. 497, 512 (2000).

132. *See* Andrew J. McClurg, *Armed and Dangerous: Tort Liability for the Negligent Storage of Firearms*, 32 CONN. L. REV. 1189, 1226 (2000) ("Virtually all cases involving negligence liability for criminal attacks, including those involving negligent firearm storage, purport to turn on the elements of 'duty' and 'proximate cause' along with their damnable, murky companion, 'foreseeability.' Duty and proximate cause have long been sources of confusion for law students and lawyers alike, who search in vein for concrete rules on which to base such determinations." (footnote omitted)).

133. *See* Lopez v. McDonald's Corp., 238 Cal. Rptr. 436, 440, 441 (Ct. App. 1987) (explaining judicial treatment of duty—which "is not sacrosanct or an immutable fact of nature, but only a shorthand expression of the sum total of public policy considerations which lead the law to protect a particular plaintiff from harm"—has "left a legacy of analytical confusion"); Craig v. A.A.R. Realty Corp., 576 A.2d 688, 692 (Del. Super. Ct. 1989) (explaining that duty is "frequently an expression by the court of evolving public policy"); *see also* Porter, *supra* note 17, at 567 (observing that "[o]ver the past few decades, duty has evolved into a more individualized, less predictable inquiry").

134. *See* Rhonda V. Magee Andrews, *The Justice of Parental Accountability: Hypothetical Disinterested Citizens and Real Victims' Voices in the Debate over Expanded Parental Liability*, 75 TEMP. L. REV. 375, 389 (2002); *see also* W. Jonathan Cardi, *Purging Foreseeability*, 58 VAND. L. REV. 739, 740–41 (2005) (noting foreseeability determination in negligence is problematic for two reasons: (1) it leads to inconsistency in judicial outcomes and (2) it usurps the role of the jury in deciding whether the defendant's conduct was reasonable).

volution of the duty analysis in the case of a mass shooting is the fact that the law has been reluctant to hold parents responsible for the violent acts of their *minor* children.[135] If the law is hesitant to hold parents liable for the violent criminal acts of their minor children, will courts be even more reticent to hold parents responsible for the violent acts of their legally emancipated offspring, who are often the offenders in mass shootings?

The third element of negligence—causation—has also been a source of puzzlement. Tort experts have long opined "[t]here is perhaps nothing in the entire field of law which has called forth more disagreement, or upon which the opinions are in such a welter of confusion."[136] Part of the confusion stems from the fact that a foreseeability determination must also be made at this stage of the negligence analysis, and some courts use duty-foreseeability cases to make causation-foreseeability determinations.[137] The foreseeability consideration for causation is different from the one used to determine whether the law should recognize a duty, however. As one scholar has noted, "[w]hile duty employs a forward-looking foreseeability analysis, proximate cause employs a backward-looking foreseeability analysis."[138] In other words, when considering foreseeability in the duty analysis, the question is whether it was generally foreseeable that the harm—or, in the third-party criminal act context, the third party's criminal act—would occur. This requires *foresight.* When considering foreseeability for the purposes of determining proximate cause, however, the relevant inquiry is whether, looking back in *hindsight* at what actually occurred, was it foreseeable that the precise harm that occurred would happen, i.e., was the injury or mass shooting a natural and probable consequence of the defendant's breach of duty?[139] Thus, foreseeability in the causation-element analysis is more specific, whereas the foresight determination in the duty analysis is more general.[140] The concept of foreseeability also appears in the analysis for the second element, breach of

135. *See* Porter, *supra* note 17, at 535 (observing "common law courts have resisted exposing parents to civil liability" and "it remains the rare case that survives summary judgment or a motion to strike").

136. *See* WILLIAM L. PROSSER, HANDBOOK OF THE LAW OF TORTS § 41, at 236 (4th ed. 1971).

137. *See* Deborah J. La Fetra, *A Moving Target: Property Owners' Duty to Prevent Criminal Acts on the Premises*, 28 WHITTIER L. REV. 409, 410 (2006) (noting courts "frequently use duty-foreseeability cases as precedent to support causation-foreseeability holdings and vice versa").

138. Rory Bahadur, *Almost a Century and Three Restatements After Green It's Time to Admit and Remedy the Nonsense of Negligence*, 38 N. KY. L. REV. 61, 86 (2011).

139. *See id.* at 87; *see also* McKown v. Simon Prop. Grp., Inc., 344 P.3d 661, 665 (Wash. 2015) (en banc) (explaining foreseeability of harm determination in causation element analysis limits scope of defendant's duty and asks "whether the harm sustained is reasonably perceived as being within the general field of danger covered by the duty owed by the defendant").

140. *See* Bahadur, *supra* note 138, at 87.

duty, and this overlap also adds to the perplexity of the negligence analysis.[141]

Notwithstanding the complexities of the negligence analysis, the rest of this Article will demonstrate that by employing general negligence principles, there are some instances where a plaintiff should be able to prove that a parent's conduct or failure to act is clearly unreasonable and negligent under common law negligence principles in mass shooting cases. There will be other cases where the parents' conduct or failure to act was certainly reasonable and no liability should attach. Finally, some parental conduct or omissions will fall squarely in the middle of the spectrum and should be left for a jury to decide.

B. *Proving a Duty Should Be Recognized in Mass Murder Cases*

It is axiomatic that

> [r]egardless of how morally, ethically or socially deplorable a defendant's conduct may be viewed by other constituencies, in the eyes of the law, the defendant may not be held to answer in negligence unless and until the court determines, as a matter of law, that the defendant owed a duty of care to the plaintiff.[142]

A "duty" is an "obligation, recognized by the law, requiring the person to conform to a certain standard of conduct, for the protection of others against unreasonable risks."[143] Thus, the first issue that must be addressed is whether the parent of an emancipated, mentally ill child has a duty to prevent that adult child from committing a mass shooting. More specifically, does the parent have a duty to supervise or control that child's behavior so that they do not have access to firearms, and if they cannot control the child, do they have a duty to warn the proper authorities (the police, the court, or mental health professionals) so they can attempt to contain the child or prevent them from harming the public?

Whether a duty exists is a threshold question of law that is decided by the court.[144] In fact, many negligence cases fail because of the court's

141. *Accord id.* at 62–64 (describing confusion over foreseeability determination that is part of test for three negligence elements—duty, breach, and proximate cause—and various ways courts have tried to distinguish the determinations by using differentiating terms like *specific foreseeability* vs. *general foreseeability* and *forward-looking foreseeability* vs. *backward-looking foreseeability*); Cardi, *supra* note 134, at 743 (noting that courts rely upon some form of foreseeability determination in ascertaining whether duty, breach, and proximate cause elements of negligence are satisfied).

142. Doe 30's Mother v. Bradley, 58 A.3d 429, 436 (Del. Super. Ct. 2012).

143. Bridges v. Parrish, 731 S.E.2d 262, 265 (N.C. Ct. App. 2012) (quoting Davis v. N.C. Dep't of Human Res., 465 S.E.2d 2, 6 (N.C. Ct. App. 1995)) (internal quotation marks omitted), *aff'd*, 742 S.E.2d 794 (N.C. 2013).

144. *See* Shelley Ross Saxer, *"Am I My Brother's Keeper?": Requiring Landowner Disclosure of the Presence of Sex Offenders and Other Criminal Activity*, 80 NEB. L. REV. 522, 526 (2001) (footnote in title omitted).

finding that there was no duty to act, either because there was no special relationship or the third party's criminal conduct was unforeseeable.[145] As previously explained, a court can find a special relationship between an adult mass shooter and his parents if the criteria specified in Part I(D) are met.

"Courts will generally find a duty where reasonable persons would recognize and agree that it exists,"[146] and it is well established that "[e]very one owes to the world at large the duty of refraining from those acts that may unreasonably threaten the safety of others."[147] But remember, parental liability is generally not premised on negligent acts affirmatively committed by the parents themselves, but rather on their failure to prevent malevolent acts by their children. Thus, how the duty issue is analyzed depends upon whether the plaintiff is alleging misfeasance or nonfeasance.[148]

1. *Nonfeasance vs. Misfeasance*

Nonfeasance refers to when a defendant fails to act for the protection of others,[149] whereas misfeasance is "active misconduct working positive injury to others."[150] Classic examples of nonfeasance in the third-party, criminal attacker context are the defendant's failure to (1) control the criminal attacker; (2) warn identifiable potential victims or the proper authorities of the assailant's dangerousness; or (3) take measures to protect the victims from the third-party attacker.[151] An example of misfeasance that could arise in the context of a mass shooting would be that the parent acted negligently by entrusting a dangerous instrumentality to a known mentally ill person.[152]

---

145. *See* Bahadur, *supra* note 138, at 69–70 (observing that judges play important role in negligence cases "as gatekeepers or screeners of negligence actions" because "[i]f the court determines that no duty of care is owed to the plaintiff or that no duty of care exists, then the court may dismiss the action as a matter of law").

146. Estate of Heck v. Stoffer, 786 N.E.2d 265, 268 (Ind. 2003).

147. Palsgraf v. Long Island R.R. Co., 162 N.E. 99, 103 (N.Y. 1928) (Andrews, J., dissenting).

148. *See* Smith v. Freund, 121 Cal. Rptr. 3d 427, 431 (Ct. App. 2011).

149. *See* Price v. E. I. DuPont de Nemours & Co., 26 A.3d 162, 167 (Del. 2011).

150. Cara McDonald, Case Note, *Torts Law: Blurred Elements: The Nebulous Nature of Foreseeability, the Confounding Quality of Misfeasance, and the Minnesota Supreme Court's Decisions*—Doe 169 v. Brandon, 41 WM. MITCHELL L. REV. 365, 369 (2015) (internal quotation marks omitted).

151. *See* Eric J. v. Betty M., 90 Cal. Rptr. 2d 549, 551, 558 (Ct. App. 1999) (recognizing failure to warn and failure to protect as nonfeasance claims).

152. *See* RESTATEMENT (SECOND) OF TORTS § 302B cmt. e, illus. 11 (1965) (listing negligent entrustment of firearms to minors as one situation where defendant can be responsible for third party's criminal act because his affirmative act—misfeasance—created the risk of harm).

Obviously, claims of misfeasance are easier to prove and will fare far better than nonfeasance allegations, because misfeasance actually involves wrongdoing on the parents' parts.[153] For example, in the case of Nancy Lanza, it would be easier to make a case that she negligently stored her firearms or entrusted them to her son, whom she knew was severely mentally ill, than to establish that she was negligent for failing to supervise or control him.[154] This is especially true because in negligent entrustment and other misfeasance cases, the plaintiff does not have to establish a special relationship between the parent and third-party criminal actor, since the gravamen of a misfeasance complaint is not based upon the third party's criminal act, but rather the parent's own affirmative act that created an unreasonable risk of harm.[155] Indeed, it could be argued that giving a mentally ill person a firearm is negligence per se in light of federal and state laws prohibiting the transfer of firearms to mentally unstable persons.[156] Accordingly, where possible, a plaintiff should frame the parents' duty as a question of misfeasance to increase their odds of success.

---

153. *See* John M. Adler, *Relying Upon the Reasonableness of Strangers: Some Observations About the Current State of Common Law Affirmative Duties to Aid or Protect Others*, 1991 WIS. L. REV. 867, 872–73 (noting if defendant's behavior is characterized as nonfeasance, in most cases courts will find no duty); Brian D. Bender, Case Note, *Torts: The Failings of the Misfeasance/Nonfeasance Distinction and the Special Relationship Requirement in the Criminal Acts of Third Persons*—State v. Back, 37 WM. MITCHELL L. REV. 390, 391–92 (2010) (acknowledging that "[i]f the defendant's act is characterized as misfeasance, a duty generally attaches to prevent the harm caused by those actions[,]" but "[i]f the defendant's act is characterized as nonfeasance, there is typically no duty to prevent the harm caused by the third person absent some special relationship").

154. Courts have no problem finding that homeowners have a duty to not entrust an inherently dangerous instrumentality like a gun to a mentally disturbed individual or to safely store such a dangerous instrumentality in a home where a mentally ill individual has access. *See e.g.*, Jupin v. Kask, 849 N.E.2d 829, 832–33 (Mass. 2006) (holding that "a homeowner who permits guns to be stored on her property and allows unsupervised access to that property by a person known by her to have a history of violence and mental instability, has a duty of reasonable care to ensure that the guns are properly secured," and that it was foreseeable that her live-in boyfriend's mentally ill son could take one of the guns and shoot someone, including a police officer). "[N]early every state has [recognized] some form of the negligent entrustment doctrine." *See* Andrew D. Holder, Comment, *Negligent Entrustment: The Wrong Solution to the Serious Problem of Illegal Gun Sales in Kansas [*Shirley v. Glass*, 241 P.3d 134 (Kan. Ct. App. 2010)]*, 50 WASHBURN L.J. 743, 747 (2011).

155. *See* Bender, *supra* note 153, at 396 (stating "courts will generally hold a person accountable for their misfeasance . . . . [because] [t]his accountability generally does not depend on whether that harm arises from an act—criminal or otherwise—of a third person[,]" and that Minnesota requires plaintiff to show special relationship only in nonfeasance cases); *see also* Price v. E. I. DuPont de Nemours & Co., 26 A.3d 162, 170 (Del. 2011) (recognizing that "[i]n cases of nonfeasance, no duty of care exists between the parties unless a 'special relationship' between them gives rise to one").

156. *See, e.g.*, Martin v. Schroeder, 105 P.3d 577, 583 (Ariz. Ct. App. 2005) (recognizing that parents violated 18 U.S.C. § 922(d) by buying a gun for their adult, eighteen-year-old son whom they knew was a drug abuser and thus commit-

Additionally, it should be noted that whether a policy consideration weighs in favor of recognizing a duty depends on the particular duty the plaintiff alleges the defendant owed to them and the specific facts of the case. For instance, the duty-policy analysis in a case alleging that parents were negligent for failing to have their child involuntarily committed to a mental institution will be quite different than one alleging the parent was negligent for failing to properly supervise their offspring to prevent him from accessing a gun, the former being much more difficult to prove because of the competing policy considerations concerned—the need to protect the public versus the mentally ill individual's freedom. Thus, it would be better to cast the duty in, broader terms—such as *the duty to supervise or control the third-party attacker*—than a more specific allegation—such as *the defendant parents had a duty to have their emancipated offspring involuntarily committed.*

Although nonfeasance duties are harder to prove than misfeasance duties, it does not follow that plaintiffs cannot successfully establish nonfeasance negligence cases against parents. Because most parental liability cases involving mass shooters will probably also involve negligence allegations based upon nonfeasance, this Section will primarily focus on the nonfeasance duties of parents' failure to control, protect, or warn.

2. *Duty-Policy Considerations*

As previously explained, in order for courts to find a special relationship-based duty to prevent a mass shooting under a nonfeasance theory, a plaintiff has to prove more than the existence of a special relationship between the parents and their emancipated offspring; they must also establish that a myriad of policy considerations weigh in favor of recognizing a duty.[157] The policy concerns that have been considered include (1) the reasonable foreseeability of the injury; (2) the degree of certainty that the plaintiff suffered injury; (3) the closeness of the connection between the defendant's conduct and the injury suffered; (4) the moral blame at-

ted negligence per se); West v. MacHe of Cochran, Inc., 370 S.E.2d 169, 173 (Ga. Ct. App. 1988) (holding licensed firearms dealer committed negligence per se by knowingly selling firearm to mentally ill individual, in violation of federal and state law); Rubin v. Johnson, 550 N.E.2d 324, 330–31 (Ind. Ct. App. 1990) (holding there was sufficient evidence that pawn shop owner violated state statute prohibiting sale or transfer of a firearm to person of "unsound mind" and thus committed negligence per se). To establish negligence per se, you have to show (1) a violation of a statute; (2) the statute was intended to protect the class of people to which the plaintiff belongs; and (3) the statute was design to prevent the type of harm that occurred. *See Rubin*, 550 N.E.2d at 329. It should be noted that a finding of negligence per se is not necessarily a finding of liability per se; the plaintiff must still prove that the defendant's negligence caused the harm. *See* Williams *ex rel.* Raymond v. Wal-Mart Stores E., L.P., 99 So. 3d 112, 116–17 (Miss. 2012).

157. *See* Smith v. Freund, 121 Cal. Rptr. 3d 427, 433 (Ct. App. 2011). *But see* Estate of Strever v. Cline, 924 P.2d 666, 670 (Mont. 1996) (dispensing with special-relationship inquiry in duty analysis and only considering foreseeability of risk of harm and weighing policy considerations to determine existence of duty).

tached to the defendant's conduct; (5) the policy of preventing future harm; (6) the extent of the burden to the defendant to prevent the harm; (7) the consequences to the community of imposing a duty to exercise care with resulting liability for breach; and (8) the availability, cost, and prevalence of insurance for the risk involved.[158] All of these policy factors will be addressed below. Because foreseeability is the most important policy consideration in the duty analysis in the majority of jurisdictions,[159] it will be addressed first, separate from the other policy concerns.

a. Foreseeability of Mass Murders

It is well settled that for an act to be foreseeable for purposes of the duty analysis, the third party's criminal conduct must be reasonably anticipated,[160] thereby excluding remote and unexpected events from the realm of foreseeability.[161] Courts disagree, however, over what test should be applied to determine whether the defendant should have foreseen the third party's conduct.

Some jurisdictions have found that the third party had to have committed a "similar" prior criminal act for the defendant to foresee a criminal act could occur again, but nevertheless treat similar prior criminal acts as if they must be identical.[162] In these jurisdictions, it will be impossible

158. These policy concerns were originally enunciated by the California Supreme Court in *Rowland v. Christian*, and many courts have subsequently adopted some or all of the *Rowland* policy concerns in its duty analysis. *See, e.g.*, Marshall v. Burger King Corp., 856 N.E.2d 1048, 1057 (Ill. 2006); Rodriguez v. Spencer, 902 S.W.2d 37, 43 (Tex. Ct. App. 1995); *see also* Tory A. Weigand, *Duty, Causation and* Palsgraf*: Massachusetts and the Restatement (Third) of Torts*, 96 MASS. L. REV. 55, 57–58 (2015) (reporting recent survey that revealed most states (1) utilize multi-policy test to make duty determinations and (2) include foreseeability in that multi-factor policy duty assessment).

159. *See* Weigand, *supra* note 158, at 58 (observing most jurisdictions include foreseeability in their duty analysis and leave duty-foreseeability determination for jury to decide); Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 44 WAKE FOREST L. REV. 1247, 1260 (2009) (noting foreseeability plays significant role in duty analysis in forty-seven states); *see also* Tarasoff v. Regents of Univ. of Cal., 551 P.2d 334, 342 (Cal. 1976) (observing "most important" policy consideration in establishing whether duty exists is foreseeability); Margaret W. v. Kelley R., 42 Cal. Rptr. 3d 519, 530 (Ct. App. 2006) (recognizing that in determining defendant's liability in third-party criminal conduct cases, foreseeability is crucial factor); Romero v. Giant Stop-n-Go, Inc., 212 P.3d 408, 410 (N.M. Ct. App. 2009) (reiterating "[f]oreseeability is a critical and essential component of New Mexico's duty analysis because no one is bound to guard against or take measures to avert that which he or she would not reasonably anticipate as likely to happen" (alteration in original) (internal quotation marks omitted)).

160. Steve Jasper, Note, *Congratulations! It's a Tort: Expanding the Scope of Duty in the Surrogacy Setting*, 67 MO. L. REV. 421, 429 (2002); Jeffrey H. Powell, Comment, Marshall v. Burger King Corp.*: Making a Mess of "Duty" for Businesses in Illinois*, 28 N. ILL. U. L. REV. 95, 112 (2007).

161. *See* Lopez v. McDonald's Corp., 238 Cal. Rptr. 436, 444 (Ct. App. 1987).

162. *See* Trammell Crow Cent. Tex. Ltd. v. Gutierrez, 267 S.W.3d 9, 11, 12, 17 (Tex. 2008) (stating that foreseeability of criminal acts of third party on one's property is established by evidence of "specific previous crimes on or near the

to establish that an adult child's mass shooting was foreseeable because, as previously mentioned, the typical mass shooter has no prior criminal history, and in the event they had committed a mass murder before, they may have already been incarcerated, serving a life sentence, or sitting on death row, consequently unable to commit the mass shooting at issue.

Other states have concluded that for an act to be foreseeable, it has to be of the same general character of the past event or harm, and they apply a totality of the circumstances test—which includes considerations such as the proximity of, frequency of, and publicity surrounding prior similar incidents—to determine whether the act is of the same general character.[163] In these jurisdictions, it will also be virtually impossible to show a mass shooting is foreseeable because rarely, if ever, does a mass shooting occur at the same locale. And while the occurrence of mass shootings overall nationwide may have increased, the frequency of mass shootings in particular areas has not.

A few jurisdictions, like California and Tennessee, treat foreseeability as a flexible, fluid concept and employ a balancing test that weighs the foreseeability and gravity of the harm against the burden of preventing the harm.[164] In these jurisdictions, if the burden of preventing the harm is great—such as hiring security guards to prevent criminal activity on a property—a heightened degree of foreseeability is required.[165] Conversely, if the burden of preventing harm is minimal—such as placing a telephone call to 911 during an ongoing criminal attack—then a lesser degree of foreseeability is required because, for example, it does not take much foresight for a defendant who is witnessing a fistfight to know someone will be seriously injured when the defendant also knows one of the participants has a knife.[166]

Depending upon the duty alleged and the facts of the case, applying this more flexible foreseeability test will make it easier to determine

---

premises" and holding prior robberies at a shopping center were insufficient to make shooting and killing of apparent robbery victim foreseeable when no one was seriously injured in previous robberies and no weapon was used on person (internal quotation marks omitted)); *see also* Craig Crawford, Comment, Delgado v. Trax Bar & Grill*: Determining the Scope of the Prior Similar Incidents Test in Terms of Efficient Resource Allocation*, 39 U.S.F. L. REV. 499, 503 (2005) (noting that in premises liability cases, "[s]ome jurisdictions interpreted the prior similar incidents test to require prior 'identical' incidents. . . . [and] imposed a duty on a landowner only where there was evidence of a prior incident of the same crime occurring in approximately the same manner").

163. *See* Glenda K. Harnad, *Effect of Foreseeability*, 53 TEX. JUR. 3D NEGLIGENCE § 59 (3d ed. & Supp. 2015); Saxer, *supra* note 144, at 527–30 (acknowledging foreseeability may be based upon prior similar conduct or totality of circumstances).

164. *See* Delgado v. Trax Bar & Grill, 113 P.3d 1159, 1166–70 (Cal. 2005); *Lopez*, 238 Cal. Rptr. at 444 (noting "[f]oreseeability is . . . an elastic factor within a somewhat flexible concept"); McClung v. Delta Square Ltd. P'ship, 937 S.W.2d 891, 901–02 (Tenn. 1996).

165. *See Delgado*, 113 P.3d at 1166–69.

166. *See* Morris v. De la Torre, 113 P.3d 1182, 1184, 1189 (Cal. 2005).

whether a mass shooting was foreseeable. For instance, if it is alleged that the defendant had a duty to supervise the adult child to prevent a mass shooting, the degree of foreseeability will be high and more difficult to meet. If the alleged duty is a duty to call the police when a known-to-be mentally ill child runs out the door with a gun or plans a mass murder, the degree of foreseeability required is low because the burden on the parents is minimal. Thus, the foreseeability requirement will be easier to satisfy in the latter case and others like it.

The foreseeability test will be easiest to satisfy in jurisdictions that also require lesser foreseeability when there are strong policy reasons for preventing the harm.[167] For example, in *Juarez v. Boy Scouts of America, Inc.*,[168] a California appellate court found that it was foreseeable that a child would fall victim to a sexual predator who was serving as an adult scoutmaster in the Boy Scouts, irrespective of the fact that the troop leader had no prior documented instances of sexual molestation.[169] In reaching this conclusion, the court emphasized that based upon legal commentary, reported cases, and books, the Boy Scouts organization had to have known the possibility existed that pedophiles would be attracted to the Boy Scouts organization because working there would give them legitimate access to young children.[170]

Additionally, the court pointed to a much earlier California case that found sexual assault by unknown third parties against young children was foreseeable because "[i]t is certain that there exists in our civilization the constant possibility that persons suffering from a lack of proper mental balance or normal decency might subject young people to sexual molestation."[171] In that earlier case, the court stressed that the constant threat of sexual abuse against children abounded, as illustrated by frequent newspaper accounts of sexual crimes against children, the many litigated criminal cases on the issue, and the legislative enactment of laws to protect children.[172] Most significantly, the court emphasized that *"[i]f injury to another . . . is likely enough in the setting of modern life that a reasonably thoughtful [person] would take account of it in guiding practical conduct[,] . . . we must label the injury reasonably foreseeable* . . . ."[173] The court further explained that a showing of foreseeability—not the more stringent standard of probability—is what is required to satisfy this aspect of the special relationship-based duty test.[174] Accordingly, the court found that the sexual as-

167. *See id.*

168. 97 Cal. Rptr. 2d 12 (Ct. App. 2000).

169. *See id.* at 31.

170. *See id.* at 30.

171. *Id.* at 31 (quoting Wallace v. Der-Ohanian, 18 Cal. Rptr. 892, 896 (Ct. App. 1962)) (internal quotation marks omitted).

172. *See Wallace*, 18 Cal. Rptr. at 896.

173. *Juarez*, 97 Cal. Rptr. 2d at 30 (second alteration in original) (emphasis added) (internal quotation marks omitted).

174. *See id.* at 31.

sault was foreseeable, and the Boy Scouts had a duty to take reasonable measures to protect the victim, such as by training the boy scouts on how to handle a sexual predator.[175]

Applying *Juarez* to the foreseeability of mass shootings in general, and Adam Lanza's and Jared Loughner's mass shootings in particular, an argument can easily be made that mass shootings are foreseeable when certain factors exist. First, similar to the sexual assaults in *Juarez*, the possibility of a random mass shooting committed by a mentally unbalanced person who exhibits certain signs has been so well-publicized in literature and the media that it should be reasonably anticipated if someone is acting in the manner described in the next two sentences, they are a candidate for committing a mass murder. For instance, the media has consistently reported that most mass shooters have no prior criminal history of violence, which makes the lack of a criminal record itself a common characteristic of a mass shooter.[176] Other common characteristics include a person (1) displaying mentally unbalanced behavior to the extent that it causes others around them to be concerned; (2) being anti-social, reclusive, or angry with the world; (3) having an unusual obsession with prior public mass shootings or firearms; and (4) manifesting an intent to kill either in writings, YouTube postings, other social media, or telling a friend or acquaintance of their murderous plot.[177]

Both Adam Lanza and Jared Loughner displayed most, if not all, of these signs, and thus their parents should have known they were severely mentally ill, potentially dangerous, and capable of shooting someone. As to Adam Lanza, he displayed clear signs of mental illness and was the epitome of a recluse. He had stopped verbally communicating with his mother except through email and refused to leave their home for any reason in the three months leading up to the shootings, even when their electricity went out due to Hurricane Sandy.[178] These actions, coupled with his lifetime history of mental health issues, should have put his mother on notice that he was not in a healthy mental state and had no business handling a firearm. Adam also had an unusual obsession with mass shootings and guns. After Adam's killing spree, police found an enormous amount of material on mass shootings and firearms, including a spreadsheet on mass murders listing details about each shooting.[179] Thus, it was foreseeable that Adam could harm anyone in the general public with a firearm.[180]

175. *See id.*

176. *See* Weber & Chang, *supra* note 124 (observing commonalities between recent mass shooters James Holmes, Adam Lanza, and Elliott Rodger—"[a]ll were young loners with no criminal history").

177. *See* Lewis, *supra* note 2, at 338–39.

178. *See* SEDENSKY, *supra* note 10, at 28.

179. *See id.* at 25–26.

180. *See* Silberstein v. Cordie, 474 N.W.2d 850, 856 (Minn. Ct. App.) ("Significantly, the duty to control, unlike the duty to warn, may arise if there is foreseeable

Regarding Jared Loughner, there is substantial evidence that his parents also should have reasonably anticipated that he would shoot and kill someone, given his mental instability and potential for violence. Evidence of his mental instability included talking to himself while at the same time refusing to communicate with his parents, making nonsensical rants, and writing incoherent statements in his journal.[181] Additionally, Jared's parents knew he had been suspended from school for exhibiting mentally unstable behavior, including making an outburst in class about "blowing up babies" and posting a YouTube video showing him prowling around his campus at night with a video camera threatening to torture students.[182] Jared's college further advised his parents that they believed Jared posed a danger to himself and others and could not return to campus until a mental health professional had determined he was no longer dangerous.[183] Accordingly, it was certainly foreseeable that someone with this mental state could harm someone if they were armed.

Indeed, not only were Adam Lanza's and Jared Lougnher's mass shootings foreseeable, but their parents' actions demonstrate that they actually foresaw their sons seriously injuring or murdering someone. As previously mentioned, Adam Lanza's mother had planned to have him involuntarily committed to a mental health facility shortly before his shooting spree,[184] which indicates she believed he was capable of committing violence because in order to have someone involuntarily committed, you have to believe the person is mentally ill and dangerous.[185] Jared Loughner's parents obviously believed he was capable of harming the public, as they confiscated the only firearm they believed he owned and disabled his vehicle every night to keep him home.[186] Also, on the morning of Loughner's shooting rampage, his father tried to see what was in the bag he was carrying, rightfully believing it was a gun, and even chased Jared in an apparent effort to seize him and his belongings, though he was ultimately unsuccessful.[187]

It could be argued, based upon *Lopez v. McDonald's*, that mass shootings of random people in general are a remote and unexpected event and thus are unforeseeable. In *Lopez*, a California court found that a mass

harm to a member of the general public."), *rev. granted in part, cause remanded by* 477 N.W.2d 713 (Minn. 1991).

181. *See* Wagner, *supra* note 110.

182. *See* Shaundra K. Lewis, *Bullets and Books by Legislative Fiat: Why Academic Freedom and Public Policy Permit Higher Education Institutions to Say No to Guns*, 48 IDAHO L. REV. 1, 2 (2011).

183. James B. Jacobs & Jennifer Jones, *Keeping Firearms Out of the Hands of the Dangerously Mentally Ill*, 47 CRIM. L. BULL. 388, 388 (2011).

184. *See* Winter, *supra* note 14, and accompanying text.

185. *See* John Glenn et al., *Persons Subject to Commitment or Confinement*, 56 C.J.S MENTAL HEALTH § 54 (2015) ("A person may not be involuntarily committed to an institution for the mentally ill unless the person is mentally ill and dangerous.").

186. *See* Martinez & Carter, *supra* note 109.

187. *See* Wagner, *supra* note 110.

murderous assault on McDonald's patrons and employees that left twenty-one persons dead and eleven others wounded was unforeseeable.[188] There, survivors and deceased victims' family members sued McDonald's for negligence, claiming the company had a duty to protect its patrons from reasonably foreseeable criminal acts of third parties—namely, a mentally disturbed gunman—and that it had breached that duty by not having security personnel to protect the patrons.[189] The plaintiffs presented evidence the shooting was foreseeable, because that particular McDonald's was in a high-crime area, and there had been some thefts, robberies, and assaults with a deadly weapon in the vicinity.[190] The appellate court found that the prior criminal activity at McDonald's and the immediate surrounding area were predominately theft-related offenses and not the type of shooting massacre that occurred in the instant case. The court further reasoned that a shooting rampage was so remote and unexpected that, as a matter of law, McDonald's nonfeasance in not providing security guards did not facilitate the shooting's occurrence.[191] Further, the court concluded that a mass killing was so unlikely to occur within the setting of modern life that a reasonably prudent business would not consider the possibility of its occurrence in trying to protect its invitees.[192]

Lopez is clearly distinguishable from Adam Lanza's and Jared Loughner's case. First, and significantly, the mass shooting in *Lopez* occurred in 1984.[193] Since then, mass shootings have occurred with much greater frequency; thus, it can no longer be said in 2016 that a mass killing is an unlikely or remote occurrence.[194] Second, having the foresight that a stranger entering a business would embark on a killing spree is quite different than anticipating someone with whom you live—and whom you know has significant mental health issues and an unusual fascination with firearms, killing, or mass shootings—may be planning a mass murder of their own. Indeed, the law in every jurisdiction reasonably anticipates that a mentally ill person with a firearm might harm someone, which is why every jurisdiction has some variation of a law intended to preclude mentally ill persons from possessing firearms.[195] Accordingly, not only were Adam Lanza's and Jared Loughner's mass killings foreseeable, but it is more generally foreseeable that any mentally ill person acting as they did could commit mass murder.

Even if a court reasoned that a mass shooting was unforeseeable in Adam Lanza's, Jared Loughner's, or a future mass shooting case, it does not mean that no duty can be recognized; foreseeability is a dominant

188. *See* Lopez v. McDonald's Corp., 238 Cal. Rptr. 436, 438 (Ct. App. 1987).
189. *See id.*
190. *See id.* at 439.
191. *See id.* at 445.
192. *See id.*
193. *See id.* at 438.
194. *See id.* at 445.
195. *See* Lewis, *supra* note 2, at 341–48.

consideration in the duty-policy analysis, but it is not the *sin qua non*.[196] A court can still find that a duty exists if the majority of other policy considerations weigh in favor of recognizing a duty. Moreover, a minority of jurisdictions have adopted the Restatement (Third) of Torts, which removes foreseeability completely from the duty calculus.[197] This is probably because no special expertise is required to determine whether something is foreseeable, since "deciding what is reasonably foreseeable involves common sense, common experience, and application of the standards and behavioral norms of the community—matters that have long been understood to be uniquely the province of the finder of fact."[198] Thus, in these jurisdictions, foreseeability will not even be a factor, much less a bar, to the recognition of a duty.

b. Other Policy Considerations

The majority of the remaining policy factors weigh in favor of recognizing a parental duty to supervise or control a mentally ill, adult child residing at home, and if the parents cannot control the child, to warn the proper authorities (the police, the court, or mental health professionals) that their adult child has a firearm, is mentally ill, and may be potentially dangerous. As previously mentioned, the other policy considerations include the degree of certainty that the plaintiff suffered injury; the closeness of the connection between the defendant's conduct and the injury suffered; the moral blame attached to the defendant's conduct; the policy of preventing future harm; the extent of the burden to the defendant to prevent the harm; the consequences to the community of imposing a duty to exercise care with resulting liability for breach; and the availability, cost, and prevalence of insurance for the risk involved.[199] Each will be addressed in turn.

i. The Degree of Certainty the Plaintiff Suffered Injury

This policy consideration weighs in favor of recognizing a parental duty to (1) supervise or control mentally ill, adult children residing at home to prevent them from accessing a firearm or (2) warn the proper authorities when they do have a gun or have evidenced an intent to commit a mass murder. Thus, this policy factor without question will always tip in favor of recognizing a parental duty.

196. *See* Harnad, *supra* note 163, § 59; *see also* Juarez v. Boy Scouts of Am., Inc., 97 Cal. Rptr. 2d 12, 31 (Ct. App. 2000) (noting "foreseeability is not coterminous with duty"). Thus, it stands to reason a finding of unforeseeability, standing alone, will not automatically mandate a finding of no duty.

197. *See* Weigand, *supra* note 158, at 64.

198. *See* A.W. v. Lancaster Cnty. Sch. Dist. 0001, 784 N.W.2d 907, 914 (Neb. 2010).

199. *See supra* note 156 and accompanying text.

ii. The Closeness of the Connection Between the Defendant's Conduct and the Injury Suffered

Second, there is a close connection between the defendant's conduct of (1) failing to supervise or control an adult, mentally ill child residing with them or (2) warning the proper authorities of a mentally disturbed child planning a mass killing or on the loose with a gun, and someone in the community randomly being shot by that individual. A recent FBI report shows that almost 70% of mass shooting incidents end before the police can respond to the scene.[200] Indeed, research on mass school shootings reveal that most foiled school rampage shootings were detected and prevented because a family member or friend alerted the proper authorities before the would-be perpetrator could execute their plan.[201] Parents who reside with their children and observe them on a daily basis are in the best position to observe their child's mental state and behavior. Failure to prevent a mentally ill child from accessing a firearm is strongly connected to a third party's injury or death. Thus, this factor weighs in favor of recognizing a parental duty in the typical mass shooting case.

iii. The Moral Blame Attached to the Defendant's Conduct

Third, at least one research study indicates that parents of a mass shooter are partially morally to blame for their child's mass shooting because, prior to its occurrence, the parents had a history of accepting and accommodating their child's pathological behavior. According to this study, parents of school shooters failed to "react to behavior that most parents would find very disturbing or abnormal" and minimized their child's erratic conduct.[202] These parents set no limits on their child's behavior and did not supervise or monitor their child's use of the Internet.[203] The parents also had a turbulent relationship with their child to the extent that the child expressed contempt for one or more of their parents and rejected their parents' role in their lives.[204] Finally, the par-

---

200. *See* BLAIR & SCHWEIT, *supra* note 2, at 8–9.

201. *See* PETER LANGMAN, *Ten Lessons Learned from School Shootings and Foiled Attacks*, *in* WHY KIDS KILL: INSIDE THE MINDS OF SCHOOL SHOOTERS (2014), *available at* https://schoolshooters.info/sites/default/files/ten_lessons_1.1.pdf [https://perma.cc/9ENM-QZ2G] (recognizing "[n]umerous students have prevented possible [school shooting] attacks because they knew enough to report what they heard to parents or school personnel); *see also* KATHERINE S. NEWMAN ET AL., RAMPAGE: THE SOCIAL ROOTS OF SCHOOL SHOOTINGS 59 (2004) (noting that in *New York Times* study of 102 rampage killers, "more than half of the cases, friends, family, and even the offenders themselves tried to get help or warn others about the impending violence"). *See generally* MARY ELLEN O'TOOLE, FBI, THE SCHOOL SHOOTER: A THREAT ASSESSMENT PERSPECTIVE 21 (2000), *available at* http://www.fbi.gov/stats-services/publications/school-shooter [https://perma.cc/N7EC-NZYQ].

202. O'TOOLE, *supra* note 201, at 21.

203. *See id.* at 22.

204. *See id.* at 21.

ents carelessly kept firearms in the home, giving their children unfettered access to these deadly weapons.[205]

The commonalities observed among parents of school rampage shooters in this study are consistent with Nancy Lanza's case. Despite expressing concern over her son's behavior, including plans to commit him, Nancy downplayed her son's behavior and told friends that she was not afraid Adam would harm her.[206] Nancy allowed her son to live in isolation, even permitting him to stay in their home without leaving once for three months.[207] Adam had a strained relationship with both his parents. He had stopped verbally communicating with them and was estranged from his only sibling.[208] Toward the end, Nancy apparently realized that her son needed emergency mental health attention and was in the process of having him involuntarily committed, but she nevertheless left him alone in their home with numerous firearms that were not securely stored while she went on vacation.[209] While she was away, her son made a practice run in his car to Newtown where Sandy Hook Elementary School was located.[210] Shortly thereafter, her son used her firearms to murder the elementary school children and educators.[211] Not only did she provide Adam with unrestrained access to her firearms, but she also encouraged his usage of firearms by taking him to shooting ranges and giving him a check to purchase his own gun for Christmas.[212] This was not only morally reprehensible; it was arguably criminally negligent. Therefore, this policy consideration would tip in favor of recognizing a duty in the case of Adam Lanza's mother.

However, it is a closer call as to whether this factor would weigh in favor of finding a duty in the case of Jared Loughner's parents. On one hand, Jared's parents were warned by Jared's community college that Jared was mentally unstable, dangerous, and needed a mental health evaluation, yet the parents never followed up or made sure that Jared saw the mental health professional. Although they advised him to see one,[213] after Jared refused, they should have given him an ultimatum similar to the one issued by Jared's college—seek help or leave.[214] Alternatively, they could have sought involuntary commitment.

---

205. *See id.*

206. *See* SEDENSKY, *supra* note 10, at 28.

207. *See id.*

208. *See id.* at 29–30.

209. *See id.* at 25, 28; Winter, *supra* note 14.

210. *See* SEDENSKY, *supra* note 10, at 28.

211. *See* Lewis, *supra* note 2, at 322.

212. *See* SEDENSKY, *supra* note 10, at 28; Cummings et al., *supra* note 11.

213. *See* Wagner, *supra* note 110.

214. *See Helpful Tips for Families*: *Living with a Mentally Ill Relative–Practical Advice*, NAT'L ALLIANCE MENTAL ILLNESS, http://www.namiyolo.org/tipsforfamilies.html [https://perma.cc/5E5K-P8GY] (last visited Mar. 1, 2016); *see also* Donald Kern, *Tough Love and Bipolar Disorder*, BIPOLAR BY CHANCE, http://bipolarbychance.blogspot.com/2009/06/tought-love-and-bipolar-disorder.html [https://perma

Although not as much information is available on the relationship between Jared and his parents, one media outlet reported that Jared had a strained relationship with his parents and claimed that they were "hassling" him.[215] Neighbors described Jared's father as angry and antagonistic towards them and insinuated that he was "crazy."[216]

On the other hand, the Loughners did try to supervise Jared and monitor his whereabouts.[217] For instance, Jared had once left home without his parent's knowledge, and his parents went looking for him.[218] The parents of Jared's friend told the Loughners Jared was staying at a hotel, and the Loughners went and retrieved him.[219] The parents also took away Jared's firearm and disabled his car every night to keep him home in the evenings,[220] although they forgot to disable his car the night before the shootings.[221] Additionally, in the months leading up to the shootings, the Loughners pleaded with their son to "see someone."[222] Thus, a court could find that this factor goes either way.

As illustrated by the Loughner and Lanza exemplars, whether the "moral blame" policy consideration weighs in favor of recognizing a duty will depend on the specific facts of each mass shooting case.

iv. The Policy of Preventing Future Harm and the Consequences to the Community of That Harm

Fourth, the policy of preventing future harm also weighs in favor of finding parents liable for failing to prevent an adult, mentally ill child from possessing a firearm. According to a recent FBI report, mass shoot-

---

.cc/TCE6-MRZZ] (last visited Mar. 1, 2016) (advocating for "tough love" to get a mentally ill family member to seek treatment, explaining that "[s]ometimes it means giving an ultimatum of going for treatment or leaving the family home"); Madison Park & Elizabeth Landau, *Warning Signs from a Troubled Mind: What Parents Should Do*, CNN (Jan. 14, 2011, 6:09 PM), http://www.cnn.com/2011/HEALTH/01/14/mental.health.loughner/index.html [https://perma.cc/KBH7-9K8T] (reporting how Dr. Charles Raison, psychiatrist at Emory University, recommends giving adult children ultimatums to take their prescribed medication or leave).

215. Amanda Lee Myers & Justin Pritchard, *Jared Loughner's Parents Devastated, Neighbor Says*, HUFFINGTON POST (Jan. 1, 2011), http://www.huffingtonpost.com/2011/01/11/jared-loughner-parents_n_807479.html [https://perma.cc/W6H9-L786] (updated May 5, 2011, 6:25 PM).

216. *See* Martin & Siegel, *supra* note 104.

217. *See id.*

218. *See* Myers & Pritchard, *supra* note 215.

219. *See id.*

220. *See* Martinez & Carter, *supra* note 109.

221. Hipolito Corella, *Tucson Shooting: No Mental Health Treatment for Loughner Before Giffords Rampage*, ARIZ. DAILY STAR, Mar. 27, 2013, 8:38 AM, http://tucson.com/news/local/crime/tucson-shooting-no-mental-health-treatment-for-loughner-before-giffords/article_70b5b03c-96f4-11e2-b7dc-0019bb2963f4.html [https://perma.cc/XP4W-C439].

222. *See* Martinez & Carter, *supra* note 109.

ings undoubtedly are on the rise.[223] It has become an epidemic. Since an individual can kill a number of people in a matter of minutes before the police even arrive on the scene, the best approach to decreasing the incidence of mass shootings is to prevent them before they start, and the best place to start is in the home. No one knows a child better than his or her parent. Parents are in the best position to evaluate the mental health and propensity for violence of their child and whether they should have a gun.

Moreover, most mass shootings are committed at schools where innocent children are victimized.[224] As one court has recognized, "[o]ur greatest responsibility as members of a civilized society is our common goal of safeguarding our children, our chief legacy, so they may grow to their full potential and can, in time, take our places in the community at large."[225] Recognizing a parental duty to control or supervise adult, mentally ill children living with them will hopefully encourage more dialogue on mental illness and what parents should do when dealing with an adult, mentally ill child—a dialogue that can save lives. Thus, this policy factor weighs in favor of recognizing a parental duty to supervise/control adult children to the extent necessary to prevent them from accessing a firearm and to warn authorities when they have a gun. Failure to do so will have dire consequences for the community.

v. The Extent of the Burden to Prevent the Harm

Fifth, the burden on parents of keeping firearms out of a mentally ill, emancipated child's hands is slight. It does not take much effort to safely store firearms, or even better, to not have them in the home where a mentally ill person lives at all. Indeed, at least one state has a statute expressly prohibiting firearms in a house where a mentally ill individual resides.[226] Alternatively, parents can take measures to ensure that their child cannot access firearms by safely storing them in a safe that is only unlocked by the parents' fingerprints. Additionally, the burden on parents of monitoring an adult child's activities living in their home is also slight. For example, parents can easily search their child's room for any weapons or writings suggesting their child is planning something sinister.

vi. The Availability, Cost, and Prevalence of Insurance for the Risk Involved

This factor can weigh in favor of or against imposition of a duty depending upon the facts of a case. If the defendant parents own a house,

---

223. *See* BLAIR & SCHWEIT, *supra* note 2, at 20.

224. *See id.* at 12–13.

225. *See* Juarez v. Boy Scouts of Am., Inc., 97 Cal. Rptr. 2d 12, 33 (Ct. App. 2000).

226. Jupin v. Kask, 849 N.E.2d 829, 838–39 (Mass. 2006) (recognizing "the financial costs of ensuring the proper securing of firearms and the more esoteric costs involved with requiring certain actions to relieve potential liability—are modest").

insurance may be available because most homeowners' insurance policies cover personal injuries caused by the insured's negligence on their property and "on the street[s]."[227] Some homeowner's insurance, however, expressly excludes criminal acts from coverage,[228] and some parents may not own a home or have any insurance whatsoever. If no insurance covers the risk of someone in the household committing a crime like a shooting, this factor would weigh against recognizing a duty.

vii. Other Policy Considerations

Some may argue that exposing parents to civil liability for the shooting rampages of their children may dissuade parents from housing their mentally ill children who already have few options in life, as they often cannot keep employment and support themselves. Even if this was true, this is not necessarily a negative. A person who is severely mentally ill and dangerous also poses a danger to their parents with whom they reside—Adam Lanza is a prime example, as he shot his mother, killing her, before unleashing his fusillade upon the twenty-six elementary school children and educators at Sandy Hook. Thus, it is in the parent's own best interest to supervise and control their mentally ill, emancipated offspring to make sure their children receive mental health assistance and avoid firearms, and if they cannot control them, to make them leave their residence or have them involuntarily committed to a mental healthcare facility. Finally, perhaps refusing to provide refuge for a severely mentally ill, adult child who will not accept mental health treatment will force that child to obtain the medical assistance he or she so desperately needs in order to have their parents continue to provide food, shelter, and the other necessities and comforts of life. Forcing the mentally ill child to seek professional treatment is also in the individual's best interest, as mass shooters often die during a shooting rampage incident, either from a self-inflicted wound, a bystander, or law enforcement trying to stop the shooting.[229]

Second, some may argue that recognizing a duty in cases involving adult, mentally unstable shooters could unfairly penalize the parents for having a mentally ill child and further stigmatize mental illness. After all, no one is advocating for parental liability in other murder cases (e.g., gang-related). This argument misses the point. Recognizing a parental duty in mass shooting cases is based on the widely accepted premise that a person with significant mental health issues has no business possessing a

227. *See Survivor's Guide to Civil Remedies Against Criminals*, ELLIOT GLICKSMAN PLLC, http://www.glicksmanlaw.com/CM/Custom/Survivors-Guide.asp [https://perma.cc/DZ4H-MG6C] (last visited Mar. 1, 2016) (stating that most homeowners policies covers personal injuries caused by insured's negligence on insured's property and "on the street[s]").

228. *See* Steven Plitt et al., *Risks and Activities Covered by Insurance Policy*, 7A COUCH ON INS. § 103:40 (3d ed. 2015).

229. *See* BLAIR & SCHWEIT, *supra* note 2, at 11.

firearm because, like a minor child, they do not have the mindset to safely handle such a dangerous instrument. Severely mentally ill individuals require the same supervision and monitoring as a minor child. Parents have no obligation to support an adult, mentally ill child, but if they do assume responsibility, they should take the same precautions with them as they would with a five-year-old child. In contrast, a parent has no ability to control an adult, legally emancipated drug dealer or gang member with no mental illness who is not financially dependent upon them.

Moreover, there are multiple safeguards embedded in the intricate negligence analysis that will prevent parental liability in mass shooting cases solely because the parents have a mentally ill son or daughter. For a parental duty to arise, the plaintiff has to do more than merely show the parents were aware of their child's mental health issues. Plaintiffs must prove that the parents also knew or had reason to know their child posed a risk to the community, and that they had the ability to control the child before a special relationship—a prerequisite to recognizing a duty and thus liability—will be found. Assuming that a special relationship can be established, the plaintiff must still show a myriad of policy considerations weigh in favor of recognizing a duty, including the policy consideration that considers the parents' blameworthiness. Thus, a parent will never be penalized for simply having a mentally ill offspring.

In sum, the majority of policy considerations tip in favor of recognizing a parental duty, and, therefore, courts should recognize such a duty under the circumstances described.

IV. PROVING A BREACH OF PARENTAL DUTY

Unlike the issue of whether a duty exists, which is a question of law, whether there was a breach of that duty is a question of fact for the jury.[230] In ascertaining whether a defendant breached any particular duty, the fact-finder is essentially asking whether the defendant acted reasonably under the circumstances; in other words, was it reasonably foreseeable that by acting (or in the case of nonfeasance, not acting), the complained of harm would result?[231]

The foreseeability determination in this stage of the negligence analysis is slightly different than the one in the duty-foreseeability test because it is not just asking whether the third-party adult child's act—the mass shooting—was foreseeable generally, but whether it was reasonably foreseeable that the parents' misfeasance or nonfeasance created a risk for the mass shooting to occur. Part of the analysis to determine whether the defen-

230. *See, e.g.*, Regions Bank v. Joyce Meyer Ministries, Inc., 15 N.E.3d 545, 549 (Ill. App. Ct. 2014).

231. *See* A.W. v. Lancaster Cnty. Sch. Dist. 0001, 784 N.W.2d 907, 916–18 (Neb. 2010) (adopting RESTATEMENT (THIRD) OF TORTS approach of removing foreseeability from duty analysis and considering it in context of whether the defendant breached his duty, i.e., acted reasonably).

dant's conduct or omission was a breach of duty or, stated differently, a failure to exercise reasonable care, is considering what a reasonable and prudent parent would have done.[232]

Applying this test to Lanzas' and the Loughners' cases, a jury could reasonably find that a breach occurred. In Nancy's case, a reasonable and prudent parent would not give her mentally ill, adult son unlimited access to a military-style assault rifle and other firearms and most certainly would not take him to the shooting range or give him funds to purchase his own gun. A reasonable and prudent parent would also not leave a son they had recently planned to have involuntarily committed alone and unsupervised for a whole weekend with her guns,[233] as it is foreseeable that he could use those weapons to harm himself or someone in the general public. What a reasonable and prudent parent would have done in Nancy's case is completely deny their mentally ill son access to their guns by safely storing the weapons or not having guns at all in the house. A reasonable and prudent person would not have left their severely mentally ill child—who had not left the house in months—alone and unsupervised. A reasonable and prudent person would have sought immediate mental health assistance for their child, even involuntary commitment, before leaving town. At the very least, Nancy could have given Adam an ultimatum, as some mental health professionals have recommended, to seek help or get out.[234]

Likewise, a jury could find that the Loughners did not exercise reasonable care. A reasonable and prudent parent in their case would have forced their child to obtain an evaluation from a mental health care professional upon their child being suspended from college for displaying mentally disturbing behavior.[235] The Loughners themselves witnessed Jared's profound, yet professionally undiagnosed, mental illness when they observed him engaging in incomprehensible and nonsensical rants and being nonverbal.[236] Despite being concerned about his erratic and angry behavior,[237] they never took the community college's advice to have Jared evaluated.[238] One could argue that after Jared refused to be evaluated, a

232. *See* Doe Parents No. 1 v. Haw. Dep't of Educ., 58 P.3d 545, 593 (Haw. 2002).

233. *See id.*

234. *See* Kern, *supra* note 214; Park & Landau, *supra* note 214.

235. *See* Martin & Siegel, *supra* note 104; *see also* Bob Orr, *Newly Released Jared Lee Loughner Files Reveal Chilling Details*, CBS NEWS (Mar. 27, 2013, 7:12 PM), http://www.cbsnews.com/news/newly-released-jared-lee-loughner-files-reveal-chilling-details/ [https://perma.cc/UXM5-R9AH].

236. *See* Dan Barry, *Looking Behind the Mug-Shot Grin*, N.Y. TIMES, Jan. 15, 2011, http://www.nytimes.com/2011/01/16/us/16loughner.html?_r=0 [perma.cc/623M-UDJA]; Loughner CBS Article, *supra* note 112.

237. *See* Loughner CBS Article, *supra* note 112 (explaining how parents were so disturbed by Jared's behavior they believed he was on drugs and tested him for narcotics but test came back negative); Orr, *supra* note 235.

238. *See* Orr, *supra* note 235.

reasonable parent would have had him involuntarily evaluated by a hospital or committed to a mental health healthcare facility, just as Nancy Lanza should have done in Adam's case. Alternatively, they could have given him an ultimatum to get help or leave because it was foreseeable that someone with untreated mental illness would harm others.

V. CAUSATION

Establishing a breach of duty is one thing, proving a causal connection between the defendant-parents' breach of duty and the resulting harm, is quite another. Like the issue of whether there was a breach, the question of whether the defendant's breach caused the complained of harm is a question of fact for the jury that calls for a foreseeability determination.[239]

Jurisdictions vary on what test should be applied to ascertain whether the defendant's negligent act caused the complained of injury. Traditionally, most states require the plaintiff to establish that the defendant's conduct was *both* (1) the cause-in-fact of the complained of injury and (2) the proximate or legal cause.[240] In ascertaining whether the defendant's conduct was the cause-in-fact of the plaintiff's injury, most jurisdictions employ the "but-for" test.[241] In other words, would the mass shooting have occurred but for the defendant-parents' negligence? "The but-for test [only] applies in cases where only one negligent act is at issue."[242] If more than one negligent act is at issue, then the "substantial factor" test applies, which considers if it is more likely than not that the defendant's conduct was a substantial factor in producing the plaintiff's injuries.[243] Some states, however, have used the substantial factor test when only one negligent act is at issue, adding to analytical confusion surrounding this element.[244]

Regardless of what test is used for cause-in-fact, a jury could find that Nancy Lanza's and the Loughners' breach of duty was the cause-in-fact for their children's mass shootings. As to Nancy Lanza, but-for her giving her son unfettered access to her firearms, the mass shooting would not have occurred. Adam used her .223-caliber Bushmaster Model XM15 semi-automatic firearm, known as a "killing machine," to commit his heinous

---

239. *See* Regions Bank v. Joyce Meyer Ministries, Inc., 15 N.E.3d 545, 549 (Ill. App. Ct. 2014).

240. *See, e.g.*, Pittway Corp. v. Collins, 973 A.2d 771, 786–87 (Md. Ct. App. 2009); Fisher v. Swift Transp. Co., 181 P.3d 601, 609 (Mont. 2008); Baggerly v. CSX Transp., Inc., 635 S.E.2d 97, 101 (S.C. 2006); s*ee also* Weigand, *supra* note 158, at 62.

241. *See* Michael Panella, *Problematic Legal Causations of Death*, 44 TENN. B.J., Feb. 2008, at 21, 22; *see also Baggerly*, 635 S.E.2d at 101.

242. *See Pittway Corp.*, 973 A.2d at 786–87 (internal quotation marks omitted).

243. *See id.*

244. *See* David Jakubowitz, *"Help, I've Fallen and Can't Get Up!": New York's Application of the Substantial Factor Test*, 18 ST. JOHN'S J. C.R. & ECON. DEV. 593, 606 (2004).

acts.[245] Under federal law, you must be at least twenty-one years old to legally purchase any firearm other than a rifle or shotgun,[246] and Adam was only nineteen years old. Thus, Adam could not have legally purchased the military-style weapon he used on most of the victims on his own. Also, but for Nancy Lanza leaving Adam unsupervised and failing to have him involuntarily committed to a mental institution, he would not have had the opportunity to make a practice run to Sandy Hook and commit the murders.

Regarding Jared Loughner, arguably but-for his parents' failure to have him evaluated and treated by a mental health professional, he would not have committed the murders. After Loughner was charged with the mass shooting, he was diagnosed with schizophrenia.[247] Studies reveal people with serious mental illnesses like schizophrenia who do not take their medication are more dangerous than the general population.[248] A jury could find the Loughners' failure to ensure their son received treatment for his mental illness was the cause-in-fact for the mass shootings.

Once a plaintiff proves that the defendant's conduct was the cause-in-fact of the injury, the fact-finder must move to the second part of the analysis—whether the defendant's conduct was the proximate or legal cause of the injuries, which is based upon whether the harm was foreseeable.[249] The foreseeability mandated here asks if, in retrospect, the injury emanating from the act or omission should have been reasonably anticipated.[250] In other words, the inquiry focuses on whether the injury was a natural

---

245. *See* Lee Ferran & Shushannah Walshe, *Newtown Massacre: What Is a Bushmaster .223?*, ABC NEWS (Dec. 17, 2012), http://abcnews.go.com/Blotter/newtown-massacre-bushmaster-223/story?id=18000884 [https://perma.cc/3PH5-MCWZ] (internal quotation marks omitted); Edgar Sandoval & Corky Siemaszko, *Inside Adam Lanza's Lair: Newtown Shooter Plotted Own Death, Obsessed with Columbine Shooting, Report Details*, DAILY NEWS (Nov. 26, 2013, 4:06 AM), http://www.nydailynews.com/news/national/newtown-shooter-planned-death-obsessed-columbine-article-1.1528626 [https://perma.cc/TW6Y-MYH7].

246. *See* 18 U.S.C. §922(b)(1) (2012).

247. *See* Katie Moisse, *Jared Loughner's Defense Team May Be Looking for Mental Illness in His Relatives*, ABC NEWS (Aug. 19, 2011), http://abcnews.go.com/Health/MindMoodNews/jared-loughner-mental-illness-runs-families/story?id=14332522 [https://perma.cc/7QL9-RMBS].

248. *See* Katherine B. Cook, Note, *Revising Assisted Outpatient Treatment Statutes in Indiana: Providing Mental Health Treatment for Those in Need*, 9 IND. HEALTH L. REV. 661, 669–71 (2012) (stating several studies suggest that one important predictor of future violence among severely mentally ill is non-adherence to medication); *see also* E. Fuller Torrey, *Violence and Schizophrenia*, 88 SCHIZOPHRENIA RES. 3 (2006) (explaining that severely mentally ill persons, like schizophrenics, who do not take their medication are more dangerous than people in general public); E. Fuller Torrey, *Violence and People with Mental Illness*, MENTAL ILLNESS POL'Y ORG. (1994), *available at* http://mentalillnesspolicy.org/consequences/violence-statistics.html [https://perma.cc/6CSQ-SLYB] (same).

249. *See* Pittway Corp. v. Collins, 973 A.2d 771, 787–88 (Md. Ct. App. 2009); Baggerly v. CSX Transp., Inc., 635 S.E.2d 97, 101 (S.C. 2006).

250. *See Baggerly*, 635 S.E.2d at 101.

and probable consequence of the defendant's conduct.[251] Stated differently, the harm or injury cannot be a "freakish" consequence of the defendant's act, making the victim of the harm unforeseeable.[252] The landmark case of *Palsgraf v. Long Island Railroad Company*,[253] clearly makes this point.

In *Palsgraf*, the court held that a passenger who was waiting for a train on the other side of the platform was not a foreseeable victim of the railway guard's act of helping a passenger board a train by pushing him.[254] Unbeknownst to the guard, the passenger was holding a package of fireworks that was covered in newspaper.[255] The passenger dropped the package, the fireworks fell on the rails and exploded, and the explosion threw down some scales on the other end of the platform that struck and injured the plaintiff.[256] The court essentially held that the woman standing far away was not a foreseeable victim; if the guard committed a wrong, it was committed against the passenger holding the package of fireworks.[257]

Applying *Palsgraf* and the other causation principles to the Lanza and Loughner cases, a jury could arguably find that random people being shot by a mentally ill child was a natural and probable consequence of giving that child a gun in Nancy Lanza's case and of leaving their child's mental illness untreated in both cases. Guns are "inherently dangerous instrumentalit[ies], the use of which is likely to produce death or serious injury."[258] Firearms are even more dangerous when left in the hands of a mentally unstable person. Indeed, that is why there are both federal and state laws aimed at precluding mentally ill persons from possessing firearms,[259] laws the United States Supreme Court has recognized place a reasonable restriction on the Second Amendment right to bear arms.[260] Thus, the shooting of innocent victims by a mentally ill, adult child is a natural and probable consequence of parents of that child failing to adequately supervise or control the child by not preventing them from accessing firearms and of not forcing their child to obtain mental health treatment.

---

251. *See id.*

252. *See* Lee S. Kreindler et al., *Foreseeability as a Factor in Proximate Cause*, 14 N.Y. PRAC. SERIES, N.Y. L. TORTS § 8:8 (West 2015).

253. 162 N.E. 99 (N.Y. 1928).

254. *See id.* at 340–41.

255. *See id.* at 341.

256. *See id.*

257. *See id.*

258. Byers v. Hubbard, 669 N.E.2d 320, 322 (Ohio Ct. App. 1995); *see also* Kitchen v. K-Mart Corp., 697 So. 2d 1200, 1202 (Fla. 1997).

259. *See* Lewis, *supra* note 2, at 341–47.

260. *See* D.C. v. Heller, 554 U.S. 570, 626 (2008) (stating "nothing in our opinion should be taken to cast doubt on longstanding prohibitions on the possession of firearms by . . . the mentally ill").

On the other hand, both parties could argue that the victims of the shootings were unforeseeable. In Nancy Lanza's case, who could predict that her son would shoot children and school personnel at an elementary school he did not attend? In the Loughners' case, who could foresee that their son would shoot Congresswoman Giffords and her constituents waiting to see her at the grocery store?[261] Again, as one court has recognized, albeit in the context of determining foreseeability in the duty context, it is foreseeable that a member of the general public could be a victim of a severely mentally ill person with a gun.[262] Moreover, this is not a case where the perpetrators committed their mass shootings hundreds of miles away from where they lived; the crimes occurred in their own communities. Therefore, a jury could find that the victims were foreseeable and the causation element is met.

Additionally, in cases involving more than one negligent act, like mass shooting cases, an extra analytical step is required. Specifically, the fact-finder must inquire if an independent, intervening negligent act relieves the defendant of liability.[263] An "independent, intervening" negligent act is "a force that comes into motion after the defendant's negligent act, and combines with the negligent act to cause injury to the plaintiff."[264] Significantly, if the intervening act was foreseeable, then it does not break the causal chain between the defendant's conduct and the injury.[265] If the intervening act is unforeseeable because it is so remote or extraordinary that it could not have been reasonably anticipated, it is a superseding cause that breaks the causal link between the defendant's conduct and the injury—therefore, absolving the defendant(s) from liability.[266] Whether an intervening act is foreseeable is also a question of fact that should be left to the jury's province.[267] Third-party criminal conduct is a superseding cause unless it should have been reasonably anticipated.[268]

In the Lanza and Loughner cases, it should have been reasonably anticipated that the adult children would shoot someone with a gun if their parents did not adequately monitor, supervise, or control their behavior to

261. After the shootings, material indicating Jared's hatred and obsession with Giffords was found in his safe, but it is unknown whether his parents were aware of this material. In fact, Jared's father said that he came across some of Jared's writing in a journal and found it indecipherable.

262. *See* Silberstein v. Cordie, 474 N.W.2d 850, 856 (Minn. Ct. App.), *rev. granted in part, cause remanded by* 477 N.W.2d 713 (Minn. 1991).

263. *See* Pittway Corp. v. Collins, 973 A.2d 771, 788–89 (Md. Ct. App. 2009).

264. Fisher v. Swift Transp. Co., 181 P.3d 601, 609 (Mont. 2008).

265. *See id.*

266. Paul M. Coltoff et al., *Proximate Causation*, 65 C.J.S. NEGLIGENCE § 223 (West 2015).

267. *See Fisher*, 181 P.3d at 610.

268. *See* Spears v. Coffee, 153 S.W.3d 103, 106 (Tex. App. Ct. 2004); *see also* RESTATEMENT (SECOND) OF TORTS § 448 (1965). It appears that an *intervening* cause is interchangeable with *superseding* causes.

prevent them from accessing firearms. Thus, a jury could find in these cases and future similar cases that the causation element was satisfied.

VI. CONCLUSION

In sum, it is clear that pursuant to general negligence principles, some parents can be found negligent and civilly liable for an act or omission that results in their adult child taking a gun and committing a mass killing. In determining whether a parent acted reasonably, common sense should prevail. Any reasonable and prudent parent knows, or should know, that their mentally ill child, regardless of their age, should not have access to a firearm, as an armed, mentally unstable person poses a serious risk to the public at large as well as to themselves. Thus, if a parent is going to assume responsibility for supervising an adult, mentally ill child, they have a legal obligation to exercise reasonable care.

Exercising reasonable care includes not having firearms in a house where a mentally ill person resides, or at the very least, safely storing firearms. It also entails monitoring your mentally ill child and being on the lookout for anything that might suggest they are planning to harm others, such as violent writings, Internet postings, manifestos, ammunition, and guns. If a parent believes their child poses an immediate threat to the public, they should contact the proper authorities.[269] Indeed, if a specific threat is made, they must contact the police upon pain of possibly being sued.[270] Additionally, a reasonable, prudent parent will also insist that their child obtain mental health assistance, and if the child refuses, the parents should do everything they can to make sure their child receives the professional help needed, including threatening to no longer support the child if they do not seek help, taking away their car keys and other privileges to encourage compliance, and initiating involuntary commitment to a mental institution when all else fails.[271]

Parenting is unquestionably the most difficult and important job on earth, which is probably why courts have historically shied away from holding parents liable for their children's actions. However, the law has long been involved in judging parents and penalizing them for poor parental decisions. For instance, courts have evaluated whether a parent's use of corporal punishment is reasonable and excessive, and if it is found to be excessive, a parent can be civilly liable or lose custody of their child.[272] A

269. *See* Hutchison, *supra* note 110 (explaining how psychiatrists advise family members to alert proper authorities if their loved one is committing worrisome behavior or has made specific threat of violence).

270. *See id.*

271. *Helpful Tips for Families*, *supra* note 214 (advising that "[i]t may be necessary to push your relative into treatment in spite of his angry response").

272. *See, e.g.*, Gonzalez v. Santa Clara Cnty. Dep't of Soc. Servs., 167 Cal. Rptr. 3d 148, 160 (Ct. App. 2014) (noting "parents are privileged to administer reasonable punishment with impunity, but the parent who exceeds that limit . . . commits a battery and is civilly liable for the consequences" (alteration in original) (inter-

mother was arrested for leaving her nine-year-old child in the park unsupervised while she worked at McDonald's.[273] Accordingly, the law should expect a certain standard of care from parents of adult children with mental illness and penalize them when they do not act reasonably with regard to that child if that child lives in the home and they have assumed a de facto guardianship over them. It not only takes a village to raise a child, but a village to stop a mass killing,[274] and the most important villager is a parent—our first line of defense.

---

nal quotation marks omitted)); *In re* T.A., 663 N.W.2d 225, 229 (S.D. 2003) (affirming trial court's decision to remove special needs child, who was spanked with belt and received bruises, from home because there was sufficient evidence child was abused and neglected).

273. *See* Kelly Wallace, *Mom Arrested for Leaving 9-Year Old Alone at Park*, CNN (July 21, 2014, 4:41 PM), http://www.cnn.com/2014/07/21/living/mom-arrested-left-girl-park-parents/ [https://perma.cc/R4DH-ETUN].

274. *See* O'TOOLE, *supra* note 201 (discussing letter from Attorney General Janet Reno noting that federal government agencies have shown that "if communities, schools, government and other key players pull together to address the roots of violence, we can make America safer for our children").

BRANDS AS COPYRIGHT

JESSICA M. KISER*

INTRODUCTION

SOME trademarks inspire love. There are people who wear rain coats covered in pictures of McDonald's Big Mac sandwich.[1] Some join clubs where members are expected to have Harley Davidson's logo tattooed on their bodies.[2] Some even choose to be buried in a coffin emblazoned with NASCAR imagery.[3] While it is possible to dismiss these individuals as both rare and absurd, this level of devotion by consumers makes it clear that some trademarks have fans for reasons that surpass the quality of the product or service offered. Numerous fan groups exist on websites and social media to provide consumers with the opportunity to share their love and commitment to a trademarked product or service.[4] Communities can develop around trademarks online or in the real world in much the same way that they develop around shared geographical or cultural interests. This development of fervent fandom around a trademark, including the creation of active trademark-related social communities, demonstrates a shift in business management such that trademarks are now only one element of a larger brand development strategy. While consumers may coalesce around the trademark as a unifying symbol of

---

* Assistant Professor of Law, Gonzaga University School of Law. J.D., Columbia Law School; B.S. and B.A., Boston University. This Article benefitted greatly from the comments and insights provided by the participants of the Rocky Mountain Junior Scholars Forum, the Inland Northwest Scholars Workshop, and the Junior Scholars Virtual Colloquium. Thanks also to Sean Wright for his tireless support and to Anjali Bhatt, Garrett Wilson, Joshua Grandinetti, and Charles Allen for their research and editorial assistance.

1. *See* Victoria Taylor, *McDonald's Launches Big Mac Fashion and Lifestyle Line*, N.Y. DAILY NEWS (Mar. 26, 2015, 10:09 AM), http://www.nydailynews.com/life-style/eats/mcdonald-launches-big-mac-fashion-lifestyle-collection-article-1.2163002 [https://perma.cc/97F2-KYDF].

2. *See* Nisha Patel, *35 Groovy Harley Davidson Tattoos*, SLODIVE, http://slodive.com/inspiration/harley-davidson-tattoos/ [https://perma.cc/5PRV-8ZG7] (last visited Jan. 11, 2016).

3. *See generally 15 Weird and Wacky Coffins*, CREMATION RESOURCE, http://www.cremationresource.org/cremation/15-weird-and-wacky-coffins.html [https://perma.cc/TG56-L996] (last visited Mar. 1, 2016) (showing several different coffins decorated with trademarks associated with Apple iPhones, Super Nintendo controllers, and Nokia cell phones); *Who Are Funerals Really For?*, TWO VIEWS, http://www.two-views.com/article_funerals.html [https://perma.cc/G7S6-YSUD] (last visited Mar. 1, 2016).

4. *See* Deborah R. Gerhardt, *Social Media Amplify Consumer Investment in Trademarks*, 90 N.C. L. REV. 1491, 1506 (2012) ("Social media have empowered consumers to use trademarks to gain more information, make a broader expressive impact and connect to a larger, more dispersed community.").

their fandom, they are really identifying with the brand story or brand personality that has developed in connection with that mark.

Because brands incorporate one or more trademarks, legal disputes involving brands have been addressed as trademark disputes. Trademark law has actually evolved and expanded to attempt to accommodate this larger concept of brands. Unfortunately for the brand owner seeking to cultivate loyal brand fans, brands have not fit smoothly into the strictures of trademark law. Trademark laws aiming to prevent consumer confusion and counterfeiting have inadvertently led to the stifling of non-competitive brand activities by consumers. As a result of this disconnect in the law, trademark owners may feel that they are required to stop consumer use of their marks, despite marketing scholarship that suggests consumer-initiated brand communities and other fan activities are the best way of building brand loyalty and future sales. For example, the official website for the LEGO Group (corporate owner of the LEGO brand line of toys and, relatedly, the LEGO trademark) enjoins consumer fans from using the LEGO trademark and insists that "the LEGO logo NEVER be used on an unofficial web site" or as part of an Internet address for such a website.[5] In 2003, the LEGO Group sent a cease-and-desist letter to just such a fan-operated website that ultimately resulted in the site being shut down.[6]

More recently, IKEA's general counsel sent a cease-and-desist letter to the operator of the website www.IKEAhackers.net.[7] The website was created in 2006 to collect and share ideas for ways to modify or "hack" products sold by IKEA, an international retailer known for affordable and minimalist furniture and home goods.[8] Submissions collected by the website operator (or submitted by users) ranged from minor embellishments of standard IKEA products to complete redesigns of an IKEA product to turn it into a more customized piece of furniture.[9] The website developed a large and enthusiastic fan base. The fact that IKEA inspired this successful consumer-created website and allowed it to flourish without corporate involvement was touted as evidence of IKEA's genius brand management strategy in marketing literature.[10] However, IKEA's recent cease-and-desist letter alleged that both the website domain name and the hacks

5. *Fair Play*, LEGO, http://www.lego.com/en-US/legal/legal-notice/fair-play [https://perma.cc/42L9-W3DN] (last visited Mar. 1, 2016).

6. *See* Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 CARDOZO L. REV. 1789, 1840 (2007) (describing cease-and-desist letter for website "located at www.ratemylego.com" (internal quotation marks omitted)).

7. *See* Jules Yap, *Big Changes Coming to IKEAHackers*, IKEAHACKERS (June 14, 2014), http://www.ikeahackers.net/2014/06/big-changes-coming-to-ikeahackers.html [https://perma.cc/2JT4-JPQ6].

8. *See About*, IKEAHACKERS, http://www.ikeahackers.net/about [https://perma.cc/TFU8-48EF] (last visited Mar. 1, 2016).

9. *See IKEA Threatens IKEAHacker Website with Cease-and-Desist Notice*, NEWS.COM.AU (June 18, 2014, 7:53 AM), http://www.news.com.au/finance/business/ikea-threatens-ikeahacker-website-with-ceaseanddesist-notice/story-fnkgdftz-1226957658920 [https://perma.cc/5K3V-HBCP].

10. *See* SARAH ROBINSON, FIERCE LOYALTY: UNLOCKING THE DNA OF WILDLY

posted on the site violated IKEA's intellectual property rights.[11] The operator of the website initially announced that the website would likely be shut down.[12] IKEA customers were livid and sent emails, letters, and tweets to IKEA's corporate offices, and demanded that it reconsider.[13] A month later, IKEA withdrew its opposition to the website, which then resumed operation.[14]

In both of the trademark-related instances described above, the websites and communities that were threatened by trademark owners were ones that were started by fans. Participation in these fan-initiated websites encouraged and supported use of the branded products sold under the respective trademarks. One might presume that trademark owners would love and support this free publicity. Advertising and marketing firms encourage such fan-initiated efforts and often call brand communities the marketing platform of the future.[15] After receiving IKEA's cease-and-desist letter, the owner of the IKEAHackers website stated, "IKEAhackers.net was set up in 2006 and truly not with the intent to exploit their mark. I was [just a] crazy fan. In retrospect, a naïve one too."[16] In a time when trademark owners are routinely called "bullies"[17] and mocked online and

SUCCESSFUL COMMUNITIES 59–61 (2012) (describing Ikea's decision to allow site to proliferate as intelligent from business standpoint).

11. *See* Drew Prindle, *Update: IKEA May Back Off on IKEAHackers Cease and Desist, Will Discuss New Solutions*, DIGITAL TRENDS (June 20, 2014), http://www.digitaltrends.com/home/ikea-issues-cease-desist-order-long-running-fan-blog-ikeahackers-net/ [https://perma.cc/XSD6-VLNU].

12. *See* Yap, *supra* note 7.

13. *See* Cory Doctorow, *Ikea Bullies Ikeahackers with Bogus Trademark Claim*, BOINGBOING (June 15, 2014, 10:26 AM), https://boingboing.net/2014/06/15/ikea-bullies-ikeahackers-with.html [https://perma.cc/N4G8-PLH5]; *see also* Kelsey Campbell-Dollaghan, *Why Ikea Shutting Down Its Most Popular Fan Site Is a Giant Mistake,* GIZMODO (June 16, 2014, 12:05 PM), http://gizmodo.com/why-ikea-shutting-down-its-most-popular-fan-site-is-a-g-1591401344 [https://perma.cc/4XEW-LCZ6]; Jennifer Karmon, *After Global Outcry, IKEA Softens Stance Against Superfan*, YAHOO! HOMES (June 18, 2014, 5:15 PM), https://homes.yahoo.com/blogs/spaces/ikea-threatens-legal-action-against-ikea-fan-s-8-year-old-site-230646533.html?nf=1 [https://perma.cc/82TS-7KG3].

14. *See* Jules Yap, *Inter IKEA Systems BV Called Me!*, IKEAHACKERS (June 19, 2014), http://www.ikeahackers.net/2014/06/inter-ikea-systems-bv-called-me.html [http://perma.cc/Q89K-ZMBJ] (discussing IKEA's withdrawal).

15. *See generally* Roger Katz, *2015 Will Be the Year of the Brand Community—Here's Why*, CLICKZ (Jan. 23, 2015), http://www.clickz.com/clickz/column/2391666/2015-will-be-the-year-of-the-brand-community-here-s-why [https://perma.cc/5UKU-L6F4]; *see also* Kerry O'Shea Gorgone, *Why Brand Communities Are the Future of Marketing: Jordan Kretchmer of Livefyre on Marketing Smarts [Podcast]*, MARKETING PROFS (Feb. 18, 2015), http://www.marketingprofs.com/podcasts/2015/27071/brand-communities-jordan-kretchmer-marketing-smarts [http://perma.cc/XF7R-WF64].

16. *See* Yap, *supra* note 7.

17. *See* Steve Baird, *Minnesota's Legislative Answer to "Trademark Bullying"?*, DUETSBLOG (Apr. 9, 2012), http://www.duetsblog.com/2012/04/articles/trademarks/minnesotas-legislative-answer-to-trademark-bullying/ [http://perma.cc/3T46-LF5A]; *see also* Leah Chan Grinvald, *Shaming Trademark Bullies*, 2011 WIS. L.

in the press for their aggressive responses to minor trademark infringements, it can be hard to understand why they would want to threaten their own fans publicly.[18] However, such actions are a result of the uneasy way in which brands have been shoehorned into trademark law. This Article argues that brands contain expressive, creative content, and this creative content encourages a response by consumer fans. As such, brands would be better served under a copyright regime. Brand fan activities and brand communities are very similar to the fanfiction and fanfiction communities created in connection with works protected under copyright. Whether they love fanfiction or hate it, copyright owners are supported by copyright law. They can stop fanfiction by challenging it as an unlicensed derivative work, or they can ignore it and remain confident that their copyright remains unscathed. "Super fans" can also structure their activities to attempt to fall under the infringement protection of copyright fair use.

Trademark law does not offer this freedom. Trademark law has expanded over the past several decades to provide increasing protection for trademark owners and few avenues for expressive uses of trademarks by non-owners. A brand-theory approach to trademark law can account for this expansion by acknowledging the substantial investment that trademark owners make in developing a brand.[19] However, that is only half of the brand story. Modern branding invites consumer involvement, and marketing literature suggests that brand communities are the future of advertising. Trademark law's focus on the brand owner apart from the consumer ignores this push for consumer involvement in branding. Because of this disconnect, trademark law requirements are ill-suited to address brand development, and this Article advocates for a copyright-centered approach to brands. Brands are creative works deserving of copyright protection. The boundaries of copyright and the discretion granted to copyright owners better address the reality of consumer involvement in the creation of derivative brand works.

REV. 625, 642; Jessica M. Kiser, *To Bully or Not to Bully: Understanding the Role of Uncertainty in Trademark Enforcement Decisions*, 37 COLUM. J.L. & ARTS 211, 218 (2014) (describing instance in which company was "shamed" via social media for alleged trademark bullying).

18. *See* Timothy Geigner, *Water Company Goes Trademark Bully on Graffiti Activists over Hashtag*, TECHDIRT (May 14, 2015, 1:44 PM), https://www.techdirt.com/articles/20150511/13595230964/water-company-goes-trademar-bully-graffiti-activists-over-hashtag.shtml [https://perma.cc/KHK7-MZSG] (discussing bullying by the water company); Trevor Little, *Monster Energy Company Heads Trademark Bullies Top 10, but List Is Not All It Seems*, WORLD TRADEMARK REV. (May 27, 2015), http://www.worldtrademarkreview.com/Blog/detail.aspx?g=7a58f3aa-6f8a-4f54-a544-c04b76539d57 [https://perma.cc/N5CV-2MZZ] (listing top ten trademark bullies and victims for 2014).

19. *See generally* Deven R. Desai, *From Trademarks to Brands*, 64 FLA. L. REV. 981 (2012) [hereinafter Desai, *From Trademarks to Brands*] (describing brand-centered approach to trademark law).

Part I of this Article explains the theoretical justifications underlying both copyright and trademark law and the evolution of trademark law over the twentieth century. Part II provides background regarding marketing research on branding, as well as how brand theory can explain the expansion of trademark law as it has attempted to recognize the substantial investments made by trademark owners engaged in branding. It also explores the ways in which trademark law conflicts with the goals of brand owners in their brand development efforts, as well as the current gap in trademark scholarship regarding the role of the consumer in the development of brands. Part III explores the similarities between brands and copyrightable works and discusses how brands may be better viewed under copyright law as something akin to creative works and their derivative fanfiction.

I. FOUNDATIONS FOR PROTECTION

In order to understand why fan activity is treated differently under copyright law and trademark law, it is crucial to trace the theoretical basis and historical development of each legal framework. While trademark law has long focused on the protection of consumers while denying explicit property rights in trademarks to their owners, copyright law began and continues as a protection granted to the copyright owner.

A. *Copyright Protection*

Federal copyright law in the United States stems from Article I, Section 8 of the United States Constitution, which states that "[t]he 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."[20] In *The Trade–Mark Cases*,[21] the United States Supreme Court defined the terms *authors* and *writings* as they pertained to the constitutional scope of copyright protection; for a work to receive copyright protection as a "writing" created by an "author," the Court determined that "originality is required."[22] It is the originality that justifies the monopolistic rights authorized under copyright law: "[W]hile the word *writings* may be liberally construed, as it has been, to include original designs for engravings, prints, [etc.], it is only such as are *original*, and are founded in the creative powers of the mind."[23]

Therefore, original and creative works are rewarded with copyright protection, but courts have clarified that "[t]he copyright law, like the patent statutes, makes reward to the owner a secondary consideration."[24] While the rights conferred by copyright are to be measured and balanced

20. U.S. CONST. art. I, § 8, cl. 8.
21. *In re Trade-Mark Cases*, 100 U.S. 82 (1879).
22. *See id.* at 94.
23. *Id.* (emphasis added).
24. *See* United States v. Paramount Pictures, Inc., 334 U.S. 131, 158 (1948).

by Congress to "assure contributors to the store of knowledge a fair return for their labors,"[25] the true focus of the constitutional mandate establishing a federal copyright law is distributing original, creative works to the public. In *Sony Corp. v. Universal City Studios*,[26] the Supreme Court stated that:

> [T]he limited grant [under copyright law] is a means by which an important public purpose may be achieved. It is intended to motivate the creative activity of authors and inventors by the provision of a special reward, and to allow the public access to the products of their genius after the limited period of exclusive control has expired.[27]

In the broadest sense, copyright law aims to create an environment of expansion and creativity for the benefit of the consumer by using limited copyright ownership as an incentive for creators.

B. *Trademark Protection*

Whereas inventors and authors receive patent and copyright protection in order to reward and encourage their contributions to the progress of "Science and [the] useful Arts,"[28] trademark owners are individuals or businesses seeking protection for commercial activities, a seemingly less noble pursuit. Jurists and scholars have long distinguished trademarks from the creativity associated with patents and copyrights by suggesting that trademarks are mere menial symbols that do not "depend upon novelty, invention, discovery, or any work of the brain" and that require "no fancy or imagination, no genius, no laborious thought."[29] Perhaps due to this perceived divide between the banal commercial world of trademarks and the creativity and innovation ascribed to works protected by copyrights and patents, trademarks have been described as "the overlooked stepchild of the world of intellectual property goods."[30] Trademarks are seen as a tool of commerce, and their protection is typically explained in terms of economic efficiency.[31] A trademark is a word or symbol used "to identify and distinguish" the goods or services of a person from those of another.[32] J. Thomas McCarthy has identified four primary functions of trademarks that justify their legal protection:

---

25. *See* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 546 (1985) (citing Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975)).

26. 464 U.S. 417 (1984).

27. *Id.* at 429.

28. *See* U.S. CONST. art. 1, § 8, cl. 8.

29. *See In re Trade-Mark Cases*, 100 U.S. at 94.

30. *See* Sonia K. Katyal, *Trademark Intersectionality*, 57 UCLA L. REV. 1601, 1613 (2010).

31. *See* WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 167–68 (2003).

32. Lanham Act § 45, 15 U.S.C. § 1127 (2012) (defining *trademark*); *see also* 1 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 3:1

(1) To identify one seller's goods and distinguish them from goods sold by others;

(2) To signify that all goods bearing the trademark come from or are controlled by a single, albeit anonymous, source;

(3) To signify that all goods bearing the trademark are of an equal level of quality; and

(4) As a prime instrument in advertising and selling the goods.[33]

A trademark serving these purposes will be protected "against confusingly similar commercial use."[34]

Scholars and jurists have long argued that trademark protection is essential to the marketplace; trademarks reduce consumer search costs because buyers and sellers can use trademarks as shorthand to refer to the source and consistent quality of goods and services.[35] For example, imagine a consumer looking to buy running shoes for the first time. That consumer may ask friends for recommendations and spend time reading reviews of the various running shoes available before deciding to purchase a pair of shoes sold under the NEW BALANCE trademark. If the consumer is pleased with the fit and quality of the shoes, the consumer may then seek out another pair of shoes sold under the NEW BALANCE mark in the future without having to incur the same research or search costs inherent in the initial purchase. It has been argued that the use of trademarks "makes effective competition possible in a complex, impersonal marketplace by providing a means through which the consumer can identify products which please him and reward the producer with continued patronage."[36] Trademark law accomplishes this task by preventing businesses from using marks that are likely to confuse consumers about the source of the goods or services being offered.

The likelihood of confusion test is the "basic test of both common-law trademark infringement and federal statutory trademark infringement."[37] When evaluating whether a defendant has infringed the plaintiff's trademark under the Lanham Act (also known as the Trademark Act of 1946), the court must ask whether the defendant's use is "likely to cause confu-

---

(4th ed. 1998 & Supp. 2015). For the purposes of this Article, the words *mark* and *trademark* will be used interchangeably to refer to all forms of marks, registered or not, including service marks, collective marks, and certification marks. It is not necessary to highlight the distinctions between types of marks for the analysis herein.

33. 1 MCCARTHY, *supra* note 32, § 3:2 (footnotes omitted).

34. *Id.*

35. *See* LANDES & POSNER, *supra* note 31, at 167–68. However, other justifications have been proposed. *See, e.g.*, Mark P. McKenna, *The Normative Foundations of Trademark Law,* 82 NOTRE DAME L. REV. 1839 (2007) (arguing that trademark law historically focused on preventing trade diversion).

36. Smith v. Chanel, Inc., 402 F.2d 562, 566 (9th Cir. 1968).

37. 4 MCCARTHY, *supra* note 32, § 23:1 (footnote omitted).

sion, or to cause mistake, or to deceive . . . ."[38] Confusion is more ambiguous than one might expect, and this ambiguity has opened the door to a wide variety of infringement suits.

C. *The Historical Development of Trademark Law*

Trademark cases in the nineteenth and early twentieth centuries were interested in consumer confusion, as it was evidence of fraudulent intent on the part of the defendant.[39] At that time, the primary concern of trademark law was the unfair competitive practice of "passing off," which consisted of the "sale of the goods of one manufacturer or vendor as those of another."[40] The Supreme Court illustrated this focus in 1918 in *United Drug Co. v. Theodore Rectanus Co.*,[41] when it explained that a trademark's "function is simply to designate the goods as the product of a particular trader and to protect his good will against the sale of another's product as his."[42] The Court noted the lack of bad faith on the part of the defendant in this case, where the two parties simultaneously used the same mark in different locations prior to an overlap in markets that led the plaintiff to bring suit.[43]

This fraudulent intent was de-emphasized over time, and the attention was "shifted from the state of mind of the defendant to the state of mind of buyers."[44] This shift of focus from the defendant to the consumer may have paved the way for the modern expansion of trademark law—which now allows redress for noncompetitive uses of a mark and other less traditional claims—including dilution, sponsorship confusion,

---

38. *See* Lanham Act § 43(a), 15 U.S.C. § 1125(a) (2012); *see also* Lanham Act § 32, 15 U.S.C. § 1114(1).

39. *See* Graeme W. Austin, *Trademarks and the Burdened Imagination*, 69 BROOK. L. REV. 827, 840 (2004) (discussing historical shift from focus on conduct of plaintiffs to evidence of fraud on part of defendants).

40. *See* Lawrence Mfg. Co. v. Tenn. Mfg. Co., 138 U.S. 537, 546 (1891); *see also* Champion Spark Plug Co. v. Sanders, 331 U.S. 125, 131 (1947) (holding that injunction rather than damages is fair remedy for unfair competition given that "there has been no showing of fraud or palming off"); Hanover Star Milling Co. v. Metcalf, 240 U.S. 403, 419 (1916) (holding that owner of trademark used in specific geographical area only cannot enjoin good faith use of trademark by third party in different area when consumers in each area are only aware of their regional trademark as source of product), *superseded by statute*, Lanham Act (codified as amended in scattered sections of 15 U.S.C.), *as stated in* Park 'N Fly, Inc. v. Dollar Park & Fly, Inc., 469 U.S. 189 (1985).

41. 248 U.S. 90 (1918).

42. *See id.* at 97.

43. *See id.* at 101.

44. 1 MCCARTHY, *supra* note 32, § 5:2; *see also, e.g.*, Aunt Jemima Mills Co. v. Rigney & Co., 247 F. 407, 410 (2d Cir. 1917). In this early case, which focused on confusion, the court permitted the plaintiff to block the use of plaintiff's trademark in connection with a different type of good because the public might be confused about a connection between the two.

initial interest confusion, and post-sale confusion.[45] Each new basis for trademark infringement or dilution liability deviates from the traditional notion of actionable confusion. Mark McKenna has made a strong argument that "only confusion that affects purchasing decisions should be relevant to trademark law."[46] However, courts now "routinely say that trademark law targets 'confusion of any kind.' "[47] This broadened view of confusion is still explained as a natural effect of focusing on trademarks as search cost optimizers.[48] McKenna argues, "Anything that can be characterized in confusion-based terms seems to raise search costs, and if search costs are the harm to be avoided, then anything that causes confusion ought to be at least prima facie actionable."[49]

For example, sponsorship confusion is premised on the idea that consumers may be confused about whether a third-party product or service is sponsored or endorsed by the trademark owner.[50] A plaintiff may sue based on this confusion about the relationship between the parties, even where the relationship, or lack thereof, would have no impact on consumer purchasing decisions.[51] In initial interest confusion cases, the consumer is temporarily confused when the consumer seeks out the plaintiff's good or service, only to find instead the good or service offered by the defendant.[52] In an initial interest scenario, if the consumer purchases the

45. *See* Mark A. Lemley & Mark P. McKenna, *Owning Mark(et)s*, 109 MICH. L. REV. 137, 140 (2010) [hereinafter Lemley & McKenna, *Owning Mark(et)s*] ("[T]he claim that consumers are injured by the defendant's use of a mark in an unrelated market is implausible except under specialized circumstances, circumstances that trademark plaintiffs should have to prove.").

46. *See* Mark P. McKenna, *A Consumer Decision-Making Theory of Trademark Law*, 98 VA. L. REV. 67, 83 (2012) [hereinafter McKenna, *Consumer Decision-Making Theory*].

47. *See id.* at 69–70 (quoting Kos Pharm. v. Andrx Corp., 369 F.3d 700, 711 (3d Cir. 2004)); *see also id.* at 70 n.6 ("The Act is now broad enough to cover the use of trademarks which are likely to cause confusion, mistake, or deception of any kind, not merely of purchasers nor simply as to source of origin." (quoting *Kos Pharm.*, 369 F.3d at 711) (internal quotation marks omitted)).

48. *See id.* at 79 ("Because the search costs narrative is bound up with consumer confusion, this view of trademark law's purpose has manifested itself primarily in courts' fetishizing confusion and feeling compelled to respond whenever mark owners can characterize a defendant's use in confusion-based terms.").

49. *See id.* at 71 (discussing search costs and confusion-based theory of trademark law). McKenna argues that "courts should find trademark infringement only when the defendant's use of the plaintiff's trademark creates a risk that consumers will be deceived into buying goods or services they otherwise would not have or refraining from buying what they otherwise would have." *Id.* at 72.

50. *See* 4 MCCARTHY, *supra* note 32, § 23:8 (explaining sponsorship confusion); *see also, e.g.*, Champions Golf Club, Inc. v. Champions Golf Club, Inc., 78 F.3d 1111, 1116 (6th Cir. 1996) (describing relevant inquiry as whether golfer would be confused about affiliation between two golf clubs using same trademark).

51. *See Champions Golf Club*, 78 F.3d at 1116.

52. *See* 4 MCCARTHY, *supra* note 32, § 23:6; *see also, e.g.*, Brookfield Commc'ns, Inc. v. W. Coast Entm't Corp., 174 F.3d 1036, 1063–64 (9th Cir. 1999) (recognizing initial interest confusion); Elvis Presley Enters. Inc. v. Capece, 141 F.3d 188, 204 (5th Cir. 1998) (noting initial interest confusion may unfairly get customers to

defendant's product instead of continuing to seek out the plaintiff's product, the consumer is not confused at the time of purchase.[53] The confusion is only brief and fleeting. Post-sale confusion occurs when the consumer purchases a product (such as a pair of jeans with pocket embellishments designed to mimic those on a more expensive brand of jeans) with full knowledge that the product was sold under the defendant's trademark and not that of the plaintiff.[54] However, liability is premised on the idea that third parties will see the products after the time of the sale (being used by the purchaser, for example) and potentially be confused about the source of the good due to similarities between the plaintiff's and defendant's trademark or trade dress.[55] McKenna clarifies that in such cases, where the potential confusion does not affect the purchasing decision, "we can think of the confusion as generating search costs only if we think the mental act of wondering is the search cost."[56]

The focus on reducing search costs cannot account for "unbranded" advertisements like Procter & Gamble's "Thank You, Mom" campaign, which celebrated the role of mothers without explicitly referencing the company's branded products.[57] The purpose of this kind of marketing is to deliberately increase search costs, a tactic that seems incomprehensible under traditional trademark dogma. Perhaps trademark law's focus on

---

enter defendant's bar and stay even though customers realize upon entering that bar is not affiliated with Elvis Presley Enterprises).

53. *See* 4 MCCARTHY, *supra* note 32, § 23:6 (defining *initial interest confusion*).

54. *See, e.g.*, Mastercrafters Clock & Radio Co. v. Vacheron & Constantin-Le Coultre Watches, Inc., 221 F.2d 464, 466 (2d Cir. 1955) (recognizing post-sale confusion for first time); *see also* United States v. Torkington, 812 F.2d 1347, 1352 (11th Cir. 1987) (discussing applicability of likelihood of confusion test in post-sale context); Rolex Watch U.S.A., Inc. v. Canner, 645 F. Supp. 484, 493 (S.D. Fla. 1986) ("The fact that an immediate buyer of a $25 counterfeit watch does not entertain any notions that it is the real thing has no place in this analysis. Once a product is injected into commerce, there is no bar to confusion, mistake, or deception occurring at some future point in time.").

55. *See Mastercrafters Clock & Radio Co.*, 221 F.2d at 466 ("This goes to show at least that some customers would buy plaintiff's cheaper clock for the purpose of acquiring the prestige gained by displaying what many visitors at the customers' homes would regard as a prestigious article.").

56. *See* McKenna, *Consumer Decision-Making Theory*, *supra* note 46, at 84. McKenna goes on to argue, "After all, trademark law regulates the commercial marketplace; it is not an all-purpose remedy for having to think. There are, of course, sometimes costs associated with being confused more generally, but these costs do not harm consumers as consumers if they do not affect purchasing behavior." *See id.* at 85 (footnote omitted); *see also* Laura A. Heymann, *Naming, Identity, and Trademark Law*, 86 IND. L.J. 381, 441 (2011) ("Name or trademark changes that make it more difficult for others to retrieve information about the person or entity are not legally prohibited, even though such changes can result in increased search costs, and even though others may have been induced to act in a way in which they would not have acted if they had known about the person's or the company's history.").

57. *See* Jay Bolling, *Take the Brand Out of the "Branded vs. Unbranded" Conversation,* PM360 (June 16, 2014), http://www.pm360online.com/take-the-brand-out-of-the-branded-vs-unbranded-conversation/ [https://perma.cc/7B5R-HFT8].

search costs obscures instances in which consumers and trademark owners benefit from increased search cost. In fact, others have noted that "[f]ar from harming a brand by increasing search costs, practices that require consumers to exert cognitive effort benefit brands by more actively engaging consumers."[58] This paradoxical effect cannot be explained by traditional trademark doctrine, but the paradox is resolved if trademarks are treated as brands and analyzed through the lens of brand theory.

II. BRANDS AS TRADEMARKS

As trademark law has expanded, the role of trademarks for businesses has also changed and evolved. Trademarks now add substantial value to companies as stand-alone assets. For example, Forbes recently named Google as the most valuable trademark, with a value estimated at $44.3 billion.[59] AT&T came in as the tenth most valuable mark, with a value of $28.9 billion.[60] Such value is due in part to the increasing roles that advertising and marketing play in both creating consumer-source associations and creating the need or desire for a particular good or service in the mind of consumers. As Laura Heymann has explained, "[w]ith the development of advertising techniques starting in the 1920s, trademarks moved from functioning primarily at the point of sale (i.e., as a heuristic for repeat customers) to having a psychological effect on consumers well before the consummation of any actual sale and a lingering effect thereafter."[61] This shift in the importance of advertising and the function of trademarks corresponds to a shift in focus from trademarks to brands (of which trademarks are only components of a larger whole).

While a trademark can be described as a symbolic indicator of the source or quality of a product, the product's brand includes both the trademark and all the other information about the product presented to the marketplace by the trademark owner, including the product's packaging and the various forms of marketing materials produced to sell the product.[62] Stephen King, the former Director of WPP Group, the world's largest communications and marketing services company, once remarked:

58. *See* McKenna, *Consumer Decision-Making Theory, supra* note 46, at 91.

59. *See The 10 Most Valuable Trademarks*, FORBES, http://www.forbes.com/pictures/eidl45jl/google/ (last visited Mar. 2, 2016).

60. *See id.*

61. *See* Laura A. Heymann, *The Public's Domain in Trademark Law: A First Amendment Theory of the Consumer*, 43 GA. L. REV. 651, 695 (2009).

62. However, some marketing theorists are so focused on the importance of the brand that they conflate the brand with the trademark. *See, e.g.*, DAVID A. AAKER, MANAGING BRAND EQUITY: CAPITALIZING ON THE VALUE OF A BRAND NAME 7 (1991) ("A brand is a distinguishing name and/or symbol . . . intended to identify the goods or services of either one seller or a group of sellers, and to differentiate those goods or services from those of competitors. A brand thus signals to the customer the source of the product, and protects both the customer and the producer from competitors who would attempt to provide products that appear to be identical.").

"A product is something that is made in a factory; a brand is something that is bought by a customer. A product can be copied by a competitor; a brand is unique. A product can be quickly outdated; a successful brand is timeless."[63] A product new to the market will likely be sold under a trademark that fits the traditional role of such a mark; it indicates the source of the product. If the trademark for the new product is inherently distinctive, the trademark will receive legal protection absent any secondary meaning (or time-developed, consumer knowledge of the source of the product) in the marketplace.

However, that new product will not yet possess a brand identity. Famous brands possess numerous markers of identity that contribute to the overall brand image—including trademarked names (COCA COLA, HARLEY DAVIDSON), logos (the NIKE swoosh, the NBC peacock), and trade dress (the COCA COLA bottle shape, the round shape of a MOBIL gas pump)—but they also possess a story or history that has been shared with consumers. Douglas Holt explains:

> A brand emerges as various "authors" tell stories that involve the brand. Four primary types of authors are involved: companies, the culture industries, intermediaries (such as critics and retail salespeople), and customers (particularly when they form communities). . . .
>
> Brand stories have plots and characters, and they rely heavily on metaphor to communicate and to spur our imaginations. As these stories collide in everyday social life, conventions eventually form. . . . A brand emerges when these collective understandings become firmly established.[64]

This brand "story" is developed first by the brand owner and its marketing efforts. In developing this story, the marketing team is likely creating a brand strategy for how the various branding efforts will be released to the consumer: through print advertisements, radio or television commercials, or even a viral or online campaign.[65] The first interactions that consumers have with a new product will likely occur either at a store when

63. *See id.* at 1.

64. DOUGLAS B. HOLT, HOW BRANDS BECOME ICONS: THE PRINCIPLES OF CULTURAL BRANDING 3 (2004).

65. Viral marketing or viral branding is a modern marketing tactic whereby consumers are invited to assist in spreading a marketing message (with the hope that it will spread from consumer to consumer like a "virus"). Viral campaigns can be beneficial in that they appear to have more authenticity and excitement for consumers. However, brand owners run the risk that the campaign will veer off in an unintended or negative direction since the brand owner is ceding significant control to consumers. *See* TILDE HEDING, CHARLOTTE F. KNUDTZEN & MOGENS BJERRE, BRAND MANAGEMENT: RESEARCH, THEORY AND PRACTICE 17 (2009).

the consumer is contemplating a purchase or through print or television marketing.[66]

The modern approach to branding is to think of and describe the brand as if it were a person.[67] Therefore, a brand can have a personality, which is typically described as the psychological personality characteristics ascribed to a brand by consumers.[68] By conducting focus groups, brand owners can uncover the brand personality perceived by consumers and explore the personality characteristics that would cause such consumers to become more personally engaged with the brand.[69] The goal is to then further develop marketing efforts that imbue the brand with these sought-after personality traits.

Brands are also described as having an identity.[70] This is "a set of associations the brand strategist seek[s] to create or maintain" in the mind of the consumer.[71] In essence, the brand identity includes the brand personality, but it also must be a more long-lasting and permanent sense of the vision and uniqueness of the brand so that it serves as the "driver of all brand-related activities."[72] Marketing teams are tasked with the goal of developing a brand identity and communicating it to consumers through various types of marketing campaigns and forms of advertising.

Developing a new brand can be a very expensive undertaking. However, this expense is essential, as brand positioning and advertising are seen as "[t]he most significant contributor to the development of a brand, beyond the product itself (in the splendor of all of its tangible and intangible elements) . . . ."[73] In 2013, for example, Unilever spent $195 million on U.S. advertising for its DOVE line of soaps, body washes, and hair care products.[74] However, in 2002, it allegedly spent $110 million alone on the

66. *See* Alexander L. Biel, *Converting Image into Equity, in* BRAND EQUITY & ADVERTISING: ADVERTISING'S ROLE IN BUILDING STRONG BRANDS 67, 74 (David A. Aaker & Alexander L. Biel eds., 1993) ("In addition to direct and indirect (e.g., word of mouth, media reports, etc.) personal experience with a brand, media advertising is an obvious source of [brand] image, both reflecting and forming the brand's gestalt.").

67. *See* Katya Assaf, *Brand Fetishism*, 43 CONN. L. REV. 83, 84 (2010) (arguing that "trademark law should serve to discourage brand fetishism, and should act to restore the original informative function of trademarks").

68. *See* HEDING ET AL., *supra* note 65, at 117–19 (describing brand personality approach).

69. *See id.*

70. *See id.* at 12 (describing formation of brand identity).

71. *See id.* (citation omitted) (internal quotation marks omitted).

72. *See id.* at 13.

73. See RICHARD D. CZERNIAWSKI & MICHAEL W. MALONEY, CREATING BRAND LOYALTY: THE MANAGEMENT OF POWER POSITIONING AND REALLY GREAT ADVERTISING 8 (1999); *see also* HOLT, *supra* note 64, at 7 ("Most iconic brands have been built through the mass media, usually with television advertising.").

74. *See* Alexandra Bruell, *Unilever to Review Massive Media Buying and Planning Account,* ADVERTISING AGE (Jan. 22, 2015), http://adage.com/article/agency-news/unilever-review-massive-media-buying-planning-account/296720/ [https://perma.cc/S4RS-PMFS].

marketing launch of its new line of DOVE hair care products.[75] Marketing executives and researchers agree that advertising is a necessary expense:

> Advertising . . . is a critically essential plank for establishing in the minds of your customers how and what to think about your offering in the absolute and relative to other products in the marketplace. Advertising contributes to the development and establishment of the brand. Really great advertising helps establish great brands and, of course, brand loyalty.[76]

In spite of the cost of developing a brand, a company benefits greatly from the brand equity that is produced. While the monetary value of trademark assets has been said to encompass the "goodwill" of a business, it may be more accurate to say that it is a shorthand reference for all of the value contained within the relevant brand. This is the brand equity: "[T]he value, usually defined in economic terms, of a brand beyond the physical assets associated with its manufacture or provision."[77] Possessing a well-developed and well-known brand name translates into numerous benefits to the brand owner, including the ability to spend less money introducing new product extensions.[78] Established brands have also been found to be more immune to competitor price fluctuations and promotions: "Brand loyalty, long a central construct in marketing, is a measure of the attachment that a customer has to a brand. . . . As brand loyalty increases, the vulnerability of the customer base to competitive action is reduced."[79]

Brand loyalty is especially important to companies now that there is significantly more competition and significantly more products and services than ever before. Marketers routinely use price promotions and sales to encourage fickle customers to try the next new product in a category. To combat this competitive environment, companies must either enter into this battle of price promotions or develop brand loyalty.[80] The varied benefits stemming from solid brand development are recognized in

75. *See Dove: Brand Profile*, ADBRANDS.NET, http://www.adbrands.net/us/dove_us.htm [https://perma.cc/NLS6-BB9J] (last visited Mar. 1, 2016).

76. CZERNIAWSKI & MALONEY, *supra* note 73, at 9.

77. *See* Biel, *supra* note 66, at 69.

78. *See* Rajeev Batra, Donald R. Lehmann & Dipinder Singh, *The Brand Personality Component of Brand Goodwill: Some Antecedents and Consequences*, *in* BRAND EQUITY & ADVERTISING: ADVERTISING'S ROLE IN BUILDING STRONG BRANDS, *supra* note 66, at 83, 83 ("Brand names are regarded among the most valuable assets owned by a company. A well-known and well-regarded brand name—one with a high level of *equity* or *goodwill*—can often be extended into new product categories, in a way that saves the extending company many of the expenses of establishing a new brand name. As a consequence, companies acquiring others pay significant asset valuation premiums for the portfolio of brand names that are acquired." (emphasis added) (citation omitted)).

79. *See* AAKER, *supra* note 62, at 39.

80. *See* CZERNIAWSKI & MALONEY, *supra* note 73, at xix.

merger and acquisition scenarios, whereby investors and purchasers will pay a premium that reflects brand equity.[81] Currently, Forbes ranks Apple as the most powerful brand: it is valued at $145.3 billion.[82] Apple's yearly advertising spending of $1.2 billion was significantly less than the $4 billion spent by Samsung (ranked as the seventh most powerful brand by Forbes), which demonstrates the financial impact of brand loyalty since Apple is said to "rel[y] on its avid fan base more than Madison Avenue to promote its products."[83]

A. *Brands and Marketing Theory*

Marketing scholars began studying brands in earnest in the 1990s, drawing on a variety of scientific disciplines, including economics, consumer research, organizational behavior, psychology, sociology, and anthropology.[84] The earliest brand research was based on economic theory and studied brands with the goal of determining how to optimize sales. Under this approach, it was believed that optimal sales would result from the correct marketing mix of the "Four Ps": price, product, place, and promotion.[85] The goal of a brand manager was to study economic theory to determine the proper balance of costs and revenue related to these four factors. In this model of branding, consumers played a passive role and were assumed to simply receive marketing messages and then respond in a rational fashion to those messages in determining whether to buy the relevant product at the set price point.[86]

Subsequent research recognized that consumers are not always so passive or so rational in making purchasing decisions. The economic approach struggled to explain why a consumer may pay a higher price for a brand name product when there is an equivalent, cheaper product sold next to it on the same shelf in the grocery store. Therefore, the next wave of branding research focused more on the relationship between the consumer and the brand.

As part of this focus on the consumer, brand identity became an essential focus, and it included both an understanding of consumer perceptions of the brand owner (or corporate identity) as well as an analysis of

81. *See* David A. Aaker & Alexander L. Biel, *Brand Equity and Advertising: An Overview, in* BRAND EQUITY & ADVERTISING: ADVERTISING'S ROLE IN BUILDING STRONG BRANDS, *supra* note 66, at 1, 1 ("Philip Morris purchased Kraft for more than six times its book value. In his 1989 keynote address to the Advertising Research Foundation, Hamish Maxwell, the man behind the acquisition, emphasized that he was buying strong brands.").

82. *See The World's Most Valuable Brands,* FORBES, http://www.forbes.com/powerful-brands/list/ (last visited Mar. 1, 2016).

83. *See* Kurt Badenhausen, *Apple and Microsoft Head the World's Most Valuable Brands 2015,* FORBES (May 13, 2015, 9:57 AM), http://www.forbes.com/sites/kurtbadenhausen/2015/05/13/apple-and-microsoft-head-the-worlds-most-valuable-brands-2015/.

84. *See* HEDING ET AL., *supra* note 65, at 4.

85. *See id.* at 37.

86. *See id.* at 31.

how consumers are personifying the brand (and developing its perceived brand personality).[87] To study this dyadic relationship between brand and consumer, researchers began to use both quantitative and qualitative techniques, which added detailed interviews and home studies to the traditional use of consumer surveys and quantitative focus group testing.[88] The underlying assumption is that a strong and well-developed brand personality can help to increase consumer attachment to a brand.[89] In order to create this strong personality and brand identity, brand owners must engage with consumers in an ongoing discourse to determine how the brand is perceived.[90] The brand owner is then able to change its marketing message in order to change consumer perceptions in a manner that will increase sales and consumer loyalty.[91]

Recent research into brand communities takes the brand–consumer relationship a step further by also incorporating the relationship between consumers.[92] Focus is now on the "brand triad."[93] This approach recognizes that other consumers can impact how a single consumer views a brand; consumers receive official messages from brand owners, as well as numerous unofficial messages about the brand from other consumers.[94] As discussed in more detail below, this is a more complex view of branding that is heralded as providing insights into the future of branding, and a robust theory of trademark law should be able to account for it.

B. *A Brand Theory of Trademark Law*

In order to better understand modern trademark law's evolution and expansion, trademark law must be viewed through brand theory. According to Deven Desai in his foundational work explaining the brand theory approach to trademark law, "trademarks have moved far beyond the commercial sphere. Trademarks have become brands; that is, they now are more about allowing corporations to protect reputation and persona than preventing unfair competition and advancing consumer protection."[95] While this statement reflects the new role of the brand as a reputational asset to a corporation, the brand encompasses more than just product or corporate reputation. Desai continues elsewhere:

> The noncorporate dimension of branding involves consumers and communities as stakeholders in brands. Consumers

87. *See id.* at 23.
88. *See id.*
89. *See id.* at 118.
90. *See id.* at 118–20.
91. *See id.*
92. *See id.* at 182.
93. *See id.* at 183.
94. *See id.* at 182–85.
95. *See* Deven R. Desai, *Speech, Citizenry, and the Market: A Corporate Public Figure Doctrine*, 98 MINN. L. REV. 455, 475 (2013) [hereinafter Desai, *Speech, Citizenry, and the Market*].

> often buy branded goods not for their quality but as badges of loyalty, ways to express identity, and items to alter and interpret for self-expression. Some consumers form brand communities which either evangelize or police brand meaning and corporate practices. From the perspective of trademark law, this behavior presents a problem; from the perspective of brand scholarship, it is both ordinary and expected.[96]

When trademark law is analyzed using brand theory, two things become clear. First, brand theory helps to explain the expansion of trademark law with regard to initial-interest confusion, post-sale confusion, and dilution. By widening one's understanding of trademarks from simple symbols for source differentiation to components of a larger brand, one can better understand the creation of these causes of action and their logical boundaries. Second, brand theory also highlights a current weakness in trademark law with regard to its view of the role of consumers: "[T]rademark law champions corporations as the sole custodians of trademark meaning. . . . Trademark law thus clings to the model of the firm as the one with the exclusive power to develop the brand and to control its meaning."[97] Instead, trademark law must recognize the active role that consumers now play in the development of brands. In doing so, it will need to allow for more interplay between the trademark owner and the consumer in order to accurately reflect modern marketing reality.

C. *Brand Theory and Trademark Expansion*

Viewing trademarks through the lens of brand theory, initial-interest confusion, post-sale confusion, and dilution can be conceptualized as developments that assist trademark owners in protecting their brand. Rather than simply preventing source confusion or reducing consumer search costs, these causes of action recognize the investments being made by trademark owners in the larger notion of the trademark-affiliated brand. For example, initial-interest confusion has been maligned by some commentators because it allows an action for infringement when the consumer is not actually confused at the time of the purchase of the good or service.[98] Initial-interest confusion is said to occur, for example, when a consumer is drawn to a store by advertisements for a well-known trademarked good. Upon arriving at the store, the trademarked good is not available, so the consumer decides to buy a competitor's product instead.

96. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 986 (footnotes omitted).

97. *See id.* at 983–84.

98. *See, e.g.*, Glynn S. Lunney, Jr., *Trademarks and the Internet: The United States' Experience*, 97 TRADEMARK REP. 931, 936 (2007) (arguing against initial-interest confusion as "muddle[d]" over-expansion of trademark law); Jennifer E. Rothman, *Initial Interest Confusion: Standing at the Crossroads of Trademark Law*, 27 CARDOZO L. REV. 105 (2005) (arguing for "pre-sale confusion" instead of initial-interest confusion).

The consumer may have been misled about the availability of the trademarked good (and may have incurred some transaction costs in seeking out this particular store), but the consumer is not mistaken about the source of the product purchased. The consumer has the option of leaving the store but opts to buy the competitor's product instead.

Under some circumstances, courts have permitted trademark owners to obtain damages for trademark infringement based solely on this initial pre-sale confusion.[99] However, a brand-centric view of trademark law recognizes that initial-interest confusion is a form of unfair competition that can directly harm the truthful advertising efforts of the trademark owner. Substantial investment in truthful advertising is necessary to the development of an iconic brand. As such, initial-interest confusion can be seen as addressing a commercial injury to that brand investment.[100]

Post-sale confusion, on the other hand, occurs when a consumer buys counterfeit merchandise. Even though the consumer is aware of the fact that the item is counterfeit and therefore is not confused as to the source of the good, courts have found infringement in post-sale confusion cases based on the fact that third parties unconnected to the actual sale may mistakenly believe that the item is a genuine, trademarked good.[101] As such, the trademark owner may be harmed by the mental association of the potentially inferior counterfeit good with the trademark, thereby lessening the perceived prestige of the trademark and its genuinely affiliated goods. Again, the actual consumer in such a scenario was not confused about the source of the good, so this cause of action does not reflect the original underpinnings of trademark law. Liability for post-sale confusion does not prevent consumer confusion or reduce consumer search costs. However, it does support branding efforts. By allowing trademark owners to sue for such counterfeiting where direct consumers are not confused, courts are again supporting the investment of business owners in the larger concept of a brand.[102]

Dilution might be the most criticized of the recent expansions of trademark law.[103] It has been called "a fundamental shift in the nature of

---

99. *See, e.g.*, Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc., 109 F.3d 1394, 1405 (9th Cir. 1997) ("Sixth, the use of the Cat's stove-pipe hat or the confusingly similar title to capture initial consumer attention, even though no actual sale is finally completed as a result of the confusion, may be still an infringement."); Mobil Oil Corp. v. Pegasus Petroleum Corp., 818 F.2d 254, 260 (2d Cir. 1987) ("As explained above, the district court's concerns focused upon the probability that potential purchasers would be misled into an initial interest in Pegasus Petroleum. Such initial confusion works a sufficient trademark injury.").

100. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 1025 ("[T]he [initial-interest confusion] doctrine has little to do with rational choice problems that traditionally animate trademark law.").

101. *See supra* note 54.

102. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 1025.

103. *See, e.g.*, David J. Franklyn, *Debunking Dilution Doctrine: Toward a Coherent Theory of the Anti-Free-Rider Principle in American Trademark Law*, 56 HASTINGS L.J. 117, 131 (2004) ("There is scant empirical evidence that multiple uses of a famous

trademark protection"[104] and may signal an expansion of trademark rights at the expense of the public domain.[105] Under the federal Trademark Dilution Revision Act of 2006, the holder of a "famous mark"[106] may now bring a dilution claim against a junior user of a mark if the junior user's use is "likely to cause dilution by blurring or dilution by tarnishment of the famous mark . . . ."[107] The statute expressly states that a dilution claim may be brought "regardless of the presence or absence of actual or likely confusion, of competition, or of actual economic injury."[108] Given the broad nature of this cause of action, critics of dilution liability note that "holders of famous marks can sue junior users even when the junior user does not compete with the mark holder, there is no likelihood of confusion, or there is no quantifiable economic harm."[109]

These criticisms all highlight the underlying problem that dilution law does not easily follow from a traditional, consumer-focused theory of trademark protection. Dilution is not about reducing search costs or increasing the amount or quality of information available to consumers. Instead, it is explicitly about protecting the trademark owner: "[D]ilution law is producer-focused rather than consumer-focused: It seeks to prevent diminution in the value of a famous mark stemming from the use of the mark by someone other than the trademark holder."[110] Congress took a brand view of trademarks when it drafted and passed this federal dilution act; such focus can be clearly seen in Congress's explanation that the law was intended to protect "the substantial investment the owner has made in the mark and the commercial value and aura of the mark itself."[111] If trademark law is reframed as a law of brand protection, dilution law becomes less incongruous. Dilution law expressly protects the owners of fa-

---

mark dilute the selling power of the mark in connection with the first class of products to which it was attached."); Mary LaFrance, *No Reason to Live: Dilution Laws as Unconstitutional Restrictions on Commercial Speech*, 58 S.C. L. REV. 709, 711 (2007) (arguing federal and state dilution states are unconstitutional restraints on commercial speech); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1698 (1999); Margaret Jane Radin & R. Polk Wagner, *The Myth of Private Ordering: Rediscovering Legal Realism in Cyberspace*, 73 CHI.-KENT L. REV. 1295, 1305 (1998) ("Owners of 'famous' marks can use this statute to capture the domain name they want, even if someone else got it first, but owners of non-famous marks seem to be out of luck.").

104. *See* Lemley, *supra* note 103, at 1698.

105. *See* Radin, *supra* note 103, at 1305.

106. A famous mark is statutorily defined as a mark that is "widely recognized by the general consuming public of the United States as a designation of source of the goods or services of the mark's owner." 15 U.S.C. § 1125(c)(2)(A) (2012).

107. *See id.* § 1125(c)(1).

108. *See id.*

109. *See* Deven R. Desai & Spencer Waller, *Brands, Competition, and the Law*, 2010 BYU L. REV. 1425, 1459.

110. *See* Clarisa Long, *Dilution*, 106 COLUM. L. REV. 1029, 1034 (2006) (presenting empirical evidence of judicial dissatisfaction with dilution law).

111. *See* Desai & Waller, *supra* note 109, at 1460 (quoting H.R. REP. NO. 104-374, at 3 (1995)) (internal quotation marks omitted).

mous marks, and those famous marks became well-known enough to justify this protection through the use of substantial branding techniques.[112] Therefore, dilution law encourages substantial investment in branding and rewards those companies that have made such an investment by allowing them to safeguard their brand from the more subtle blurring or tarnishment injuries that may be caused by non-competitors.

D. *Brand Theory and Consumer Involvement*

While brand theory can be used to better understand the legal shifts in modern trademark law, it also highlights the fact that trademark law turns a blind eye to the changing relationship between brands (and thus trademarks) and consumers. Traditionally, trademark law placed consumers in the role of passive parties needing protection from confusing advertising tactics. The law's focus on the likelihood of confusion highlights this approach. However, brand owners are increasingly asking consumers to engage with their brands, including by participating in market research to help develop boundaries on the brand's perceived "personality," as well as by actively spreading marketing information through viral or word-of-mouth campaigns.[113] Noting this growing role for consumers in branding efforts, Laura Heymann argues, "If trademark law recognized the active work that consumers do in engaging with trademarks, it would incorporate a theory of the consumer that sees him as capable of engaging with these trademark associations without the law's interference."[114]

While it is true that a brand concept is initially developed by the brand owner and that initial communications regarding the brand will come from that owner, consumers will have a variety of interactions with and influence on that brand once the product and its brand image are made public. At the very least, the consumer must be recognized as someone who "not only perceives the trademark as a source identifier but who also can call to mind (and then accept or reject) the various associations the mark comprises."[115]

---

112. *See id.* ("In other words, criticisms that dilution is far removed from trademark law's search-cost and consumer-focused foundations are accurate but miss the point that trademark law has already imported a brand perspective into its doctrine. Dilution, like the other brand-based extensions of trademark law in recent times, can be seen as merely the most obvious iteration of that view.").

113. *See, e.g.*, TOMI T. AHONEN & ALAN MOORE, COMMUNITIES DOMINATE BRANDS: BUSINESS AND MARKETING CHALLENGES FOR THE 21ST CENTURY 220–22 (2005) (citing examples of consumer involvement in marketing of Ford Escape and Firefox web browsers); *see also* Luis V. Casaló, Carlos Flavián & Miguel Guinalíu, *Promoting Consumer's Participation in Virtual Brand Communities: A New Paradigm in Branding Strategy,* 14 J. MARKETING COMM. 19, 19–20 (2008) ("Indeed, more and more firms are starting to use several online tools (e.g. chats, forums, etc.) to contact their consumers and to allow interaction among them.").

114. *See* Heymann, *supra* note 61, at 655.

115. *See id.* at 654.

Deven Desai notes the many ways that a brand may be used by these intersecting stakeholders:

> A company uses brands to provide product information to consumers, but it also uses brands to enhance the overall corporate image as the company pursues a full range of business goals. Consumers may, of course, use brands to find products. But consumers may simultaneously use brands as expressions of individuality and identity as they take a brand and alter it to match what they see as the meaning of the brand and how that meaning relates to their self-image or message. Communities may also engage with a brand as a symbol about which they wish to comment and share both positive and negative information.[116]

This connection between brands and consumers can be valuable for the consumer as well as for the brand owner.

Some customers develop intense emotional connections with brands and with the trademarks connected to those brands. Douglas Holt has conducted research on iconic brands and found that:

> Customers value some products as much for what they symbolize as for what they do. For brands like Coke, Budweiser, Nike, and Jack Daniel's, customers value the brand's stories largely for their *identity value*. Acting as vessels of self-expression, the brands are imbued with stories that consumers find valuable in constructing their identities.[117]

In this way, consumers find emotional and psychological value in associating with certain brands. Some consumers even join brand communities to further cement these connections within the larger community of consumers. A brand community is defined as "a specialized, non-geographically bound community, based on a structured set of social relationships among admirers of a brand."[118] Frequently these communities form online as virtual communities.[119] Branded communities are unique when compared to more traditional communities founded based on shared geography or shared affinities, because at the center of the community is a

---

116. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 988–89.

117. *See* HOLT, *supra* note 64, at 3.

118. *See* Albert M. Muniz, Jr. & Thomas C. O'Guinn, *Brand Community*, 27 J. CONSUMER RES. 412, 412 (2001); *see also* Marcelo Royo-Vela & Paolo Casamassima, *The Influence of Belonging to Virtual Brand Communities on Consumers' Affective Commitment, Satisfaction and Word-of-Mouth Advertising: The ZARA Case*, 35 ONLINE INFO. REV. 517, 519 (2011).

119. *See* Royo-Vela & Casamassima, *supra* note 118, at 518 ("A virtual community is defined as the integration of a group of individuals using the internet to maintain social relations around a common interest." (citation omitted)). *See generally* HOWARD RHEINGOLD, THE VIRTUAL COMMUNITY: HOMESTEADING ON THE ELECTRONIC FRONTIER (1993).

branded good or service.[120] However, the psychological factors that bring the members together to form the community, the foundational principles required to identify such a group as a community, and the benefits received by members[121] are all consistent with other, more traditional types of communities: "Like other communities, [a brand community] is marked by a shared consciousness, rituals and traditions, and a sense of moral responsibility."[122]

The first indicator of community, "[c]onsciousness of kind[,] is the intrinsic connection that members feel toward one another, and the collective sense of difference from others not in the community."[123] It is a shared sense of belonging.[124] In their 2001 ethnographical study that laid the foundation for modern brand community research, Albert Muniz Jr. and Thomas O'Guinn investigated three brand communities through interviews conducted with community members located in a Midwestern town.[125] They found evidence of all three community characteristics, listed above, in the brand communities studied.[126] With regard to consciousness of kind, the investigators noted that community members felt a strong connection to the brand but an even stronger connection to one another.[127] Members felt that they could describe other members without having met them.[128]

This statement encompasses ideas of "legitimacy" and "oppositional brand loyalty"—two ways in which members establish themselves as insiders versus outsiders.[129] In the brand community formed around Saab cars in the study mentioned above, members felt that "legitimate" members were ones who purchased the car for the right reasons: in appreciation of the durability of the car.[130] More recent purchasers, those who purchased based on trends or current advertising, were not "making a long-term commitment" and thus were deemed illegitimate.[131] Additionally, oppositional brand loyalty was one of the strongest ways in which community members defined themselves.[132] Members of the Macintosh computer brand community were fervently opposed to Microsoft and its products.[133] Community websites would frequently lambast and demonize Bill Gates

---

120. *See* Muniz & O'Guinn, *supra* note 118, at 412.

121. *See* HOLT, *supra* note 64, at 4 ("Conventional branding models largely ignore how brands buttress consumer identities.").

122. *See* Muniz & O'Guinn, *supra* note 118, at 412.

123. *See id.* at 413.

124. *See id.*

125. *See id.* at 415.

126. *See id.* at 426.

127. *See id.* at 418.

128. *See id.*

129. *See id.* at 419–20; *see also* HEDING ET AL., *supra* note 65, at 189.

130. *See* Muniz & O'Guinn, *supra* note 118, at 419.

131. *See id.*

132. *See id.* at 420.

133. *See id.*

and Microsoft as evidence of the website developer's (and the community's) dedication to the community's core beliefs.[134] If a community member violates this oppositional brand loyalty by purchasing a competing product, the member's legitimacy as part of the community will be questioned.[135]

The second indicator of community is the existence of shared rituals and traditions. Rituals and traditions "perpetuate the community's shared history, culture, and consciousness."[136] Rituals also "help to reproduce and transmit the community's meaning in and outside of the community."[137] The rituals in brand communities may be small undertakings, such as honking or flashing one's lights when a Saab brand community member encounters another Saab driver on the road.[138] However, these are shared rituals, and they work to reinforce community values and strengthen the consciousness of kind discussed above.[139] Another common ritual among brand communities is repetitive story-telling regarding the brand's history or origin. Saab brand community members were found to often repeat the same basic origin story (focused on the fact that Saab first developed airplanes) at both meetings and on the numerous community-created websites.[140] Members show evidence of their legitimacy by reciting the agreed-upon history of the brand, where such a story may have been gleaned from the original community members or may result from the brand owner's own marketing efforts.[141]

The third and final indicator used to define a community is a perceived sense of moral responsibility, which manifests as "a felt sense of duty or obligation to the community as a whole, and to its individual members."[142] A shared sense of moral responsibility helps communities to recruit new members and retain current members. In brand communities, "[m]oral responsibility also includes looking out for and helping other members in their consumption of the brand."[143] Members of the Saab and Macintosh communities indicated that they felt a responsibility to help other users of Macintosh computers or Saab cars—whether they were current members or not—when their computer or car was malfunctioning.[144] This responsibility included an obligation to share information and resources related to the brand, such as details about newly released

134. *See id.*
135. *See id.* at 420–21.
136. *See id.* at 413.
137. *See* Royo-Vela & Casamassima, *supra* note 118, at 520.
138. *See* Muniz & O'Guinn, *supra* note 118, at 422.
139. *See id.*
140. *See id.*
141. *See id.*
142. *See id.* at 413.
143. *See id.* at 425.
144. *See id.*

products or opinions regarding the best repair businesses.[145] This moral responsibility provides a virtually free way for companies to distribute information among customers and to encourage consumer engagement.

Moreover, marketing research indicates that companies should consider setting up brand communities as a way to enhance consumers' levels of satisfaction and commitment to brands.[146] In studies of brand communities, including online virtual brand communities, simply registering for the community boosts consumer satisfaction, stated levels of affective commitment to the brand, and the likelihood of positive word-of-mouth activities by the registered consumer.[147] Active participation in the brand community brings higher levels of satisfaction and commitment than merely passive belonging.[148] This is an especially exciting finding for brand owners, as other studies have found that consumers are increasingly likely to mistrust and dislike traditional forms of advertising.[149] Perhaps this research underlies the current enthusiasm among marketing professionals regarding brand communities, with some marketing websites calling 2015 "the year of the brand community."[150]

Numerous studies have shown that participation in a brand community influences participants' behaviors. For example, members feel obligated to disparage other similar products by competitors.[151] Participation increases consumer loyalty toward the brand around which the community is developed, and that loyalty can result in a greater likelihood of repeat purchases for the brand owner.[152] Additionally, a high level of satisfaction stemming from brand community participation has been shown to lead to an increase in the amount of positive word-of-mouth ad-

145. *See id.*

146. *See* Royo-Vela & Casamassima, *supra* note 118, at 538.

147. *See id.*

148. *See id.*

149. *See* Michael Trusov, Randolph E. Bucklin & Koen Pauwels, *Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site*, 73 J. MARKETING 90, 90 (2009) (noting that, when comparing surveys of consumer perceptions from 2002 to 2004, "40% fewer people agree that advertisements are a good way to learn about new products, 59% fewer people report that they buy products because of their advertisements, and 49% fewer people find that advertisements are entertaining").

150. Katz, *supra* note 15; *see also* Kerry O'Shea Gorgone, *Why Brand Communities Are the Future of Marketing: Jordan Kretchmer of Livefyre on Marketing Smarts [Podcast]*, *supra* note 15.

151. *See* John W. Schouten & James H. McAlexander, *Subcultures of Consumption: An Ethnography of the New Bikers,* 22 J. CONSUMER RES. 43, 53 (1995) (noting that community of Harley-Davidson owners bond over shared sense of heritage or tradition that leads to criticism of Japanese motorcycles because of "perceived disdain of Japanese manufacturers for tradition, as demonstrated by the frequent introduction of new models and the extinction of others after only a few years. Harley-Davidson, in contrast, emphasizes a continuity that connects its newest motorcycle in a direct line of ancestry to its earliest prototype").

152. *See* Casaló et al., *supra* note 113, at 31; *see also* Royo-Vela & Casamassima, *supra* note 118, at 520.

vertising that is shared by community members about the product or company.[153] Word-of-mouth (WOM) advertising has been shown to be more effective at persuading purchasers and to have a significant positive effect on post-purchase product perceptions.[154] Overall, word-of-mouth communication among brand community members has been found to be much more effective than traditional advertising activities in stimulating sales.[155]

Additionally, some research suggests that brand owners would receive more of the benefits described above if they allowed consumers to develop their own brand communities independent of the brand owner. Past consumer behavior research suggests that consumers tend to view corporations' marketing strategies (including corporate-sponsored brand community development) in a cynical fashion and assign exploitative motivation to even purely altruistic promotions.[156] However, research has also shown that consumer-created online brand communities engender higher intrinsic motives of altruism than a similar marketer-created online brand community.[157] By allowing consumer-created brand communities to develop and flourish, brand owners can receive numerous benefits without suffering from the negative connotations consumers ascribe to traditional, corporate-generated marketing.[158]

E. *The Trademark–Brand Disconnect*

If consumer-initiated branding is as beneficial to brand owners as the marketing research suggests, it seems incongruous that brand owners like IKEA would take actions to stop it. However, trademark law contains en-

153. *See* Trusov et al., *supra* note 149, at 90; *see also* Ángel Millán & Estrella Díaz, *Analysis of Consumers' Response to Brand Community Integration and Brand Identification,* 21 J. BRAND MGMT. 254, 255 (2014).

154. *See* Royo-Vela & Casamassima, *supra* note 118, at 522.

155. *See, e.g.*, Trusov et al., *supra* note 149, at 91 ("[W]e find that WOM referrals strongly affect new customer acquisitions and have significantly longer carryover than traditional forms of marketing used by the firm (21 days versus 3 to 7 days). We estimate a long-term elasticity for WOM of .53—approximately 20–30 times higher than the elasticities for traditional marketing."); *see also id.* at 95–96 ("[T]he WOM effect on sign-ups remains significantly different from zero for approximately three weeks. In contrast, the effects of media and events . . . lose significance within just a few days. Compared with traditional marketing activities, the WOM referrals induce both a larger short-term response and a substantially longer carryover effect.").

156. *See, e.g.*, Doohwang Lee, Hyuk Soo Kim & Jung Kyu Kim, *The Impact of Online Brand Community Type on Consumer's Community Engagement Behaviors: Consumer-Created vs. Marketer-Created Online Brand Community in Online Social-Networking Web Sites*, 14 CYBERPSYCHOLOGY, BEHAV. & SOC. NETWORKING 59, 60 (2011).

157. *See id.* at 61.

158. *See id.*; *see also id.* at 62 ("[M]arketers may have to employ a passive role when facilitating online brand communities such as being merely a sponsor of consumer-created online brand communities. It is certainly plausible that corporations can sponsor charity events or activities for members in consumer-created communities.").

meshed duties and burdens placed on trademark owners that cause them to fear the loss of their trademark rights if they fail to take action against use of their mark by any third parties, including the trademark owner's own fans. These risks include the potential for genericide and abandonment for failure to police a mark or failure to maintain quality control if the mark is licensed. Moreover, consumer-initiated brand activities are increasingly likely to be characterized as "confusing" (and thus infringing) because of the way trademark confusion has been broadened to protect brand owner interests in other contexts.

1. *Genericide and the Duty to Police*

While trademark law may be expanding in an attempt to protect the investment interests of brand owners, it is still ill-suited to address the push for greater consumer involvement in branding and in brand community development. Trademark owners have long been advised to monitor and "police" third-party uses of trademarks identical to or similar to their own. This obviously would include consumer-initiated activities involving the owner's trademarks. The advice to trademark owners to police such use is intended to prevent genericide and abandonment and to stop infringement.[159] Trademark owners feel that they must remain vigilant in this regard in order to avoid a determination by a court or by the United States Patent and Trademark Office (USPTO) that they have abandoned their trademark.

A trademark will be deemed "abandoned" under Section 1127 of the Lanham Act when "its use has been discontinued with intent not to resume such use" or when "any course of conduct of the owner, including acts of omission as well as commission, causes the mark to become the generic name for the goods or services on or in connection with which it is used or otherwise to lose its significance as a mark."[160] When broken down into its bare elements, Section 1127 indicates that abandonment can occur in three ways:

> (1) a trademark owner intends to abandon the mark and thus discontinues its use, (2) a trademark owner, through overt acts or omissions, allows the mark to become the generic name for the good or service with which it has been used, or (3) a trademark owner, through overt acts or omissions, allows the mark to lose its significance as a mark (such that it no longer identifies the source of the good or service to a consumer).[161]

---

159. *See* William T. Gallagher, *Trademark and Copyright Enforcement in the Shadow of IP Law*, 28 SANTA CLARA COMPUTER & HIGH TECH. L.J. 453, 490 (2012) ("The interviewed lawyers often cited a need to 'police' their clients' trademarks and copyrights. They explained that the failure to do so on any particular occasion could lead to difficulties in enforcing rights against other targets in the future.").

160. 15 U.S.C. § 1127 (2012).

161. *See* Kiser, *supra* note 17, at 225.

The resulting "duty to police" third-party trademark usage that was spawned from these grounds for abandonment is now considered a bedrock principle of trademark law.[162] Some scholars have argued that the uncertain boundaries of trademark law's duty to police have led to aggressive and exaggerated responses by some trademark owners to minor, perceived threats by third parties—activity that has been called "trademark bullying" in trademark scholarship and the popular media.[163]

When brand owners attempt to stop consumer-initiated brand activities, they are often called trademark bullies.[164] IKEA was barraged with intense criticism from both consumers and the press when it sent its cease-and-desist letter to the owner of the IKEAhackers website.[165] However, brand owners often describe their actions as a necessary response dictated by trademark law's duty to police.[166] Setting aside brand owners intending to abandon a mark under Section 1127, brand owners are concerned with consumer activities that may lead to genericide or may cause the mark to lose its source-identifying significance.[167] *Genericide* is a term used to describe the process through which a distinct trademark loses its ability to signify the source of the good or service product with which it has been previously connected because the mark is now known by consumers as a generic term.[168] When genericide occurs, the trademark comes to be

162. *See generally* 6 MCCARTHY, *supra* note 32, § 31:38.

163. *See* Irina D. Manta, *Bearing Down on Trademark Bullies*, 22 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 853, 869, 866–71 (2012) ("A significant amount of trademark bullying stems from bullies' impression that to maintain a mark, it is the owner's duty to aggressively police it. Thus, many bullying situations involve mark owners who have taken this perceived duty to extreme levels."); *see also* Grinvald, *supra* note 17, at 628–30; Kiser, *supra* note 17, at 225; Jeremy N. Sheff, *Fear and Loathing in Trademark Enforcement*, 22 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 873, 873 (2012); Sara Marie Andrzejewski, Note, *"Leave Little Guys Alone!": Protecting Small Businesses from Overly Litigious Corporations and Trademark Infringement Suits*, 19 J. INTELL. PROP. L. 117, 118 (2011); Angus Loten, *New Tool in Trademark Fights*, WALL ST. J., Feb. 23, 2012, http://online.wsj.com/article/SB10001424052970203358704577237473534179392.html [https://perma.cc/5S5X-47KN] (discussing small business strategy in making these matters public and the desired effects of shaming).

164. *See* Manta, *supra* note 163, at 854 ("[A] trademark bully is usually a large company that seeks to put an end to behavior by individuals and small businesses that it perceives as a danger to its own intellectual property even though its legal claims against these other parties are spurious or non-existent."); *see also, e.g.*, Loten, *supra* note 163 (discussing instance of trademark bullying).

165. *See* Campbell-Dollaghan, *supra* note 13; Doctorow, *supra* note 13; Karmon, *supra* note 13.

166. *See, e.g.*, Gallagher, *supra* note 159, at 490.

167. *See* Kiser, *supra* note 17, at 239 ("[T]rademark law imposes limitations on what bullies feel are acceptable settlement terms, as they [ ] must also protect against genericide."); Manta, *supra* note 163, at 858 ("Trademark bullying is a touchy problem because the law does require owners to police their marks if they want to maintain exclusive rights in their marks and prevent so-called 'genericide.'").

168. *See* Desai & Rierson, *supra* note 6, at 1790.

known as the generic term used to describe the whole category of goods in which that particular product exists.[169]

The history of the Aspirin trademark is one of the most commonly cited examples of genericide. Aspirin was once used by Bayer Company as a fanciful trademark in connection with its acetyl salicylic acid pain reliever.[170] When a defendant in an infringement suit argued that the mark had become the generic term used by consumers to refer to all such pain relievers produced by any manufacturer, the court in *Bayer Co. v. United Drug Co.*[171] held that the trademark Aspirin had lost its source-identifying significance.[172] That holding meant that *aspirin* had become a generic term available in the public domain to any of Bayer's competitors producing acetyl salicylic acid.[173]

In addition to the risk posed by genericide, a mark can also be deemed abandoned under the Lanham Act if it loses its significance as a mark.[174] A mark that has lost such significance can no longer identify the source of a particular good or service to consumers. It is not the generic label for the product class, as is the case for marks that have become generic, but it fails to serve the foundational purpose of a trademark, which is to provide the shorthand link between the product and its source.[175] The direct link between product or service and its source has been severed. Trademark significance, as it pertains to abandonment, has been explained by the Court of Appeals for the Federal Circuit:

> Without question, distinctiveness can be lost by failing to take action against infringers. If there are numerous products in the marketplace bearing the alleged mark, purchasers may learn to ignore the "mark" as a source identification. When that occurs, the conduct of the former owner, by failing to police its mark, can be said to have caused the mark to lose its significance as a mark.[176]

---

169. *See id.* at 1789–90.

170. *See generally* Bayer Co. v. United Drug Co., 272 F. 505 (S.D.N.Y. 1921).

171. 272 F. 505 (S.D.N.Y. 1921).

172. *See id.*

173. *See id.*

174. *See* 15 U.S.C. § 1127 (2012).

175. *See* Nat'l Football League v. Governor of Del., 435 F. Supp. 1372, 1389–90 (D. Del. 1977). The "loss of significance" grounds for abandonment is often confused with the defense of acquiescence (a defense that can be raised when a plaintiff was aware of the defendant's infringing use of a mark and failed to take action for an unreasonable period of time). *See id.* Treatise author Thomas McCarthy once explained the distinction as follows: "Acquiescence should not be confused with abandonment of trademark rights. The defense of abandonment results in a loss of rights as against the world, while the defense of acquiescence merely results in a loss of rights as against one defendant." *Id.* at 1389 (quoting 6 MCCARTHY, *supra* note 32, § 31:14).

176. Wallpaper Mfrs., Ltd. v. Crown Wallcovering Corp., 680 F.2d 755, 766 (C.C.P.A. 1982).

This duty to police serves as a justification for bully-like behavior by trademark owners.[177] However, the risk of losing one's trademark as a result of insufficient policing of the mark is quite low. Courts have frequently held that failure to police third-party use of a trademark is not sufficient to cancel protection of the mark absent genericide or substantial loss of trademark significance.[178] The dispositive factor in cases where courts have terminated trademarks is not a lack of policing, but whether the trademark owner demonstrated intent to abandon the mark or allowed the mark to be used so freely that it no longer was a distinct source indicator.[179] Courts must find facts sufficient to meet a "high burden of proof" when deciding whether abandonment has occurred, and that burden is rarely met.[180]

However, trademark owners and their counsel are likely aware of the few instances in which genericide and loss of significance abandonment have occurred.[181] Those instances, while incredibly rare and based on unique factual circumstances, remain salient. Given the risk-averse nature of attorneys and the cognitive biases that may amplify perceptions of such risk, many brand owners are hesitant to allow consumer-initiated brand activities, even though marketing research advocates their encouragement.[182]

This disconnect between perceived duties under trademark law and the marketing interests of brand owners is evidence of the uneasy manner in which trademark law has attempted to accommodate the growth of brands in the modern economy. When consumers associate themselves

177. *See, e.g.*, Gallagher, *supra* note 159, at 490.

178. *See, e.g.*, *Wallpaper Mfrs.*, 680 F.2d at 766.

179. *See, e.g.*, Saxlehner v. Eisner & Mendelson Co., 179 U.S. 19 (1900) (confirming Second Circuit's holding that plaintiff could no longer enforce her rights in mark HUNYADI for bottled water because mark had become generic in eyes of consumers); Acme Valve & Fittings Co. v. Wayne, 386 F. Supp. 1162, 1167 (S.D. Tex. 1974) (finding intent to abandon due to discontinuance of manufacture, selling off of all inventory, and failure to renew trademark registration).

180. *See* STX, Inc. v. Bauer USA, Inc., No. C 96–1140 FMS, 1997 WL 337578, at *10 (N.D. Cal. June 5, 1997) ("A party seeking to prove abandonment has a high burden of proof."); *see also* Citibank, N.A v. City Bank of S.F., No. C 79 1922, 1980 WL 30239, at *16 (N.D. Cal. Mar. 23, 1980) ("Abandonment places a strict burden of proof upon the party seeking to prove abandonment. The party seeking to prove abandonment must prove an intent to abandon on the part of the trademark owner.").

181. *See* Kiser, *supra* note 17, at 230 ("Between the date of registration of the mark in 1905 and the start of the cancelation proceeding on June 25, 1924, the defendant trademark owner in this case had allowed numerous competitors to sell high quantities of the same basic product under the 'Milk of Magnesia' mark. The Second Circuit determined that the defendant 'had taken no steps whatever to assert its rights, and had really ignored the extensive use by others.' Such blatant disregard for the trademark allowed it to become abandoned, because the mark 'no longer indicat[[ed] [sic] the origin of the goods sold under it.'" (footnotes omitted) (citing McKesson & Robbins, Inc. v. Charles H. Phillips Chem. Co., 53 F.2d 342, 344–45 (2d Cir. 1931))).

182. *See id.* at 225.

with a brand community or engage in brand-related activities (like hosting brand fan websites), they are not detracting from the original source identifying nature of the mark; rather, the mark is still being used by the brand owner to identify itself as the source of the good or service. Often, consumer-initiated activity supports the brand owner's desired source identification because brand fans are quick to highlight the benefits of that specific source and to direct other consumers to the brand owner's official resources.[183] In this way, consumer-initiated brand activity builds the strength of the brand and the source-related associations made in the mind of new consumers or new fans of the brand.

Additionally, the risk of genericide for brand owners that allow brand communities and other consumer-initiated activity is likely the same as the risk faced by more conservative businesses. Genericide has long been seen as a risk faced by companies that become popular or famous.[184] If you are an early entrant into a product category (like Xerox) or control the majority of the market share in that category (like Google), consumers may conflate your trademark with that product class. These scenarios posed a risk long before the advent of modern branding, and that risk remains relatively unchanged. However, brand owners may be able to use brand communities and loyal brand fans to combat this risk. Currently, trademark owners use marketing and informational campaigns to combat this problem.[185] In such campaigns, trademark owners routinely ask customers for help in preventing genericide. They explain the proper and improper ways to use the company's marks. This seems like a ripe opportunity for enlisting the help of one's brand community filled with loyal consumers willing to help lessen the risk of genericide. While logic and marketing literature would suggest that consumer involvement in branding may strengthen a company's trademarks, trademark law's abandonment rules and duty to police are not aligned with this aspect of branding.

2. *Criticism of Trademark Expansion*

While trademark law can be described as having expanded to protect the investments trademark owners are making into their brands, there are numerous reasons to question whether this expansion was in the best interest of branding or of trademark law in general. Numerous scholars have argued that the expansion of trademark law to allow remedies for dilution, initial-interest confusion, sponsorship confusion, and post-sale

---

183. For a discussion of Saab and Macintosh owners who perform this function, see *supra* notes 125–44 and accompanying text.

184. *See* Kiser, *supra* note 17, at 225–26 (discussing how cognitive biases may lead to overly aggressive litigation tactics of trademark owners).

185. *See* Megan Garber, *'Kleenex Is a Registered Trademark' (and Other Desperate Appeals)*, ATLANTIC (Sept. 25, 2014), http://www.theatlantic.com/business/archive/2014/09/kleenex-is-a-registered-trademark-and-other-appeals-to-journalists/380733/ [https://perma.cc/B66Z-DLCE] (containing examples of these sorts of print advertisements).

confusion has broadened the idea of confusion to absurd levels.[186] As federal trademark law gets its congressional authority through the Commerce Clause, trademarks should be protected as a means of preventing trade diversion and consumer deception (which was the historical goal in preventing "passing off").[187] However, as trademark law has grown to protect the property-like investment interests of brand owners, it has veered away from this focus. Trademark law's likelihood of confusion standard now allows trademark owners to sue for any confusion, even when it is irrelevant to the purchasing consumer, because any confusion could potentially increase search costs for that consumer.[188] Some scholars have argued that this paternalistic approach to trademark law, which protects consumers from all manners of possible confusion, is not supported by current cognitive science research on consumer decision-making.[189] Consumers simply do not need to be coddled in this fashion.[190]

---

186. *See, e.g.*, Lemley, *supra* note 103, at 1708 ("But when trademark law reaches beyond that—when it precludes a haberdasher from selling a hat with the 'Cowboys' logo, even when the circumstances preclude a finding of consumer confusion—it has left its theoretical foundations."); Mark A. Lemley & Mark McKenna, *Irrelevant Confusion*, 62 STAN. L. REV. 413, 414 (2010) [hereinafter Lemley & McKenna, *Irrelevant Confusion*] ("[T]rademark law has taken the concept of confusion too far."); William McGeveran, *Rethinking Trademark Fair Use*, 94 IOWA L. REV. 49, 51 (2008) [hereinafter McGeveran, *Rethinking Trademark Fair Use*] ("Many observers warn that this increased scope of trademark protection threatens free speech, including both dissemination of useful commercial information and discussion, critique, or parody about famous brands and the culture they embody."); McKenna, *Consumer Decision-Making Theory*, *supra* note 46, at 70 ("Thus, over the course of the last century, we have moved from a system in which confusion was actionable only insofar as it related to the particular end of trade diversion to one in which confusion itself defines the cause of action. Trademark law, in other words, now abstracts away from consumer decisions and targets confusion 'in the air.'").

187. *See* Lemley, *supra* note 103, at 1708 ("The point of trademark law has never been to maximize profits for trademark owners at the expense of competitors and consumers.").

188. *See* Lemley & McKenna, *Irrelevant Confusion*, *supra* note 186, at 414 ("[C]ourts expanded the range of actionable confusion beyond confusion over the actual source of a product—trademark law's traditional concern—to include claims against uses that might confuse consumers about whether the trademark owner sponsors or is affiliated with the defendant's goods."); *see also* McKenna, *Consumer Decision-Making Theory*, *supra* note 46, at 72 (arguing that trademark liability should only be permitted based on consumer confusion that would affect purchasing decision).

189. *See, e.g.*, Austin, *supra* note 39, at 829–31 (arguing that "ordinary prudent consumer" in trademark cases needs to be reimagined to better reflect actual prudent consumer); Barton Beebe, *Search and Persuasion in Trademark Law*, 103 MICH. L. REV. 2020 (2005) (analyzing inconsistent ways in which consumer is being used and defined as either sovereign or fool in trademark cases); Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 TEX. L. REV. 507 (2008) (arguing against cognitive model of dilution law).

190. *See* Beebe, *supra* note 189, at 2024 ("Trademark apologists, primarily scholar-practitioners or academics of the law and economics tradition, have countered that the consumer is not nearly so gullible.").

Another effect of this expansion of trademark confusion is that brand owners are often able to sue for trademark infringement when they have suffered little to no harm from the defendant's actions.[191] Mark McKenna and Mark Lemley have argued that there is little empirical support for the argument that trademark owners require expanded confusion liability in order to prevent "free riding" by third parties.[192] Instead, the data indicate that the alleged free rider may benefit from a non-competitive association with another party's trademark, but often that trademark owner does not suffer any harm from this association.[193] As such, trademark owners are being given the ability to block others from a commercial benefit without actually proving any harm underlying their infringement claim. Trademark owners are, in essence, being granted trademark monopolies with all of the "unjustified and inappropriate market power" that entails.[194]

Branding is now a snake eating its own tail. By expanding trademark law to protect the investments made by businesses into their brands, trademark law has made it easier to determine that a third party's activity is infringing. This development means that consumer-initiated brand activities will more likely fall under the umbrella of confusion and be prohibited or deterred, even though consumer activities rarely rise to the level of "passing off." That outcome goes against marketing research that indicates consumer-initiated branding is especially effective at developing brand loyalty, increasing word-of-mouth advertising, and thus increasing sales.

3. *Quality Control and the Licensing Dilemma*

If unauthorized use of a brand owner's marks can cause abandonment issues, one may wonder why the consumer-involved uses are not simply authorized in the first place. Can licensing save a brand owner from the duty to police and abandonment problems? This approach is an option available to brand owners. When IKEA changed direction and allowed the continued operation of the IKEAhackers website, it is likely that IKEA entered into some form of contractual licensing arrangement such that the website owner would be required to operate within certain parameters regarding IKEA's intellectual property. In this way, IKEA can show

191. *See* Lemley & McKenna, *Owning Mark(et)s*, *supra* note 45, at 141 (proposing "'trademark injury' requirement [in trademark law] is akin to the 'antitrust injury' requirement currently used to weed out undeserving anti-trust plaintiffs"). *See generally* Mark P. McKenna, *Testing Modern Trademark Law's Theory of Harm*, 95 IOWA L. REV. 63 (2009) (arguing that there should be no presumption of harm from uses of marks for non-competing goods but presumption of harm in context of market expansion may be justified).

192. *See* Lemley & McKenna, *Owning Mark(et)s*, *supra* note 45, at 156–65.

193. *See id.*

194. *See* Lunney, *supra* note 98, at 373 (discussing property justification for trademark protection as pertains to evolution of trademark law and rise of anti-competitive trademark monopolies).

evidence that it is exercising some control over this now-authorized third-party use of its marks. However, this option is hard to manage on a large scale. Due to the quality control requirements imposed under trademark law, it would be inefficient and potentially unmanageable for a brand owner to allow large numbers of consumers to make licensed uses of the brand owner's marks.

Licensing consists of an arrangement, typically contractual, under which a third party is authorized to use the trademark owner's marks under specific conditions. This arrangement may occur when the trademark owner wants to outsource production of the trademarked good to a cheaper, third-party manufacturer, rather than manufacture the product themselves. It can also occur when a trademark owner wants to allow a third party to create a new (and likely related) line of goods under the established trademark, such as when the brand owner of OREO cookies decides to authorize a third party to utilize the OREO mark in connection with its line of ice cream flavors. Corresponding with the growth of modern branding, some companies also enter into license agreements to produce merchandise like clothing and keychains that bear the licensed trademark. In these instances of promotional trademark licensing, the consumer is seeking products that announce the consumer's affinity for and association with the brand. Trademark law in the United States has evolved from a time when licensing of a trademark was not permitted at all. It was prohibited under common law and the Trademark Act of 1905 because licensing was thought to be misleading in that it hid the true origin of a good or service.[195] As such, it went against the trademark's primary role of indicating the source of a good or service for a consumer. As licensing increasingly became a business necessity, the Lanham Act (the Trademark Act of 1946) finally allowed the licensing of trademarks, but contained several limitations on the practice.

For instance, Section 5 of the Lanham Act states that a registered mark may be used "legitimately by related companies" where the use shall "inure to the benefit of the registrant" and such use "shall not affect the validity of such mark . . . provided such mark is not used . . . to deceive the public."[196] This provision indicates that licensing is permitted if the parties are "related companies" and consumers are not misled. The Lanham Act further defines a "related company" in Section 45 as "any person whose use of a mark is controlled by the owner of the mark with respect to the nature and quality of the goods or services on or in connection with which the mark is used."[197] Based on the language of these provisions, courts have indicated that a trademark owner must exercise control over a

195. *See* FRANK I. SCHECHTER, THE HISTORICAL FOUNDATIONS OF THE LAW RELATING TO TRADE-MARKS (1925); Sidney A. Diamond, *The Historical Development of Trademarks* (pt. 1), 65 TRADEMARK REP. 265 (1975).

196. 15 U.S.C. § 1055 (2012).

197. *Id.* § 1127.

licensee in order to monitor the quality of the product being produced under the licensed mark so that consumers will not be deceived.[198]

However, the statute does not specify the amount of control necessary or the manner in which that control must be exercised. If trademark owners fail to exercise this quality control over licensees, then they have entered into an invalid "naked license" that can lead to consumer confusion.[199] Naked licensing and insufficient quality control could potentially lead to abandonment and forfeiture of the mark if it is determined that the mark has lost its significance and ability to serve its source-identifying function.[200] Irene Calboli has made a persuasive argument that this quality control requirement has been greatly minimized by courts that will rarely order a mark to be forfeited.[201] Courts have found minimal control to be sufficient or have upheld licenses upon proof that product quality remained consistent.[202] Calboli advocates for a revision of the Lanham Act that would allow licensing "with or without 'quality control'" in light of the modern realities of the business of trademark licensing.[203] However, Calboli acknowledges that despite this growing acceptance of minimal or de facto licensing by the courts, "the traditional view that lack of control will result in naked licensing has nonetheless continued to be included in the language of most judicial decisions, proving courts are generally reluctant to abandon quality control as the theoretical standard for valid licensing."[204] As such, the actual licensing requirements under the Lanham Act are uncertain and ambiguous.

Because failure to meet this uncertain quality control requirement could potentially lead to forfeiture of a mark (and has done so on several occasions[205]), trademark owners are routinely advised to include extensive

198. *See* 3 MCCARTHY, *supra* note 32, § 18:42; *see also* Dawn Donut Co. v. Hart's Food Stores, Inc., 267 F.2d 358, 367 (2d Cir. 1959).

199. *See* 3 MCCARTHY, *supra* note 32, § 18:42; *see also Dawn Donut Co.*, 267 F.2d at 367.

200. *See* 3 MCCARTHY, *supra* note 32, § 18:42; *see also Dawn Donut Co.*, 267 F.2d at 366–67.

201. *See* Irene Calboli, *The Sunset of "Quality Control" in Modern Trademark Licensing*, 57 AM. U. L. REV. 341, 366–74 (2007).

202. *See, e.g.*, Land O'Lakes Creameries, Inc. v. Oconomowoc Canning Co., 330 F.2d 667, 670 (7th Cir. 1964) (finding sufficient quality control based on parties' long relationship and lack of consumer complaints about quality); Embedded Moments, Inc. v. Int'l Silver Co., 648 F. Supp. 187, 194 (E.D.N.Y. 1986) (upholding license based on history of trouble-free manufacture by licensee); Hurricane Fence Co. v. A-1 Hurricane Fence Co., 468 F. Supp. 975, 987–89 (S.D. Ala. 1979) (finding sufficient control based on the fact that licensor and licensee were brothers); Joseph Bancroft & Sons Co. v. Shelley Knitting Mills, Inc., 212 F. Supp. 715, 740 (E.D. Pa. 1962) (holding licensor exercised sufficient quality control such that consumers were unlikely to be deceived).

203. *See* Calboli, *supra* note 201, at 396.

204. *See id.* at 376.

205. *See, e.g.*, Ritchie v. Williams, 395 F.3d 283, 290 (6th Cir. 2005) (holding band promoter abandoned trademark by granting naked license); Barcamerica Int'l USA Trust v. Tyfield Imps., Inc., 289 F.3d 589, 596 (9th Cir. 2002) (finding

quality control provisions in all license agreements.[206] A licensor may be advised to add quality control language in the license agreement that would include specific guidelines for use of the mark, detailed specifications for the production of the product, product inspection and approval requirements and procedures, a right on behalf of the licensor to inspect the manufacturing facilities, and penalties for violating these terms. For example, a licensee could be obligated to submit to random inspections by the licensor and to send product samples to the licensor every few months or each time a new "batch" of the product is produced.

This is a burden on both parties. Monitoring the quality of products produced by licensees can be an onerous and time-consuming prospect when imposed on traditional business-to-business transactions. However, the same uncertain quality control requirements would apply to any license that a brand owner enters into with its consumers. IKEA is not terribly burdened by having to monitor the use of its trademarks on the IKEAHackers website, but this is not a feasible solution to allow widespread consumer involvement in branding. If IKEA were to enter into individual licenses with each brand fan seeking to engage in brand-related activities, it would incur substantial initial transaction costs and then the more burdensome costs of monitoring all of those licensees. Entering into a blanket "fan license" would make monitoring trademark use more difficult given the fact that the brand owner would not know the name, location, and type of uses made under all of these de facto authorizations. Such a blanket license would open the door to numerous future defendants raising naked licensing as a defense in any infringement actions brought by the brand owner.

Even if the company were to win on a finding of sufficient quality control, the transaction costs incurred make this option untenable. It is simply easier for a brand owner to comply with current trademark law by sending numerous cease-and-desist letters to discourage consumers from

trademark for wine abandoned due to lack of contractual or actual quality control by licensor); Halo Mgmt., LLC v. Interland, Inc., 308 F. Supp. 2d 1019, 1028–30 (N.D. Cal. 2003) (holding license was naked due to lack of contractual right to monitor or enforce quality); Stanfield v. Osborne Indus., 839 F. Supp. 1499, 1504–05 (D. Kan. 1993) (finding naked license due to lack of actual quality control).

206. *See, e.g.*, KENNETH L. PORT, JAY DRATLER, JR., FAYE M. HAMMERSLEY, TERENCE P. MCELWEE, CHARLES R. MCMANIS & BARBARA A. WRIGLEY, LICENSING INTELLECTUAL PROPERTY IN THE INFORMATION AGE 303 (2d ed. 2005). *See generally* R. CHARLES HENN JR., ALICIA GRAHN JONES, LAUREN SULLINS RALLS & LAUREN A. LINDER, INTELL. PROP. DESK REFERENCE, TRADEMARK LICENSING BASICS 69 (2015), *available at* https://clients.kilpatricktownsend.com/IPDeskReference/Documents/Trademark%20Licensing%20Basics.pdf [https://perma.cc/NS4L-RVR6]; Timothy J. Kelly, *Trademark Licensing Consideration Checklist*, 69 TRADEMARK 475 (2005), *available at* http://www.pli.edu/emktg/toolbox/TM_Licensing14.pdf [https://perma.cc/PQA4-38ER]; Monique L. Ribando & Michael S. Pavento, *Quality Control in Trademark Licensing*, 15 METRO. CORP. COUNSEL, Aug. 2007, at 1, *available at* http://www.kslaw.com/library/publication/mcctrademarklicensing_ribandopavento.pdf [https://perma.cc/5P2R-VB79].

engaging with the owner's marks. Additionally, licensing a consumer's brand-related activity may make it "official" or "authorized" in the minds of other consumers. As discussed above, this could result in more consumer cynicism about such activity and less of the brand loyalty benefits realized by purely consumer-initiated efforts.

4. *Brands and Trademark Fair Use*

Finally, it must be stated that the First Amendment is not irrelevant to the world of trademarks. However, it is rarely helpful to those consumers looking to create brand communities or make non-parody uses of trademarks.[207] Robert Denicola once reasoned that limits on commercial speech were exempt from constitutional scrutiny because "[t]he information conveyed through the use of a trademark generally relates not to the momentous philosophical or political issues of the day, but rather to the details of prospective commercial transactions—the source or quality of specific goods or services."[208] However, it is clear that this statement is overly simplistic in light of the modern world of brands: "Corporations no longer exist in a purely commercial world."[209] Given the highly creative nature of brand development, it follows that the rights of brand owners may conflict at times with the expressive interests of consumers wanting to align themselves with or comment somehow upon that creative content. The consumer's increasing role in brand development is just one part of this puzzle. If a brand owner is unwilling to allow consumer expression related to the brand, then one must look to the First Amendment safe-

207. *See generally* Stacey L. Dogan & Mark A. Lemley, *Parody as Brand*, 47 U.C. DAVIS L. REV. 473 (2013) (arguing trademark fair use may excuse brand parodies from liability in some cases); William McGeveran, *The Imaginary Trademark Parody Crisis (and the Real One)*, 90 WASH. L. REV. 713 (2015) (arguing that reforms aimed at unsubstantiated and threatening pre-litigation tactics would ensure protection for trademark parodies better than trademark doctrine reform). More free speech protection is generally granted to parodies under both trademark and copyright law. *See generally* Dogan & Lemley, *supra*. Parodies utilize some amount of a copyrighted work or trademark in order to comment on or criticize the protected work or mark. *Id.* For this reason, parodies are generally described as being closer to the heart of the First Amendment and the underlying purpose of the fair use defense. *Id.*

208. *See* Desai, *Speech, Citizenry, and the Market*, *supra* note 95, at 480 (quoting Robert C. Denicola, *Trademarks as Speech: Constitutional Implications of Emerging Rationales for the Protection of Trade Symbols*, 1982 WIS. L. REV. 158, 158–59) (internal quotation marks omitted); *see also* Heymann, *supra* note 61, at 696 ("[T]rademarks have recently taken on yet another communicative function: as an element of cultural discourse. Trademarks are now used as a linguistic shorthand in addition to an economic one, as a way of describing something more efficiently or creating a shared discourse through a common cultural referent. When we hear about something being as difficult as 'nailing Jell-O to the wall' or refer to something as the 'Cadillac of its class,' those of us who are familiar with the product and its advertising persona understand what the speaker is saying." (footnote omitted)).

209. *See* Desai, *Speech, Citizenry, and the Market*, *supra* note 95, at 456 (arguing that corporations should be treated as public figures when analyzing corporate speech under First Amendment).

guards built into trademark law to see if they provide adequate avenues for the balancing of these competing interests.

Fair use, under both trademark and copyright law, is often described as a means of integrating those doctrines with the broad expression and speech rights guaranteed by the First Amendment. However, the trademark fair use doctrine has been described as another example of "trademark law's inability to handle the consumer dimension of brands."[210] While it is generally conceded in the abstract that purely expressive, non-competitive uses of trademarks should be permitted as fair uses of the marks, case law indicates that this is often not true in practice.[211] The Ninth Circuit went so far as to state, "Trademarks are part of our common language, and we all have some right to use them to communicate in truthful, non-misleading ways."[212] However, courts struggle to determine what constitutes an "expressive use of a mark and what constitutes fair use" in this "unclear and unstable area of trademark law."[213] Additionally, the expansion of trademark rights in favor of brand owners may have encouraged aggressive tactics by brand owners faced with third-party uses; as a result, much expressive use of brands is stopped before a suit is ever filed.[214]

On its face, the fair use defense to trademark infringement falls into one of two categories: descriptive fair use or nominative fair use. Descriptive fair use applies when the defendant uses the plaintiff's trademark in a descriptive fashion in good faith in order to describe its own goods or services.[215] In *Zatarains v. Oak Grove Smokehouse*,[216] the court held that competitors were permitted to use the term "fish fry" to describe a type of seasoned breading for use when making fried fish recipes under the doctrine of descriptive fair use, despite the fact that Zatarains possessed a trademark for the phrase "Fish Fri" for use with its own products.[217]

---

210. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 1031.

211. *See, e.g.*, Anheuser-Busch, Inc. v. Balducci Publ'ns, 28 F.3d 769, 774–78 (8th Cir. 1994); Deere & Co. v. MTD Prods., Inc., 41 F.3d 39, 43–45 (2d Cir. 1994); Dallas Cowboys Cheerleaders, Inc. v. Pussycat Cinema, Ltd., 604 F.2d 200, 204–05 (2d Cir. 1979); Am. Dairy Queen Corp. v. New Line Prods., Inc., 35 F. Supp. 2d 727, 727–35 (D. Minn. 1998).

212. *See* Toyota Motor Sales U.S.A., Inc. v. Tabari, 610 F.3d 1171, 1185 (9th Cir. 2010).

213. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 1031.

214. *See generally* McGeveran, *Rethinking Trademark Fair Use*, *supra* note 186 (arguing that even when courts reach right decision on fair use, trademark law's uncertainty in area could still chill expressive speech).

215. *See* 15 U.S.C. § 1115(b)(4) (2012). *See generally* KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111 (2004); Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786 (5th Cir. 1983), *abrogated by KP Permanent Make-Up*, 543 U.S. 111 (2004).

216. 698 F.2d 786 (5th Cir. 1983), *abrogated by KP Permanent Make-Up*, 543 U.S. 111 (2004).

217. *See id.* at 796.

Nominative fair use allows a party to use a trademark owned by another when that use is needed to refer to the trademark owner or its products.[218] In *New Kids on the Block v. News America Publishing, Inc.*,[219] for instance, the Ninth Circuit determined that a newspaper's use of the trademarked name of the plaintiff band was a non-infringing nominative fair use, as the newspaper was using the trademark in a truthful manner to conduct a survey that required making a reference to the trademark.[220] In *Toyota Motor Sales U.S.A., Inc. v. Tabari*,[221] Chief Judge Kozinski explained that when the nominative fair use defense is raised in the Ninth Circuit, a court must ask whether "(1) the product was 'readily identifiable without use of the mark; (2) defendant used more of the mark than necessary; or (3) defendant falsely suggested he was sponsored or endorsed by the trademark holder."[222] This test is intended to balance the rights of the defendant to be able to make expressive or descriptive uses of a trademarked word or phrase with the interests of the trademark owner in preventing consumer confusion.[223] If the defendant used the mark when it was unnecessary in violation of prong one, used the mark more than necessary in violation of prong two, or confused consumers into thinking there was a sponsorship relationship in place between the defendant and the trademark owner in violation of prong three, then the defendant does not receive the benefit of the fair use defense.

The Third Circuit applies a modified version of this test with the same basic first and second prongs, but the third prong asks instead whether the "defendant's conduct or language reflect[s] the true and accurate relationship between plaintiff and defendant's products or services."[224] At its heart, the nominative fair use defense, in either formulation, is still concerned first and foremost with trademark law's purported goal of preventing consumer confusion. Whenever a consumer creates a brand community or uses a mark in a similar non-competitive fashion, there will always be the possibility that some unaffiliated consumers will be confused as to the brand community's sponsorship or affiliation with the brand owner.[225] Therefore, given the broad allowance for confusion by most

218. *See* 4 MCCARTHY, *supra* note 32, § 23:11.

219. 971 F.2d 302 (9th Cir. 1992).

220. *See id.* at 310.

221. 610 F.3d 1171 (9th Cir. 2010).

222. *See Toyota Motor Sales U.S.A., Inc.*, 610 F.3d at 1175–76.

223. *See id.* at 1176–77 ("This test 'evaluates the likelihood of confusion in nominative use cases.' It's designed to address the risk that nominative use of the mark will inspire a mistaken belief on the part of consumers that the speaker is sponsored or endorsed by the trademark holder." (citation omitted)).

224. *See* Century 21 Real Estate Corp. v. Lendingtree, Inc., 425 F.3d 211, 222 (3d Cir. 2005).

225. Parodies regarding brands may be the exception. *See* Dogan & Lemley, *supra* note 207, at 508 ("If, as we believe, trademark parodies are a unique and non-replicable form of speech about trademark holders, then they resemble the sorts of uses that nominative fair use seeks to protect. They involve speech about

modern courts, courts will rarely permit consumers to utilize trademarks for the community-building and identity-reinforcing activities inherent in brand communities once an action by a trademark owner has been initiated. The current boundaries on trademark fair use simply ignore "what brand management literature acknowledges and exploits: people use brands in ways that are beyond source identification and beyond legal conceptions of speech."[226]

III. BRANDS AS COPYRIGHTS

Because brands serve creative and expressive purposes beyond the source identification function of trademarks, brands are better conceptualized under the framework of copyrights. Copyright protection is founded on the idea of balancing the interests of the creator of works with the expressive and informational interests of the public. As such, it provides a strong foundation for understanding the full potential of brands as creative works that invite the expressive involvement of consumers.

A. *Brands as Copyrightable Works*

Brands, as creative undertakings, fit neatly under copyright law. Copyright protection automatically attaches to "original works of authorship fixed in any tangible medium of expression."[227] In *Feist Publications v. Rural Telephone Service*,[228] Justice O'Connor stated that "[t]he *sine qua non* of copyright is originality."[229] However, it is not difficult to meet this originality requirement; the work must be independently created by the author and possess a minimal amount of creativity.[230] Again in *Feist*, the Court

---

the trademark holder, and they cannot serve their inherent function—indeed, they can't be a parody—without borrowing from the original.").

226. *See* Desai, *From Trademarks to Brands*, *supra* note 19, at 1032.

> Community dimensions of brands fare no better under trademark law than consumer dimensions and are possibly in worse shape. By their nature, community brand situations involve a group of consumer enthusiasts who take it upon themselves to define the brand. This group may engage in one or more activities, including building Web sites, holding meetings, writing polemics, creating artwork, and producing branded merchandise. The mark in question will be prominently displayed, discussed, and distributed. In many cases the mark will appear as or near how it was originally displayed precisely because of the power of that context. Trademark law, however, asks whether these acts are likely to cause confusion, and the nature of the test for that question does not provide room for this sort of community action.

Id. (footnotes omitted).

227. 17 U.S.C. § 102(a) (2012).

228. 499 U.S. 340 (1991).

229. *See id.* at 345; *see also* L. Ray Patterson & Craig Joyce, *Monopolizing the Law: The Scope of Copyright Protection for Law Reports and Statutory Compilations*, 36 UCLA L. REV. 719, 763 & n.155 (1989) ("The originality requirement is *constitutionally mandated* for all works.").

230. *See Feist*, 499 U.S. at 345; *see also* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 547 (1985).

clarified, "To be sure, the requisite level of creativity is extremely low; even a slight amount will suffice. The vast majority of works make the grade quite easily, as they possess some creative spark, 'no matter how crude, humble or obvious' it might be."[231]

This originality requirement is so minimal that the Supreme Court has said that "[c]reation of a nonfiction work, even a compilation of pure fact, entails originality."[232] However, that compilation of pure fact would require originality in the selection, coordination, or arrangement of the non-protectable factual information.[233] Therefore, courts have held that the creators of photographs may claim copyright in their work,[234] as can the creator of a taxonomy listing dental insurance billing codes,[235] and the author of real estate maps of a Texas county.[236] While it is obvious that the originality standard presents only a small hurdle for those seeking copyright protection, such protection has been denied if this minimal requirement is not met. The phone book at issue in *Feist* was denied copyright protection based on a lack of originality due to the typical, alphabetical manner in which it presented factual information.[237] Similarly, in *Southco, Inc. v Kanebridge Corp.*,[238] copyright was denied for a listing of codes used in a replacement parts catalog that used a conventional organizational system.[239]

While it has been said, as mentioned above, that trademarks do not "depend upon novelty, invention, discovery, or any work of the brain" and require "no fancy or imagination, no genius, no laborious thought,"[240] this statement is simply not true regarding brands (and may have never been entirely true of trademarks). Advertisers and brand owners view brand efforts as storytelling. Through advertisements in print, online, on television, and other media, the company tells the story of the brand as they want it to be seen by consumers. These brand stories have all of the elements needed to justify copyright protection.

First, brand stories are based on independent marketing campaigns that are created by the marketing firms or in-house marketing depart-

231. *See Feist*, 499 U.S. at 345 (quoting 1 M. NIMMER & D. NIMMER, COPYRIGHT § 1.08(C)(1) (1990)).

232. *See Harper & Row*, 471 U.S. at 547.

233. *See Feist*, 499 U.S. at 345.

234. *See* Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884) (establishing precedent that photographs are original works capable of receiving copyright protection).

235. *See* Am. Dental Ass'n v. Delta Dental Plans Ass'n, 126 F.3d 977, 981 (7th Cir. 1997).

236. *See* Mason v. Montgomery Data, Inc., 967 F.2d 135, 142 (5th Cir. 1992) (establishing precedent that real estate maps are capable of receiving copyright protection).

237. *See Feist*, 499 U.S. at 362–63.

238. 390 F.3d 276 (3d Cir. 2004).

239. *See id.* at 285.

240. *See In re Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

ments, which are employed by the brand owner.[241] Each print advertisement, commercial, or website is therefore independently created and, as creativity is a low bar, they each possess the minimal amount of creativity required for copyright protection. If copyright can be obtained for a photograph or a taxonomic listing of billing codes, then it follows that print advertisements and television or radio commercials would also possess the requisite creativity. However, this was not always so settled. In 1891, the Supreme Court in *Higgins v. Keuffel*[242] held that commercial writings like product labels would not be eligible for copyright protection if they could have "no possible influence upon science or the useful arts," based on the wording of the Constitution granting federal authority over the copyright system.[243] The Court asserted that copyright protection was available only for works that had value "as a composition" apart from the product which it advertised or described.[244] For copyright protection to attach, the writing must serve a purpose other than that of a "mere advertisement."[245]

However, the Court's view of the copyrightability of advertising changed course shortly thereafter. In 1903, the Supreme Court upheld copyright protection for an illustration used in an advertisement for a circus in *Bleistein v. Donaldson Lithographing Co.*[246] The Court reiterated that the pictorial works in question were original and creative enough to justify copyright protection, as they contained more originality than "directories and the like, which may be copyrighted."[247] The use of the article does not change its nature as a creative work. Therefore, "[a] picture is none the less a picture and none the less a subject of copyright that it is used for an advertisement."[248]

Subsequent cases recognized that this was a significant change in the treatment of advertising: "The *Bleistein* Case established a new and liberal standard with respect to the originality or artistic merit required to entitle illustrated advertising matter—now frequently referred to as 'commercial

241. It is not necessary to go into a detailed discussion of ownership of the resulting work for the purposes of this Article. Absent a contract assigning ownership of the advertising work to the brand owner, it is likely that the outside marketing firm would be considered an independent contractor with ownership of the resulting work. However, under the works made for hire provision of the Copyright Act, advertising created by in-house marketing employees would automatically be owned by the brand owner's employer. 17 U.S.C. § 201(b) (2012).

242. 140 U.S. 428 (1891).

243. *See id.* at 431; *see also* Lisa P. Ramsey, *Intellectual Property Rights in Advertising*, 12 MICH. TELECOMM. & TECH. L. REV. 189, 216 (2006) (questioning utilitarian justifications of copyright protection for advertising).

244. *See Higgins*, 140 U.S. at 431.

245. *See id.*

246. *See* Bleistein v. Donaldson Lithographing Co., 188 U.S. 239, 252 (1903).

247. *See id.* at 250.

248. *See id.* at 251; *see also* SHL Imaging, Inc. v. Artisan House, Inc., 117 F. Supp. 2d 301, 311 (S.D.N.Y. 2000) ("That the photographs were intended solely for commercial use has no bearing on their protect[a]bility.").

art'—to the protection of copyright statutes."[249] Based on this low originality requirement and new liberal approach to advertising materials, the court in *Ansehl v. Puritan Pharmaceutical Co.*[250] upheld the validity of a copyright on an advertisement for cosmetic and toilet articles.[251] This more liberal standard was implemented again when the Ninth Circuit in *Drop Dead Co. v. S.C. Johnson & Son, Inc.*[252] upheld copyright protection for the label of a commercial cleaning product. The eligibility of copyright protection of advertising is widely accepted now.[253] In *Ets-Hokin v. Skyy Spirits, Inc.*,[254] the Ninth Circuit went so far as to grant copyright protection to photographs of a vodka bottle taken for advertising purposes while refusing to extend such protection to the underlying bottles themselves, based on its determination that the bottle was a purely utilitarian "useful article" incapable of copyright protection.[255]

While courts are now willing to accept print and television advertisements as original works of authorship deserving of copyright protection, those advertisements must still be "fixed in any tangible medium of expression . . . ."[256] Fixation is sufficient if the work "can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[257] The work must be capable of being "perceived, reproduced, or otherwise communicated for a period of more than transitory duration."[258]

Again, this is not a difficult requirement to meet for most advertising. Printed advertisements, whether they are contained in newspapers, magazines, or even standalone pamphlets, are all clearly fixed on paper. Television and radio commercials (that are not broadcast live) are fixed as a "motion picture" or "sound recording" (as those terms are used in the

249. *See* Ansehl v. Puritan Pharm. Co., 61 F.2d 131, 134–35 (8th Cir. 1932) (emphasis added).

250. 61 F.2d 131 (8th Cir. 1932).

251. *See id.* at 136.

252. 326 F.2d 87 (9th Cir. 1963).

253. *See, e.g.*, Comptone Co. v Rayex Corp., 251 F.2d 487, 488 (2d Cir. 1958) (extending copyright protection to sunglass display cards and leaflets); Griesedieck W. Brewery Co. v Peoples Brewing Co., 56 F. Supp. 600, 606 (D. Minn. 1944) (extending copyright protection to portion of beer can label); Advertisers Exch., Inc. v Bayless Drug Store, Inc., 50 F. Supp. 169, 169 (D.N.J. 1943) (extending copyright protection to advertisements).

254. 225 F.3d 1068 (9th Cir. 2000).

255. *See id.* at 1080. Under the Copyright Act, a useful article "is an article having an intrinsic utilitarian function that is not merely to portray the appearance of the article or to convey information." 17 U.S.C. § 101 (2012) (defining *useful article*). A useful article is only eligible for copyright protection "if, and only to the extent that, such design incorporates pictorial, graphic, or sculptural features that can be identified separately from, and are capable of existing independently of, the utilitarian aspects of the article." *Id.* (defining *pictorial, graphic, and sculptural works*).

256. *See* 17 U.S.C. § 102(a).

257. *See id.*

258. *See id.* § 101 (defining *fixed*).

Copyright Act) in the same manner as other works falling under these definitions.[259] Even websites and online advertisements are fixed in electronic computer memory and are thus protectable.[260]

However, this requirement may exclude less structured forms of advertising from protection. For example, live promotional events and campaigns would not be fixed in a tangible medium of expression. This has long held true for live musical and dramatic performances, but such performances may receive copyright protection in other ways. If the live performance is based on a written script, then the script is fixed and protectable, which indirectly protects the performance.[261] Brand owners could also use this strategy. Additionally, "[a] work consisting of sounds, images, or both, that are being transmitted" will be considered fixed under the Copyright Act if "fixation of the work is being made simultaneously with its transmission."[262] Thus, recording live promotional activities would allow brand owners to receive copyright protection. This will be of special concern to brand owners engaged in nontraditional or viral marketing.

While copyright protection is denied to advertising activities that fail to meet the fixation requirement, it will also not prevent third parties from using and repeating some of the information contained in protected advertisements. Copyright protection will not be afforded to purely factual information or to the ideas contained within a work. In *Harper & Row, Publishers v. Nation Enterprises*, the Court stated with certainty that "[n]o author may copyright his ideas or the facts he narrates."[263] This is codified in Section 102(b) of the Copyright Act, which states:

> In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.[264]

---

259. *See id.* (defining *motion pictures* and *sound recordings*). Motion pictures and sound recordings are both listed as works of authorship expressly falling under the subject matter of copyright. *See id.* § 102(6)–(7).

260. These would likely be classified under the Copyright Act as "audiovisual works," which are defined as "works that consist of a series of related images which are intrinsically intended to be shown by the use of machines, or devices such as projectors, viewers, or electronic equipment, together with accompanying sounds, if any, regardless of the nature of the material objects, such as films or tapes, in which the works are embodied." *See id.* § 101 (defining *audiovisual works*).

261. *See id.* § 106(5). Public performance of written work (i.e., "literary, musical, dramatic, and choreographic works") is one of the exclusive rights granted under copyright. *Id.*

262. *See id.* § 101.

263. *See* Harper & Row, Publishers, Inc. v. Nation Enters., Inc., 471 U.S. 539, 556 (1985).

264. 15 U.S.C. § 102(b).

Therefore, factual information contained in an advertisement, such as facts pertaining to product specifications, qualities, or prices, will not be protected and may be freely repeated by consumers and competitors.

Additionally, copyright protection only extends to the expressive or creative aspects of the advertisement and would not protect the ideas contained therein. In *Reed-Union Corp. v. Turtle Wax, Inc.*,[265] the Seventh Circuit refused to extend copyright protection to the general idea or theme of polishing an old car and the idea of making claims based on laboratory testing, as such ideas were used in similar, competing car wax commercials.[266] In *Perma Greetings, Inc. v. Russ Berrie & Co.,*[267] the court refused to extend copyright protection to items displaying the general ideas of friendship, sunshine, and flowers where the defendant sold products containing similar images and wording to convey the same themes.[268] It follows that a brand owner cannot prevent a competitor from, for example, using the general idea of a cartoon character as a brand mascot in advertising for children's cereal.[269]

Furthermore, copyright cannot be used to grant monopolies over commonly used phrases, slogans, or short expressions lacking sufficient originality, or over any specific word necessary to express a particular point. The Copyright Office has specifically clarified that certain types of works are excluded from copyright eligibility, including "[w]ords and short phrases such as names, titles, and slogans . . . ."[270] In *Perma Greetings*, the court also addressed this issue and denied copyright protection for the short phrases, like "hang in there," that the plaintiff had included on its mug-type coasters.[271] Based on these copyright limitations, brand owners would not be able to rely on copyright protection for short phrases and slogans used in branding materials. However, such short phrases might be eligible for trademark protection if they were used to signify the source of the good or service being offered.

Additionally, the judicially created "merger doctrine" denies copyright protection to any work where the wording of the work represents the only way, or one of but a few ways, of expressing the idea underlying the work.[272] In such cases, the idea and its expression are said to merge to-

---

265. 77 F.3d 909 (7th Cir. 1996).

266. *See id.* at 913–14.

267. 598 F. Supp. 445 (E.D. Mo. 1984).

268. *See id.* at 449.

269. This is clearly an example of an unoriginal idea, given all of the cartoon mascots currently being used for children's cereal, including, for example, Trix Rabbit, Tony the Tiger, Boo Berry, Franken Berry, Lucky the Leprechaun, Count Chocula, Sonny the Cuckoo Bird, Dig 'Em Frog, Snap, Crackle, Pop, and Toucan Sam.

270. *See* 37 C.F.R. § 202.1(a) (2015).

271. *See Perma Greetings*, 598 F.Supp. at 448–49 (internal quotation marks omitted).

272. *See* Morrissey v. Procter & Gamble Co., 379 F.2d 675, 678–79 (1st Cir. 1967).

gether; therefore, the First Amendment interests of those seeking to convey that particular idea would be harmed by the protection of the merged expression.[273] Similar reasoning led to the denial of copyright protection for the baked goods preparation and serving directions at issue in *Kitchens of Sara Lee, Inc. v. Nifty Foods Corp.*[274] There are very few ways to express the underlying idea that one should thaw a frozen cake and then slice it into individual servings.[275] Therefore, routine phrases and formulaic wording, like that used by brand owners in promotional descriptions or sweepstakes rules, would not be protected due to the merger doctrine.[276]

While each fixed advertising work is copyrightable, subject to the limitations discussed above, each work is part of a larger brand campaign or overall strategy. They are all part of the larger brand story that is being told in pieces, in various media, to consumers. As these pieces are not explicitly created as "serial works" (such that they lack the requisite sequential numbering that one would find in a daily edition of a newspaper or a single issue of a comic book series), they cannot be registered as serial works under the Copyright Act.[277] However, courts can still consider all of the pieces of the brand story as a whole in an infringement analysis. Applying such a big picture approach to advertising would be novel, but this approach has been taken with regard to television programs or books written as a series. For example, the Second Circuit decided to treat the eighty-six copyrighted episodes of the *Seinfeld* television series as a single work for the purposes of determining infringement in *Castle Rock Entertainment, Inc. v. Carol Publishing Group, Inc.*[278] The court in *Warner Brothers Entertainment, Inc. v. RDR Books*[279] followed the reasoning established in *Castle Rock* when it analyzed infringement of the seven *Harry Potter* books as a whole rather than analyzing the infringement of each individual novel.[280]

---

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

274. 266 F.2d 541, 545 (2d Cir. 1959).

275. *See id.*

276. *Compare Morrissey*, 379 F.2d at 678–79 (employing merger doctrine), *with* Cont'l Cas. Co. v. Beardsley, 253 F.2d 702, 706 (2d Cir. 1958) (granting insurance forms thin copyright protection against purely duplicative copying).

277. *See Copyright Registration for Single Serial Issues*, CIRCULAR 62.0811 (U.S. Copyright Office, Wash., D.C.), Aug. 2011, at 1, *available at* http://www.copyright.gov/circs/circ62.pdf [https://perma.cc/QVY8-43ZZ]. Serial works are "works issued or intended to be issued in successive parts bearing numerical or chronological designations and intended to be continued indefinitely." *Id.* This would include works such as newspapers, journals, and bulletins. In some instances, serial works may be registered as a group (rather than individually) thereby saving the author some registration costs.

278. 150 F.3d 132, 138 (2d Cir. 1998).

279. 575 F. Supp. 2d 513 (S.D.N.Y. 2008).

280. *See id.* at 535 n.14 (analyzing "the amount of expression copied from the *Harry Potter* series in the aggregate, rather than from each individual novel in the series" and finding that defendant's encyclopedia of *Harry Potter* terminology and characters infringed copyrighted works).

However, one could also view all of the branding messages produced by a brand owner as interconnected pieces rather than a unitary whole. Subsequent brand activities that follow an initial branding effort could be viewed as derivative works of that first copyrightable work. For an example of how derivative branding could play out, one could look to Apple's now iconic "Get a Mac" campaign, in which numerous commercials contrasted the personification of a PC-type computer, portrayed by comedian and author John Hodgman, with Apple's Mac computer, portrayed by actor Justin Long.[281] The first "Get a Mac" commercial aired in 2006 and was intended to communicate a clear branding story to consumers: Mac computers are fun, innovative, and user-friendly, and Mac users are young, energetic, and hip, in contrast to the traditional, stuffy, older users of Microsoft (i.e., PC) computers.[282] This same message was continued through the other sixty-six commercials in the campaign, which ran until 2009.[283] Each of these subsequent commercials was likely registered under an individual copyright, but each could also be seen as a derivative work of that first commercial. The Copyright Act defines a derivative work as

> a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a "derivative work[.]"[284]

This distinction between an independently copyrighted work and a copyrightable derivative work may not have much practical effect when the works at issue all arise from the brand owner directly. In both cases, the brand owner owns any possible copyrights and controls use of the works. However, the distinction becomes much more important when the derivative work is consumer-initiated. Such is the case with websites like the IKEAHackers site, which was created by an IKEA customer for the benefit of the website owner and other IKEA brand fans.

---

281. *See* Tim Nudd, *Apple's 'Get a Mac,' the Complete Campaign: Hodgman and Long, Ad Characters for the Ages,* ADWEEK (Apr. 13, 2011, 6:49 PM), http://www.adweek.com/adfreak/apples-get-mac-complete-campaign-130552 [https://perma.cc/MCN7-J7X7]; *see also* Lindsay Kolowich, *Mac or PC? A Brief History of Apple & Microsoft's Ad War,* HUBSPOT BLOGS (Jan. 9, 2015, 8:00 AM), http://blog.hubspot.com/marketing/mac-pc-advertising-history [https://perma.cc/236P-H9C7].

282. *See* Nudd, *supra* note 281; *see also* Eric Jaffe, *How Apple's Famous "I'm a Mac" Ads Branded Fanboys for Life,* FAST CO. (Feb. 26, 2014, 8:00 AM), http://www.fastcodesign.com/3026521/evidence/apples-famous-ads-created-an-im-a-mac-effect [https://perma.cc/3MWL-AP46].

283. *See* Nudd, *supra* note 281.

284. 17 U.S.C. § 101 (2012).

While it is generally true that the Copyright Act "grants the author of a derivative work copyright protection in the incremental original expression he contributes" to the original work,[285] copyright protection will only attach to that original expression if the derivative work was created with the consent of the original copyright owner.[286] This means that "[t]o be copyrightable, a derivative work must not be infringing."[287] The owner of the copyright to the original, underlying work possesses the exclusive rights to (1) reproduce the work; (2) prepare derivative works based upon the work; (3) distribute copies of the work to the public; (4) perform the work publicly; (5) display the work publicly; and (6) perform the work publicly by means of a digital audio transmission in the case of sound recordings.[288] Violating any of these exclusive rights would constitute infringement of the copyright; therefore, it constitutes infringement if a fan creates a derivative work based on copyrighted brand material without the brand owner's consent. For example, in *Anderson v. Stallone*,[289] the plaintiff was denied copyright protection for his screenplay for *Rocky IV*.[290] The court held that the plaintiff lacked the consent of the owner of the copyright to the original three *Rocky* films and therefore could not receive copyright protection for the original elements he added in his derivative work.[291]

This is an important aspect of copyright law from the perspective of brand owners. Consumers will not possess a copyright on any of their derivative works created using the copyrightable brand materials or, as argued herein, the copyrightable expression contained within the larger brand story. Thus, brand owners need not worry about genericide and risks of abandonment in the same way that they worry when consumer involvement in branding is viewed under the umbrella of trademark law. Consumers are not creating competing copyrightable content based on the owner's trademark and brand. If such consumer-initiated branding

---

285. Schrock v. Learning Curve Int'l, Inc., 586 F.3d 513, 518 (7th Cir. 2009).

286. *See id.* at 523 ("This means the author of a derivative work must have permission to make the work from the owner of the copyright in the underlying work."); *see also* Gracen v. Branford Exch., 698 F.2d 300, 302 (7th Cir. 1983) ("[E]ven if Miss Gracen's painting and drawings had enough originality to be copyrightable as derivative works *she* could not copyright them unless she had authority to use copyrighted materials from the movie.").

287. *See Schrock*, 586 F.3d at 522; *see also* Pickett v. Prince, 207 F.3d 402, 406–07 (7th Cir. 2000) (stating plaintiff did not have permission to make derivative work based on defendant's copyrighted design and was therefore infringing).

288. *See* 17 U.S.C. § 106.

289. No. 87–0592 WDKGX, 1989 WL 206431 (C.D. Cal. Apr. 25, 1989).

290. *See id.* at *5–6. Additionally, this case reiterates that individual characters within a work may be afforded copyright protection. This is relevant to brand since well-defined characters in advertising campaigns could potentially be copyrightable. *See* Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (noting copyright protection can be granted to a character if the character is developed with enough specificity so as to rise to level of protectable expression).

291. *See Anderson*, 1989 WL 206431, at *5–6.

occurs without the express consent of the brand owner, then the brand owner has the option under copyright law to bring suit for infringement of the derivative work right. In this way, consumer branding activities can be viewed and treated in the same way that copyright owners currently treat fanfiction.

Fanfiction has been characterized as non-professional writing by a fan that uses as its source an identifiable cultural work such as a novel or a movie.[292] An author has the freedom to bring a copyright suit against the creator of fanfiction based on the author's work or to turn a blind eye to such activities altogether without the risks of genericide and abandonment built into trademark law.[293] Neil Gaiman, author of numerous best-selling books, explicitly allows fanfiction based on his books explaining, "I don't believe I'll lose my rights to my characters and books if I allow/fail to prevent/turn a blind eye to people writing say Neverwhere fiction, as long as those people aren't, say, trying to sell books with my characters in."[294] He sees the benefit to his fans: "[G]iven how much people enjoy it, it's obviously doing some good. It doesn't bother me."[295] However, other copyright owners take a different view of fanfiction, which constitutes an unlicensed derivative work of their own creation. Author Anne Rice explained, "I do not allow fan-fiction. The characters are copyrighted. It upsets me terribly to even think about fan-fiction with my characters. I advise my readers to write your own original stories with your own characters. It is absolutely essential that you respect my wishes."[296] Diana Gabaldon, author of the *Outlander* series, thinks fanfiction is immoral and has said that she "want[s] to barf whenever [she has] inadvertently encountered some of it involving [her] characters."[297]

Copyright owners have the discretion to allow fanfiction. As fanfiction and branding both entail consumers engaging with a product in an

---

292. *See* Rebecca Tushnet, *Legal Fictions: Copyright, Fan Fiction, and a New Common Law,* 17 LOY. L.A. ENT. L.J. 651, 655 (1997) ("'Fan fiction,' broadly speaking, is any kind of written creativity that is based on an identifiable segment of popular culture, such as a television show, and is not produced as 'professional' writing."). *See generally* Stacey M. Lantagne, *The Better Angels of Our Fanfiction: The Need for True and Logical Precedent,* 33 HASTINGS COMM. & ENT. L.J. 159 (2011).

293. However, copyright owners who ignore fanfiction creations based on their work could run the risk of having a future claim against the fanfiction author dismissed under the equitable defense of laches. This defense can be raised when a plaintiff is said to have unreasonably delayed in asserting a claim, and a court may hold that delay unfairly prejudices the defendant. *See* 30A C.J.S. EQUITY § 138 (2015).

294. *See* Neil Gaiman, *Neil Gaiman's Opinion on Fanfiction,* TUMBLR (Apr. 24, 2012, 6:46 PM), http://neil-gaiman.tumblr.com/post/21746253134/neil-gaimans-opinion-on-fanfiction [https://perma.cc/G2RY-SCPV].

295. *See id.*

296. *See* George R.R. Martin, *Book Post: How Authors Feel About Fan-fiction,* ONTD (Apr. 19, 2012, 12:04 PM), http://ohnotheydidnt.livejournal.com/68332629.html#ixzz3HrLzXDJr [https://perma.cc/55KM-8FB8] (internal quotation marks omitted).

297. *See id.* (internal quotation marks omitted).

expressive fashion, and both could lead to increased sales for the underlying copyright or brand owner, it seems logical that the discretion about when to stop such activity should be reserved for the copyright or brand owner. Branding is an expressive activity entitled to copyright protection, so it follows that brand owners should be able to exercise discretion regarding any infringement of their copyrightable brands. If consumer-initiated brand activities occur without the brand owner's consent, then brand owners have the choice under copyright law to turn a blind eye to the activity, to bring suit for infringement of the derivative work right, or to enter into a blanket or individual license with such consumers on terms acceptable to the brand owner.

B. *Applying Copyright Fair Use to Brands*

Where brand owners deny consumers the ability to use branded materials, consumer-initiated branding activities would be a violation of copyright law. Copyright infringement occurs whenever any one of the exclusive rights of a copyright owner is violated. However, some violations are expressly excused in the interest of the First Amendment and the public domain. The constitutional mandate that a federal system of copyright protection be created to promote the progress of science and the arts has long instructed courts that "copyright is intended to increase and not to impede the harvest of knowledge."[298]

Therefore, there is a natural tension that exists between increasing knowledge and hindering it. This tension is played out in the balance that is struck between the Copyright Act and the First Amendment. The Supreme Court has acknowledged "that some restriction on expression is the inherent and intended effect of every grant of copyright . . . ."[299] However, the refusal of copyright law to protect facts and ideas, as well as the statutory fair use defense, are recognized as copyright law's limitations that balance the restriction on expression granted with a copyright monopoly with the free speech requirements of the First Amendment.[300] The Supreme Court routinely praises this balance: "All reproductions of the work, however, are not within the exclusive domain of the copyright owner; some are in the public domain. Any individual may reproduce a copyrighted [work] for a 'fair use'[;] the copyright owner does not possess the exclusive right to such a use."[301] The fair use defense was codified in

298. *See* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 545 (1985).

299. Golan v. Holder, 132 S. Ct. 873, 876 (2012).

300. *See* Eldred v. Ashcroft, 537 U.S. 186, 219 (2003); *see also Golan*, 132 S. Ct. at 890.

301. *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 433 (1984); *see also Harper*, 471 U.S. at 549 ("Fair use was traditionally defined as 'a privilege in others than the owner of the copyright to use the copyrighted material in a reasonable manner without his consent.'").

the Copyright Act and states that courts analyzing the defense should consider the following factors:

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
> (2) the nature of the copyrighted work;
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> (4) the effect of the use upon the potential market for or value of the copyrighted work.[302]

In a fair use analysis, the court must first determine "the purpose and character of the use" of the copyrighted work.[303] The purpose of the use refers to whether the new work was created for commercial or for nonprofit reasons.[304] "If a new work is used commercially rather than for a nonprofit purpose, its use will less likely qualify as fair."[305] Additionally, courts must determine the character of the use by asking whether the nature of the work has changed, meaning whether the alleged infringer has created something fundamentally different from the original work.[306] The fair use doctrine has long been thought to preclude a use that would simply supersede the use of the original.[307] Courts have stated that "the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use."[308] To be transformative, the new work must add "something new, with a further purpose or different character, altering the first [work] with new expression, meaning or message."[309] This first factor will entail a case-by-case factual analysis if applied to consumer branding activities and brand communities. Many brand communities, like the community formed around the IKEAhackers website, are not created as commercial undertakings.

However, the ease of adding revenue-generating advertisements to a brand community website may result in some commercial aspect to these otherwise non-profit groups. For example, the IKEAHackers website creator eventually added advertising to the site in order to compensate for the time commitment of running the community once it became larger and

302. *See* 17 U.S.C. § 107 (1)–(4) (2012).

303. *See id.* § 107(1).

304. *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584 (1994).

305. Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 198 (3d Cir. 2003).

306. *See Campbell*, 510 U.S. at 579 (discussing degree of transformation and how it factors into analysis).

307. *See* Folsom v. Marsh, 9 F. Cas. 342, 344–45 (C.C.D. Mass. 1841) (No. 4,901); *accord* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 551 (1985).

308. *See Campbell*, 510 U.S. at 579.

309. *See id.*

more complex. Commercial activity such as this may tilt the scales against a finding of fair use. However, this one factor alone would not be dispositive, so courts would be asked to undertake a more nuanced analysis.

The same uncertainty is present regarding whether consumer-branding activities could be considered transformative. If they are determined to be transformative, that would lessen the effect of any commercial aspect of their use. It is certainly possible that some consumers would build upon a brand's trademarks and brand story to create "something new," which could include a new direction for the brand to take. In the example of the Saab brand community discussed above, the community members actively displayed the Saab trademark and circulated Saab promotional and educational materials. This use of copyrighted materials seems very straightforward and not very transformative in the traditional sense. However, the infringing use could certainly entail a new purpose or meaning. As brand communities have been studied and shown to possess the same social characteristics of other types of communities, the purpose underlying such communities could be described as social and psychological, rather than the sales-oriented commercial purpose motivating the brand owner. It is unclear whether this would be enough to call brand communities transformative. If so, this would lessen the fair use impact of any commercial revenue generated and cause this factor to weigh in favor of the fair use defense.

Under an analysis of the second statutory fair use factor, the court must take into consideration the nature of the copyrighted work.[310] An examination of this factor recognizes the reality that "some works are closer to the core of intended copyright protection than others."[311] For example, *Feist* held that creative works are entitled to more stringent copyright protection than factual works because there is a presumptively greater need to disseminate facts to the public.[312] However, simply because a work is creative in nature does not preclude the fair use defense. Additionally, "the scope of fair use is narrower with respect to unpublished works" in order to protect the author's exclusive right to first publication.[313] As discussed above, when analyzing the copyrightability of branding activities taken by the brand owner, most brand activities can be seen as creative in nature. As they are produced specifically to inform and persuade consumers, it is clear that they are not subject to the increased protection granted to unpublished works. However, the protectable, creative expression contained in branding works would be entitled to the enhanced protection for creative works. Therefore, this factor would consistently tilt in favor of the brand owner over the infringing consumer.

310. *See* 17 U.S.C. § 107(2) (2012).

311. *See Campbell*, 510 U.S. at 586.

312. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 363 (1991) (holding copyright protects works with *de minimus* level of creativity).

313. *See* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 564 (1985).

Analysis of the "amount and substantiality of the portion used," under the third fair use factor, requires consideration of both the quantity as well as the quality and importance of the portion of the copyrighted work used "in relation to the copyrighted work as a whole . . . ."[314] There is no bright-line rule specifying exactly how much copying is "too much" to be considered fair use. Courts have found that taking a few seconds from several minutes of video footage was substantial copying,[315] yet the use of an entire copyrighted Barbie doll in artistic parody photographs was deemed permissible.[316]

In *Harper*, taking only a small number of quotations from an autobiographical book was found to be unfair due to the fact that the portions taken captured the commercial heart or essence of the copyrighted work.[317] Therefore, courts must determine in each instance whether the use made by the defendant was reasonable and fair in terms of both quantity and quality. This will also need a fact-intensive, case-by-case analysis. In the case of the IKEAHackers website, which included some reference to IKEA's trademarks and goods for sale, most of the content was user-generated, so the extent of copyrighted content used might be considered rather small. Conversely, the Saab brand community actively circulated copyrighted materials produced by the brand owner for informational and community membership purposes, so their level of use might be considered more extensive.

The fourth fair use factor is "the effect of the use upon the potential market for or value of the copyrighted work."[318] In *Harper*, the court said "This last factor is undoubtedly the single most important element of fair use."[319] In order to determine the effect on the market for a plaintiff's copyrighted work, the court must consider the extent of the market harm caused by the alleged infringer as well as the magnitude of harm that could be caused if such infringement was widespread.[320] The extent of harm refers not only to the obvious harm caused if an infringing use takes some of the original market, but also harm to the market for derivative works that could have been rightfully produced by the copyright holder.[321] Courts must attempt to "strike a balance between 'the benefit

314. *See* 17 U.S.C. § 107(3); *see also Harper*, 471 U.S. at 564–65.

315. *See* L.A. News Serv. v. CBS Broad., Inc., 305 F.3d 924, 940–42 (9th Cir. 2002).

316. *See* Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 804 (9th Cir. 2003).

317. *See Harper*, 471 U.S. at 566; *see also* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 587 (1994) ("[T]aking the heart of the original and making it the heart of a new work [i]s to purloin a substantial portion of the essence of the original." (internal quotation marks omitted)).

318. *See* 17 U.S.C. § 107(4).

319. *See Harper*, 471 U.S. at 566.

320. *See Campbell*, 510 U.S. at 602–03.

321. *See Harper*, 471 U.S. at 541.

the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied.' "[322]

Since this fourth factor has been called the most important factor in the fair use analysis, it is noteworthy that consumer-initiated branding has been shown to have positive effects on the market for the brand owner's products. Analysis of this factor is imperfect, as brand-related copyrighted works generally do not have value in and of themselves. A commercial or branded website may cost the brand owner a substantial amount to create, and that cost is not recouped by sales of the commercial or through website memberships in the way that might apply to sales of copyrighted books or films. For the purposes of branding works, the relevant market value is the marketing value. The brand owner is looking for a return on its branding investment in the form of increased sales of the affiliated product. If consumer-initiated branding activities assist in this effort to increase product sales, then this factor should weigh heavily in favor of a finding of fair use.

In any copyright fair use analysis, there is bound to be uncertainty about how the various factors will balance out in the eyes of a court. However, consumer-initiated brand uses seem to have a greater likelihood of success under copyright fair use than under trademark fair use. The multi-factor test under copyright fair use allows for a more nuanced assessment of whether a consumer's use of a brand story is unfairly taking advantage of the underlying work or likely to replace the market for the underlying work. In many instances, the branded work would receive a benefit from the consumer's derivative work, and fair use should prevail. A wise brand owner would only challenge these consumer derivative works in rare instances; therefore, the concerns of trademark scholars regarding the potential of trademark fair use to chill free speech would be minimized as well.

IV. CONCLUSION

A copyright framework for brands offers a number of benefits to brand owners, consumers, and those seeking to reign in the expansion of trademark law. Because brands are a modern phenomenon, there will always be some inconsistencies when trying to accommodate brand development under established legal paradigms.[323] Copyright law may need time and modifications to adjust to brands, but this adjustment should be less disruptive than the expansion being forced upon trademark law. As it stands, copyright law is well-suited to recognize the original and creative work that is invested in the development of a brand. Individual advertis-

322. *See* Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 804–05 (9th Cir. 2003) (quoting Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc., 109 F.3d 1394, 1403 (9th Cir. 1997)).

323. In a forthcoming project, I explore whether an entirely new "brandright" regime would better address the needs of brand owners and respect the established boundaries of trademark and copyright law.

ing works already receive some protection under copyright law, so incorporating branding more generally is a logical next step. Additionally, reframing brands as copyrighted works allows for the treatment of consumer-initiated brand activities in a fashion similar to fanfiction. Brand owners recognize the value of consumer involvement in their marketing efforts and desire the ability to utilize brand communities to further develop brand loyalty. Copyright law allows brand owners to encourage consumer involvement in branding without the fear of losing one's rights, which brand owners currently experience under trademark law. Copyright law allows brand owners to have discretion over the kinds of consumer activities that will be permitted or ignored, while also allowing some freedom for unauthorized expressive uses of a brand story through copyright fair use. This is a better approach for both brand owners and consumers.

Reframing brands as copyrightable works is also better for trademark law. Brand theory explains how the modern expansion of trademark law, which has been maligned by numerous trademark scholars and practitioners, is an effort to accommodate brands under the rubric of trademark. However, trademark law has adopted a conception of brands that is dated and that does not recognize the high level of consumer involvement in modern branding efforts. Trademark law now faces the paradoxical dilemma of protecting brand owners' investment in developing brands while stifling the ability of those owners to pursue consumer involvement in branding, which has been described in marketing research as the required next step for companies seeking to respond to a cluttered marketplace and increasing consumer cynicism. Trademark law's established requirements that prevent genericide, police third-party trademark use, and monitor quality control in licenses all reflect the role of trademarks in identifying the source of goods and protecting consumers from deception in the marketplace. However, these rules are ill-suited to address the fundamental needs of brand owners. Brands are not simply large trademarks. They are creative works that invite consumer participation and encourage consumers to engage in expressive uses of the brand. Marketing scholarship has already recognized this and noted that the future of branding will entail brands as the center of communities and cultural discourse. It is time for trademark law to release its grip on brands and allow them to be adopted under the umbrella of copyright law.

STUDENT LOAN DERIVATIVES: IMPROVING ON INCOME-BASED APPROACHES TO FINANCING LAW SCHOOL

BENJAMIN M. LEFF* & HEATHER HUGHES**†

I. INTRODUCTION

EXTENSIVE, public discussion surrounds the high cost of legal education and student debt levels, yet very few critics of degree-cost show creativity in thinking about the optimal mechanism for funding a legal education. In recent years, numerous policy-makers, investors, and academics have been searching for ways to tie students' financing obligations to post-graduate income rather than interest rates.[1] The federal government currently offers income-based repayment options for student loan borrowers.[2] Private investors and policymakers advocate for so-called income-share agreements: agreements under which investors provide capital to students in exchange for a share of students' future incomes. These

---

* Professor, American University, Washington College of Law.

** Professor, American University, Washington College of Law.

† The authors would like to thank John R. Brooks, Ted Janger, Omri Marian, Shu yi Oei, Michael Simkovic, David Snyder, and the participants at the Junior Tax Scholars Workshop, the Washington College of Law Business Law Workshop, the National Tax Association Annual Meeting, the AALS Annual Meeting, the Georgetown Tax Law and Public Finance Workshop, and the College of William and Mary School of Law Faculty Scholarship Series.

1. For example, Senator Marco Rubio and Representative Tom Petri proposed legislation in 2014 that would make what they call "income share agreements" enforceable. *See* Investing in Student Success Act, H.R. 4436, 113th Cong. (2d Sess. 2014). Similar legislation was introduced in the House in 2015 by Representatives Todd Young and Jared Polis that modified some provisions of the Rubio/Petri legislation. *See* Investing in Student Success Act, H.R. 3432, 114th Cong. (1st Sess. 2015).

2. In 1993, Congress enacted a program under which student borrowers could limit their loan repayments to a fixed percentage of their income, with the ability for any remaining balance on their loans to be forgiven after twenty-five years. *See* Omnibus Budget Reconciliation Act of 1993, 20 U.S.C. §§ 1087a, 4021 (2012). The current version of this program, called Pay As You Earn (PAYE), permits borrowers to limit repayment to 10% of income after an exemption of 150% of the federal poverty level. *See* Income-Contingent Repayment Plans, 34 C.F.R. § 685.209(a) (2016). Graduates who have been employed by nonprofit or government employers can have the remaining balance of their loans forgiven after ten years. *See* Public Service Loan Forgiveness Program, 34 C.F.R. § 685.219 (2012). Graduates who have a remaining balance on their loans left after twenty years can have it forgiven regardless of the character of their employer. *See id.* § 685.209(a)(6). *See generally* JASON DELISLE & ALEXANDER HOLT, NEW AM. EDUC., ZERO MARGINAL COST: MEASURING SUBSIDIES FOR GRADUATE EDUCATION IN THE PUBLIC SERVICE LOAN FORGIVENESS PROGRAM 3 (2014) [hereinafter DELISLE & HOLT, ZERO MARGINAL COST], *available at* https://static.newamerica.org/attachments/759-zero-marginal-cost/ZeroMarginalCost_140910_DelisleHolt.pdf [https://perma.cc/9YJE-JLW8]; John R. Brooks, *Income-Driven Repayment and the Public Financing of Higher Education*, 104 GEO. L.J. 229 (2016).

existing approaches to income-based finance are problematic. The government programs create incentives for students to over-borrow and schools to overcharge by focusing government funds on those students who borrow the most,[3] and some fear the government programs may not continue to be available.[4] Private "income-share agreements"—which were traditionally called "human capital contracts"[5]—face legal and practical problems as investors make initial investments to fund students' education that they then must recover from the student's income under contracts that may be practically difficult—and legally problematic—to enforce.[6]

This Article introduces a new and innovative approach to financing law school: the Income-Based Repayment Swap (IBR Swap). The IBR Swap combines structural and financial advantages of derivatives with the appeal of income-based approaches to financing law school.[7] The IBR Swap uses a derivative structure to enable income-based payment for education that does not (1) rely on taxpayer subsidies or (2) implicate the practical and legal impediments associated with human capital contracts.[8] IBR Swaps coordinate with existing, traditional student loans such that the investors who provide the IBR Swap to students do not have to provide up-front capital as part of the transaction. Instead, every month an institutional counterparty makes a fixed payment to a student that the student uses to pay off student loans; the student makes a reciprocal payment to the institutional counterparty of a percentage of income.

Because the IBR Swap does not involve an initial disbursement to the student, it solves two sets of problems that plague human capital contracts.[9] First, because no capital is initially placed at risk, the cost of collec-

---

3. *See, e.g.*, DELISLE & HOLT, ZERO MARGINAL COST, *supra* note 2, at 2.

4. *See, e.g.*, Josh Mitchell, *Student-Debt Forgiveness Plans Skyrocket, Raising Fears over Costs, Higher Tuition*, WALL ST. J., Apr. 22, 2014, http://www.wsj.com/articles/SB10001424052702303887804579503894256072308 [https://perma.cc/N7AM-JBVW]. Others are more optimistic about governmental IBR programs' contribution to the education-financing problem. *See, e.g.*, Kevin Carey, *A Quiet Revolution in Helping Lift the Burden of Student Debt*, N.Y. TIMES, Jan. 24, 2015, http://www.nytimes.com/2015/01/25/upshot/a-quiet-revolution-in-helping-lift-the-burden-of-student-debt.html?_r=0 [https://perma.cc/B324-TT8N].

5. *See, e.g.*, Miguel Palacios, *Human Capital Contracts: "Equity-Like" Instruments for Financing Higher Education*, POL'Y ANALYSIS (Cato Inst., Wash., D.C.), Dec. 16, 2002 [hereinafter Palacios, *Human Capital Contracts*], at 1, 3–4, *available at* http://object.cato.org/sites/cato.org/files/pubs/pdf/pa462.pdf [https://perma.cc/B5Y4-RDFN].

6. For a further discussion of human capital contracts, see *infra* notes 12–29 and accompanying text.

7. Readers may find the IBR Swap concept relevant to higher education contexts beyond law school; the authors find it best to begin to think through the IBR Swap using the institutional context with which we are most familiar.

8. *See infra* notes 113–72 and accompanying text.

9. In this Article, we use the term *human capital contract* to describe a conventional income-share agreement, in which one party disburses funds to a student, and the student pays a percentage of future income to the disbursing party for

tion and risk of nonpayment decrease dramatically, reducing costs for all parties involved. Investors are not in the position of trying to enforce an unsecured payment obligation for return on a sizable initial investment. Rather, if students default on IBR Swaps, they are still obligated to lenders. Second, the unique structure of the IBR Swap reduces the legal uncertainty associated with human capital contracts. Because of the lack of initial disbursement, an IBR Swap is not debt and therefore is not subject to the regulations on debt or student loans. For the same reason, an IBR Swap is not "equity" in a person, as some have claimed about human capital contracts.[10] Because swaps are a recognized financial instrument category under current law, it is possible that IBR Swaps could become a reality without new legislation or other law reform.

But like other income-share agreements, IBR Swaps raise a host of serious concerns about differential treatment of prospective students based on criteria related to assumptions about earning power. Many object to human capital contracts on the grounds that they encourage a market which, if unregulated, might provide capital to men on more favorable terms than women and to students from more privileged backgrounds on more favorable terms than students from less privileged backgrounds, simply because investors believe that the more privileged student is a better investment. In addition, income-share agreements potentially obscure the non-financial value of higher education to the individual and the collective value of higher education to society.[11]

The IBR Swap does not alleviate these concerns. If anything, it compounds them by introducing a derivative structure that adds efficiency to the income-share agreement concept, potentially generating a market in financial instruments backed by student payment obligations. This Article presents the IBR Swap in order to incite critical consideration of both the potential and the limitations of private-market mechanisms for law school finance. As such, it should inspire critical assessment of what is the best and most impactful role for the government in ensuring access to law school.

Under current law, the IBR Swap structure removes some regulatory uncertainty and clears the way for a market in income-share agreements, even without new legislation. But ethical, distributional, and other questions surrounding the IBR Swap concept remain and warrant careful consideration. This Article merely begins the (daunting) task of thoroughly addressing all of the various regulatory tradeoffs that are triggered by moving from a debt-based system of financing higher education to one that interacts meaningfully with students' post-graduate income. Nonetheless, we do identify and address several important regulatory challenges, and

---

some period of time. We use the term *income-share agreement* to describe both human capital contracts and our innovation, the IBR Swap.

10. *See infra* note 146 and accompanying text.

11. *See, e.g.*, Daniela Kraiem, *The Cost of Opportunity: Student Debt and Social Mobility*, 48 SUFFOLK U. L. REV. 689 (2015).

discuss one in some detail: the differential pricing of income-share agreements.

This Article proceeds in four parts. Part II introduces human capital contracts and presents the IBR Swap concept. It explains the IBR Swap and how it is structurally similar to existing swaps in the market. Part III presents the benefits of IBR Swaps, explaining how these contracts can—better than the current student loan model—align costs with benefits, generate information, discipline schools with respect to costs, and potentially enable the government to advance education accessibility at a lower cost to the taxpayer. Part IV considers the current legal state of affairs. Part V examines important issues that should be addressed in any future regulation of IBR Swaps and other income-share agreements.

II. HUMAN CAPITAL CONTRACTS AND INCOME-BASED REPAYMENT SWAPS

A. *Human Capital Contracts*

For more than half a century, economists have dreamed of a novel way to fund higher education. Instead of using traditional debt, private investors could provide capital to students in exchange for a percentage of each student's future earnings. A "human capital contract" is a kind of "equity-like" interest in a person's future earnings.[12] In a human capital contract, "a student who wants to attend college, but does not have the resources to do so, signs a contract with an investor in which he commits to pay [a percentage] of his income for [a period of time] after graduation in exchange for [an upfront payment] received today to pay for tuition fees and living expenses."[13]

Economists point to a footnote in a 1945 article by Milton Friedman and Simon Kuznets as the origin of the idea to use human capital contracts to finance higher education.[14] Friedman and Kuznets complained that "investment in training is not governed by the usual profit incentives" because investors could not capture the expected return on an investment in the education of stranger.[15] Because future income is so variable, advancing a student money to finance education is risky. The student may earn enough to pay you back with interest, but may well *not* earn enough and default on the obligation. If that happens, the lender has very little recourse, since the student has no property to offer as security in the case of default. Because of the variability of future income, debt is an ill-suited

12. MIGUEL PALACIOS LLERAS, INVESTING IN HUMAN CAPITAL: A CAPITAL MARKETS APPROACH TO STUDENT FUNDING 1 (2004) [hereinafter PALACIOS, INVESTING IN HUMAN CAPITAL].

13. *Id.*

14. MILTON FRIEDMAN & SIMON KUZNETS, *Income in the Professions and in Other Pursuits*, *in* INCOME FROM INDEPENDENT PROFESSIONAL PRACTICE 62, 90 n.20 (1945).

15. *Id.* at 89.

mechanism for financing education.[16] When business ventures are risky, and when they have little property to offer as collateral, it is common for businesses to raise capital by offering investors an ownership interest in exchange for their investments. That way, while the risk of loss is greater than in the case of debt, it is balanced by the ability to receive higher returns if the business venture is successful. Friedman and Kuznets's footnote describes "an analogy that at first blush may seem fantastic"—"if individuals sold 'stock' in themselves, i.e., obligated themselves to pay a fixed proportion of future earnings, investors could 'diversify' their holdings and balance capital appreciations against capital losses."[17] In other words, if investors could participate in the upside gain of financially successful students, they could use that money to offset losses caused by financially unsuccessful students, and the result should be a lower cost of capital for the average student.

Friedman was not done with the idea in 1945. He returned to it in 1955, in an article that was influential in Congress's later decision to offer federal guarantees in order to foster a student loan market.[18] This time, Friedman argued more fully that this problem with fixed-rate debt as the sole source of capital for education would not exist if it were possible for investors to take an "equity interest" in the future earnings of a student.[19] So why do students not find such "equity" investors? Friedman concedes that "[t]here seems no legal obstacle to private contracts of this kind, even though they are economically equivalent to the purchase of a share in an individual's earning capacity and thus to partial slavery."[20] But Friedman recognized a host of difficulties with such a contract. Primary among them, according to Friedman, is the fact that they might be hard to enforce, since there is no property to use as collateral and the person subject to the contract may well move around to avoid collection.[21]

But he also commented:

> I have never been able to persuade myself that a major role has not also been played by the cumulative effect of such factors as the novelty of the idea, the reluctance to think of investment in human beings as strictly comparable to investment in physical assets, the resultant likelihood of irrational public condemnation

16. *See generally* Bas Jacobs & Sweder J.G. van Wijnbergen, *Capital-Market Failure, Adverse Selection, and Equity Financing of Higher Education*, 63 FINANZARCHIV 1 (2007); Palacios, *Human Capital Contracts*, *supra* note 5.

17. FRIEDMAN & KUZNETS, *supra* note 14, at 90 n.20.

18. *See* Milton Friedman, *The Role of Government in Education*, *in* ECONOMICS AND THE PUBLIC INTEREST 123 (Robert A. Solo ed., 1955). *But see* J.R. Walsh, *Capital Concept Applied to Man*, 49 Q.J. ECON. 255, 278–84 (1935) (purporting to show that there was no under-investment in "vocational" education in medicine, engineering, academics, or masters degrees, and that law was one area in which it appeared that there was under-investment).

19. Friedman, *supra* note 18, at 137–38.

20. *Id.* at 138.

21. *Id.*

> of such contracts . . . and legal and conventional limitation on the kind of investments that may be made by the financial intermediaries that would be best suited to engage in such investments . . . .[22]

In other words, it may be that the impediments to these human capital contracts are not inherent in the market, but are external, caused either by "irrational" conventional attitudes or laws.

Even if the concept of human capital contracts has a long history, actual human capital contracts have only begun to appear in the market recently.[23] But their modest appearance has been accompanied by a significant amount of attention by investors, academics,[24] and (most recently) lawmakers.[25] While observers of the contemporary explosion of

22. *Id.* at 138–39.

23. Companies that have recently begun to offer investments that are arguably human capital contracts include Fantex, Upstart, Lumni, Pave, the recently dissolved My Rich Uncle, 13th Avenue Funding, Base Human Capital, and Cumulus Funding. *See About Lumni*, LUMNI, http://www.lumni.net/about/ [https://perma.cc/SFZ3-2ZX7] (last visited Feb. 23, 2016); *About Pave*, PAVE, https://www.pave.com/about [https://perma.cc/SY3B-7D5D] (last visited Feb. 23, 2016); *Home*, 13TH AVENUE FUNDING, http://www.13thavenuefunding.org/ [https://perma.cc/KDQ5-BCNB] (last visited Feb. 23, 2016); *How It Works*, BASE HUM. CAP., http://www.base-hc.com/#option_title [https://perma.cc/8DCT-KARK] (last visited Feb. 23, 2015); *MyRichUncle*, WIKIPEDIA, http://en.wikipedia.org/wiki/MyRichUncle [https://perma.cc/R7XM-5HZ3] (last visited Feb. 23, 2016); *Our Product*, CUMULUS FUNDING, http://cumulusfunding.com/our-product/ [https://perma.cc/J7EF-KU45] (last visited Feb. 23, 2016); *Overview*, FANTEX, https://fantex.com/how-it-works [https://perma.cc/HE2A-BDKG] (last visited Feb. 23, 2016); UPSTART, https://www.upstart.com/ [https://perma.cc/EK28-826S] (last visited Feb. 23, 2016) (no longer offering human capital contract-like loans).

24. *See, e.g.*, PALACIOS, INVESTING IN HUMAN CAPITAL, *supra* note 12; MIGUEL PALACIOS, TONIO DESORRENTO & ANDREW P. KELLY, AM. ENTER. INST., INVESTING IN VALUE, SHARING RISK: FINANCING HIGHER EDUCATION THROUGH INCOME SHARE AGREEMENTS (2014) [hereinafter PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK], *available at* https://www.aei.org/wp-content/uploads/2014/02/-investing-in-value-sharing-in-risk-financing-higher-education-through-inome-share-agreements_083548906610.pdf [https://perma.cc/QH93-DC6H]; Jacobs & van Wijnbergen, *supra* note 16; Michael C. Macchiarola & Arun Abraham, *Options for Student Borrowers: A Derivatives-Based Proposal to Protect Students and Control Debt-Fueled Inflation in the Higher Education Market*, 20 CORNELL J.L. & PUB. POL'Y 67 (2010); Shu-Yi Oei & Diane Ring, *Human Equity? Regulating the New Income Share Agreements*, 68 VAND. L. REV. 681 (2015) [hereinafter Oei & Ring, *Human Equity?*]; Shu-Yi Oei & Diane M. Ring, *The New "Human Equity" Transactions*, 5 CALIF. L. REV. CIR. 266 (2014) [hereinafter Oei & Ring, *The New "Human Equity" Transactions*]; Palacios, *Human Capital Contracts*, *supra* note 5; Jeff Schwartz, *The Corporatization of Personhood*, 2015 U. ILL. L. REV. 1119; Ritika Kapadia, Note, *A Solution to the Student Loan Crisis: Human Capital Contracts*, 9 BROOK. J. CORP. FIN. & COM. L. 591 (2015); Matthew Soldner, *The Potential Market for Income Share Agreements Among Low-Income Undergraduates*, ISSUE BRIEF (Am. Inst. for Research, Wash., D.C.), Sept. 2015, *available at* http://www.air.org/sites/default/files/downloads/report/Income-Share-Agreements-ISAs-Potential-Among-Low-Income-Undergraduates-Sept-2015.pdf [https://perma.cc/YH78-TYB5].

25. *See* H.R. 3432, 114th Cong. (1st Sess. 2015); H.R. 4436, 113th Cong. (2d Sess. 2014).

human capital contract-like financial innovation disagree about many things, they all appear to agree that the current legal and regulatory environment creates significant uncertainty and is therefore an impediment to the development of a market for human capital contracts or similar financial products.[26]

Perhaps the most vocal, recent advocate of human capital contracts is Miguel Palacios, a professor of finance at Vanderbilt University's business school and a founder of one of the first human capital contract providers, Lumni.[27] Lumni provides capital upfront in exchange for a percentage of future income for students who show unusual promise. Lumni purports to be currently financing students in the United States, but Lumni started in Chile, where apparently the legal environment is less uncertain, and Palacios has complained that legal uncertainty is a significant barrier to the development of a market for human capital contracts in the United States.[28] Palacios is not alone in believing the time is ripe to develop a market for human capital contracts or other similar financial instruments. Nobel Prize-winning economist Robert Shiller touted what he calls "income-linked loans" in a 2003 book.[29] Recently, members of both houses of Congress have noticed the potential for what they call *income-share agreements*, proposing legislation in both the House and Senate to "provide the legal framework necessary for the growth of innovative private financing options for students to fund postsecondary education . . . ."[30] The premise of the legislation is that various types of income-share agreements are potentially beneficial, but that the current legal environment does not permit them or is uncertain enough to raise their cost unnecessarily. In this Article, we use the term *income-share agreement* to refer to the whole range of financial instruments that resemble human capital contracts, including the IBR Swap.

B. *The IBR Swap*

This Article presents a relatively simple innovation to the human capital contract, which, to the authors' knowledge, has never been proposed before.[31] We call this innovation an *Income-Based Repayment Swap* (IBR

26. For example, the Upstart blog, in explaining why Upstart discontinued its income-based funding operations, stated, "while many regulatory and policy efforts are underway to facilitate the development of the market, these efforts will likely take many years." Oei & Ring, *The New "Human Equity" Transactions*, *supra* note 24, at 270 (internal quotation marks omitted).

27. *Miguel Palacios*, VAND. U., http://www2.owen.vanderbilt.edu/miguel.palacios/ [https://perma.cc/P6YP-3CEL] (last visited Feb. 23, 2016).

28. *Our Story*, LUMNI, http://www.lumniusa.net/about/our-story [https://perma.cc/98MP-SU9Y] (last visited Feb. 23, 2016).

29. *See* ROBERT J. SHILLER, THE NEW FINANCIAL ORDER: RISK IN THE 21ST CENTURY 140 (2003).

30. H.R. 4436; H.R. 3432.

31. While no one has proposed a swap structure for human capital contracts, Michael Macchiarola and Arun Abraham proposed a financial instrument that has

Swap). It solves a surprising number of both practical problems and legal uncertainty associated with human capital contracts.

An IBR Swap, as we present it, is a contract between a student in a three-year juris doctor (J.D.) program at some United States law school (the student counterparty) and a fund or financial institution (the institutional counterparty) under which the student pays a percentage of income and the institution covers the costs of the student's law school loans in return. The institution could be any unrelated third-party investor.[32]

The student borrows money to pay for education from the government, just as students do now, taking out the same type of loan as any other student seeking funding for graduate education.[33] Let us imagine that a student borrows $150,000 to pay for three years of law school. Those loans come with a federally guaranteed fixed interest rate of 6.8% and a repayment schedule of ten years.[34] Repayment is scheduled to begin six months after graduation. The student's monthly loan repayments would be $1,726. The student takes that $150,000 and pays it to the school in tuition, just as any other student would.

But simultaneously with taking out loans and paying tuition, the student enters into an IBR Swap agreement with an institutional counterparty. The institutional counterparty agrees to pay the student $1,726 per month for 120 months (ten years) starting on the same day their student-loan obligations begin. In other words, the institutional counterparty agrees to pay the student the exact amount needed to pay back their loans for exactly the same term as the their loans.

With the institutional counterparty agreeing to pay the student counterparty $1,726 per month for 120 months, the student agrees to pay the institution 15% of their income for the same 120 months. So, at some

an "option" feature. *See* Macchiarola & Abraham, *supra* note 24. Miguel Palacios proposed using an option structure in a slightly different way. *See* PALACIOS, INVESTING IN HUMAN CAPITAL, *supra* note 12, at 81–101.

32. One possibility, of course, is for the institutional counterparty to be a law school, a fund created by a number of law schools, or some sort of charitable investor. These possibilities introduce a number of unique questions and concerns, however, and so for simplicity's sake, we assume at the outset that the institutional counterparty is a generic institutional investor.

33. Under current law, any student attending an accredited law school in the United States can borrow from the government any amount up to a school's total cost of attendance as unsubsidized graduate loans. *See Subsidized and Unsubsidized Loans*, FED. STUDENT AID, U.S. DEP'T OF EDUC., https://studentaid.ed.gov/sa/types/loans/subsidized-unsubsidized [https://perma.cc/NSS2-2WMP] (last visited Feb. 23, 2016).

34. For simplicity's sake, we use the 6.8% rate that has applied to most government direct graduate or professional school student loans for several years, even though under current law the interest rate for new loans is tied to the ten-year Treasury Note. For loans disbursed between July 1, 2014, and June 30, 2015, that rate is 6.21%, and for loans disbursed between July 1, 2015, and the present, the rate is 5.84%. *See id.*; *Interest Rates and Fees*, FED. STUDENT AID, U.S. DEP'T OF EDUC., https://studentaid.ed.gov/types/loans/interest-rates [https://perma.cc/73CJ-EAEC] (last visited Feb. 28, 2016).

set day in each month, the institution pays the student $1,726, and the student pays the institution 15% of the student's income. If the student is making $120,000 per year ($10,000 per month), then the student would owe the institution $1,500 each month. In its basic structure, this is a simple swap transaction.

A swap is a transaction in which the parties, called "counterparties," exchange cash flows or obligations.[35] A swap is a derivative—an agreement to transfer risk, the value of which is derived from the value of an underlying asset.[36] The underlying asset in derivatives transactions may be any tradable instrument: an interest rate, a commodity, a currency, etc.[37] The parties trade payment obligations (or cash flows) at specified payment dates during the agreed-upon term of the transaction.[38] The payment obligations that the parties exchange are called the "legs" of a swap.[39] Many different kinds of swaps exist in the market.[40] The IBR Swap involves exchanging "fixed" payments—contractual obligations to a lender—for "floating" payments—a contractual obligation to pay a percentage of income.[41]

The payments that swap counterparties exchange are based on a "notional amount"—a stipulated "principal" amount.[42] As the International Swaps and Derivative Association (ISDA) states, "the notional amount[ ] of a derivative contract is a hypothetical underlying quantity upon which

35. *See Product Descriptions and Frequently Asked Questions*, INT'L SWAPS & DERIVATIVES ASS'N, http://www.isda.org/educat/faqs.html [https://perma.cc/4YQL-HPNM] (last visited Feb. 23, 2016) [hereinafter ISDA FAQs]; *see also* JOHN C. HULL, OPTIONS, FUTURES, AND OTHER DERIVATIVES 757 (6th ed. 2006); Saul S. Cohen, *The Challenge of Derivatives*, 63 FORDHAM L. REV. 1993, 1997–2000 (1995); *Times Topics: Derivatives*, N.Y. TIMES, http://topics.nytimes.com/top/reference/timestopics/subjects/d/derivatives/index.html [https://perma.cc/XDQ4-J59U] (last updated Feb. 11, 2016) (explaining what derivatives are and referring readers to various recent articles about derivatives).

36. *See* ISDA FAQs, *supra* note 35.

37. *Id.*

38. *Id.*

39. *Id.*

40. For example, there are, among others: currency swaps (in which the parties exchange principle and interest payments in one currency for principle and interest payments in a different currency); commodity swaps (in which parties exchange a fluctuating market price for a fixed price of some commodity over a designated time period); and credit default swaps (in which one party pays a fee to the other party in return for compensation for default, however defined, by some reference entity). *See generally id.*

41. The student-loan leg of the swap is fixed in the sense that it is denominated by contract, not in the sense that the payment amounts necessarily are fixed; they may be fixed or variable depending on whether the student borrows on a fixed or floating rate basis from lenders.

42. In an IBR Swap, the counterparties could stipulate the "notional amount" to reflect the student's full obligation to law school lenders, or this amount also could be some portion of the student's law school loans.

interest rate or other payment obligations are computed."[43] Furthermore, the reciprocal payments may or may not be netted.[44]

There are several different ways to classify derivatives.[45] Some classify derivatives based on the underlying assets or metrics involved.[46] Others focus on whether a derivative is exchange-traded or over-the-counter (OTC),[47] along with whether it is booked with a clearinghouse or not.[48] The IBR Swap would be an OTC (as opposed to an exchange-traded) swap. Simply put, an OTC derivative is done directly between the counterparties, without involvement of an exchange.[49]

For purposes of understanding the IBR Swap, this Section draws on recent legal scholarship that classifies derivatives according to counterparty motivation.[50] We can classify derivatives among those in which (1) both counterparties are hedging, (2) one counterparty is hedging and the other is speculating, and (3) both counterparties are speculating.[51] The IBR Swap involves a student counterparty that is hedging a preexisting risk, and an institutional counterparty that is speculating on students' future income, thus it is in the speculator-hedger category.[52] To explain, in some contexts, derivatives serve a hedging purpose: they can offset or hedge against preexisting risks, as does insurance.[53] The ability to acquire insurance can offset loss; hence, buying insurance reduces risk.

---

43. ISDA FAQs, *supra* note 35. In a swap, the counterparties do not exchange the notional amount. *See id.*; Timothy E. Lynch, *Derivatives: A Twenty-first Century Understanding*, 43 LOY. U. CHI. L.J. 1, 5 n.12 (2011) [hereinafter Lynch, *Derivatives*].

44. As ISDA explains: "Payment netting reduces payments due on the same date and in the same currency to a single net payment. Payment netting is essentially identical to the legal concept of *set-off*." ISDA FAQs, *supra* note 35.

45. *See generally* Lynch, *Derivatives, supra* note 43; Timothy E. Lynch, *Gambling by Another Name; The Challenge of Purely Speculative Derivatives*, 17 STAN. J.L. BUS. & FIN. 67 (2011) [hereinafter Lynch, *Gambling*]; Jeffrey Manns, *Insuring Against a Derivative Disaster: The Case for Decentralized Risk Management*, 98 IOWA L. REV. 1575 (2013).

46. *See* Lynch, *Derivatives*, *supra* note 43, at 34–38.

47. *See* Lynch, *Gambling, supra* note 45, at 75–76.

48. *Id.*

49. *See* I PHILIP MCBRIDE JOHNSON & THOMAS LEE HAZEN, DERIVATIVES REGULATION 74–75 (I PHILIP MCBRIDE, THOMAS LEE HAZEN & CARY C. BOSHAMER, COMMODITIES REGULATION (3d ed. 1998), successor ed. 2004); *see also* ISDA FAQs, *supra* note 35. Note that some OTC derivatives are booked with a clearinghouse. *Id.* An IBR Swap would not be subject to regulatory clearinghouse requirements. *See infra* notes 126–31 and accompanying text.

50. *See* Lynch, *Gambling*, *supra* note 45; Lynn A. Stout, *Why the Law Hates Speculators: Regulation and Private Ordering in the Market for OTC Derivatives*, 48 DUKE L.J. 701 (1999) [hereinafter Stout, *Why the Law Hates Speculators*].

51. *See* Lynch, *Gambling*, *supra* note 45, at 71.

52. The possibility that a student would not have incurred risk but for the existence of an IBR Swap program does not affect this designation. *See id.* at 79–82.

53. *See id.* at 71, 75–82; *see also* I JOHNSON & HAZEN, *supra* note 49, at 28–29; M. Todd Henderson, *Credit Derivatives Are Not "Insurance"*, 16 CONN. INS. L.J. 1, 1–5 (2009); Manns, *supra* note 45.

In other contexts, derivatives serve a speculative purpose: they can create payment obligations based on the parties' (contrasting) predictions of future events, which creates risk that did not previously exist.[54]

The literature further divides derivatives that fall into the speculator-hedger category in which the speculator assumes risk for a premium—the insurance model[55]—and those in which speculators expect to leverage superior information or predictive capacity—the information arbitrage model.[56] IBR Swaps could follow either an insurance model or an infor-

54. Commentators often describe derivatives as bets, in a literal sense. They are agreements under which one party will pay the other, depending on whether certain events occur. Lynn A. Stout, *Derivatives and the Legal Origins of the 2008 Credit Crisis*, 1 HARV. BUS. L. REV. 1, 6 (2011) [hereinafter Stout, *Derivatives*] (stating that "[t]he value of a derivative agreement is 'derived' from the performance of the underlying financial phenomenon, just as the value of a betting ticket at the racetrack is 'derived' from the performance of a horse in a race"). Betting can serve very different purposes in different market contexts. Regulatory and market challenges surrounding derivatives concern the fact that speculative betting can reduce welfare by exposing market actors to new risks without a compensating increase in returns. *See id.* at 8; Lynch, *Gambling*, *supra* note 45, at 93–94. Bets that are hedging, or serve an insurance function, in contrast, involve transfer of risk for a premium. "In the parlance of economic theory," Stout explains, "speculative derivatives trading is a form of *rent-seeking*—trying to acquire wealth not by creating it, but by taking existing wealth from someone else." Stout, *Derivatives*, *supra*, at 9. However, many defend speculative derivatives on grounds that they provide other benefits such as greater liquidity and price discovery. *See* Lynch, *Gambling*, *supra* note 45, at 74.

55. Under an insurance model, counterparties agree to swap payments based on a formula that slightly favors the risk purchaser—paying, in essence, a premium for transferring the risk. In the case of a derivative that follows the insurance model, the terms of the derivative will favor, to some degree, the speculating counterparty. Such speculators earn profits in aggregate, over time, by entering many contracts and then allowing the favorable terms, combined with probabilities and events over time, to yield a return. *See* Lynch, *Gambling*, *supra* note 45, at 80–82. The speculating counterparty does not receive a "price" embedded in the contract, like an insurance premium, necessarily. Rather, the counterparties may calculate the amounts of payments to be swapped, or other terms of the swap, using formulas designed to slightly favor the speculating counterparty. *See id.* at 79 n.49.

56. Under an information arbitrage model, risk purchasers invest in generating better information than their counterparties have. In the case of a speculator-hedger derivative that follows the information arbitrage model, the speculating party may not assume risk for a "price." Rather, this counterparty may be privy to superior information about the direction or future value of the underlying asset or may have better tools or predictive skills. This kind of speculator counterparty will enter into a derivative contract with market (not favorable) terms and then wait for its predictive skills and informational advantages to produce a return. *See id.* at 79–81. This counterparty may invest in research and collect and analyze data in a way that the hedger does not. The hedger counterparty gets to mitigate risk at current market rates. The speculator counterparty gets to leverage the value of its information and analysis vis-à-vis a certain type of underlying asset with the numerous hedgers seeking to transfer risk. *Id.* Derivatives in which a speculating counterparty follows the information arbitrage model can contribute to price discovery, helping the market to determine accurate market prices for the underlying asset.

mation arbitrage model. Either way, students benefit from the fact that the institutional counterparty is able to assume risk and diversify its exposure to student-counterparty earning trajectories.

The pricing of an IBR Swap turns on a prediction of the student's future earnings over the term of the swap. A swap priced to "break even" would be one in which the predicted average earnings of the student produce payments to the swap counterparty that exactly match the payments that the swap counterparty agrees to make to the student. So, for example, if the institutional counterparty has agreed to make monthly payments of $1,726 to the student for ten years (exactly the amount needed to service a loan for $150,000 at 6.8% interest), then the parties would break even if the swap was "priced" at 15% and the student earned $138,080 per year on average over the ten-year term of the swap. In that scenario, the student would pay on average $1,726 per month to the institutional counterparty and the institutional counterparty would pay the student the same $1,726 per month.

But, of course, parties presumably would not price a swap at a perfect "breakeven price." First of all, in an IBR Swap, the two counterparties—the student counterparty and the institutional counterparty—assume credit risk associated with the willingness and ability of the other counterparty to perform. The student counterparty assumes the risk that the institutional counterparty could default on the student's obligations to lenders. The institutional counterparty assumes risk that the student counterparty could default on making the income-based payments. If either party defaults, the result is that the student is liable for any payments owed on outstanding student loans. This credit risk would have to be incorporated into the price of the swap. In addition, institutional counterparties would presumably want to be compensated for their costs of capital; their payments are likely to be bigger than the student's corresponding payment in the beginning of the swap's term. In addition, institutional counterparties will want to price in some profit for themselves.

If the student counterparty earns more than expected, there is an incentive for the student to default on their payment obligations, but the IBR Swap would be designed to prevent student default in this circumstance, giving institutional counterparties all the tools that lenders have to ensure compliance.[57] Conversely, if the student counterparty earns a relatively low income, the IBR Swap has a negative value for the institutional counterparty and a positive value for the student counterparty. IBR Swaps may require regulation—such as capital adequacy requirements and pen-

57. *Cf.* Patricia Brown, *Tax Consequences of Interest Rate Swaps: Characterization by Function, Not Prejudice*, 6 BERKELEY J. INT'L L. 122, 131 (1988) (citing U.C.C. § 2-718(1) (1978)).

alties for breach—to prevent institutional counterparties from defaulting on swaps that have a negative value to them.[58]

III. BENEFITS OF THE IBR SWAP

The current interest in income-share agreements accompanies widespread anxiety about the rising costs of higher education and the recognition that traditional student loans are significantly burdening students.[59] In the law school context, where students can borrow the total cost of attendance at their institutions, which may be as high as $230,000 for three years of attendance,[60] the situation for graduates is even more pronounced than in other areas.[61] Supporters of income-share agreements argue that market-based financial innovations, like human capital contracts, offer several benefits over traditional debt.

These same benefits apply equally to IBR Swaps. But there are additional benefits that are unique to IBR Swaps. This Section first discusses benefits that IBR Swaps share with human capital contracts and other similar instruments. It includes both benefits over traditional debt and potential benefits over the current governmental income-based repayment programs. It then explains those benefits that arise from unique attributes of the IBR Swap.

A. *Benefits Shared with Other Income-Share Agreements*

1. *Alignment of Costs with Benefits*

The primary benefit of IBR Swaps and other income-share agreements is that they do a better job than debt of aligning the costs of an

58. In addition, contractual terms or regulation may be necessary in IBR Swaps to reduce moral hazard, as they are in some other derivatives contexts. *See* Lynch, *Gambling*, *supra* note 45, at 107; Frank Partnoy & David A. Skeel, Jr., *The Promise and Perils of Credit Derivatives*, 75 U. CIN. L. REV. 1019, 1034–36 (2007).

59. *See* Palacios, *Human Capital Contracts*, *supra* note 5, at 2.

60. *See, e.g.*, DELISLE & HOLT, ZERO MARGINAL COST, *supra* note 2, at 3 ("[G]raduate students have been able to use the program to finance the entire cost of their educations . . . without limit since 2006." (citing Deficit Reduction Act of 2005, Pub. L. No. 109-171, 20 U.S.C. § 8005 (2006))).

61. Some commentators have focused on the law school context because the volume of student debt is so large there, and the earnings of law school graduates are so uneven. *See, e.g.*, JASON DELISLE & ALEX HOLT, NEW AM. FOUND., SAFETY NET OR WINDFALL?: EXAMINING CHANGES TO INCOME-BASED REPAYMENT FOR FEDERAL STUDENT LOANS 10 (2012) [hereinafter DELISLE & HOLT, SAFETY NET OR WINDFALL?], *available at* https://static.newamerica.org/attachments/2332-safety-net-or-windfall/NAF_Income_Based_Repayment.18c8a688f03c4c628b6063755ff5dbaa.pdf [https://perma.cc/44HG-WXKX]; BRIAN Z. TAMANAHA, FAILING LAW SCHOOLS, at x (2012) [hereinafter TAMANAHA, FAILING LAW SCHOOLS]; Macchiarola & Abraham, *supra* note 24, at 77–79; Brian Z. Tamanaha, *The Problems with Income Based Repayment, and the Charge of Elitism: Responses to Schrag and Chambliss* (Wash. U. St. Louis Sch. of Law Research Paper Series, Paper No. 13-06-01, 2013) [hereinafter Tamanaha, *The Problems with Income Based Repayment*].

education to the financial benefit to the student.[62] A debt arrangement defines the cost to students before they know how much money they will earn post-graduation. If a student loan program is going to break even, the interest rate must be high enough for regular borrowers to subsidize those borrowers who default because they do not earn enough to service their loans. Thus, all students who earn enough to avoid default must subsidize those students who earn too little and default. But the large majority of students who earn enough to avoid default pay the same amount, no matter how much they earn. For low earners who nonetheless earn enough to avoid default, the rate can be a very high percentage of their income.

The IBR Swap solves the problem of variable earnings by retrospectively tailoring the cost of education, so to speak, based on the amount the graduate actually earns. Instead of having a single rate that everybody pays, the rate of repayment is adjusted based on the student's individual ability to pay after the fact. This tailoring of the repayment rate to post-graduation income means that graduates who earn above the mean income subsidize students who earn below the mean. Thus, the cost for the median earner goes down, since above average returns can be captured from high earners, unlike in a traditional debt arrangement in which every non-defaulting borrower pays the same (relatively high) rate.

This decrease in cost for median earners could be substantial, depending on the default rate in the traditional loan program. But for non-defaulting *low* earners, the decrease in cost is even more substantial. Thus, the primary benefit of income-share agreements is that they dramatically decrease the cost of education for graduates who end up not earning high incomes. If it is true that there is some inherent uncertainty about which students will be high earners and which will be low earners, this reduction in the downside risk of low earning is a huge benefit. It enables students to decrease the downside financial risk that their educations will not result in earnings high enough to justify the cost.[63]

In the law school context, the variation of earnings is extreme.[64] Currently, this problem of extreme variation in earnings among graduates is

62. SHILLER, *supra* note 29, at 139.

63. If there is no real uncertainty from the student's perspective, and the student's actions largely determine earnings (through effort, for example), then the benefits of insuring against downside risk are dramatically reduced. In this case, so-called moral hazard could reduce the value of the IBR Swap's structure. *See infra* notes 143–45 and accompanying text.

64. According to the National Association of Law Schools (NALP), the distribution of starting salaries of law school graduates is bimodal, with one peak (representing about 16% of students) making approximately $160,000 and a second peak (representing about 51% of students) making between $40,000 and $65,000. *See Class of 2013 Bimodal Salary Curve*, NALP (July 2013), http://www.nalp.org/class_of_2013_bimodal_salary_curve [https://perma.cc/6AXE-MRJ7]. This data excludes graduates working part-time and graduates who are not employed. Michael Simkovic and Frank McIntyre survey available data on *lifetime* earnings by law school graduates, which shows a similar wide disparity in earnings. *See* Michael

at least partially addressed by an expansion of income-based repayment programs by the Obama Administration.[65] But these programs, which are designed to enable those borrowers who have the most debt and the lowest income to avoid default, do not do as good a job aligning the costs of education with the financial benefits derived from it as an IBR Swap.

A governmental income-based repayment option prevents very low-income students from defaulting, but it does not reduce the cost for median students. It also does nothing for students who only borrow part of the costs of their education, because they do not have enough debt to trigger the entrance requirements for the program.[66] An IBR Swap, on the other hand, adjusts the amount repaid for one's education dollar for dollar based on the ex post financial value of that education, as reflected in one's overall income. Thus, an IBR Swap performs a more precise matching of ability-to-pay with financial benefit than government income-based repayment and loan forgiveness programs.

Some proponents of income-share agreements support them because they believe that income-share agreements have the potential to decrease the government's involvement in higher education finance.[67] These commentators believe that existing government programs, especially governmental IBR expanded by the Obama Administration, commit taxpayer funds in ways that can have perverse effects on costs and can potentially subsidize students whose families could afford to pay without assistance.[68] Because IBR Swaps and other income-share agreements subsidize the risk of low earning graduates with earnings from high-earning graduates, the need for taxpayers to subsidize low-earning graduates is potentially diminished.[69] And because income-share agreements permit low-earning students to pay less, the need for government-subsidized insurance against

Simkovic & Frank McIntyre, *The Economic Value of a Law Degree*, 43 J. LEGAL STUD. 249, 254 (2014). For example, the United States Bureau of Labor Statistics has data for all lawyers, which shows that in 2012, the 75th percentile compensation was about $160,000 (at all levels) and that the 25th percentile was about $70,000. *See Occupational Employment Statistics*: *Lawyers*, U.S. DEP'T OF LABOR (May 2012), http://data.bls.gov/cgi-bin/print.pl/oes/2012/may/oes231011.htm [https://perma.cc/B9EB-EFY6] (last modified Mar. 29, 2013).

65. *See, e.g.*, Carey, *supra* note 4.

66. This feature of the government's IBR programs creates strong incentives for students to over-borrow. It could be corrected in the government program by tailoring the eligibility requirements for the income-based repayment programs to the amount borrowed. *See, e.g.*, DELISLE & HOLT, ZERO MARGINAL COST, *supra* note 2, at 21–23.

67. *See, e.g.*, Palacios, *Human Capital Contracts*, *supra* note 5, at 3.

68. *See, e.g.*, DELISLE & HOLT, SAFETY NET OR WINDFALL?, *supra* note 61, at 10; DELISLE & HOLT, ZERO MARGINAL COST, *supra* note 2, at 21.

69. The government subsidies are currently almost exclusively to be found in the government's IBR programs, since the current system of federal student loan financing (excluding the loan forgiveness programs) currently operates at a profit for the government.

low earnings in the form of income-based repayment or loan forgiveness is dramatically reduced.[70]

But even for those who believe that higher education *should* be subsidized by taxpayers, IBR Swaps and other income-share agreements offer an opportunity to clarify the discourse around governmental support for education and redirect resources where they will do the most good.[71] If the existence of a market for IBR Swaps would diminish the problem of low-earning graduates being unable to pay for their higher education, then that would enable the government to focus its resources on providing educational opportunities to those who need them the most. For example, the government could redirect its efforts and funds to provide grants for students without means, or to fund state educational institutions or other programs that are targeted to increase access to higher education for those students who have barriers to access. In the law school context, a market for IBR Swaps would enable the government to withdraw from its current loan forgiveness programs (at least for future students) and refocus its resources on making legal education better, cheaper, and more accessible to low-income and minority students—if those are the policy objectives the government chooses to pursue. Alternatively, the government could focus its resources on making higher education, and even law school, cheaper and more accessible for all students.

2. *Information for Students*

The second most often touted potential benefit of income-share agreements is their ability to communicate information to students seeking an education.[72] Advocates for human capital contracts emphasize in-

70. Some argue that the current governmental IBR represents a significant subsidy in the law school context. *See, e.g.*, DELISLE & HOLT, SAFETY NET OR WINDFALL?, *supra* note 61, at 1; Tamanaha, *The Problems with Income Based Repayment*, *supra* note 61. For example, a law student who borrows $125,000 will qualify for the government's income-based repayment and loan forgiveness programs even if the student earns up to $172,620 per year. (The math here is slightly simplified—$125,000 at 6.8% for ten years is $17,262 per month; a student qualifies for PAYE if payments under the standard ten-year payment option exceed 10% of annual income.) However, qualifying for income-based repayment does not necessarily represent a subsidy from the federal government, since loan forgiveness comes after ten years only for students who work for that whole period in public interest jobs. For all others, loan forgiveness only comes after twenty years, after which time many students will have paid back the full amount of their loans.

71. *See, e.g.*, LAUREN ASHER, DIANE CHENG & JESSICA THOMPSON, INST. FOR COLL. ACCESS & SUCCESS, SHOULD ALL STUDENT LOAN PAYMENTS BE INCOME-DRIVEN?: TRADE-OFFS AND CHALLENGES 10–18 (2014), *available at* http://ticas.org/sites/default/files/pub_files/TICAS_IDR_White_Paper.pdf [https://perma.cc/J9S8-24AT]; Stephen Crawford & Robert G. Sheets, *Managing Risk, Reaping Reward: The Case for a Comprehensive Income-Based Loan System*, *in* REINVENTING FINANCIAL AID: CHARTING A NEW COURSE TO COLLEGE AFFORDABILITY 171, 171–90 (Andrew P. Kelly & Sara Goldrick-Rab eds., 2014).

72. *See* SHILLER, *supra* note 29, at 133; *see also* Palacios, *Human Capital Contracts*, *supra* note 5, at 1, 5–6.

formation-generating benefits of income-share financing, as the pricing of individual income-share agreements would communicate information to students about the institutional counterparty's assessment of a student's potential and of the value of the programs or careers the student is pursuing.[73]

However, the extent to which this "differential pricing" of income-share agreements can generate useful information for students may be far less than advocates indicate. In order to generate information about the value of a given program, vis-à-vis another program, pricing would have to control for other attributes of the student, concentrating differentials on program differences. Current models for income-share agreements do not do this. In addition, as discussed below, while many describe this differential pricing as a "feature" of income-share agreements, it is also a significant "bug."[74] In any event, to the extent that human capital contracts produce useful information, the IBR Swap would do so as well.

Palacios describes information benefits of human capital contracts this way:

> The pricing of human capital contracts will be based on the investor's expectations of a student's future income during the repayment period. Those expectations will depend on the school that the student is attending, the student's field of study, and other factors considered relevant to the student's future earnings. Thus, by observing the price of these contracts, comparisons of earnings expectations will be possible in an easy, straightforward manner.[75]

There is significant outcry that such information is not currently available to students, especially law students, and that much of the information provided is skewed or even fraudulent.[76] Thus, a law-school financing mechanism that could incentivize investors to spend the time and money to collect the highest quality information available to predict probable earnings of students graduating from specific schools, or based on other characteristics, would be welcome by many.

IBR Swaps, in theory, would incentivize investment in information gathering.[77] The nature of that information, however, and its usefulness

73. *See, e.g.*, Crawford & Sheets, *supra* note 71.

74. *See infra* Part IV(A).

75. Palacios, *Human Capital Contracts*, *supra* note 5, at 5.

76. TAMANAHA, FAILING LAW SCHOOLS, *supra* note 61, at 143.

77. *Price discovery* is the impounding of new information into asset prices, through trading. In theory, the concept of price discovery can refer broadly to the capture and aggregation of market wisdom about future events. *See* DON M. CHANCE & ROBERT BROOKS, AN INTRODUCTION TO DERIVATIVES AND RISK MANAGEMENT 11–12 (7th ed. 2007); S.L. GUPTA, FINANCIAL DERIVATIVES (THEORY, CONCEPTS AND PROBLEMS) 16–17, 35, 90 (2006); Joel Hasbrouck, *One Security, Many Markets: Determining the Contributions to Price Discovery*, 50 J. FIN. 1175, 1175–77 (1995); Lynch, *Gambling*, *supra* note 45, at 108–18. Some contend that derivatives are use-

to students would remain to be seen. To understand how the "price" of an IBR Swap or other income-share agreement could communicate information to a student, it is first necessary to understand what "price" means in this context. Palacios explains it this way: "Let's define the price of a human capital contract as the percentage of income that a student agrees to pay back to the investor *per* dollar provided."[78] As was discussed above, a "breakeven price" could be identified for each student based on the amount the counterparty will be obligated to pay each month,[79] the term of the IBR Swap, and the projected earnings of the student over the term of the IBR Swap. In the example described above, the counterparty paid the student $1,726 per month ($20,712 per year), the term was ten years, and the projected average annual earnings of the student over those ten years was $138,080, resulting in a "price" of 15%.[80]

However, that 15% "price" is derived from the other variables and is therefore not fixed. If differential pricing were permitted, then institutional counterparties might offer IBR Swaps (or other income-share agreements) to different students at different prices. For example, while one student (Ben) might be expected to earn on average $138,080 per year, justifying a 15% price for receiving $150,000 upfront, another student (Heather) might be expected to earn more. If Heather was expected to earn $207,129, for example, she may be offered a contract with a price of 10% of her income over the same period in exchange for the same

---

ful to predict information, such as creditworthiness. *See* Mark J. Flannery, Joel F. Houston & Frank Partnoy, *Credit Default Swap Spreads as Viable Substitutes for Credit Ratings*, 158 U. PA. L. REV. 2085 (2010). In the literature on derivatives, scholars debate whether a robust futures market can provide beneficial "price discovery," focusing on the relationship between futures or options prices and spot prices in exchange-based markets. Some contend that the price discovery benefits of speculative derivatives justify their risks (or offset the net reduction in welfare associated with rent-seeking). *See* Kenneth D. Garbade & William L. Silber, *Price Movements and Price Discovery in Futures and Cash Markets*, 65 REV. ECON. & STAT. 289 (1983); Robert W. Kolb, James V. Jordan & Gerald D. Gay, *Futures Prices and Expected Future Spot Prices*, 2 REV. RES. FUTURES MARKETS 110 (1983). Others contend that the information that derivatives markets yield outpace spot prices by such short intervals that the benefits do not offset speculative derivatives' costs. *See* Lynch, *Gambling, supra* note 45, at 116–18; Lynn A. Stout, *Betting the Bank: How Derivatives Trading Under Conditions of Uncertainty Can Increase Risks and Erode Returns in Financial Markets*, 21 J. CORP. L. 53, 55–59 (1995); Lynn A. Stout, *Regulate OTC Derivatives by Deregulating Them*, REG., Fall 2009, at 30, 31–33 [hereinafter Stout, *Regulate OTC Derivatives*]. This literature may be only marginally relevant to the IBR Swap, the "price discovery" functions of which may be quite limited.

78. *See* Palacios, *Human Capital Contracts*, *supra* note 5, at 5.

79. *See id.*

80. The price of 15% is based on the fact that the student received $150,000 up front. However, the price should actually be expressed as a function of how much was received. So, for example 15% for $150,000 is really 1% for each $10,000 received. Therefore, if the student has some savings and so chooses only to obtain $100,000 through a human capital contract, they will only have to pay back 10% of their income over twenty years.

$150,000.[81] Even though they both pay for their entire education with a human capital contract, and even though their educations cost the same amount, Ben will have to pay back one and one-half as much *as a percentage of his future income* as Heather. The difference in price reflects the fact that the investor predicted that Heather's future income over the relevant period is likely to be higher than Ben's future income. They will pay the same amount as each other if they earn the amount predicted, which, if the price is the "breakeven price," will be the same amount that either would pay under a traditional loan.[82] If either one earns more or less than predicted, they will pay the commensurate amount more or less according to the "price" of their contract.

Palacios and others assume that investors would take into account (1) school, (2) field of study, and (3) other factors when setting an appropriate price for a human capital contract.[83] In the law school context, investors might charge more—as a percentage of income—to a student attending a low-ranked law school than a student attending a highly-ranked law school, based on projected income. If they were permitted, investors might charge students who attended the same school less or more depending on their Law School Admission Test (LSAT) scores or undergraduate GPAs, if they determined that these factors were predictive of earnings.[84] Finally, if it was legal to do so, investors might charge a student more if that student was a woman than if that student was a man, again based on projected income, since women (including female lawyers) earn less on average than men.[85] Some may find this differential pricing deeply disturbing, and we discuss tradeoffs in regulating differential pricing below in Part V. But it is important to recognize that there is no production of information to students without some kind of differential pricing.

Currently, law schools are required by the American Bar Association (ABA) to survey their students to determine how many of them are employed nine months after graduation and what type of employment they

81. For Ben, the calculation is as follows: 20,712 ÷ 138,080 = .15; for Heather, it is as follows: 20,712 ÷ 207,129 = .10.

82. Just to be clear, the cost of financing a law school education is likely to be about the same under fixed debt and an income-share agreement if the student earns the amount that they are predicted to earn. If the student earns less than predicted, then their cost will be higher under debt financing than under an income-share agreement. If they earn more than predicted, their cost will be higher under an income-share agreement than under debt financing.

83. *See* Palacios, *Human Capital Contracts*, *supra* note 5, at 5.

84. It appears that the few companies currently purporting to offer human capital contracts are making individualized pricing decisions based on a variety of potentially intangible factors, given that they appear to be choosing a very few, very promising students to receive the first contracts. They are, in effect, "cherry picking" students. The authors know of no study that supports an inference that law school graduates with higher LSAT scores or undergraduate GPAs make more money than their peers, once other factors are controlled.

85. *See infra* Part V(A).

have. Some commentators have criticized this employment information as untrustworthy, misleading, insufficient, or even fraudulent.[86] An investor in IBR Swaps would need to have much more robust information about the earnings of those students in which they invested. In order to "price" the contract adequately, an investor would have to assess the long-term earnings prospects of each participant.[87] In theory, as a market for IBR Swaps developed, investors would compete to develop the most predictive models of future earnings, providing students with a range of opinions about their future earning capacity.

Some believe that the information-providing benefit of differentially priced IBR Swaps may introduce pressure on what is perceived as excessive costs of education, especially law school education. The extent to which higher education costs are in fact excessive (rather than simply high), their relationship to inflation, the rates at which they rise, and their capacity to respond to downward pressure, are subjects of a rich literature.[88] Though assessing cost is complex, an income-based financing program like the IBR Swap might create some downward pressure if it provided information to students about the relationship between cost and future earnings that make the cost seem unjustified.[89]

Because the current education-financing system provides financing on the same terms no matter what a student studies and where, and because students have such thin or misleading information about earnings projections for graduates of various schools, commentators believe that tuition costs are relatively insensitive to price pressures.[90] In other words, because students are unable to evaluate how much different types of educations are worth, from a financial point of view, they are prevented from making good choices about where to go and what to study.[91] In the law school context, some commentators believe that law schools with low-earn-

86. *See, e.g.*, TAMANAHA, FAILING LAW SCHOOLS, *supra* note 61, at 143.

87. If an investor was more interested in communicating rosy information about employment prospects of law school graduates than in making money or breaking even, that investor could "distort" the signal provided by predictive data. That might be a reason to favor independent investors as an institution providing a human capital contract or IBR Swap over a law school.

88. *See generally* ROBERT B. ARCHIBALD & DAVID H. FELDMAN, WHY DOES COLLEGE COST SO MUCH? (2010); WILLIAM J. BAUMOL, THE COST DISEASE: WHY COMPUTERS GET CHEAPER AND HEALTH CARE DOESN'T (2012); WILLIAM ZUMETA ET AL., FINANCING AMERICAN HIGHER EDUCATION IN THE ERA OF GLOBALIZATION (2012); W.J. Baumol & W.G. Bowen, *On the Performing Arts: The Anatomy of Their Economic Problems*, 55 AM. ECON. REV. 495 (1965); Brooks, *supra* note 2, at 11–15.

89. Of course, if the effect instead shows the opposite, then the price pressure could be upwards.

90. *See, e.g.*, Michael Simkovic, *Risk-Based Student Loans*, 70 WASH. & LEE L. REV. 527, 557 (2013) (discussing students' lack of information in college context).

91. *See, e.g.*, *id.* at 567 ("[S]kewed incentives and information asymmetries have increasingly shifted educational resources away from human capital investment and toward present consumption.").

ing graduates are unsustainably expensive.[92] Students are willing to pay the high cost of education at least partially because the structure of the current educational financing system makes it easy to overburden oneself with debt in programs that are unlikely to provide employment opportunities commensurate with their price.

Governmental IBR currently exacerbates this problem. Governmental IBR is structured as a ceiling on loan repayments for students who participate. But this ceiling is the same no matter how much a student borrows. So, a student who borrows $30,000 for college will be able to take advantage of governmental IBR only if their income is less than $55,215.[93] A student who borrows $125,000 for law school, in addition to the $30,000 they borrowed for college, will be able to take advantage of governmental IBR until their income exceeds $210,240.[94] If they expect their income to be below that amount, they can, in effect, continue to borrow without their cost going up at all. No matter how much the student borrows, they pay a flat 10% of their income in repayment obligations. Some commentators think this situation creates bad incentives for students, because they have an incentive to continue attending school beyond when it is useful to them and also to borrow as much as permitted to fund their educations.[95] It also creates bad incentives for schools, which can increase tuition without any additional cost to students who are already borrowing so much that they are virtually guaranteed to be covered by governmental IBR.[96]

Because the percentage a graduate owes under an IBR Swap depends on how much they borrow, the perverse incentives of governmental IBR go away. In other words, rather than owing a flat 10% ceiling no matter how much a student borrows, under an IBR Swap (or other income-share

92. *See, e.g.*, Gregory Crespi, *Will the Income-Based Repayment Program Enable Law Schools to Continue to Provide "Harvard-Style" Legal Education?*, 67 SMU L. REV. 51, 60 (2014); *see also* TAMANAHA, FAILING LAW SCHOOLS, *supra* note 61, at 122–25.

93. This calculation is for the PAYE program, under which a student may limit their monthly payments to 10% of "discretionary income," which is defined as 150% of the DHHS Poverty Guideline for the year. In 2014, the 150% of the Poverty Guideline for a single person living in the continental United States was $17,655. Assuming that the student had total eligible debt of $30,000 at 4.66% interest rate (Stafford rates for 2014), they would owe $313 per month (or $3,756 per year) under the standard (ten-year) repayment plan. Therefore, once the student's income exceeded $55,215, their payments under the PAYE program would be less than their payments under the standard program, and they would be eligible to participate in PAYE.

94. The slightly simplified math is as follows: under the standard plan, $3,756 for the undergraduate loan, plus $17,268 ($125,000 at 6.8%, see *supra* note 34, under the standard ten-year payment plan) equals $21,024. Therefore, when the student's income reached $227,895, the amount of discretionary income would be $210,240 (227,895 minus 17,655), their annual payment under PAYE would be $21,024, and so they would at that point be eligible for PAYE.

95. *See* DELISLE & HOLT, ZERO MARGINAL COST, *supra* note 2, at 10 (describing what Delisle and Holt call "zero marginal cost threshold").

96. *See id.*

agreement), the student would owe a percentage tied to the amount they borrowed. In that case, the cost of further borrowing always increases the future cost of repayment, removing the perverse incentives in the current government programs.[97]

B. *Benefits Unique to IBR Swaps*

While income-share agreements have a number of benefits, Milton Friedman had already identified problems with them over fifty years ago. There are two problems with human capital contracts that are *not* solved by IBR Swaps: adverse selection[98] and moral hazard.[99] IBR Swaps, however, do largely solve the three other biggest problems associated with human capital contracts. First, because the institutional counterparty provides no money up front to the student, the costs of collection and risk of default should dramatically decrease as compared to human capital contracts. Second, IBR Swaps integrate with existing government programs to help students finance their educations, while human capital contracts compete with such programs. Finally, IBR Swaps are less risky for institutional counterparties than human capital contracts, because there is no need for the institutional counterparty to put significant amounts of capital at risk. These benefits make it possible for a widespread market in income-share agreements to become a reality for students.

1. *IBR Swaps Reduce Default Risk and Collection Costs*

There is an inherent collection problem in all student loans, since such loans are unsecured. Students do not generally have property to serve as collateral for their loans, and therefore one of the primary mechanisms for reducing risk of non-payment in large lending markets is absent in the student loan context. This was one of the factors that induced the federal government to enter the student loan business in the first place.

Risk of non-payment exists in the human capital contract market as well as in the student loan market. Under any type of income-share agreement, the investor shares some of the student's risk of under-earning and is compensated for this risk-sharing by distributing the risk of low earning broadly among a pool of students. But in addition to the risk of under-earning by students, there is a risk that a student who has received an upfront payment from an investor and used it for their education would then refuse to repay the investor, even if they are able. This risk exists every time one person provides something of value upfront in exchange for another promising something of value in the future.[100] An investor

97. The government could also remove these perverse incentives by tying the percentage owed under their IBR programs to the amount borrowed.

98. *See infra* Part IV(B).

99. For a variety of reasons we think moral hazard is not a significant problem in the IBR Swap context. *See infra* notes 143–45.

100. *See, e.g.*, Arthur Allen Leff, *Injury, Ignorance and Spite—The Dynamics of Coercive Collection*, 80 YALE L.J. 1, 1 (1970) ("Whenever one person does something

incurs costs when students do not pay what they owe, both because some amount will never be collected and because monitoring the student's earnings and enforcing their obligation to pay has costs.

On one hand, human capital contracts may have a lower cost of collection than traditional debt. Because low-earners pay very little under the terms of the contract, income-share agreements shift the focus of collection for non-payment from low-earning graduates (the perennial collection problem for traditional student loans) to high-earning graduates. Low-earning graduates owe less, so they are more able to make their required payments. High-earners, as their income goes up, have more incentive to try to avoid repayment. If it is true that it is harder to collect money from someone who has very little, focusing collection on high earners may decrease at least some collection costs.

On the other hand, costs of collection may be high for human capital contracts, since default becomes more attractive as a student's post-graduate income rises. As a high earner's income grows, they may compare the amount they owe under a human capital contract to the amount they would have paid under a traditional debt instrument. As this difference grows, the high earner might feel justified reneging on their agreement, even if they would not have felt justified in the case of a traditional loan. And the fact that the student is a high earner means that they may have access to the means, like legal counsel, to press a claim to avoid repayment.

Whether high-earning or low-earning, collection-evaders are actually more expensive to pursue is an empirical question beyond the scope of this Article. It is difficult to add up these disparate and speculative costs of collection, but they are likely substantial. Some believe that factors like these were instrumental in the demise of Yale's so-called "Tuition Postponement Option," a voluntary income-based tuition program that was available for students at Yale College from 1971 to 1978.[101]

The primary practical benefit of an IBR Swap over a human capital contract is that the IBR Swap should have significantly lower costs of collection across the board. This decrease in costs of collection comes from the fact that the institutional counterparty provides no money up front to the student. Rather, the student gets all the money they need for their education from a lender, who is likely the federal government. Therefore, the student's payment obligation is split between the lender and the inves-

in the expectation that another will then do something else, there arises, given the nature of people and time, a potential problem: the other person might not.").

101. Robert Shiller has an informative description of the program, in which he concluded that it "was a wonderful idea, but it . . . [among other things] affronted then-current individual impressions of fairness." SHILLER, *supra* note 29, at 143.

tor. This bifurcation of payment obligation reduces the cost of collection for the investor.[102]

As discussed above, the student has an obligation to pay a fixed amount to the lender no matter how much they earn. They and the institutional counterparty then make reciprocal payments that are either exactly equal, result in a net positive for the student, or result in a net positive for the institutional counterparty. Remember, the amount of income the student earns at which the reciprocal payments are exactly equal is called the *break-even point*.

If the student earns less than the break-even point, then they are receiving more from the institutional counterparty than they are paying it. In that situation, obviously, the student has a strong incentive to perform. Every time the reciprocal payments do not occur, they lose money. So, if the reciprocal payments only occur when the student both reports information to the institutional counterparty and makes the appropriate payment to the institutional counterparty, then the institutional counterparty's costs of collection should be quite low. The counterparty presumably never needs to chase down low-earning students or expend significant sums to monitor them. It does have to make sure it has a system in place to keep track of which students have made payments to it, so it does not make payments to students who have not made their appropriate reciprocal payments. It also needs some system to ensure that students are not providing false or fraudulent information about their income, since a student's reported income determines the amount of money the student pays the institutional counterparty.[103] But the counterparty has a pretty big stick that it can swing: each month it pays the student more than the student pays it.

This situation is dramatically different from a human capital contract, in which the investor provides a bulk payment up front. Once the student receives the upfront capital and spends it on tuition, the entire obligation flows the other way, from student to investor. Therefore, it is in the student's interest to avoid payment or even disappear if possible, even if their income is below what would be the break-even amount in an IBR Swap. The only "sticks" the institutional investor wields are the enforcement provisions that lenders use to enforce unsecured debts.[104] Since in the case of an IBR Swap the institutional counterparty has provided nothing up

102. Note that the total cost of collection shared between the investor and the lender may be the same or even greater than the cost of collection associated with a human capital contract. We focus on the cost of collection for the investor alone, who is able to "piggy back" on the collection efforts of the lender and thereby save money.

103. There is a wide range of contractual options available to enforce payment and income-monitoring, including required submission of student's federal income tax forms and required information sharing from student loan lenders.

104. In this context, legal uncertainty about whether human capital contracts constitute debt or are dischargeable in bankruptcy becomes important. *See infra* Part IV(C) & (D).

front, the student depends on it to make ongoing monthly payments. It is in the student's interest to maintain the relationship and provide what is needed to receive those ongoing reciprocal payments.

When students earn more than the break-even amount, the situation is reversed, and it is in the student's interest to avoid payment. But, as compared to a human capital contract, an institutional counterparty providing an IBR Swap is in much better shape. That is because the amount owed to the IBR Swap institutional counterparty is a fraction of the amount that would be owed to a human capital contract investor, as the student's total payment obligation is split between the student's lender and the swap counterparty.

Take as an example the student described above who earns $160,000. The student owes $1,726 per month on law school loans. The student also has a reciprocal obligation under which they pay the institutional counterparty $2,000, and it pays them $1,726, which represents a net payment from the student to the institutional counterparty of $274. If the student stops paying the IBR Swap institutional counterparty the $2,000 they owe, the institutional counterparty will stop paying the student the $1,726 it owes them.[105] This default results in a monthly benefit to the student of only $274, while their monthly benefit would be the full $2,000 if they defaulted on a human capital contract with the same terms.

If the student wants to avoid payment altogether, they have to default not only on their payments under the IBR Swap, but also to their lender. Obviously, if the benefit of defaulting on the student's obligation to the institutional counterparty alone is only $274 instead of $2,000 per month, their incentive to default is lower. If they choose to default on their entire obligation—$274 to the institutional counterparty and $1,726 to the lender—then they will have *both* the lender *and* the institutional counterparty seeking to enforce the obligation. In that case, presumably, there would be at least some collection synergies that would reduce the cost of collection for the institutional counterparty. And if they both fail, the institutional counterparty will only lose $274 per month. The lender takes the bigger loss from the default.

In addition, the fact that a student's income is likely to go up over the course of their post-graduate career means that they are likely to have

105. Reciprocal payments could be reduced to a single net payment. Under this system, the payment obligation of the student counterparty would be netted against the payment obligation of the institutional counterparty; one counterparty would make a net payment. If the institutional counterparty owed the student $1,726, and the student owed the institutional counterparty $2,000, there would be a single payment of $274 from the student to the institutional counterparty. In another alternative, the institutional counterparty could make its full payment directly to the student's lender and collect the full amount the student owes directly. Under this scenario, the institutional counterparty would ensure that the student was not defaulting on its loan and keeping the institutional counterparty's payments, but the institutional counterparty would be required to collect the whole amount from the student.

started their career receiving a net benefit from the institutional counterparty. In that case, they may have developed a personal "culture of compliance," in which they have gotten into the habit of making their payments to the institutional counterparty, and this culture of compliance may raise the personal, psychic costs of defaulting for the student. In any case, while collection costs do not disappear in an IBR Swap, limiting collection problems to high-earning students, and even decreasing them substantially, should dramatically decrease collection costs.

But more importantly, the fact that IBR Swaps coordinate with governmental loan programs means that the government is still collecting a significant portion of the student's repayment obligation. The government can use collection mechanisms that are superior to those available to ordinary lenders or human capital contract providers. It currently administers $1 trillion of student loans,[106] and thus economies of scale drive down costs of collection. In addition, once a student has defaulted, the government has the ability to withhold tax refunds, garnish paychecks, and take a portion of social security benefits—all without a court order.[107] It also has made student loans non-dischargeable in bankruptcy.[108] If human capital contracts were entered into with private investors, none of the collection benefits of governmental student loans would apply. The private parties would be left to their own devices to enforce collection, and these costs could represent a significant cost per agreement, driving up the price. In the case of an IBR Swap, however, because the payment obligation is split between the government lender and the IBR provider, the collection methods available to the government are not lost, but can be used to collect a portion of the student's repayment obligation.

2. *IBR Swaps Coordinate with Government Student-Loan Programs*

Unlike human capital contracts, the IBR Swap complements, rather than competes with, existing governmental student loan programs. The benefit of this coordination is apparent in reducing collection costs, but it also has additional benefits. Under current law, the federal government provides loans to students for attending law school at a fixed rate. There is some controversy currently about whether the rate is too high, but it is inarguably lower than could be obtained from private lenders. Because the federal government still provides the upfront capital (in the form of a student loan) for a student's education when the student enters an IBR Swap, the student can benefit from this favorable interest rate.

106. *See* Rohit Chopra, *Student Debt Swells, Federal Loans Now Top a Trillion*, CONSUMER FIN. PROTECTION BUREAU (July 17, 2013), http://www.consumerfinance.gov/newsroom/student-debt-swells-federal-loans-now-top-a-trillion/ [https://perma.cc/manage/create/].

107. *See Federal Loans*, STUDENT LOAN BORROWER ASSISTANCE, http://www.studentloanborrowerassistance.org/collections/federal-loans/ [https://perma.cc/TPA5-HDQW] (last visited Feb. 27, 2016).

108. *See* 11 U.S.C. § 523(a)(8) (2012).

Human capital contracts, on the other hand, compete with federal student loans. A student who wishes to enter into a human capital contract must forego student loans in favor of the contract.[109] Therefore, any subsidies that are provided by the federal government to the student loan industry by guaranteeing or directly offering student loans cannot be used by students who obtain their financing through human capital contracts. In addition, any loan-forgiveness programs offered or subsidized by the government will not be available to human capital contract holders. In this way, human capital contracts must not only provide terms that are more attractive than those offered by traditional lenders, they must offer terms that are more attractive than those offered by the government.

IBR Swaps are designed specifically to be used in conjunction with federal student loan programs, and so any subsidy available through the federal student loan system is also available to students with IBR Swaps. If the government should ever increase the subsidies provided through the student loan system, this would negatively impact a market for human capital contracts. IBR Swaps, on the other hand, act in the opposite way. They complement governmental loan programs and automatically integrate any benefits provided by government loans into themselves. The price of an IBR Swap is directly tied to the amount a student needs to pay back their loans, and it therefore automatically incorporates any subsidies or benefits provided by the government into the Swap. If the government loan rate goes down, then the amount of reciprocal payments a student needs to cover their loan payments goes down, and the price of an IBR Swap would also go down.[110]

This Article does not assess how the IBR Swap would or should interact with government income-based repayment programs. That question is a complex one, and the objective here is to present the basic IBR Swap concept. For example, understanding how an IBR Swap program would interface with a government income-based repayment program would require discussion of which counterparty should capture benefits of the government subsidy. Also, the mechanics of how the Swap would interact

109. Students are presumably not required to completely forego student loans, but each dollar they acquire under a human capital contract is a dollar they do not acquire through a student loan.

110. If the government changes the repayment rate on loans that have already been disbursed, the IBR "price" has already been fixed, and so the student's obligation to the institutional counterparty would presumably not change. But in this case, the student still automatically receives the benefit of a reduced student-loan repayment obligation, because the reciprocal payment from the institutional counterparty to the student would be more than the student needs to repay their loans. Absent any contractual provision to the contrary, they could keep the difference, or it would be incorporated into the calculation of the single net payment made from the student to the institutional counterparty or from the institutional counterparty to the student.

with a student election of government-sponsored, income-based payments would require explication.[111]

3. *IBR Swaps Do Not Put Capital at Risk*

Finally, IBR Swap institutional counterparties avoid a part of the costs associated with providing upfront capital to invest in a student's education. Under an IBR Swap, the student borrows the upfront capital from a lender, and accordingly the IBR Swap institutional counterparty does not have to take on the cost of such capital. Of course, the capital still has a cost, which is the interest rate the student owes the lender, and the IBR Swap must be priced to compensate the student for the cost of capital the student obtained through a traditional loan. But the institutional counterparty does not have to take on much additional debt of its own.[112]

Again, a human capital contract does not have this feature. A provider of human capital contracts has to invest significant capital up-front on the promise of repayment in future years. Because providers of human capital contracts must provide capital upfront to students, because they won't receive it back for many years, and because they are likely to receive less of it back in the early stages of the contract, providers of human capital contracts have the costs associated with acquiring the capital to provide to students.

---

111. One possibility would be for the institutional counterparty's obligations to be unaffected by a student's participation in a governmental IBR program and all the benefit to be received by the student. So, if a student had an IBR Swap under which the student paid the institutional counterparty 10% of their income in exchange for the counterparty's payment of $1,726 per month, both parties payments would remain the same whether the student made use of a governmental IBR program or not. If the student's income was $50,000, they would pay the institutional counterparty $417 per month (10% of their income), and the institutional counterparty would pay the student $1,726 per month—enough to make the payments on their student loan. But the student would also qualify to make payments on their loan of $417 per month (slightly simplified math) under a governmental IBR program, and if the student chose to take advantage of it, they could use the $1,726 they received from the institutional counterparty to pay the government $417, and still have $1,309 in their pocket. Of course, the other possibility would be for the institutional counterparty to receive the benefit of the government IBR program. In either case, the chance of the benefit would be incorporated into the "price" of the IBR Swap, decreasing the overall price the student pays if the benefit were to be captured by the institutional counterparty. Of course, whichever party was contractually permitted to capture the benefit of government IBR would, in effect, be bearing the risk that the government will discontinue or change its IBR programs.

112. The institutional counterparty will probably have to take on some debt, since it is predictable that students' earnings will be lower in the earlier part of their careers than the latter parts, and therefore the IBR Swap institutional counterparty will have to borrow money to pay out more than it is receiving in the first several years.

IV. CURRENT LEGAL REGIME

The practical benefits unique to IBR Swaps discussed above are not the only benefits that make the IBR Swap superior to other income-share arrangements. IBR Swaps, because they are a recognizable financial instrument, do not face as much legal uncertainty as human capital contracts face under current law.

Proponents of human capital contracts have argued that legal uncertainty is a significant impediment to the development of a market for human capital contracts.[113] That uncertainty arises primarily from the fact that a human capital contract is not a preexisting (and therefore recognizable) category of financial instrument. Is it an "equity investment in human beings," as Milton Friedman suggested over half a decade ago?[114] Is it a form of debt? Is it a kind of insurance? Is it a partnership or joint venture between the student and the contract provider?[115] In order to know what law would apply to a human capital contract, one needs an answer to these questions. To remove some of this uncertainty, Senator Marco Rubio and Representative Tom Petri introduced legislation to clarify the legal treatment of income-share agreements in 2014, and in 2015, Representatives Todd Young and Jared Polis introduced similar legislation.[116]

For IBR Swaps, the legal and regulatory landscape is quite different from that of other income-share agreements like human capital contracts. The design of IBR Swaps makes the question of how they would be regulated under existing law much easier to answer. Because they are a preexisting category of financial instruments—a derivative—there is no question about whether they are an "equity investment in a human being." They are not. Just like other swap transactions, the agreement is legally

113. *See, e.g.*, PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK, *supra* note 24, at 12 ("Significant legal uncertainty exists regarding the treatment of [human capital contracts]. . . . this legal uncertainty has made it very difficult for any kind of market to develop on a larger scale. . . . A major impediment to the growth of [a human capital contract] industry is regulatory uncertainty: not only are some of the rules uncertain, but even the *source* of any future rules is also uncertain . . . . [I]s [a human capital contract] more like a loan, an investment contract, or a hedging instrument?"); *see also* Jacobs & van Wijnbergen, *supra* note 16, at 5 (noting "legal problems prevent the execution of both equity and insurance contracts by the private sector in the case of education financing," then listing as causes "contract enforceability," illegality of "slavery and indentured labor," which makes it illegal "to sell claims on future incomes," and "bankruptcy laws" not covering equity-like financing); *id.* ("These legal limitations effectively preclude financial contracts that are contingent upon the returns of human-capital investment."); *id.* at 6 ("[G]iven the legal limitations, private contracts are currently prohibitively costly to execute. Attempting to eliminate legal limitations to facilitate trade in equity-type contracts requires far-reaching changes in the legal system that are also very costly.").

114. *See* Friedman, *supra* note 18, at 140.

115. *See* Oei & Ring, *Human Equity?*, *supra* note 24, at 723–25.

116. *See* H.R. 3432, 114th Cong. (1st Sess. 2015); H.R. 4436, 113th Cong. (2d Sess. 2014).

enforceable.[117] Because nothing of value is provided upfront to the student in exchange for a future promise of repayment, they are not debt, and are therefore not subject to fair lending laws or laws that apply specifically to student loans.[118] For the same reason, they are not "securities" subject to securities regulation; the investor makes no disbursement in exchange for a speculative return.[119] As discussed below, they are not insurance under current law. Even their treatment in bankruptcy[120] and how they would be taxed[121] is more certain than human capital contracts.

But the fact that regulatory treatment of IBR Swaps is relatively certain under existing law does not mean that such treatment is right, from a normative perspective. Recently, scholars have begun to examine what

117. Jacobs & van Wijnbergen, *supra* note 16, at 5 ("[I]n general, contract enforceability of private equity contracts is problematic because legal frameworks are not yet adapted to protecting the rights of investors who provide the funds for the investment.").

118. *Id.* at 12; *see also* PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK, *supra* note 24, at 12 ("[P]olicymakers should make clear that the total cost of [a human capital contract] should not be used retrospectively to impute an interest rate for usury purposes . . . .").

119. The income-share obligation of the student counterparty is not a security, nor is the IBR Swap transaction itself. Legal scholars have analyzed whether human capital contracts are securities for regulatory purposes, focusing on the possibility that investing in a person, taking an "equity" interest in future earnings, could fall within the catchall category of "investment contract" for SEC purposes. *See* 15 U.S.C. § 77b(a)(1) (2012); *id.* § 78c(a)(1). In order to be an investment contract, a transaction must involve (1) investment of money, (2) in a common enterprise, (3) to earn a profit, (4) solely from the effort of others. *See* Schwartz, *supra* note 24, at 1156–63. In an IBR Swap, the institutional counterparty does not make an investment of money in the student. The IBR Swap involves a contractual commitment to pay a loan obligation in exchange for payment of a share of income. There is not an investment or disbursement on which the institution seeks a return. If the student fails to remit income, the institution ceases to make loan payments. The payments are simultaneous. The IBR Swap itself is also not a security. *See* Securities Act of 1933 § 2A, 15 U.S.C. § 77b(a)(1) (providing definition of security does not include non-security-based swap agreement as defined in section 206C of Gramm-Leach-Bliley Act). The IBR Swap would not be a securities-based swap. *See Final Rules and Interpretations i) Further Defining "Swap," "Security-Based Swap," and "Security-Based Swap Agreement"; ii) Regarding "Mixed Swaps"; and iii) Governing Books and Records for "Security-Based Swap Agreements*, U.S. COMMODITIES & FUTURES TRADING COMM'N, *available at* http://www.cftc.gov/idc/groups/public/@newsroom/documents/file/fd_factsheet_final.pdf [perma.cc/NAW7-J24C] (last visited Feb. 27, 2016); *see also* Jacobs & van Wijnbergen, *supra* note 16, at 5 ("Since human-capital contracts are not legally acknowledged as securities, trade in claims on human capital is legally obstructed.").

120. PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK, *supra* note 24, at 13; *see also* Jacobs & van Wijnbergen, *supra* note 16, at 5 ("[B]ankruptcy laws do not generally feature provisions for graduates who declare themselves bankrupt to avoid dividend payments to financiers, whereas bankruptcy laws do cover provisions for debt contracts." (citation omitted) (citing Palacios, *Human Capital Contracts*, *supra* note 5)).

121. PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK, *supra* note 24, at 14 ("[P]articipants face some uncertainty regarding tax treatment of payments [under a human capital contract] . . . .").

rules *should* apply to income-share agreements like human capital contracts.[122] Since IBR Swaps are different from other human capital contracts in critical ways, we address regulatory questions and opportunities raised by IBR Swaps.

A. *Swaps: Federal Derivatives Regulation and State Insurance Laws*

Two different regulatory regimes may pertain to a contract involving transfer of loan default risk in exchange for a speculative return: derivatives regulation and insurance regulation.[123] Subsection 1 below explains that an IBR Swap would not be subject to federal regulations pertaining to over-the-counter (OTC) derivatives.[124] The IBR Swap, though, serves an "insurance" or hedging function for the student counterparty.[125] Subsection 2 discusses the extent to which IBR Swaps could, in theory, fall within the ambit of state insurance regulation. It considers the purpose and design of insurance regulation, the identities of IBR Swap counterparties, and the fitness of state insurance regulators (as opposed to other intermediaries), for addressing regulatory concerns that the IBR Swap could raise. This Subsection argues that the IBR Swap should not be subject to state insurance regulation. However, it identifies the possibility that a regulator could claim that an IBR Swap program falls within its jurisdiction.

1. *Federal Derivatives Regulation*

Simply put, the IBR Swap as presented above would be exempt from federal regulations pertaining to OTC derivatives.[126] Derivatives regulation centers on clearinghouse requirements—rules requiring that certain swaps are confined to a derivatives clearing organization registered with

---

122. *See* Oei & Ring, *Human Equity?*, *supra* note 24, at 729–36; Schwartz, *supra* note 24, at 1165–74.

123. As mentioned above, the IBR Swap would not be a security. *See supra* note 119 and accompanying text.

124. *See infra* notes 126–31 and accompanying text.

125. There are insurance policies designed to cover risk of inability to meet loan obligations. These kinds of policies, however, tend to have a maximum coverage of twenty-four months and are designed to cover loan payments upon occurrence of some event, such as unemployment, that temporarily affects the insured's ability to pay. In addition to policies especially for payment protection, other insurance policies and workplace benefit plans can cover temporary inability to pay loans, such as disability or unemployment benefits. *See* Ana Gonzalez Ribeiro, *Is Loan Protection Insurance Right for You?*, INVESTOPEDIA, http://www.investopedia.com/articles/pf/08/loan-protection-insurance.asp [https://perma.cc/5B64-ESQL] (last visited Feb. 27, 2016). The IBR Swap offers a very different kind of hedging arrangement, obviously, than insurance policies that protect borrowers for a limited time if they cannot pay.

126. Title VII of the Dodd-Frank Act amends the Commodities Exchange Act (CEA) to require that OTC derivatives trade through clearinghouses, but provides exceptions. *See* Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) § 723, 7 U.S.C. § 2 (2012); Commodities and Exchange Act, 7 U.S.C. §§ 2(h)1, 2(h)7.

the Commodity Futures Trading Commission (CFTC).[127] However, OTC derivatives in which one counterparty is not a financial entity—like the IBR Swap—are exempt from clearinghouse requirements.[128] Derivatives in which one counterparty is hedging or mitigating commercial risks are also exempt.[129] Because student counterparties would meet the end-user exception for non-financial entities, we need not assess whether student counterparties to IBR Swaps would be mitigating "commercial risks" as defined by the CFTC.

Clearinghouses serve enforcement functions that mitigate risks and market effects of counterparty defaults.[130] Though IBR Swaps would not be regulated derivatives, it may be interesting to consider whether IBR Swap programs might want to use a private clearinghouse system.[131]

2. *State Insurance Laws*

Insurance regulation is a state enterprise that primarily addresses consumer protection concerns.[132] However, Dodd-Frank does create federal regulatory oversight for "systemically important" insurers, which addresses the systemic risk associated with failure of the largest insurers, such as AIG.[133] Whether an IBR Swap program would fall within the ambit of insurance regulation depends upon factors such as the nature of the institutional counterparty and the existence of intermediaries (such as schools or a private clearinghouse) that can protect the interests of student counterparties.[134] Robust discussion has surrounded the question of

127. *Id.* §§ 2(h)(1), 2(h)(7).

128. Dodd-Frank Act § 723; *see also* CFTC *End-User Exception to the Clearing Requirement for Swaps Rule,* 17 C.F.R. § 39.1 (2016).

129. 17 C.F.R. § 39.6(a)(i).

130. *See* Stout, *Derivatives*, *supra* note 54, at 34–35.

131. *See generally* Mark J. Roe, *Clearinghouse Overconfidence*, 101 CALIF. L. REV. 1641 (2013) (describing clearinghouse system and questioning its efficacy in financial regulation).

132. *See* McCarran-Ferguson Act of 1945, 15 U.S.C. §§ 1011–1015 (2012) (stating that regulation of insurance by states is in public interest and that no federal law of general applicability shall preempt state laws regulating business of insurance); Daniel Schwarcz & Steven L. Schwarcz, *Regulating Systemic Risk in Insurance*, 81 U. CHI. L. REV. 1569, 1579–84 (2014). *See generally* KENNETH S. ABRAHAM & DANIEL SCHWARCZ, INSURANCE LAW AND REGULATION 2–5 (6th ed. 2015); THE FUTURE OF INSURANCE REGULATION IN THE UNITED STATES (Martin F. Grace & Robert W. Klein eds., 2009); John Patrick Hunt, *Rating Dependent Regulation of Insurance*, 17 CONN. INS. L.J. 101 (2010); Jonathan R. Macey & Geoffrey P. Miller, *The McCarran-Ferguson Act of 1945: Reconceiving the Federal Role in Insurance Regulation*, 68 N.Y.U. L. REV. 13 (1993); Susan Randall, *Insurance Regulation in the United States: Regulatory Federalism and the National Association of Insurance Commissioners*, 26 FLA. ST. U. L. REV. 625 (1999).

133. *See* Schwarcz & Schwarcz, *supra* note 132, at 1589–93; Daniel Schwarcz, *Regulating Insurance Sales or Selling Insurance Regulation?: Against Regulatory Competition in Insurance*, 94 MINN. L. REV. 1707, 1770–71 (2010); *see also* Hunt, *supra* note 132; Manns, *supra* note 45.

134. Substantively, insurance laws establish (1) licensing requirements for insurers, (2) fiduciary duties to policy holders, (3) capital reserves, (4) disclosure of

whether derivatives should be regulated as insurance.[135] In order for a contract to be "insurance," the protection buyer must have an insurable interest—a property interest or other risk of real loss that the contract covers.[136]

An IBR Swap would fall within the class of derivatives in which a hedging counterparty has an insurable interest. The student counterparty is in privity of contract with law school lenders and, by entering into the Swap, is hedging risk of default on those loans. Because of this, state regulators could claim that IBR Swaps are subject to insurance laws. Despite this possibility, however, IBR Swaps do not necessarily raise the concerns that insurance regulation is designed to address. This Subsection now presents these concerns, discussing the extent to which IBR Swaps implicate them.

While the major, overarching concern of insurance law is consumer protection, the need for consumer protection stems from two different sources. The first is lack of sophistication among consumers and consumers' general vulnerability in procuring essential insurance products. The second is solvency risk due to the nature of insurance firms, namely, the facts that (1) they have an inverted production cycle that detaches contracting and pricing from customers' receipt of the product[137] (payment on claims which may happen years later or never) and (2) they have diffuse creditors, or policyholders, that do not assert control to discourage excessively risky decision-making when firms encounter distress.[138]

---

insurers' financial data to regulators, (5) approval of form contracts, and (6) restrictions on prices that insurers can charge consumers. Henderson, *supra* note 53, at 42–48.

135. *See, e.g.*, Thomas Lee Hazen, *Disparate Regulatory Schemes for Parallel Activities: Securities Regulation, Derivatives Regulation, Gambling, and Insurance*, 24 ANN. REV. BANKING & FIN. L. 375 (2005); Thomas Lee Hazen, *Filling a Regulatory Gap: It is Time to Regulate Over-the-Counter Derivatives*, 13 N.C. BANKING INST. 123 (2009) [hereinafter Hazen, *Filling a Regulatory Gap*]; Henderson, *supra* note 53; Stout, *Regulate OTC Derivatives*, *supra* note 77; *see also Hearing to Review the Role of Credit Derivatives in the U.S. Economy: Hearing Before the H. Comm. on Agric. Comm.*, 110th Cong. (2008) (statement of Eric Dinallo, Superintendent, Ins. Dep't, N.Y.), *available at* http://agriculture.house.gov/testimony/110/h91120/Dinallo.pdf [https://perma.cc/6MFD-N3S5].

136. The insurable interest doctrine can be complex; scholars and policymakers debate what constitutes an insurable interest for purposes of insurance laws' jurisdiction. *See* Hazen, *Filling a Regulatory Gap*, *supra* note 135, at 420–26; Michael J. Henke, *Corporate-Owned Life Insurance Meets the Texas Insurable Interest Requirement: A Train Wreck in Progress*, 55 BAYLOR L. REV. 51, 53–54 (2003) (discussing Texas's "insurable interest" doctrine); Hunt, *supra* note 133; Roy Kreitner, *Speculations of Contract, or How Contract Law Stopped Worrying and Learned to Love Risk*, 100 COLUM. L. REV. 1096, 1099–1100 (2000); *see also* GRAYDON S. STARING, LAW OF REINSURANCE § 6:1 (1993) ("In limited space we can talk around insurable interest but never talk it through. A standard text confesses that '[i]t is very difficult to give any definition of an insurable interest,' and then discusses it for about 70 pages" (alteration in original) (quoting 1 M. MUSTILL & J. GILMAN, ARNOULD ON THE LAW OF MARINE INSURANCE AND AVERAGE §§ 331–410 (16th ed. 1981))).

137. *See* Henderson, *supra* note 53; Hunt, *supra* note 132.

138. *Id.*

With respect to consumer vulnerability, an IBR Swap program would not necessarily raise the same concerns as insurance products. Law students are a specific class of consumer, and participation in the program would not be mandatory. This is *not* to say that student counterparties need no protection, are sophisticated, or enjoy even bargaining positions with institutional counterparties. It is only to say that regulation of such interactions in the IBR Swap context may raise issues different from those that arise in various insurance markets and so-tailored regulation may be appropriate. For example, intermediaries such as school advisors and financial aid office professionals may be in a better position to protect students' interests than state insurance regulators. Protection of the "protection buyer"—the student counterparty to an IBR Swap—involves both explanation of the programs' terms and conditions and also protection from unfair pricing. Private intermediaries familiar with legal education, law school graduates' career trajectories, and federal lending programs would likely be better suited than state officials to advise protection buyers in the IBR Swap context.

Regulatory concerns that stem from dangerous incentives of an inverted production cycle would most likely not arise in an IBR Swap program. Unlike other businesses, insurance firms contract with customers who pay in the form of premiums in advance of product delivery—payment on claims which may happen years later or not at all. Because of this, insurance firms lack the discipline that comes with having to spend revenues on market products and services deliverable contemporaneously with customer contracts.[139] Insurance firms can fall into a model that is not unlike a Ponzi scheme, where they solicit investment from new customers, using that revenue to pay claims to prior customers because they lost revenues from prior customers on risky investments.[140] Insurance regulation imposes capital requirements and financial disclosure requirements on insurance firms to avoid this result.[141]

An IBR Swap program would not have the inverted production cycle associated with insurance firms. The institutional counterparty begins paying on the student counterparty's obligations to law school lenders as soon as they become due. The institutional counterparty's performance obligation is certain in amount and contemporaneous with the student counterparty's payments. Furthermore, incomes generally rise over time, both because of the effects of inflation on wages and because more experienced lawyers tend to earn more than less experienced ones. Because of the likelihood that early earnings would be lower than later earnings, the institutional counterparty is likely to owe more than it collects in the early years of each contract and overall. Thus, it will need some source of capital at the outset, but its need to accumulate capital is the inverse of the

139. *See* Henderson, *supra* note 53.

140. *Id.*

141. *Id.*

production cycle that is dominant in the insurance industry, in which the insurer collects money over time and is required to pay out later because of an insured event.

With respect to concerns arising from the capital and governance structure of insurance firms, the extent to which an IBR Swap program would raise these concerns depends upon the identity of the institutional counterparty. In other kinds of firms, capitalization often comes from a large number of diffuse shareholders, along with a small group of creditors, banks, or other lenders, with monitoring capacity. When the firm faces distress, creditors often can exercise monitoring and control functions through loan covenants, other contractual obligations, and collateral obligations, preventing excessively risky behavior.[142] An insurance firm, however, has a diffuse group of creditors—policyholders—who are not in a position to monitor and exert control like institutional creditors do. This leaves insurance firms more prone than other businesses to excessive risk-taking in hard times. Again, capital adequacy, disclosures, and other risk protections imposed by insurance laws address this risk.

Counterparties to derivatives are often financial institutions that do not share the same governance and capital structure as insurance firms. If the institutional counterparty to an IBR Swap were such a financial institution, then the state insurance law requirements designed to address these concerns would not be necessary. Many swap counterparties are also not exclusively in the business of entering into one kind of swap. They are firms with multiple kinds of investments, speculative, and hedging positions across product types and even industries. It is possible that the institutional counterparty to an IBR Swap could be an entity dedicated to entering into these swaps with a pool of student counterparties and a capital structure analogous to that of an insurance firm. If this were the case, then some regulatory requirements to control excessive risk-taking in the event that the pool of swaps leans towards negative value would be desirable.

It is not clear, though, that state insurance regulation would be the best mechanism to protect student counterparties in this situation. Often, clearinghouse requirements address counterparty risk and capital adequacy concerns in the derivatives context. A private clearinghouse for IBR Swaps could be a better solution to counterparty risk for students than subjecting the IBR Swap program to state insurance regulation. However, clearinghouse requirements could raise the transaction costs of IBR Swaps, making the Swaps less advantageous for students. Given this con-

142. There is ample literature on the efficacy of creditors' monitoring functions. *See, e.g.*, Robert E. Scott, *A Relational Theory of Secured Financing*, 86 COLUM. L. REV. 901 (1986); Joanna M. Shepherd, Frederick Tung & Albert H. Yoon, *What Else Matters for Corporate Governance?: The Case of Bank Monitoring*, 88 B.U. L. REV. 991 (2008); Frederick Tung, *Leverage in the Boardroom: The Unsung Influence of Private Lenders in Corporate Governance*, 57 UCLA L. REV. 115 (2009). This simple explanation is just to compare typical insurers to possible IBR Swap counterparties.

cern, regulators might require institutional counterparties to have diversified portfolios, and perhaps subject them to other capital adequacy requirements as regulated banking institutions.

In conclusion, the IBR Swap should not be subject to state insurance regulation. IBR Swaps do not raise the same concerns for protection sellers and buyers that insurance policies and firms do. Additionally, to the extent that IBR Swaps do warrant protection for student counterparties or attention to the nature of institutional counterparties, private intermediaries would be better suited to address concerns than state insurance regulators would be. That said, the IBR Swap does involve hedging of risk by a student-counterparty with an insurable interest. As such, a state insurance regulator could potentially claim that IBR Swaps fall within insurance regulators' jurisdiction.

To the extent that the IBR Swap serves an insurance function for the student counterparty, some may wonder if moral hazard among student counterparties could undermine the efficacy of an IBR Swap program. Moral hazard is often identified as a central problem with insurance products.[143] Moral hazard "is a form of ex post opportunism" that arises in insurance markets when the existence of insurance reduces the insured's incentives to avoid the insured loss.[144] In the case of human capital contracts, some commentators fear that owing a percentage of one's income to investors might decrease a person's incentives to earn income. The existence of moral hazard in this context is an empirical question, and available evidence suggests that we should not be overly concerned about it.[145]

---

143. *See, e.g.*, SHILLER, *supra* note 29, at 113. For livelihood insurance, Shiller proposes that individual livelihood insurance only cover 50% of the decline in income "since the person is only reimbursed for half of his or her own income drop, he or she still has an incentive to work hard, reducing, if not eliminating, the moral hazard problem." *Id.* If the insurance only covers 15% of a student's income, the moral hazard problem is presumably even less.

144. FRANCESCO PARISI, THE LANGUAGE OF LAW AND ECONOMICS: A DICTIONARY 187, 188 (2013); *see also* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 50–52 (3d ed. 2000) (stating that moral hazard arises when behavior of insuree changes after purchase of insurance so that probability of loss or size of loss increases).

145. Scholars have been interested in the question of how paying a percentage of one's income affects labor participation for decades. That is because the income tax functions just like a human capital contract in that each dollar a taxpayer earns is reduced by a percentage that is paid to a third party. It is axiomatic that the disincentives for labor participation that an income tax produces are a significant potential flaw in any income tax regime. However, the empirical evidence suggests that relatively low income tax rates have modest effects on labor participation. *See, e.g.*, Robert McClelland & Shannon Mok, *A Review of Recent Research on Labor Supply Elasticities* (Cong. Budget Office, Working Paper No. 2012-12, 2012), *available at* http://www.cbo.gov/sites/default/files/cbofiles/attachments/10-25-2012-Recent_Research_on_Labor_Supply_Elasticities.pdf [https://perma.cc/AA6L-WZ3U].

Also, a market for human capital contracts or IBR Swaps might reduce a variety of moral hazard problems that exist in the current student debt regime. For example, some commentators have decried the moral hazard that arises when law

B. *Enforceability of the Contract*

Commentators have expressed concern that a human capital contract might not be enforceable.[146] If human capital contracts are unenforceable, then a student would be permitted under the law to walk away from these agreements without sanction. The drafters of the proposed legislation regarding income-share agreements thought enforceability was enough of an issue that they included a section that stated: "Any income share agreement that complies with the requirements of [this law] shall be a valid, binding, and enforceable contract notwithstanding any State law limiting or otherwise regulating assignments of future wages or other income."[147] The history of wage assignments and legal restrictions on them is long, and while no one has made a strong case that human capital contracts would be unenforceable under current law in any state, some legal uncertainty presumably remains.

But whatever uncertainty surrounds the enforceability of human capital contracts; it does not affect IBR Swaps.[148] An IBR Swap does not involve any sort of assignment of wages; it is simply a derivative that uses future earnings to measure the payment obligation of one of the counterparties. The institutional counterparty has no claim directly against the student's employer. There is no wage garnishment.[149] The institutional counterparty cannot demand specific performance. The Swap is not an unenforceable wager[150] because, as discussed above, the

---

schools can charge unsustainably high tuition and the students can avoid paying back their loans by making use of current federal income-based repayment and loan forgiveness programs. *See, e.g.*, Steven J. Harper, *Bankruptcy and Bad Behavior the Real Moral Hazard: Law Schools Exploiting Market Dysfunction*, 23 AM. BANKR. INST. L. REV. 347 (2015).

146. The most extreme claim is that human capital contracts may be illegal or unconstitutional as a form of slavery, indentured servitude, or peonage. Jeff Schwartz, for example, argues that an exchange of money for future income is "equity in a person" and therefore a form of "ownership in people," and so, it could be argued that "they should be outlawed on constitutional or policy grounds." *See* Schwartz, *supra* note 24, at 1121. He ultimately concludes, however, "human-equity investing passes [constitutional] muster." *See id.* at 1122, 1135–38; *see also* Jacobs & van Wijnbergen, *supra* note 16, at 5. Some commentators make vague statements about the enforceability of human capital contracts if students choose to breach. *See, e.g.*, Jacobs & van Wijnbergen, *supra* note 16, at 5 ("It is not generally possible to sell claims on future incomes. For example, some states in the U.S. do not allow this." (citation omitted) (citing Palacios, *Human Capital Contracts*, *supra* note 5)); *see also supra* note 117.

147. H.R. 4436, 113th Cong. § 101(b) (2d Sess. 2014).

148. For discussion of enforceability of a "settlement amount" in case of breach, see *supra* note 58.

149. Although, if there was, some states would require the term to be shorter than the likely term of an IBR Swap. *See* U.S. DEP'T OF LABOR, WAGE & HOUR DIV., FACT SHEET #30: THE FEDERAL WAGE GARNISHMENT LAW, CONSUMER CREDIT PROTECTION ACT'S TITLE 3 (CCPA) (rev. July 2009), *available at* http:www.dol.gov/whd/regs/compliance/whdfs30.pdf [https://perma.cc/GR4H-6YFR].

150. At common law, wagering agreements were unenforceable. *See* Stout, *Derivatives, supra* note 54, at 1. However, a derivative like a swap agreement is not

student counterparty is hedging risk pertaining to an insurable interest.[151] In short, an IBR Swap is a fully enforceable contract under current law.

C. *Debt Treatment*

Proponents of human capital contracts also worry that such agreements could be classified as "debt," which would make compliance with a variety of laws difficult at best. For example, if human capital contracts constituted debt or a "private education loan" they might be regulated under the Truth in Lending Act of 1968.[152] These federal laws require lenders to clearly disclose interest charges as an annual rate. But for a human capital contract, calculating such a rate is impossible because the amount the "borrower" is obligated to repay to the "lender" varies depending on the income of the borrower.

Furthermore, if human capital contracts are loans, they might be subject not only to the disclosure requirements, but also to limits on the maximum amount that could be charged.[153] Many states have limits on the total amount of interest that can be charged to a borrower, and these limits, if applicable, would mandate relatively low ceilings on the maximum amount a high earner could pay under a human capital contract. If these limits were interpreted to apply to human capital contracts, they would remove most of the benefit of such structures, since they would prevent

---

an unenforceable wagering agreement because the student counterparty is hedging risk. *See supra* notes 52–54 and accompanying text.

151. *See supra* notes 50–53 and accompanying text.

152. "Private education loans" are subject to numerous reporting requirements under Section 128(e) of the Truth in Lending Act. *See* 15 U.S.C. § 1638(e)(1) (2012). These disclosures fit poorly with (or they are impossible to comply with) human capital contracts. For example, the lender must report "(A) the potential range of rates of interest applicable to the private education loan; . . . [and] (C) limitations on interest rate adjustments, both in terms of frequency and amount, or the lack thereof, if applicable." *Id.*

153. The extent to which characterizing human capital contracts as loans would limit their pricing—and the limits that would apply—depends on a variety of factors surrounding usury laws. State usury laws restrict the amount of interest that lenders can charge on consumer loans. Though these laws can appear to impose straightforward rate limits, this is often not the case. For example, in some cases usury limits are preempted by federal law (such as home equity loans). Federal law also permits federally insured financial institutions to charge the highest interest rate limits permitted among the various states in which they are located, undermining the implementation of state usury laws with lower limits. In addition, (1) state usury laws contain numerous exceptions, such as for retail installment loans and loans issued by certain types of institutions (e.g., credit unions); (2) interest rate ceilings may be higher than they appear due to special rules for compounding fees, calculating balances and rates; (3) remedies for usury law violation may be narrow; and (4) contracting parties may avoid usury limits by drafting price terms in a way that obscures disbursement versus interest components of the transaction. *See generally* Richard M. Hynes & Eric A. Posner, *The Law and Economics of Consumer Finance* (U. Chi. Law & Econ., Olin Working Paper No. 117, 2001), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=261109 [perma.cc/YGC3-JE29].

the contract from capturing much, if any, upside gain from high-earning graduates.

There is no statutory definition of debt in either the Truth in Lending Act or the Higher Education Act (HEA), but an essential characteristic of debt is that it involves a transfer of funds from one party, the lender, to another party, the borrower, in exchange for a promise of future repayment, usually with interest.[154] In the case of the IBR Swap, there is no initial transfer of funds from the institutional counterparty to student, and so the relationship cannot be characterized as debt.[155] Thus, the IBR Swap removes ambiguity about whether lending disclosure regimes, rate ceilings, or usury laws apply.

D. *Bankruptcy*

Several commentators have mentioned that uncertainty about how human capital contracts would be treated in bankruptcy creates regulatory uncertainty.[156] For example, if a human capital contract constitutes a "qualified education loan," then it is not typically dischargeable in bankruptcy unless the student can show "undue financial hardship," which is generally difficult to show.[157] Even ignoring issues specific to student loans, some argue that it would be difficult to figure out how to treat a human capital contract in bankruptcy. In a bankruptcy proceeding, the obligations of the bankrupt party are prioritized, with low priority obligations being subject to discharge. There is some fear that a bankruptcy court would extinguish future obligations under a human capital contract unless it was classified as a private student loan under the HEA. Since a human capital contract represents a long-term obligation, and the risk of under-earning likely is highest early in the term of the agreement, extinguishing the obligation in bankruptcy early in the term of the agreement would be problematic. That uncertainty poses some risk to the investor[158] and may serve to increase the cost of the human capital contract.

Unlike a human capital contract, an IBR Swap is not a liability for a low-earning student, but an *asset.* Remember, a student who earns less than the "break-even" income receives a net benefit from the institutional counterparty. The amount that the institutional counterparty pays to the student is *more* than the student pays to the institutional counterparty, and

154. *See, e.g.*, Brown, *supra* note 57, at 144 ("If a swap is a debt instrument, it must have been issued either for money . . . or for property . . . .").

155. *See, e.g.*, *id.* at 156 (stating that payments made under interest rate swaps are not interest because there is no underlying debt (citing H.R. Rep. No. 3838, 99th Cong., 1st Sess. 457 (1985))).

156. *See, e.g.*, PALACIOS ET AL., INVESTING IN VALUE, *supra* note 24, at 13; Jacobs & van Wijnbergen, *supra* note 16, at 5.

157. *See* 11 U.S.C. 523(a)(8) (2012).

158. Since a human capital contract adjusts by its terms to the investee's ability to pay, it is not as substantial a burden on a low-earning student, and thus the need to discharge it in bankruptcy is dramatically diminished. But, presumably, there is still some possible risk.

so every month the IBR Swap makes the student richer than they would be without the IBR Swap. Thus, a student whose income is so low as to result in bankruptcy would not want to have their IBR Swap discharged. Far from it, the IBR Swap is the one thing that is enabling the student to make their presumably non-dischargeable student loan debt payments possible. Thus, the IBR Swap may avoid the possibility of discharging the student's obligation because of bankruptcy.

Literature on treatment of derivatives in bankruptcy generally concerns Chapter 11 re-organization and the status of derivatives with two institutional counterparties.[159] In the IBR Swap context, the effect of student counterparty bankruptcy may be significantly different from that of institutional counterparty bankruptcy. In the event a student counterparty enters bankruptcy, the fact that an IBR Swap is an asset rather than a liability makes its treatment different from human capital contracts, but that does not dispose of all questions surrounding bankruptcy treatment. Because the IBR Swap represents a novel transaction, such a question would require jurisdiction-specific analysis by attorneys preparing documentation and counseling prospective institutional counterparties.

The possibility that an institutional counterparty could be undercapitalized and default or even seek bankruptcy protection to avoid payments to student counterparties also warrants careful consideration. Regulation to assure capital adequacy or otherwise protect students against institutional counterparty credit risk would need to accompany any implementation of an IBR Swap market.[160]

E. *Tax*

While bankruptcy occurs only in extreme cases, every investor and student-party to an income-share agreement has to decide how to treat the investment for tax purposes. Some commentators identify tax uncertainty as an important impediment to human capital contracts,[161] and the authors of H.R. 4436 consider it significant enough to propose definitive tax treatment.[162] The tax treatment of the IBR Swap is more certain than that of a human capital contract. That said, it is probably not especially favorable treatment.

---

159. *See, e.g.*, Douglas G. Baird & Robert K. Rasmussen, *Antibankruptcy*, 119 YALE L.J. 648 (2010); Mark J. Roe, *The Derivatives Market's Payment Priorities as Financial Crisis Accelerator*, 63 STAN. L. REV. 539 (2011).

160. *See infra* notes 180–83 and accompanying text.

161. *See, e.g.*, Oei & Ring, *Human Equity?*, *supra* note 24, at 744–46.

162. *See* H.R. 4436, 113th Cong. § 201 (2d Sess. 2014) (providing that payments made from investor to student are not includible in student's income, and payment of future income to investor constitute tax-free recovery of capital for investor until full amount of investment is recovered, after which time they are taxable income).

There are many possible options for how to treat a human capital contract for tax purposes. If it is debt, then the receipt of the initial payment from the investor to the student is not a taxable event for either party—the student does not treat it as income, and the investor cannot deduct the payment from its income for tax purposes. When the student makes percentage-of-income payments back to the investor, those payments are partially a tax-free "return of capital" and partially "interest." Return of capital has no tax implications, but interest is income to the investor, and, importantly, it is income at the "ordinary" rather than at the lower "capital gains" rate.[163] But if a human capital contract is debt for tax purposes, a method must be employed for separating the return of capital from the interest. Just as was the case when we discussed lending law, the structure of a human capital contract makes it impossible to calculate a "rate," and so there is no easy way to distinguish "interest" from "return of capital."[164]

If the human capital contract is not treated as debt, it could be treated in a number of other ways, each of which provides a different solution to the problem of how to distinguish the tax-free return of capital from taxable income. At least one commentator has raised the possibility that the best way to view at least some human capital contracts is as a partnership or joint venture, in which case the student's income would be allocated between the student and the investor for tax purposes, with one (but not both) paying tax on all of it.[165] This treatment would be extremely complicated, although it might be favorable to the parties collectively.[166]

More likely, a human capital contract would be taxed like some similar investment vehicle. For example, if it was considered more "insurance-like," it might be taxed as an "annuity."[167] It could also be taxed pursuant to the "open transaction" doctrine, in which the first money received by the investor from the student is all return of capital, and only after all is

163. For the student, the interest may be deductible under certain circumstances. For example, if the human capital contract were incurred in the course of the student's existing trade or business, then the interest would be deductible. However, if it were incurred for education, even professional education, it would likely not be deductible, unless it was deductible as student loan interest, which is subject to numerous limitations and exemptions.

164. In addition to being uncertain, this treatment would probably not be preferred by taxpayers, because there is no opportunity to defer the taxation of income and interest is taxed at the full ordinary income rate.

165. *See* Oei & Ring, *Human Equity?*, *supra* note 24, at 723–25.

166. The student's earnings would somehow be treated as partnership income subject to allocation between the parties pursuant to Subchapter K of the Internal Revenue Code. Payments from the student to the investor would be distributions, as would payments from the "partnership" into the student's own personal bank account.

167. 26 U.S.C. § 72 (2012). Under annuity treatment, capital is allocated pro rata over the years of the term of the annuity, and so each year an equal amount would be tax-free return of capital. Any amount exceeding this amount would be income to the investor.

received is the rest taxed as income.[168] This uncertainty creates tax problems, though it also creates tax opportunities, since some of the potential treatments of the transaction would be favorable from a tax perspective for the investors at least.[169]

In the case of an IBR Swap, the tax treatment is more certain (even if not particularly favorable). That is because in an IBR Swap, there is no upfront payment from one party to the other. Instead, each party agrees to make reciprocal payments to each other. The proposed methods of taxing payments made under a human capital contract discussed above—debt/interest treatment, annuity treatment, open transaction doctrine, among others—are all methods of distinguishing tax-free return of principal from taxable income. But when there is no up-front investment in the transaction, none of these methods apply. Furthermore, when there is an upfront investment in a financial product, there is a question of whether the gains that accrue to that product are ordinary income or capital gains. When there is no upfront investment, ordinary-income treatment makes the most sense.

The tax treatment of an IBR Swap, however, is still subject to some uncertainty. Congress has expressly provided for the tax treatment of most swap transactions. Section 446 of the Internal Revenue Code governs the taxation of so-called notional principal contracts. While the definition of a "notional principal contract" at first blush would appear to include an IBR Swap, in fact, the IBR Swap is probably not a notional principal contract under the Code.[170] To qualify as such, each "leg" of the

168. *See* Burnet v. Logan, 283 U.S. 404, 413–14 (1931).

169. Unsurprisingly, H.R. 4436 seeks to clarify the tax treatment of qualifying income-share agreements. Under H.R. 4436, the treatment is essentially the same as it would be under the open transaction doctrine, which is the most favorable treatment for the investor. The initial payment from the investor to the student is excluded from income of the investee, just as it would be if it was debt, and the return payments are tax-free return of capital until the whole invested amount is paid back. After that, payments are income to the investor. H.R. 4436, 113th Cong. § 201(a) (2d. Sess. 2014).

170. *See* 26 C.F.R. § 1.446-3(c)(1)(i) (2016) ("A notional principal contract is a financial instrument that provides for the payment of amounts by one party to another at specified intervals calculated by reference to a specified index upon a notional principal amount in exchange for specified consideration or a promise to pay similar amounts."). The IBR Swap is a financial instrument in which one party (the institutional investor) agrees to pay amounts at specified intervals (probably monthly) to another party (the student) by reference to a specified index (probably a fixed percentage) upon a notional amount (the amount the student borrowed in student loans) in exchange for a promise to pay. The question is whether the promise is to a pay a "similar amount." And while the payments from the student to the institutional investor are paid at specified intervals, they are not calculated by reference to a specified index, because the regulations state explicitly that "a specified index" is "an index that is based on objective financial information . . . ." *Id.* § 1.446-3(c)(2)(iii). And "objective financial information" cannot be "within the control of any of the parties to the contract." *Id.* § 1.446-3(c)(4)(ii). Because the student's income is (more or less) within the student's control, an index based on that information cannot be a "specified index" under the regula-

swap has to be calculated based on a measurement that is not "within the control of any of the parties to the contract."[171] Because the student's income is (more or less) within the student's control, the IBR Swap would not likely qualify as a notional principal contract.

When a contract to exchange reciprocal payments fails to qualify as a notional principal contract, it is presumably taxed according to general principles of tax law. The most likely treatment would be that the reciprocal payments over the course of the tax year would be "netted" so that one party has an aggregate annual payment that is positive, and the other party has an exactly corresponding negative annual aggregate.[172] The party with the positive annual payments would pay tax on that positive amount as "ordinary income." The party whose payments were negative would presumably have an "ordinary loss."

V. CONSIDERING REGULATORY ISSUES

The unique structure of the IBR Swap reduces the regulatory uncertainty that undermines human capital contracts and other income-share agreements under current law. However, the fact that the legal treatment of IBR Swaps under current law is more certain than that of human capital contracts does not exempt IBR Swaps from implicating regulatory and ethical issues. On the contrary, IBR Swaps raise a host of problematic issues that deserve careful consideration. No study to date has systematically examined the regulatory tradeoffs inherent in designing a system of higher education financing that includes various types of income-share agreements.[173] Nor do we attempt such a systematic study here. Instead, in this Section we identify three obvious regulatory concerns: the issue of disclosure requirements to protect student counterparties, the issue of institutional counterparty solvency or credit risk, and the issue of differential pricing and adverse selection.

As discussed above, those who are in favor of human capital contracts should also be in favor of IBR Swaps. But if you are alarmed (or even horrified) by the possibility of a world in which human capital contracts are the primary mode for financing higher education, then you probably oppose IBR Swaps. Thus far, this Article has discussed how IBR Swaps are like human capital contracts, but "better" in the sense that they are more efficient and do not suffer from the same regulatory impediments. We have thus far deferred a discussion about potential drawbacks of income-share agreements that warrant regulation. The Rubio–Petri Legislation clears some of the legal hurdles that impede the development of a market

---

tions, and the instrument that is based on it (the IBR Swap) is not a notional principal contract for tax purposes.

171. *Id.*

172. *See generally* Brown, *supra* note 57, at 162–65.

173. Although several scholars have made good starts. *See, e.g.*, Oei & Ring, *Human Equity?*, *supra* note 24; Schwartz, *supra* note 24.

for income-share agreements, but it does little to regulate the market it purports to create.[174] Drawbacks to a market for income-share agreements potentially include risk that income-share agreements could exacerbate income and other forms of inequality, diminish access to education for some qualified students, raise the cost of education for some students, accelerate the withdrawal of public support for education,[175] and generally lead to a society that is less egalitarian, less educated, and all around worse. It is possible, though, that well-constructed regulations can manage each potential drawback such that the benefits of an IBR Swap program outweigh the detriments.

Law students entering IBR Swaps are consumer counterparties transacting with more sophisticated institutional actors. Unless a public or private governing body standardizes contract forms or specifies disclosure requirements for IBR Swap transactions, student counterparties could be at risk of committing to contract terms that they have not contemplated and do not like. Disclosure requirements are common regulatory tools in a variety of market contexts where consumer protection is a concern.[176] Along with disclosure, standardization of contract terms is also a common regulatory strategy to reduce information costs. If an IBR Swap market arose, law schools, consumer advocates, and institutional counterparties would want to determine the best regulatory body to develop disclosure rules. In some markets, federal bureaus regulate disclosure,[177] in others state or local laws do so,[178] and in still others this responsibility is met by private industry-specific organizations.[179]

---

174. Note that the current version of the Bill (1) prevents investors from controlling investees' actions, (2) requires that income below $18,000 does not trigger repayment, (3) provides a ceiling on the percentage of income that can be charged, (4) provides an aggregate limit on the percentage of income any one student can commit, (5) provides a sliding limit on how long a repayment period can be required, and (6) requires that ISAs include a series of disclosures. *See* Investing in Student Success Act of 2015, H.R. 3432, 114th Cong. §§ 103(a)(1), (3)–(5), 103(b), 103(c) (1st sess. 2015).

175. *See, e.g.*, ASHER ET AL., *supra* note 71 (arguing that income-driven repayment programs may take pressure off governments or institutions to make education more affordable).

176. For example, as discussed *supra* note 152, student loans are subject to disclosure requirements under the Truth in Lending Act.

177. For example, the Consumer Financial Protection Bureau implements numerous disclosure requirements pertaining to home mortgages and student loans, pursuant to Title XIV of Dodd-Frank. *See* Dodd-Frank Act, 15 U.S.C. § 5511 (2012); Truth in Lending Act, 15 U.S.C. § 1638; *Regulatory Implementation*, CONSUMER FIN. PROTECTION BUREAU, http://www.consumerfinance.gov/regulatory-implementation [https://perma.cc/9X7Q-XY2Y] (last visited Feb. 28, 2016).

178. For example, the market for rental housing is frequently regulated by local authorities that require disclosure of terms and tenants' rights. *See, e.g.*, *D.C. Municipal Regulations and D.C. Register*, SECRETARY D.C., OFF. DOCUMENTS & ADMIN. ISSUANCES, *available at* http://www.dcregs.dc.gov/Gateway/ChapterHome.aspx?ChapterNumber=14-3 [https://perma.cc/Q2CT-SFS5] (last visited Feb. 28, 2016).

179. Real estate brokers' associations are one example of this. *See, e.g.*, Greater Boston Real Estate Bd., Standard Form Purchase & Sale Agreement (Form

A second obvious concern that regulation should address is the risk to student counterparties of institutional counterparty default. Student counterparties may make career decisions by relying on an IBR Swap transaction. If an institutional counterparty becomes insolvent or otherwise defaults, a student counterparty could be left without capacity to meet student loan obligations. Regulations designed to protect certain types of parties to particular transactions from counterparty default risk are common. In derivatives markets generally, regulations requiring that transactions take place on an exchange or through a clearinghouse address counterparty credit risk concerns.[180] The exchange or clearinghouse meets swap obligations, absorbing risk of counterparty default. In insurance markets, capital adequacy requirements protect policyholders from underwriter insolvency.[181] In many markets, assignments of collateral secure counterparty performance, enhancing credit.[182] In the IBR Swap context, again, interested parties and policy-makers would need to determine the appropriate regulatory body and the nature of credit enhancement requirements. IBR Swaps could be traded on a private clearinghouse established by an association of law schools and institutional counterparties.[183] Or, state or federal regulators could impose capital adequacy requirements on institutional counterparties. Formulating and enforcing regulations to protect students from institutional counterparty credit risk, like disclosure requirements, questions of political will. The requirements themselves are not exceedingly difficult to design or enforce.

The remainder of this Section discusses in greater detail the issue of "differential" or "discriminatory" pricing, one of the aspects of income-share agreements that is both a feature and a bug. While regulating to protect student counterparties from obtuse or disadvantageous terms and from institutional counterparty credit risk is fairly straightforward (and faces hurdles primarily of political will), differential pricing and adverse selection present more complex regulatory challenges. Here, we briefly examine the regulatory tradeoffs surrounding differential pricing that could constrain potential harms of IBR Swaps.

ID: RA700) (Mar. 2006), *available at* http://thehomebuyingmentors.org/files/2015/01/GREB-Standard-PS-Agreement.pdf [https://perma.cc/CS9S-8CAK].

180. *See* Dodd-Frank Act § 723; CEA, 7 U.S.C. §§ 2(h)1, 2(h)7 (2012).

181. *See* Henderson, *supra* note 53; Hunt, *supra* note 132.

182. *See generally Collateral*, INT'L SWAPS & DERIVATIVES ASS'N, http://www2.isda.org/functional-areas/infrastructure-management/collateral [https://perma.cc/2WMC-NNV2] (last visited Feb. 10, 2016) ("contain[ing] [c]ollateral information and certain informational documents for the Standard Credit Support Annex (SCSA®), which seeks to standardize market practice regarding embedded optionality in current CSAs . . . and align the mechanics and economics of collateralization between the bilateral and cleared OTC derivative markets").

183. This option may not be feasible; for example, a clearinghouse structure may involve margin requirements that would be unworkable for student counterparties. *See supra* note 131 and accompanying text.

A. *Differential or Discriminatory Pricing*

The primary variable in imagining a market for income-share agreements is the range of factors involved in pricing the agreements. The student-loan market is divided into two submarkets that operate very differently with respect to differential pricing. The bulk of loans for higher education are government loans, and these loans have interest rates fixed by statute that are the same for all qualified borrowers. But there is also a significant market for private loans for higher education, and these loans carry rates that are priced differently depending on a variety of factors, including the creditworthiness of the borrower or guarantor, the loan default rate of the school the borrower is attending, and other factors. In the private student loan context, students are protected against discrimination on the basis of race and other factors by the Equal Credit Opportunity Act (ECOA).[184] But, as discussed above, unless the law is changed, IBR Swaps would not be considered student loans and so would not be subject to ECOA. Furthermore, because IBR Swaps could be used in conjunction with government loans, a vibrant IBR Swap market could introduce differential pricing into the existing market for government loans as well as avoid the regulation of discriminatory lending that currently applies to the private student loan market.

There is nothing inherent in IBR Swaps that requires them to be differentially priced. A governmentally controlled IBR Swap program could offer some of the benefits of ISAs without any differential pricing. But all commentators assume that private income-share agreements would not be available to all students on the same terms the way the government's income-based payment plans are.[185] Rather, a market would develop in which income-share agreements are differentially priced based on some criteria that investors believe predict high-earning graduates. As discussed above, most of the commentators thus far have assumed that income-share agreements for undergraduate education would be differentially priced based on undergraduate major, with "high earning" majors like engineering commanding lower priced income-share agreements, while "low earning" majors like English would have to pay higher prices for their income-share agreement. In other words, an English major may be able to borrow $10,000 in exchange for 2% of their income, while a petroleum engineering major may be able to borrow the same $10,000 in exchange for only 1% of their income. It is beyond the scope of this Article to discuss the

184. 15 U.S.C. § 1691(a) (2012) ("It shall be unlawful for any creditor to discriminate against any applicant, with respect to any aspect of a credit transaction—(1) on the basis of race, color, religion, national origin, sex or marital status, or age (provided the applicant has the capacity to contract); (2) because all or part of the applicant's income derives from any public assistance program . . . ."). ECOA is implemented by Regulation B (12 C.F.R. pt. 202) and enforced by the Consumer Financial Protection Bureau, which also has the authority to promulgate implementing regulations for it. *See id.* § 1691b.

185. *See supra* note 24.

merits of this particular assumption, but we are both skeptical of the evidence that supports it and unenthusiastic about the effects a market would have if college major was the primary driver of differential pricing.[186]

But there is no reason that course of study would necessarily be the dominant factor that would determine the price of an income-share agreement. Choice of school could be a significant factor in differential pricing, especially in the law school context.[187] If investors determined that graduates of highly-ranked law schools on average made more money than graduates of low-ranked law schools, then they could charge more to students of low-ranked law schools for their income-share agreements.[188] If they estimated that Ben, a student of a low-ranked law school was likely to earn half as much as Heather, a student of a high-ranked law school, they might charge Ben 2% of his future income for every $10,000 borrowed, while they charge Heather only 1% for the same $10,000. Remember, that does not mean that Ben will pay more than Heather for his education. If the investors are predicting correctly, and if Ben and Heather earn the average amount for their schools, they will each pay the same amount for their education. It will just be a higher percentage of income for Ben than for Heather.[189] The same is true of debt—if they earn the average amount for their schools, then Ben pays a higher percentage of his income than Heather, even though they pay the same amount of money. But, obviously, the investors might not predict accurately or Heather and Ben might not be average. If Ben is above average *for his school* then he will pay more for his education than someone who is average *for their school*. If he is below average *for his school* then he will pay less. If Ben and Heather earn the same amount as each other, then Ben, who is dramatically outperforming the other students from his school, will pay much more than Heather, who is earning the average for her school. That may seem unfair. It may even contribute to social inequality, since students from low-income backgrounds may be more likely to attend lower ranked schools.

186. Avoiding a robust discussion of exactly this issue was one of the reasons we chose to focus on law school financing in this Article. For a critique of differential pricing in the undergraduate context, see Jonathan D. Glater, *The Unsupportable Cost of Variable Pricing of Student Loans*, 70 WASH. & LEE L. REV. 2137 (2013).

187. In the private student loan market, a version of this school-based differential pricing operates to raise the price of loans for students attending certain schools that have high aggregate default rates on student loans.

188. The rankings mentioned in this hypothetical are the graduate school rankings published annually by the U.S. News and World Report. It is widely believed that graduates of higher ranked law schools earn more on average than graduates of lower ranked law schools.

189. Remember, that is how it currently works with fixed-rate debt. If Ben earns half Heather's salary, he pays twice as much for his education as she does, *as a percentage of his income*. *See supra* note 82. Of course, Ben and Heather are paying the same amount in absolute terms if they both borrow the same amount of fixed-rate debt, unless one or the other defaults or takes advantage of a governmental repayment program.

But the possibilities of differential pricing do not end there. It is likely that school will not be the only meaningful pricing factor for an investor in income-share agreements. Absent any rule or law to the contrary, an investor could use more personal factors, such as undergraduate GPA or an LSAT score. These "scores-based" factors would have a very different impact than schools-based factors. Even worse than scores-based differential pricing is differential pricing based on other personal factors. For example, we know that women on average earn less than men, and so, absent any law preventing it, one could imagine a world in which Heather and Ben have identical LSATs and GPAs and go to the same school, but Heather's income-share agreement costs more than Ben's just because she is a woman.[190] Similarly, we know that children of high-earning parents are more likely to be high-earners themselves than children of low-earning parents, even when other factors are controlled. One could imagine a world in which income-share agreements are priced largely based on the earnings history of the student's parents. One could imagine a world in which race is a factor.[191] Again, these pricing strategies would mean that students who ended up earning the same amount would pay different amounts based on sex or parental income or race. And the difference

190. As discussed above, *supra* note 184, ECOA prevents discrimination on the basis of sex in the private student loan context, but without legislative or regulatory change, ECOA would not apply to income-share agreements. It would be relatively easy to prevent discrimination on the basis of sex or race just by including income-share agreements in the definition of student loans for the purposes of ECOA. While there is room for debate on this subject, the Rubio–Petri Legislation arguably accomplishes this by providing that an income-share agreement that meets the requirements of the bill is defined as "qualified education loan" under Section 221(d) of the Internal Revenue Code. *See* H.R. 3432, 114th Cong. § 301(a) (1st sess. 2015). Once the instrument is described as a "qualified education loan, it is presumably "credit" which is subject to ECOA. *See* 15 U.S.C. § 1691a(d) (2012) ("The term 'credit' means the right granted by a creditor to a debtor . . . to incur debts and defer its payment . . . ."); 12 C.F.R. § 202.2(j) (2016) ("Credit means the right granted by a creditor to an applicant to . . . incur debt and defer its payment . . . ."). With respect to the IBR Swap, which does not constitute debt absent some law or regulation to the contrary, the Rubio–Petri Bill could change the debt analysis for any IBR Swap that qualified as an ISA under the Bill.

191. Race, like sex, is covered under ECOA. *See* 15 U.S.C. § 169a(d). It is worth mentioning in this context, that discrimination can be found under ECOA under either a theory of disparate treatment or disparate impact. *See* CONSUMER FIN. PROT. BUREAU, CONSUMER LAWS AND REGULATIONS: EQUAL CREDIT OPPORTUNITY ACT 1 (2013), *available at* http://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf [https://perma.cc/J54E-NDNH] ("Disparate impact occurs when a creditor employs facially neutral policies or practices that have an adverse effect or impact on a member of a protected class unless it meets a legitimate business need that cannot reasonably be achieved by means that are less disparate in their impact."); *see also* Rodriguez v. SLM Corp., No. 07cv1866 (WWE), 2009 WL 598252, at *3 (D. Conn. Mar. 6, 2009) ("In light of the early stage of this action and the recent decisions that [*Smith v. City of Jackson*, 544 U.S. 228 (2005)] does not preclude ECOA disparate impact claims as recognized in pre-*Smith* precedent, the Court will deny the motion to dismiss on this ground.").

would undermine egalitarianism in access to education: the children of high earners would pay *less* than the children of low earners.[192]

B. *Regulating Differential Pricing*

Congress could create a regulatory regime that would control which factors could be used in pricing income-share agreements. Lawmakers could mandate that the price of an income-share agreement not vary based on sex or race or parental income.[193] They could mandate that the price be based solely on school attended or even course of study, with students of all law schools paying the same amount.

The problem with such regulation is that it represents a tradeoff in two ways. First, the information-gathering potential of income-share agreements depends on differential pricing. When the investors set a different price for students with different factors, it communicates to the student the market's estimation of the relative importance of those factors to the students' future earnings. If the government prevents differential pricing based on LSAT or undergraduate GPA, for example, it prevents students from learning anything about their own earning potential based on their LSATs or undergraduate GPAs. After all, either thing may be irrelevant or highly relevant. The dream of many income-share agreement supporters is that permitting people to make money from accurately predicting students' earnings would fuel investment into exactly these questions, and the price would communicate the findings.

The second tradeoff in regulating differential pricing is so-called regulatory adverse selection. *Adverse selection* is the term used to describe a central problem in insurance markets. Traditional adverse selection arises when insurers and insureds have "asymmetric information" about the risks posed to the insured. For example, an insurer providing life insurance may know that the overall probability of a forty-six-year-old man dying in the next year is 2%. If the only information he had about the insured was his age and sex, then he would price the one year of life insurance based on that 2% chance of the insured dying in the coming year. But the chance of a forty-six-year-old man who is diagnosed with advanced lung cancer dying in the next year may be as high as 50%. If the insured knows that he has cancer, but the insurer does not, then there is an informational asymmetry. If the insurer prices the insurance based on its knowledge of the general population of forty-six-year-old men, then the insurance will be a bargain for the man with cancer.

Assuming that the insurer cannot discover which people are diagnosed with cancer, it will have to price its insurance slightly higher than it

192. Remember, privileged students would pay less *as a percentage of their income*, but they would pay the same amount if both parties earned the amount they were projected to earn.

193. The easiest way to protect against discrimination in ISA pricing would be to include ISAs in the instruments covered by ECOA. *See supra* note 190.

otherwise would, because it knows that people with cancer are more likely to purchase it than people without cancer. As the insurer increases the price of the insurance to take into account those people who know they have cancer, the insurance becomes less attractive to people who do not have cancer, and they purchase less, driving the price even higher. That tendency for asymmetric information to encourage high-risk individuals to acquire more insurance, and thereby drive low-risk individuals out of the market, is called "adverse selection." In extreme cases, adverse selection can cause a "death spiral" as low-risk insureds opt out of the market, driving up the price, thereby causing more low-risk insureds to exit the market.

In education financing, adverse selection potentially exists whether financing is provided in the form of debt or an income-share agreement. In the case of income-share agreements, any information that accurately predicts earnings—and that is known to students but not investors—is relevant to the adverse selection issue. For example, if a student knows that they want to work in public interest law and that salaries are very low there, then they present a greater-than-normal risk of low earnings. Similarly, if a student has a personal contact at a high-paying law firm, they potentially present a greater-than-normal chance of high earnings. If the low-earning student opts in to the human capital contract market, and the high-earning student opts out, the adverse selection problem may be acute. That is because personal commitments and personal contacts are private information held by the student and not available to the investors.

But adverse selection problems arise not only through asymmetric information, but also when both parties have access to information. However, the insurer is prohibited by law to take that information into account in setting rates. This kind of adverse selection is sometimes called *regulatory adverse selection* to emphasize the fact that it arises out of the legal regime, rather than out of asymmetrical information.[194] So, for example, the fact that federal law prohibits insurers from considering preexisting conditions when providing or pricing health insurance creates the potential for regulatory adverse selection.[195] In the case of income-share agreements, any regulatory regime that limits the factors that investors can consider in pricing the contracts could create or exacerbate the adverse selection problem. Even banning only the most disturbing forms of price discrimination could potentially cause regulatory adverse selection. For example, if the government banned sex discrimination in pricing income-share agreements, women (who are at greater risk of low earnings) might opt in to the program, while men (who have a greater likelihood of high earnings) opt out. In this scenario, the average price would go up.

194. Ronen Avraham, Kyle D. Logue & Daniel Schwarcz, *Understanding Insurance Antidiscrimination Laws*, 87 S. CAL. L. REV. 195, 204 (2014).

195. The so-called individual mandate, which requires almost all persons to purchase health insurance, is an attempt to overcome at least some of the potential adverse selection caused by the prohibition.

In the income-share agreement market, regulatory adverse selection could also occur if investors were constrained from taking into account any other individual factor, like family wealth, grades, LSAT scores, or anything else. It is conceivable that adverse selection could be prevalent enough in a market for human capital contracts to cause a "death spiral." In that case, no market could develop for income-share agreements, and it would be necessary for the government to intervene, either with some sort of "individual mandate" or by subsidizing income-share agreements enough to counter the costs of adverse selection.

Ultimately, the question of whether asymmetrical information or regulatory adverse selection would destroy a voluntary market for income-share agreements is an empirical one. There are some reasons to believe that the adverse selection problem may not be as bad as one might fear, at least if students are required to make their choice about whether to participate before they start their law school career. As for information asymmetries, the consensus appears to be that at the point students enter school, they have very little reliable information about their future earning capabilities at graduation, at least in the law school context. Remember, for adverse selection to be a concern, students would have to know more about their individual earning potential than the investors.[196] Most evidence suggests that at least prospective law students have very little private information about themselves that would enable them to make better predictions about their earning potential than investors.[197] Even the classic example of the student who plans to go into a low-earning field like public interest appears to be largely a myth. Most students do a bad job at the outset of law school predicting what kind of law they will practice, and so-called public interest careers are not the main cause of low earning among law school graduates.

Regulatory adverse selection may also pose less of a problem than one might assume if individual characteristics are not very predictive of future earnings. Transaction costs already limit the individual factors that insurers use to price insurance, and only those whose predictive ability justifies the price of collecting the information are likely to be used. It may well be that the factors that are prohibited—even if they include scores or parental income—are just not significant enough to cause a death spiral in the market. After all, we had a robust health insurance market prior to the Affordable Care Act with a significant number of antidiscrimination provi-

196. *See* Macchiarola & Abraham, *supra* note 24, at 110–13.

197. *See* PALACIOS ET AL., INVESTING IN VALUE, SHARING RISK, *supra* note 24, at 11 (giving good general introduction to adverse selection problem faced by human capital contracts).

> Given that students are often not very accurate in projecting their future income and typically have low levels of knowledge about the labor market, however, investors will most likely have better information than students about their future economic prospects in particular courses of study at particular institutions.

Id.

sions. It presumably was the ban on pricing based on preexisting conditions that legislators thought would be the straw that broke the camel's back to necessitate the individual mandate.

In regulating income-share agreements, legislators would have to make a similar judgment about which pricing factors could be prohibited without destroying the market and which would be too much. If skepticism about the predictive power of various characteristics is warranted, then regulatory arbitrage may not be such a significant problem. On the other hand, if no factor does a good job of predicting earnings, then the "information-gathering" power of income-share agreements will probably be less powerful than some commentators hope. These two issues—regulatory adverse selection and differential pricing—are inherently connected.

In summary, the point here is that the creation of a market for income-share agreements might necessitate new regulation, and this new regulation would implicate a series of policy trade-offs. We leave a systematic examination of these policy tradeoffs to another day.

VI. Conclusion

This Article has introduced a novel financial instrument that has both strengths and weaknesses. One of the benefits of a financial innovation like the IBR Swap is how flexible it is to adapt to a variety of circumstances. One example should illustrate how thinking about the IBR Swap may help policy-makers to see higher education finance in new ways, making creative policy reform possible.

In July 2013, the Oregon legislature approved a pilot program under which students can attend state universities for free in exchange for a promise to pay a percentage of their future income.[198] The program, called "Pay It Forward," has been stalled for over two years largely because legislators are unsure of how to fund it. Because Oregon state schools would have to forego tuition revenue for any student participating in the program, the cost would have to be made up elsewhere. Legislators have proposed issuing bonds, directing revenue from the state's lottery, and other sources of revenue to fund the program. Pay It Forward is in essence a human capital contract between the state of Oregon and some of the students at its state schools.

If it wanted to, Oregon could solve its revenue problem by structuring the Pay It Forward program as a series of IBR Swaps. Instead of attending Oregon state schools for free, students would borrow from the federal government the cost of tuition and pay that tuition to the school they attend. Then, Oregon could agree that it would pay the students' loan payments on their behalf, so long as the students pay Oregon a percentage of their future incomes. Oregon has estimated that four years of public college

198. *See* H.B. 3472, 77th Leg. Assemb. Reg. Sess. (Or. 2013), *available at* http://gov.oregonlive.com/bill/2013/HB3472/ [https://perma.cc/7Q5Y-2DR5].

would cost 3–5% of a student's income for twenty years. Therefore, if students borrowed enough to pay for four years, Oregon would enter into IBR Swaps with them in which it paid off their student loans, and they paid Oregon 3–5% of their incomes for twenty years. From the student's perspective, a Pay It Forward program structured around an IBR Swap is almost identical to the currently proposed Pay It Forward program. From Oregon's perspective, the primary benefit of an IBR Swap model is that it does not need to come up with upfront capital to fund its Pay It Forward Program.

The ways that the IBR Swap opens possibilities for Oregon and others is just the beginning of the conversation, however. It is crucial that we bring both creativity and detailed attention to the possibilities that income-based education finance presents. The IBR Swap combines structural and financial advantages of derivatives with the appeal of income-based approaches to paying for law school. But the benefits of the IBR Swap from a financial engineering perspective only make starker the potential social problems that could arise from an unregulated market for income-share agreements. The IBR Swap concept should inspire reexamination of the best role for government in higher education finance and the importance of channeling public funds to their best uses. Now is the time to explore innovations like the IBR Swap, as legal education strives to better match costs and capital to educate the next generation of lawyers.

HEALTHCARE AND THE BALANCE-BILLING PROBLEM: THE SOLUTION IS THE COMMON LAW OF CONTRACTS AND STRENGTHENING THE FREE MARKET FOR HEALTHCARE

GEORGE A. NATION III*

INTRODUCTION

COURTS across the country are beginning to understand that hospital bills based on list or chargemaster prices are exorbitant and unfair, because they reflect prices that are set to be discounted and not paid.[1] As

* Professor of Law and Business, Lehigh University.

1. *See, e.g.*, Howell v. Hamilton Meats & Provisions, Inc., 257 P.3d 1130, 1137 (Cal. 2011) (citing Melone v. Sierra Ry., Co, 91 P. 522, 523 (Cal. 1907)) (explaining medical expenses actually paid by or on behalf of patient, not billed charges based on chargemaster rates, is proper measure of medical expenses); Greenfield v. Manor Care, Inc., 705 So. 2d 926, 930–31 (Fla. Dist. Ct. App. 1997) (holding healthcare agreement to have implied covenant to charge reasonable fee), *receded from by* Beverly Enters.-Fla., Inc. v. Knowles, 766 So. 2d 335 (Fla. Dist. Ct. App. 2000); Payne v. Humana Hosp. Orange Park, 661 So. 2d 1239, 1241 (Fla. Dist. Ct. App. 1995) ("A patient may not be bound by unreasonable charges in an agreement to pay charges in accordance with 'standard and current rates.'"); Victory Mem'l Hosp. v. Rice, 493 N.E.2d 117, 119 (Ill. App. Ct. 1986) (permitting jury question as to whether charges presented by hospital were reasonable); Butler v. Ind. Dep't of Ins., 904 N.E.2d 198, 202 (Ind. 2009) (explaining medical expenses actually paid by or on behalf of patient, not billed charges based on chargemaster rates, is proper measure of medical expenses); *In re* Adoption of N.J.A.C., 979 A.2d 770, 785 (N.J. Super. Ct. App. Div. 2009) (presuming regulation of physician fee schedule was reasonable and valid); Kastick v. U-Haul Co., 292 A.D.2d 797, 798–99 (N.Y. App. Div. 2002) (agreeing with defendants that plaintiff could not recover damages in amount she was never obligated to pay for medical services); Nassau Anesthesia Assocs. P.C. v. Chin, 924 N.Y.S.2d 252, 255 (N.Y. Dist. Ct. 2011) (agreeing reasonable value of medical services is average amount that provider would have accepted as full payment from third-party payers such as private insurers and federal healthcare programs (citing Temple Univ. Hosp., Inc. v. Healthcare Mgmt. Alternatives, Inc., 832 A.2d 501, 510 (Pa. Super. Ct. 2003))); Moorhead v. Crozer Chester Med. Ctr., 765 A.2d 786, 789 (Pa. 2001) (holding patient's recovery for medical expenses in malpractice suit was limited to amount paid and accepted for services, rather than fair and reasonable market value of services), *abrogated by* Northbrook Life Ins. Co. v. Commonwealth, 949 A.2d 333, 337 (Pa. 2008); Hosp. Inc. v. Healthcare Mgmt. Alts., Inc., 832 A.2d 501, 510 (Pa. Super. Ct. 2003) (chargemaster rate "bears no relationship to the amount typically paid for those services"); Doe v. HCA Health Servs., Inc., 46 S.W.3d 191, 198–99 (Tenn. 2001) (affirming lower court holding that patient was obligated to pay reasonable charges for medical services and fair value of goods furnished); Haygood v. De Escabedo, 356 S.W.3d 390, 397 (Tex. 2011) (explaining medical expenses actually paid by or on behalf of patient, not billed charges based on chargemaster rates, is proper measure of medical expenses); Daughters of Charity Health Servs. v. Linnstaedter, 226 S.W.3d 409, 411 (Tex. 2007) (finding chargemaster prices are *not* value of medical services in context of hospital lien statute); *cf.* Holland v. Trinity Health Care Corp., 791 N.W.2d 724, 728 (Mich. Ct. App. 2010) (finding chargemaster rates are usual and customary).

a result, courts are becoming aware of the fact that the list prices or chargemaster rates that hospitals claim are usual and customary are instead exorbitant amounts, arbitrarily set by hospitals, as a starting point for negotiating huge discounts with insurers.[2] Much attention has been focused by commentators on the ironic unfairness of the fact that uninsured patients—that is, those least likely to be able to afford to pay for healthcare even at a reasonable price—are often expected to pay these exorbitant rates in full.[3] However, while this attention is certainly well-deserved, focusing only on the uninsured misses the fact that it is not just the uninsured that are burdened by obscenely high chargemaster rates.

A large and growing group of insured patients is also being unfairly burdened by hospitals' exorbitant chargemaster prices.[4] The burden is brought to bear on these patients through a process known as balance billing.[5] When a patient's insurance company has not negotiated a con-

2. *See, e.g.*, Stanley v. Walker, 906 N.E. 2d 852, 857 (Ind. 2009) ("Thus, based on the realities of health care finance, we are unconvinced that the reasonable value of medical services is necessarily represented by either the amount actually paid or the amount stated in the original medical bill."); Cape Reg'l Med. Ctr. v. Sanchez, No. CPM DC 109-11, at *9 (N.J. Super. Ct. Law Div. Mar. 26, 2012) (noting that most patients, upon entering hospital, sign "Authorization for Treatment," "Statement of Financial Responsibility," or another similarly open-ended agreement pursuant to which patient purports to agree to pay for all medical goods and services provided by hospital at hospital's list (chargemaster) prices; in reality, however, this type of agreement amounts to blank check given by patients to hospitals with amounts to be unilaterally filled in by hospitals later). However, not all courts have yet come to this realization, especially in balance-billing context. *See, e.g.*, Allen v. Clarian Health Partners, Inc., 980 N.E.2d 306, 310–11 (Ind. 2012) (rejecting central premise of this Article that hospitals' chargemaster rates should *not* be used as basis for pricing on contracts for healthcare services).

3. *See, e.g.*, Gerard F. Anderson, *From 'Soak the Rich' to 'Soak the Poor': Recent Trends in Hospital Pricing*, 26 HEALTH AFF. 780 (2007) (stating hospitals often bill uninsured patients full list charges); George A. Nation III, *Obscene Contracts: The Doctrine of Unconscionability and Hospital Billing of the Uninsured*, 94 KY. L.J. 101, 101–05 (2005) [hereinafter Nation, *Obscene Contracts*]; Uwe E. Reinhardt, *The Pricing of U.S. Hospital Services: Chaos Behind a Veil of Secrecy*, 25 HEALTH AFF. 57, 58 (2006) [hereinafter Reinhardt, *U.S. Hospital Services*] (explaining that self-payers usually forced to accept full charges set by hospital); Christopher P. Tompkins, Stuart H. Altman & Efrat Eilat, *The Precarious Pricing System for Hospital Services*, 25 HEALTH AFF. 45, 52 (2006) (same); Lucette Lagnado, *Anatomy of a Hospital Bill*, WALL ST. J., Sept. 21, 2004, http://www.wsj.com/articles/SB109571706550822844 [https://perma.cc/9Q6V-FTZS]; *cf.* Glenn A. Melnick & Katya Fonkych, *Hospital Pricing and the Uninsured: Do the Uninsured Pay Higher Prices?*, 27 HEALTH AFF. 116 (2008) (asserting that uninsured pay more than some but less than others).

4. *See generally* Melnick & Fonkych, *supra* note 3 (observing that insured individuals suffer from chargemaster rates).

5. *See, e.g.*, Uwe E. Reinhardt, *The Many Different Prices Paid to Providers and the Flawed Theory of Cost Shifting: Is It Time for a More Rational All-Payer System?*, 30 HEALTH AFF. 2125 (2011) (discussing high prices faced by insured patients who receive care outside of their insurers network); Steven Brill, *Bitter Pill: Why Medical Bills Are Killing Us*, TIME, Mar. 4, 2013, *available at* http://www.claims.org/pdf/time_article.pdf [https://perma.cc/DE7X-HWD7] (discussing outrageous bills imposed on *insured* patients who receive care from providers who are outside of patient's insurance network); Caroline Chen, *Surprise Medical Bills Lead to Protection*

tract with the hospital that provides services to the patient, the patient is considered out of network—"OON," in hospital speak—and as a result, the discounts that the hospital has negotiated with other insurers do not apply to the OON patient.[6] The patient's insurer pays the hospital the amount that the insurance company is obligated to pay for the services received, but this amount, being reasonable, is always far less than the unreasonably high list price set by the hospital.[7] Because the OON patient's insurer has no contract with the hospital, the hospital is not obligated to accept the payment from the insurance company as full payment, and therefore the hospital is permitted to bill the patient for the balance that is the difference between the obscenely high hospital list price and the reasonable amount that the insurance company paid. Moreover, for a variety of reasons, hospital networks are becoming narrower as hospital systems contract with fewer insurers, and as a result, more and more patients are receiving balance bills.[8]

In addition, not only do the price and collection limitations included in the Affordable Care Act (ACA or Act)—Obamacare—not prevent balance billing, the Act allows for the sale of narrow network health insurance, enshrines exorbitantly high chargemaster rates, and encourages balance billing.[9] Finally, the practice of balance billing puts upward pressure on healthcare prices in general.[10] That is, this practice leads to

---

Laws: Health, BLOOMBERG (Apr. 4, 2014), http://www.bloomberg.com/news/articles/2014-04-04/surprise-medical-bills-lead-to-protection-laws-health [https://perma.cc/QTG5-5ZJU] ("It's important to protect the consumer now [since the passage of the Affordable Care Act], because there's a little more chaos in the system and a lot more people, [and] balance-billing is the most-common payment problem seen at her nonprofit advocacy organization, which handled 65,000 healthcare cases last year." (quoting Elisabeth Benjamin, Vice President of Health Initiatives, Cmty. Serv. Soc'y of N.Y.) (internal quotation marks omitted)). *See generally* JACK HOADLEY, KEVIN LUCIA & SONYA SCHWARTZ, CAL. HEALTHCARE FOUND., UNEXPECTED CHARGES: WHAT STATES ARE DOING ABOUT BALANCE BILLING (2009), *available at*, http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20U/PDF%20UnexpectedChargesStatesAndBalanceBilling.pdf [https://perma.cc/5Z7G-KG55].

6. *See* HOADLEY ET AL., *supra* note 5, at 3–5.

7. *See, e.g.*, George A. Nation III, *Determining the Fair and Reasonable Value of Medical Services: The Affordable Care Act, Government Insurers, Private Insurers and Uninsured Patients*, 65 BAYLOR L. REV. 425, 434–35 (2013) [hereinafter Nation, *Determining Fair and Reasonable Value*] (providing actual hospital pricing for hypothetical gall bladder surgery that would have list price of $14,000; $5,600 price for HMOs; price of $4,700 for Blue Cross/Blue Shield; price of $5,000 for Aetna; price of $2,590 for Medicare; and $1,260 for Medicaid—for same exact services).

8. *See, e.g.*, Reed Abelson, *More Insured, but the Choices Are Narrowing*, N.Y. TIMES, May 12, 2014, http://www.nytimes.com/2014/05/13/business/more-insured-but-the-choices-are-narrowing.html?_r=0 [https://perma.cc/F4BS-SH8A] (noting that because many health insurance policies sold on Affordable Care Act (ACA) exchanges use narrow networks, out-of-network care and balance-billing are becoming more frequent).

9. *See infra* notes 24–83 and accompanying text.

10. *See* Ge Bai & Gerard F. Anderson, *Extreme Markup: The Fifty US Hospitals with the Highest Charge-to-Cost Ratios*, 34 HEALTH AFF. 922, 925 (2015) (noting that

higher prices across the board for the uninsured, the out-of-network insured, and even the in-network insured.[11]

I have written about the balance-billing problem in other work and have suggested there the adoption of government regulations directed at hospitals that would both address the balance-billing problem and improve the functioning of the free market for healthcare by providing price transparency.[12] A few states have attempted to address this problem legislatively with mixed success.[13] The focus of this Article is on the practice of balance billing, what courts can do now under existing law to address this problem, and the type of legislation that will provide a long-term solution to the broader problem of market failure regarding the sale of healthcare. This Article argues that more government price-fixing is not the solution. The primary goal of healthcare policy in the United States should not be to increase access and control the price of healthcare; rather, it should be to develop the best healthcare in the world at the lowest price.[14] The only

---

high markups [chargemaster rates] may add to private insurance premiums and play role in rise of overall healthcare spending); Robert Murray, *Hospital Charges and the Need for a Maximum Price Obligation Rule for Emergency Department & Out-of-Network Care*, HEALTH AFF. BLOG (May 16, 2013), http://healthaffairs.org/blog/2013/05/16/hospital-charges-and-the-need-for-a-maximum-price-obligation-rule-for-emergency-department-out-of-network-care/ [https://perma.cc/JWC4-XDLE]. Murray states as follows:

> Charges Do Matter—They Matter a Great Deal.
>
> Counter to the belief of both hospital industry representatives and many of my colleagues, hospital charge levels and rapidly escalating charges matter a great deal. While individual states and the Affordable Care Act (ACA) have instituted limits on the amounts low-income uninsured patients pay hospitals, insured patients that receive care at hospitals that are "Non-Par" or "out-of-network" are still victims of hospital's exorbitant charging practices. When patients receive emergency services at an out-of-network hospital, the patient and/or insurance company (depending on insurer cost sharing for out-of-network care) pay full charges.
>
> High and increasing hospital charges, combined with increasing proportions of cases admitted through the hospital Emergency Department (ED), are major factors behind the ever-declining negotiating leverage of private health insurers. This situation, coupled with the increased pricing power of the ever-more-concentrated provider industry, will be a major contributor to the almost certain rapid escalation in total U.S. health care costs in coming years.

Id.

11. *See supra* note 10 and accompanying text.

12. *See generally*, George A. Nation III, *Hospital Chargemaster Insanity: Heeling the Healers*, 43 PEPP. L. REV. (forthcoming 2016) [hereinafter Nation, *Chargemaster Insanity*]; Nation, *Determining Fair and Reasonable Value*, *supra* note 7; Nation, *Obscene Contracts*, *supra* note 3.

13. *See, e.g.*, HOADLEY ET AL., *supra* note 5, at 6–9. (noting that nine states have enacted legislation in attempt to deal with balance billing and problem of these efforts have met with mixed results).

14. *See* George A. Nation III, *Non-Profit Charitable Tax-Exempt Hospitals—Wolves in Sheep's Clothing: To Increase Fairness and Enhance Competition in Health Care All Hospitals Should Be For-Profit and Taxable*, 42 RUTGERS L.J. 141, 198–209 (2010) [hereinafter Nation, *Wolves in Sheep's Clothing*] (citation omitted within title) (argu-

mechanism we have for doing that is the free market.[15] While there is additional government regulation that would be helpful in strengthening the free market for healthcare,[16] I argue here that we do not have to wait for such regulation; the common law of contracts contains the tools necessary to allow courts to provide a free market solution to the balance-billing problem now.[17]

Part I provides background concerning the problem of balance billing. Part II provides an analysis of the problem and suggests that part of the solution to the lack of competition in the sale of healthcare (of which the balance-billing problem is a symptom) requires, *inter alia,* government regulation designed to require meaningful price disclosure by hospitals, along with other steps to further strengthen the free market for healthcare in the United States. Part III discusses what courts can do now with the tools currently available to them to solve the problem of balance billing in a way consistent with a strong free-market for healthcare. Part IV concludes.

I. BACKGROUND: THE PROBLEM

In a recent newspaper article a journalist recounted the story of a fifty-year-old construction worker who experienced chest pains and was admitted to St. Francis Hospital in Bartlett, Tennessee, in 2014.[18] His wife said they were not told either at the time of admission or during their visit that the hospital did not accept their health insurance.[19] The couple received a bill from the hospital for $22,945.[20] As the article points out, under the ACA, a family's out-of-pocket expenses (this would include charges for things like co-pays, co-insurance, etc.) for 2014 were capped at $12,700.[21] However, this limitation under the ACA does not apply to non-emergency room charges provided by an out-of-network hospital.[22] That is, the limitation does not protect patients from balance billing, and thus the family received a bill for $22,945. In this case, the family was lucky; they appealed the charges, and the hospital eventually reduced their bill

---

ing that definition of word *better*—as in better healthcare—is key to policy formulation).

15. *See id.* at 181–85 (arguing that non-profit, tax-exempt business model is not appropriate for hospitals and that only free market competition can accomplish healthcare goals).

16. *See infra* notes 98–125 and accompanying text.

17. *See infra* notes 84–148 and accompanying text.

18. *See* Stephanie Armour, *Surprise Bills for Many Under Health Law*, WALL ST. J., June 11, 2015, http://www.wsj.com/articles/surprise-bills-for-many-under-health-law-1434042543 [https://perma.cc/Y6EP-2N9F].

19. *Id.*

20. *Id.*

21. *Id.*

22. *Id.*

to $600—but only after the bill had been sent to a collection agency, which the family worries will hurt their credit rating.[23]

A. *Narrow Networks*

In 2015, under the ACA, out-of-pocket costs were capped at $6,600 for an individual and $13,200 for a family.[24] But again, these caps do not apply to out-of-network providers who charge patients for the portion of their bills that their insurance does not pay.[25] Moreover, this balance-billing problem is getting worse because networks are becoming narrower.[26]

A network consists of the hospitals with which an insurer has contracted.[27] Pursuant to those contracts, hospitals agree to dramatically discount their list prices.[28] If a provider is not in-network, that means that the insurance company has no contract with the hospital, and patients who are insured by that company are not entitled to the huge discounts and instead are, according to the hospital, responsible for the fully undiscounted, obscenely high, list price for the services they receive.

A network is narrow if it only includes a few hospitals.[29] A wide network, which includes many hospitals, gives patients more choice as to

23. *Id.*

24. *Id.*

25. *Id.*

26. *See id.* (noting that plans with narrow networks make up about half of all health law exchange networks and about two-thirds of networks in large cities); *see also* Abelson, *supra* note 8 (acknowledging balance-billing becoming more frequent).

27. *See* HOADLEY ET AL., *supra* note 5, at 3–5.

28. *See* Nation, *Chargemaster Insanity*, *supra* note 12, at *19–23 (noting that self-pay patients billed chargemaster rates are asked to pay are at least 2.5 times amount paid by health insurers for same exact care); Nation, *Determining Fair and Reasonable Value, supra* note 7, at 429–30.

> Another important characteristic of healthcare is that chargemaster or list prices are not fair or reasonable. They are grossly inflated because they are set to be discounted rather than paid. Hospitals, in general, do not expect to recover these inflated prices, but . . . they are very reluctant to reduce them for self-pay patients. Nevertheless, hospitals and other providers maintain that the grossly inflated list prices contained in their chargemasters are "reasonable and customary," in part because every patient, insured or uninsured, receives a detailed itemized bill reflecting chargemaster prices. As a result, hospitals sometimes claim that all patients are billed at chargemaster rates. However, while all patients are billed chargemaster rates, all patients are not expected to pay the billed charges. . . . [F]or insured patients, the billed (chargemaster based) amount is dramatically (at least 50%) discounted. Thus, while hospitals claim that the chargemaster rates reflect their usual and customary *charge* for services, they certainly do not represent the usual price actually *paid* for the listed goods and services.

Id. (footnotes omitted).

29. *See* Armour, *supra* note 18 (explaining health plans offered by employers also have been reducing number of doctors and hospitals in their networks, but what have come to be known as narrow networks are more prevalent in plans offered on ACA exchanges).

where to seek care. A poll conducted by the Kaiser Family Foundation found that more than half of Americans believe that it is important to make sure that health plans have sufficient networks to provide a wide choice of doctors and hospitals.[30] However, according to a 2015 report by McKinsey & Company, plans with narrow networks make up half of all insurance networks offered through the ACA, and narrow networks make up about two-thirds of the insurance networks offered through the ACA in the largest cities.[31] As insurance networks become narrower, more patients are burdened with exorbitant hospital debt pursuant to balance billing.[32]

The reason that networks are becoming narrower is the desire on the part of hospital systems to increase profits.[33] For example, in some cases, insurance companies cannot afford the reimbursement levels being

30. *Id.*

31. *Id.*

32. *Id.*

33. *See* Murray, *supra* note 10.

> More importantly, the already astronomical and rapidly escalating hospital charge levels also have a less obvious impact on the rise in overall health care costs. High and increasing charges fundamentally undermine the negotiating leverage of private payers relative to hospitals, both big and small. This dynamic, which has been playing out in negotiations between private insurers and hospitals for years, goes something like this:
>
> . . . .
>
> When hospitals negotiate with health plans they have one of two options: 1) they can take a lower negotiated rate (around 135 percent of cost, which is the average payment level nationally as shown by the AHA statistics) and receive higher volumes of patients by virtue of being "in-network"; or 2) they can decline to be in-network and receive an average profit of 220 percent of costs on smaller patient volumes admitted through their EDs. The higher the profit on ED patients that pay out-of-network rates, the stronger the incentive for the hospital to drive hard bargains with insurers over negotiated prices.
>
> Recent analyses of private-sector pricing trends show stronger-than-average growth in hospital prices for Emergency Department services. The Health Care Cost Institute (HCCI), which monitors spending trends by private insurers, found that from 2009 to 2011, unit prices for ED services increased by 16.3 percent, compared to 9.9 percent and 8.1 percent increases in prices for inpatient and ancillary services, respectively. The profit-making opportunity to raise prices for services with highly inelastic demand curves is clearly not lost on the hospital industry.
>
> . . . .
>
> However, even under Scenario 1, with markups at 320 percent or higher, the hospital has relatively little incentive to negotiate with a health plan that cannot promise substantial volumes. The bottom line conclusion, then, is that high markups and heavy and growing use of the ED as a source of admission act to substantially reduce insurer market power, even for providers with relatively small market share. Those who negotiate on behalf [of] commercial insurers are well aware of how the ability of hospitals to raise charges completely undermines their own negotiating leverage.

Id.

demanded by hospital systems in order to become in-network.[34] Some hospital systems choose to purposely limit the size of their networks because they feel that this strengthens their financial bargaining position and allows them to recoup higher payments from insurers and patients.[35] An important cause of this is the increased concentration that has occurred on the provider side of the market.[36] As hospitals consolidate, more large healthcare systems are created, and these dominant systems do not feel any competitive pressure to contract with insurance companies at reasonable reimbursement rates.[37]

---

34. *Id.*

35. *Id.*

36. *Id.*

> It is a well-documented fact that provider consolidation—which research shows leads to higher prices—is already extreme and once again on the rise.
>
> . . . .
>
> As Barak Richman from the Duke School of Law has discussed, health care providers with market power enjoy substantially more pricing freedom than monopolists in other industries because of the presence of U.S. style health insurance, which largely insulates consumers from the full implications of monopoly pricing. This dynamic results in much greater potential for revenue generation and much greater distribution of wealth than would result from monopoly power in markets where consumers face the prices and price increases directly.
>
> Thus, the prospects for cost control are greatly diminished as long as providers are allowed to exercise their monopoly power, particularly where they face a highly inelastic demand curve—namely for emergency department services. The ability to hold a gun to the head of private insurers in this fashion is a by-product of provider consolidation, the enhanced pricing flexibility of health care monopolies, and the increasing proportions of admissions through hospital EDs.

Id. (citations omitted).

37. *See id.*; Elisabeth Rosenthal, *As Hospital Prices Soar, a Stitch Tops $500*, N.Y. TIMES, Dec. 2, 2013, http://www.nytimes.com/2013/12/03/health/as-hospital-costs-soar-single-stitch-tops-500.html?pagewanted=all&_r=0 [https://perma.cc/VU7N-LWP9?type=image] (quoting Glenn Melnick, Professor of Health Econ., Univ. of S. Cal.) (regarding California Pacific Medical Center, which is owned by Sutter Health Inc., whose chargemaster rates are 5.5 to more than 10 times the Medicare reimbursement rate).

According to Professor Melnick, "Sutter is a leader—a pioneer—in figuring out how to amass market power to raise prices and decrease competition." *Id.* (internal quotation marks omitted). Research shows that today's hospital mergers tend to drive up prices. For example in the case of Sutter, it operates the only hospital in some California cities. As a result, employers have limited ability to fight back against Sutter's high fees. Professor Melnick notes that hospitals set prices to maximize revenue, and they raise prices as much as they can. In addition, Professor Melnick notes that chargemaster prices are basically arbitrary, not connected to underlying cost or market prices; hospitals can set them at any level they want. There are no market constraints. Hospitals are the most powerful players in the healthcare system and there is little or no price regulation in the private market. *See* MARTIN GAYNOR & ROBERT TOWN, ROBERT WOOD JOHNSON FOUND., POLICY BRIEF NO. 9, THE IMPACT OF HOSPITAL CONSOLIDATION—UPDATE 2 (2012), *available at* http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2012/rwjf73261

B. *No Notice, No Control, and Unfair Surprise*

Even patients who are aware of the risks of balance billing and who, given their medical condition, are in a position to make a choice regarding where to seek treatment, find it difficult, if not impossible, to prevent balance billing.[38] This is because it is often extremely difficult for a patient to determine whether the provider from whom they are seeking medical services is in-network or out-of-network.[39] Moreover, the mere fact that a patient seeks medical services from an in-network hospital does not ensure that the doctors treating the patient are also in-network.[40] That is, it is very common for in-network hospitals to employ physicians who are not in that network.[41] As a result, patients treated at an in-network hospital may receive balance bills from the out-of-network physicians who treated them.[42] This is confusing and unfair to patients.

C. *Exorbitant, Obscenely High Charges*

The problem of balance billing would not be of nearly as much concern if the balance bills were not so outrageously high.[43] That is, if hospitals and other healthcare providers set their list prices at a fair and reasonable level to begin with, the balance bills would not represent a crippling financial burden for patients.[44] Rather, they would simply represent the difference between a reasonable list price and a likewise reasonable reimbursement amount set by the insurance company. In this economically sensible world (one that a properly functioning free market would create), balance bills would often be zero or a minimal amount. Unfortunately, this does not represent the current reality of hospital billing, as illustrated by the fact that in the case cited above, once the hospital reduced its charges, the bill went from $22,945 down to $600.[45]

What makes the problem of balance billing so pernicious is that the bills not only surprise patients, but the total cost of the bills is often financially devastating.[46] I have written before about outrageously high charge

[https://perma.cc/45J8-QNNU] (identifying hospital bargaining leverage as main determinant of relative expensiveness within same hospital market).

38. *See, e.g.*, Armour, *supra* note 18; Chen, *supra* note 5. *See generally* HOADLEY ET AL., *supra* note 5.

39. *See* Chen *supra* note 5 ("It's a pretty good bet that if you're hospitalized or having any kind of surgery, somebody along the way who touches you or your slides or films will not be in network[.]" (quoting Karen Pollitz, Senior Fellow, Menlo Park, California-based Kaiser Found.) (internal quotation marks omitted)).

40. *Id.*

41. *Id.*

42. *Id.*

43. *See, e.g.*, Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 427.

44. *Id.*

45. *See supra* notes 18–23 and accompanying text.

46. *See e.g.*, Melissa B. Jacoby & Mirya Holman, *Managing Medical Bills on the Brink of Bankruptcy*, 10 YALE J. HEALTH POL'Y L. & ETHICS 239, 247 (2010) (arguing medical debt makes it difficult to get further healthcare); Melissa B. Jacoby & Eliza-

master prices that hospitals insist are usual and customary, and I do not wish to repeat that work here.[47] It is sufficient to note that the amounts reflected on balance bills, when based on chargemaster prices, are outrageously high, set to discounted and not paid, bear no relationship to the hospital's cost, and, if they are paid, yield truly enormous profits to the hospital.[48]

D. *Balance Billing Increases the Overall Cost of Healthcare*

Because the profit maximizing conduct of hospitals, both for-profit and not-for-profit, is unrestrained by competitive market forces, the overall cost of healthcare in the United States is inflated.[49] The lethal combination of exorbitantly high chargemaster prices and the practice of balance billing combine to put upward pressure on prices for healthcare across the board.[50] These practices not only directly increase the cost of healthcare for the uninsured and out-of-network patients; they also indirectly increase the cost of healthcare for everyone.[51] While government insurers who pay for more than half of the healthcare provided in the United States no longer set reimbursement rates based directly on chargemaster rates, higher chargemaster rates do indirectly put upward pressure on government reimbursement amounts.[52] However, more importantly and the focus of this Article is the fact that the combination of obscenely high chargemaster rates and the practice of balance billing in-

---

beth Warren, *Beyond Hospital Misbehavior: An Alternative Account of Medical-Related Financial Distress*, 100 NW. U. L. REV. 535, 548 (2006) (asserting about 46%–56% of personal bankruptcies were caused by medical reasons); Christopher Tarver Robertson, Richard Egelhof & Michael Hoke, *Get Sick, Get Out: The Medical Causes of Home Mortgage Foreclosures*, 18 HEALTH MATRIX 65, 66–68 (2008) (noting 23% of home foreclosures were caused by "unmanageable medical bills").

47. *See generally* Nation, *Chargemaster Insanity*, *supra* note 12; Nation, *Determining Fair and Reasonable Value*, *supra* note 7; Nation, *Obscene Contracts, supra* note 3; Nation, *Wolves in Sheep's Clothing*, *supra* note 14.

48. *See, e.g.*, Reinhardt, *U.S. Hospital Services*, *supra* note 4, at 63 (explaining chargemaster prices "would yield truly enormous profits" if paid).

49. *See* Nation, *Chargemaster Insanity*, *supra* note 12, at *26 ("[H]igh chargemaster prices lead to [overall] higher prices for healthcare."); Nation, *Wolves in Sheep's Clothing*, *supra* note 14, at 154 ("[T]he primary reason that the non-profit model has dominated the hospital industry is that it provides camouflage and autonomy for the real profit seeking motives and/or elitist wealth transfer motives of those in de facto control, and it affords a tax deduction that enhances profits.").

50. *See supra* note 10.

51. *See supra* note 10.

52. *See* Nation, *Chargemaster Insanity*, *supra* note 12, at *26–30 (noting exorbitant chargemaster prices cause higher overall prices for healthcare); Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 454 (discussing main reason chargemaster prices are so high is that higher chargemaster prices lead to higher revenues, though not dollar-for-dollar, for hospitals from government and private insurers as well as from self-pay patients).

crease the cost of healthcare for both uninsured and privately insured patients.[53]

It is obvious why uninsured and out-of-network patients pay more—because the dysfunctional market for healthcare allows hospitals to set unreasonably high chargemaster rates and insist on balance billing out-of-network, uninsured, and other self-pay patients.[54] What is less obvious is why this also increases the cost of healthcare for in-network patients. Remember that patients are considered to be in-network if their insurance company has entered into a contract with the hospital providing medical services, and as a result, in network patients typically cannot be balance billed.[55]

However, the negotiation of the contract that makes a patient "in-network" is affected directly by exorbitant chargemaster rates and the practice of balance billing. For example, when a hospital or hospital system and an insurance company negotiate reimbursement rates, the hospital system's bargaining power is increased by the fact that if the insurance company fails to agree to the reimbursement rates desired by the hospital system, then all of the insurance company's customers are balance billed at chargemaster rates.[56] This threat—*agree to our reimbursement rates or your insureds will face huge charges for healthcare*—is strengthened each time the hospital raises its chargemaster rates.[57] This threat-based bargaining power is irresistible in the case of a hospital or hospital system that is dominant in its market.[58] Insurers simply cannot sell health insurance policies if those who buy them will be punished with exorbitant balance bills for receiving care from the dominant provider in the market.[59] As a result, many insurers have no option but to agree to the high reimbursement rates requested by the hospital system and to pass these costs along to their customers/insureds in the form of higher prices for health insurance.[60]

53. *See supra* note 10.

54. *See supra* notes 24–48 and accompanying text.

55. *See supra* notes 5–11 and accompanying text.

56. *See* Bai & Anderson, *supra* note 10, at 923 (noting that high chargemaster rates motivate insurers "to include hospitals in their networks to reduce the likelihood of having subscribers pay high out-of-network prices"); Murray, *supra* note 10 (noting that high chargemaster rates "undermine the negotiating leverage of private [insurers] relative to hospitals").

57. *See* Murray, *supra* note 10 (discussing hospitals' options when negotiating with insurers).

58. *See, e.g.*, Nation, *Chargemaster Insanity*, *supra* note 12, at *21–22 (discussing California Pacific Medical Center—owned by Sutter Health—and its amassed market power allows it to charge "5 1/2 to over 10 times the Medicare reimbursement rate").

59. *See id.* at *28 (discussing how private insurers, even these with significant market power, are forced to agree to high contractual reimbursement rates with must have hospitals in their market).

60. *See id.* at *26 (discussing how "insanely high chargemaster prices lead to over all higher prices" for insured patients as well as "self-pay patients").

The price of insurance (what insured patients pay for healthcare) consequently goes up.[61]

An alternative for insurance companies is to simply sell narrow network policies to uninformed customers and let these patients be shocked and surprised by the balance bills they receive. Narrow network policies are of course cheaper, and the low price often attracts customers who do not fully understand the risks posed by narrow network policies.[62] As noted, the majority of policies sold on ACA exchanges are narrow network policies.[63] In addition, several newspaper articles have focused on the frustration, shock, and surprise of patients who receive huge balance bills.[64]

Make no mistake, even non-profit, tax-exempt, so-called charitable hospitals act exactly like their for-profit competitors when it comes to setting obscenely high chargemaster prices and balance billing their patients.[65] I have written before about the very uncharitable conduct of so-called charitable hospitals, and there is no need to repeat that work here; it is sufficient to note that there is no meaningful difference between the conduct of non-profit and for-profit hospitals when it comes to conducting their financial affairs, except of course that the non-profits make more money because they do not pay taxes.[66] Moreover, while the pricing and collection limitations included in the ACA apply only to non-profit hospitals, these provisions do not solve the balance-billing problem even in the context of non-profit hospitals as discussed in the next Section.[67] Thus, there is little to be gained from simply applying the ACA's ineffective price and collection limitations to for-profit hospitals.

61. *See id.* at *28 ("[H]igher chargemaster rates do indirectly produce higher hospital revenue from all private insurers.").

62. *See* Armour, *supra* note 18 (discussing that customers who purchased coverage through ACA are surprised by balance bills).

63. *See id.* (discussing how "plans sold on ACA exchanges have limited networks").

64. *See supra* note 5.

65. *See* Nation, *Wolves in Sheep's Clothing*, *supra* note 14, at 174–79 ("The many instances in which non-profit hospitals seem to place the pursuit of profit over charity care has led the Commissioner of the IRS to observe that there is now very little difference between for-profit hospitals and not-for-profit hospitals."); *cf.* Bai & Anderson, *supra* note 10, at 924 (finding that 98% of top 50 hospitals with highest chargemaster rates were for-profit). These findings are misleading with respect to the difference between for- and non-profit hospitals. The sample is small at fifty, most of the top fifty were owned by just two for-profit systems, but most importantly the average chargemaster rate for *all* hospitals was 3.4 times Medicare allowable cost. *Id.* at 923.

66. *See generally* Nation, *Wolves in Sheep's Clothing*, *supra* note 14, at 170–79 (discussing issues with non-profit hospitals).

67. *See infra* notes 68–83 and accompanying text.

E. *The ACA Actually Encourages Balance Billing*

The ACA started out with laudable goals, however, much got lost in its translation into legislation. As noted above, the ACA contains limits on what patients may be asked to pay out-of-pocket if they are covered by a qualified health insurance policy, but these limits do not apply to non-emergency charges of out-of-network providers.[68] In other words, the out-of-pocket limits established by the ACA do not apply to balance billing.[69] The ACA also establishes limits on the amount that indigent, uninsured patients may be charged for healthcare and the type of collection techniques that may be used to recover healthcare debt.[70] It is important to note, however, that these limitations apply only to not-for-profit, tax-exempt hospitals; they do not apply to for-profit hospitals.[71] For-profit hospitals make up approximately 20% of the hospitals in the United States, and, while not directly relevant here, I have argued elsewhere that all hospitals should be for-profit and taxable.[72]

In any event, an indigent patient eligible for a (not-for-profit, tax-exempt) hospital's financial assistance policy (FAP) may not be charged more than the hospital's generally billed amount (GBA).[73] GBA is a reasonable amount established under the ACA and may be based on either the average amount the hospital bills private insurance and Medicare for the services provided or the prospective Medicare reimbursement rate alone.[74] Moreover, these hospitals and their collection agencies are forbidden from using extraordinary collection techniques to collect hospital debt from FAP-eligible patients.[75]

However, these provisions fail to solve the balance-billing problem even for not-for-profit, tax-exempt hospitals for several reasons. First, these hospitals are free to define who is eligible for their financial assis-

---

68. *See* Patient Protection and Affordable Care Act (ACA), Pub. L. No. 111-148, 124 Stat. 119, 855–858 (2010) (codified as amended at 26 U.S.C. § 501 (r)(5)–(6) (2012)). Emergency out-of-network charges are covered to some extent under the ACA. *See* ACA, §§ 1001, 10101(h), 42 U.S.C. 300gg-19a (2012) (amending Public Health Service Act § 2719A); 29 C.F.R. § 2590.715–2719A (2016); 45 C.F.R. § 147.138(b) (2016).

69. *See* ACA § 1302(c) (codified at 42 U.S.C. § 18022(c)); Bai & Anderson, *supra* note 10, at 923 ("The ACA requires nonprofit hospitals to provide discounts to eligible uninsured patients. However, the same provision lets individual nonprofit hospitals determine their own eligibility standards, does not address the levels of the markup faced by out-of-network patients and casualty and workers' compensation insurers, and does not apply to for-profit hospitals.").

70. *See supra* note 67.

71. *See supra* note 67.

72. *See generally* Nation, *Wolves in Sheep's Clothing*, *supra* note 14.

73. *See* I.R.C. § 501(r)(5)(A) (2012), *amended by* Pub. L. No. 114-113, 129 Stat. 2242 (2015).

74. *See* 26 C.F.R. § 1.150(r)–1 (2016) (defining amounts generally billed).

75. *See* I.R.C. § 501(r)(6).

tance programs.[76] Many hospitals define FAP eligibility according to income levels based on the Federal Poverty Guidelines (FPG). The problem is that, in addition to the FPG limits, most of these hospitals also limit eligibility for their FAPs to those who are uninsured.[77] Obviously, this offers no protection for patients subject to balance billing, who, by definition, are insured but have received care outside of their network.

Second, the recently finalized regulations implementing the ACA's limitations on hospital charges for FAP-eligible patients specifically allow balance billing of patients who qualify for financial assistance.[78] That is, if, under a specific hospital's FAP, an insured patient is eligible for financial assistance and therefore the amount the hospital may charge is limited to the hospital's GBA, the hospital is specifically permitted to recover this amount from both the insurance company *and* the patient.[79] The hospital may balance bill the FAP-eligible patient for an amount up to the GBA amount even though the hospital has already collected the GBA amount—an amount deemed to be the reasonable value of the services provided according to the ACA, from the insurance company.[80]

Finally, the ACA specifically refers to a hospital's "gross" charges,[81] clearly indicating their chargemaster rates. While the ACA does this with good intentions, specifically requiring not-for-profit, tax-exempt hospitals to charge less than their list prices to FAP-eligible patients (the ACA does not say how much less) nevertheless references to chargemasters effectively require that chargemasters—with their exorbitant prices—stay in existence. In addition, hospitals continue to have the same incentives to continually raise their chargemaster rates.[82] As noted above, high

76. *See id.* § 501(r)(4)(A); Additional Requirements for Charitable Hospitals, 77 Fed. Reg. 38148, 38149 (June 26, 2012) (codified at 26 C.F.R. pt. 1) ("Neither the [ACA] nor these proposed regulations establish specific eligibility criteria that a [financial assistance policy] must contain.").

77. *See, e.g.*, Cleveland Clinic, Summary of Financial Assistance (PWO 13998), *available at* https://my.clevelandclinic.org/ccf/media/Files/Patients/financial-assistance-app.pdf?la=en [https://perma.cc/9BMS-JBPJ] (last updated Jan. 2013) ("[W]e provide financial assistance . . . if you are a resident of the state in which you are seeking care . . . *do not have insurance*, and your family income does not exceed four times the FPG." (emphasis added)).

78. *See* 26 C.F.R. § 1.501(r)–5(b)(2) (stating it is no violation of regulations if the total amount paid by individual and health insurer exceeds AGB, so long as individual's portion—including co-payments, co-insurance, and deductibles—does not exceed the AGB).

79. *See id.*

80. *See id.*

81. *See id.* § 1.501(r)–1(b)(16) ("Gross Charges, or the chargemaster rate, means a hospital facility's full, established price for medical care that the hospital facility consistently and uniformly charges patients before applying any contractual allowances, discounts, or deductions.").

82. Higher chargemaster rates mean greater revenue, though not dollar-for-dollar. *See supra* note 49.

chargemaster rates contribute significantly to the severity of the balance-billing problem.[83]

II. Analysis: The Solution

It is important to recognize that our healthcare goals are not only to control prices and to provide access to all Americans to basic healthcare, but first and foremost, to develop the best healthcare in the world.[84] The best healthcare is that which can cure or treat the greatest number of diseases and ailments with the greatest success.[85] A thriving free market for healthcare will accomplish our goals with regard to pricing and the development of the best healthcare.[86] The point is that the issue of how to provide access to basic healthcare for low-income populations is separate and distinct from the issue of how to develop the best healthcare.[87] To develop the best healthcare and to control costs, we must enhance the free market for healthcare.[88] Also, a robust free-market economy is the best remedy for poverty and its effects, including lack of access to healthcare.

A. *More Price Fixing Will Not Help*

There is no question that when hospitals seek to collect their chargemaster rates, they are acting unreasonably and, especially in the case of charitable hospitals, unfairly.[89] But hospitals are also responding predictably—which is to say somewhat reasonably, though certainly not charitably—to existing market forces.[90] The solution is to change the market conditions, such as the current lack of price transparency, that are preventing the free market for healthcare from functioning properly.[91] Price-fixing does not strengthen the free market; it destroys it and replaces it with central planning. Not only does prior experience suggest that this will ultimately fail to control prices, it will also destroy quality.

---

83. *See supra* notes 43–48 and accompanying text.

84. *See* Nation, *Wolves in Sheep's Clothing*, *supra* note 14, at 151–55 (discussing how United States can better its healthcare system).

85. *See id.* at 151–52 & n.33 (discussing difference between "healthcare system" versus "healthcare").

86. *See id.* at 152 n.34 (discussing why free-market system is better for hospitals). I have written before about the dangers associated with incorporating the access goal too tightly with the goal of developing the best healthcare, and I do not want to repeat that work here. *See generally id.*

87. *See id.* at 207–09 ("[T]he issue of how to pay for medical care for the poor, while it is not completely unrelated, is best treated as separate and distinct from the issue of how to develop the best medical care at the lowest cost." (footnote omitted)).

88. *See id.* at 154–55 (discussing what changes should be made to healthcare market).

89. *See supra* notes 42–47 and accompanying text.

90. *See* Nation *Chargemaster Insanity*, *supra* note 12, at *6–18 (discussing evolution of higher chargemaster rates).

91. *See, e.g.*, Bai & Anderson, *supra* note 10, at 925 (discussing policy implications to market failure caused by lack of price transparency).

The United States healthcare market already has too much price-fixing. Government insurers such as Medicaid and Medicare are price fixers.[92] Moreover, they represent a large share of the U.S. healthcare market.[93] Price-fixing, whether in the form of a single-payer system or in the form of full-blown government healthcare, is not the solution because it ignores the real goal: to develop a properly functioning free market for healthcare, because this is the only way to develop the *best* healthcare.

B. *Enhancing the Free Market for Healthcare*

Enhancing the free market for healthcare will require the government to work to prevent and reduce market concentration and ensure relatively equal bargaining power of both providers and consumers (or insurance companies as the consumers' representatives).[94] In addition, a properly functioning free market requires price and quality transparency so that consumers can actively choose the best value in healthcare[95]—that is, the best healthcare at the cheapest price. When consumers have this information, providers are forced to compete in quality and price. This is the essence of a free market, and it is from this that all of its benefits flow.

Other than observing that certain provisions of the ACA are, unfortunately, encouraging concentration in an already overly concentrated provider side of the market, I focus in this Part on how to increase price transparency.[96] It is also important to note that this Article seeks to strengthen the free market for healthcare, and that necessarily includes recognizing that providers should be free to set their prices at any level they wish, with the exception of services provided in the emergency department.[97] However, a properly functioning free market will force them to set reasonable prices.

C. *Increasing Price Transparency Using the Common Law of Contracts*

One of the most significant problems in the market for healthcare is the lack of price transparency.[98] Patients do not know at the time of contracting how much they are agreeing to pay for the services that will be

92. *See, e.g.*, Reinhardt, *U.S. Hospital Services*, *supra* note 3, at 60 (discussing price setting by Medicare).

93. *See* Nation, *Wolves in Sheep's Clothing*, *supra* note 14, at 186 n.237 (noting that "the [g]overnment is now responsible for paying for more than fifty percent of U.S. healthcare via various programs, principally Medicare and Medicaid").

94. *See supra* notes 35–36 and accompanying text (discussing how hospitals hold more financial bargaining power).

95. *See, e.g.*, Bai & Anderson, *supra* note 10, at 924–25 (discussing lack of transparency in market failure).

96. *See infra* notes 99–123 and accompanying text (discussing how to increase price transparency).

97. *See supra* note 67. For a discussion of contract prices, see *infra* notes 124–33.

98. *See, e.g.*, Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 426–29 (discussing lack of price transparency in healthcare contracts).

provided to them.[99] As a result, it is virtually impossible for patients to compare hospitals based on price.[100] This, in turn, means that hospitals are not forced to compete on price. As mentioned above, it is also important for hospitals to compete on quality, but in the current market environment, it is much easier for patients to get an idea of the relative quality of hospitals than it is for them to get an idea of how much they will be responsible for paying for the services that they receive.[101] My focus here is on increasing price transparency.

Not only is price transparency critical for the proper functioning of the free market, but basic principles of contract law also require that the parties clearly establish the terms of their agreement or, if the court is convinced the parties intended a contract but did not clearly establish some terms, including the price, the court will imply a reasonable price to the contract consistent with the parties' presumed intent.[102] At the heart of the common law requirements for the creation of a contract is its recognition of the freedom of individuals to knowingly and freely enter into enforceable agreements.[103] The problem with healthcare contracts entered into directly with patients is that they are not knowingly and freely entered into with respect to price.[104]

As discussed in more detail in the next Section, courts can now use the common law principles of contracts to rein in the abusive balance-billing practices of hospitals.[105] In this Section, I renew a suggestion I have made previously for regulation that will enhance the functioning of the free market by increasing price transparency.[106] In addition, I recommend here specific patient disclosures designed to ensure healthcare contracts entered into directly with patients are knowingly and freely entered into with respect to price.[107] My recommendation regarding price disclo-

---

99. *See id.* (discussing how patients are not aware of full costs of healthcare).

100. *See id.* at 429 ("[A]s long as hospitals use á la carte pricing based on chargemasters, consumers will not be able to effectively negotiate price.").

101. There are many ranking services available regarding the relative quality of healthcare. *See, e.g.*, *U.S. News Best Hospitals 2015–16*, U.S. NEWS & WORLD REP., http://health.usnews.com/best-hospitals/rankings [https://perma.cc/6L2M-2YWV] (last visited Mar. 3, 2016).

102. *See infra* notes 166–70 and accompanying text (discussing creation of contracts between parties).

103. *See infra* notes 151–70 and accompanying text (discussing common law of contracts).

104. *See supra* notes 98–100 and accompanying text.

105. *See infra* notes 225–32 and accompanying text (discussing how common law contract principles can be used in healthcare).

106. *See* Nation, *Chargemaster Insanity*, *supra* note 12, at *36–44 (discussing proposed solution for cost of healthcare).

107. *See infra* notes 108–23 and accompanying text (proposing recommendation for price disclosures by hospitals).

sure are similar to Regulation Z[108] (Regulation Z is issued by the Board of Governors of the Federal Reserve System to implement the federal Truth in Lending Act, which is contained in Title I of the Consumer Credit Protection Act[109]), which applies to consumer lending and ensures that consumers know and understand the credit terms of the loan they are agreeing to.[110] Similarly, in the context of healthcare contracts, the government's role should be to require disclosure of the relevant pricing information.

Requiring the following would be helpful to provide patients with the relevant information they need regarding price to enter into an enforceable contract with a hospital regulation: (1) all providers should be required to adopt the same pricing system, but not the same prices[111] and (2) a number of disclosures including the following need to be made to the patient prior to the creation of the contract.[112] First, the patient should be told the total amount that the hospital will charge (including amounts charged to the patient's insurance or any other third-party payer) for the care to be provided. Second, the patient should be told the maximum amount, in dollars and cents, that the patient will be expected to pay for the services for which they are contracting.[113] Third, the hospital must disclose the amount, in dollars and cents, that the hospital receives from Medicare for the same services and the average amount the hospital receives from private insurers for the same services.[114] Fourth, hospitals must tell the patient explicitly that the hospital accepts these amounts, from private and government insurers, as full payment for these services. Fifth, the hospital must disclose explicitly, again in dollars and cents, how much *more* the hospital is asking the individual patient to pay for these

108. Regulation Z is issued by the Board of Governors of the Federal Reserve System to implement the federal Truth in Lending Act, Title I of the Consumer Credit Protection Act, as amended at 15 U.S.C. §§ 1601–1616 (2012).

109. 15 U.S.C. § 1601.

110. *See id.*; *see also* Keith T. Peters, *What Have We Here? The Need for Transparent Pricing and Quality Information in Health Care: Creation of an SEC for Health Care*, 10 J. HEALTH CARE L. & POL'Y 363, 364 (2007) (noting that with proper information healthcare consumers can make rational decisions).

111. This is a recommendation that others and I have made before. *See* Bai & Anderson, *supra* note 10, at 927 (regarding legislation requiring all reimbursements to be based on the same system such as diagnostic-related groups); Nation, *Chargemaster Insanity*, *supra* note 12, at *38.

112. *See* Bai & Anderson, *supra* note 10, at 925–26 (proposing required price disclosures by hospitals).

113. I have previously argued that a reasonable price for medical services is no more than 115% of the Medical reimbursement rate. *See* Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 460–65.

114. *See* Nation, *Obscene Contracts*, *supra* note 3, at 118–19 (discussing how hospitals usually do not collect full amount of charges from government and private insurers).

exact same services.[115] Sixth, the hospital should also inform the patient that it negotiates rates with private insurers and, if the hospital in fact does so, with self-pay patients on a case-by-case basis. Finally, a prominent space should be provided on the disclosure form for the total negotiated price the patient and the hospital have agreed to and the amount of discount received by the patient.[116]

These regulations should be written, similar to Regulation Z, so that a single, consistent, and easy to read and understand form is used by all hospitals to convey this pricing information to the patient. Not only will this type of regulation ensure that patients are knowingly and freely entering into contracts with hospitals, but it would also facilitate the comparison of prices between hospitals by patients. This in turn will result in price control via market forces.

If these disclosures regarding price are not made prior to contracting for any reason, then the court must decide if the parties intended to contract, which would clearly be the case if the patient received any treatment from the hospital, and if a contract was created, the court must provide a reasonable price term for the parties.[117] For example, a hospital may continue to use its price-ambiguous "Authorization for Treatment," "Statement of Financial Responsibility," or some other similar open-ended agreement pursuant to which the patient purports to agree to pay for all medical goods and services provided by the hospital at the hospital's chargemaster prices.[118] But if these open-price agreements were used, the patient would not be liable to the hospital for any more than a reasonable amount as decided by the court.[119] That is, whenever the provider does not make the required disclosures, the court will apply a reasonable price to the contract based on the average price paid for the same services by private insurers plus a modest percentage between 1% and 15%.[120]

As discussed more fully in the next Section, even in the absence of legislation requiring these price disclosures, courts should not interpret typical admission forms signed by patients—such as an "Authorization for Treatment," "Statement of Financial Responsibility," or some other similar open-ended agreement, which purport to establish a formula based on the hospital's chargemaster to arrive at the price the patient has agreed to pay for the services the hospital may provide—as establishing a definite price.[121] Rather, these types of forms have been compared "to a blank

115. *See id.* at 119 ("However, while all patients and payors are billed the 'full charges,' the only ones actually expected to pay these charges are those patients without medical insurance.").

116. This is similar to the disclosure box required by Regulation Z. *See supra* notes 108–09.

117. *Cf.* U.C.C. § 2-207(3) (2012).

118. *See infra* notes 121–25 and accompanying text.

119. *See supra* note 111.

120. *See supra* note 111.

121. *See* Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 426–27.

check given by the patient to the hospital with the amount to be unilaterally filled in by the hospital at a later date."[122] This is not at all consistent with common law contract principles that require parties to knowingly and freely enter into contracts.[123] As a result, the court should not find that these writings are binding with respect to establishing the price for the goods and services provided by the hospital to the patient.[124] Simply put, under the common law of contracts, the hospitals that fail to fully disclose relevant pricing information, as discussed above, may recover no more than the fair market value of the goods and services provided.[125]

D. *Price Setting Freedom in a Free Market*

One of the hallmarks of a free market is freedom of contract.[126] That is, individual market participants are free to enter into contracts on any terms that they knowingly and freely agree to within the parameters established by the common law.[127] These parameters include requirements such as capacity and legality, and compliance with normal contract policing tools such as fraud, duress, undue influence, and unconscionability.[128] Thus, in a free market, businesses are typically free to establish their prices at any levels they see fit, and potential customers are of course free to reject prices that they consider to be too high. In order to come to an agreement, the parties negotiate the terms—including the price—of the contract.

Because the purpose of the recommendations made here are to strengthen the operation of the free market, this Article takes the position that providers of medical goods and services must be free to set their prices as they see fit. Direct government price controls are not only inconsistent with the idea of freedom of contract that underlies the free market; they also disrupt the functioning of the free market and prevent it from achieving its ultimate goal of efficient resource allocation.[129] Moreover, a properly functioning free market will control prices.[130] In addition, as discussed in the next Section, I argue that the common law of contracts

122. *Id.* at 427.

123. *See infra* note 151 and accompanying text.

124. *See infra* notes 225–32 and accompanying text (discussing application of contract principles to healthcare).

125. *See infra* notes 228–34 and accompanying text (discussing how prices should be set in contracts whether its by parties or courts).

126. *See infra* notes 151–70 and accompanying text (discussing freedom of contract).

127. *See infra* notes 151–70 and accompanying text (same).

128. *See infra* notes 234–41 and accompanying text.

129. If prices are set too high, the excess profits attract more sellers and in turn the increased supply cause prices to come down. As a result, the proper level of resource allocation is achieved and maintained.

130. *See supra* note 128.

includes sufficient flexibility to allow judges to apply its principles to ensure that hospitals cannot engage in unfair pricing practices.[131]

The common law of contracts also leaves hospitals and other providers sufficient pricing flexibility to charge extremely high rates to patients who are able and willing to pay them. Commentators as well as hospital administrators often recount the "Arab Sheik" scenario.[132] This scenario involves an Arab Sheik, invariably extremely wealthy, arriving at the hospital seeking medical services; the Sheik is supposedly able and willing to pay any price that the hospital demands.[133] Hospital administrators, clearly misunderstanding the common law of contracts, fear that without their exorbitant chargemasters, they would not be able to charge the Sheik an exorbitant, overpriced amount for the hospital's services.[134] Why it is acceptable to overcharge a wealthy Arab Sheik, especially in the case of a so-called charitable hospital (which is an implicit assumption in this scenario) is unclear to me. Freedom of contract and the rules established by the common law of contracts allow hospitals to charge any patient any amount the hospital wishes as long as the patient and the hospital freely and knowingly agree.[135] The frequency with which one encounters the Arab Sheik scenario would seem to suggest that it is quite common, however, even without the advantage of empirical evidence, I am tempted to conclude that it may represent more wishful thinking than reality. In any event, to maintain such a ridiculous and pernicious chargemaster pricing system in the vein hope for the arrival of an Arab Sheik certainly seems to be an example of the tail wagging the dog.

E. *The "Just Say No" Approach*

Some companies that self-insure for health insurance and their third-party administrators have created health insurance plans that dispense with networks altogether.[136] Under these direct-pay plans, contracts are not entered into with any providers.[137] Employees insured under these plans are free to go to any provider they like.[138] When the bill is presented, the company pays an amount determined to be reasonable

131. *See infra* notes 151–241 and accompanying text (discussing how courts can apply common law contract principles to healthcare).

132. *See, e.g.*, Rosenthal, *supra* note 37.

133. *See id.*

134. *See id.*

135. *See infra* notes 151–72 and accompanying text.

136. *See, e.g.*, Jay Hancock, *Radical Approach to Huge Hospital Bills: Set Your Own Price*, KAISER HEALTH NEWS (May 13, 2015), http://khn.org/news/radical-approach-to-huge-hospital-bills-set-your-own-price/ [https://perma.cc/8EGM-DJV7] ("When hospitals send invoices with [jaw-dropping] charges . . . [ELAP Services] tells its clients (generally medium-sized employers) to just say no.").

137. *See id.* ("Under ELAP's main [approach], neither employers nor their claims administrators sign contracts with hospitals.").

138. *See id.* (discussing how individuals can use any provider they choose because there is no network).

based on reference to the amounts paid by other payers.[139] For example, the company may reimburse the hospital at the Medicare reimbursement rate plus a percentage, such as 25% or 30%. Hospitals are not required to accept this amount as full payment, because they have not entered into a contract with the company/insurer. Moreover, the hospital may balance bill the patient/employee for the difference between the hospital's chargemaster amount and the amount paid by the company/employer.[140]

However, the companies that offer such plans and their third-party administrators typically fight such balance-billing efforts on behalf of the employee/patient.[141] The few companies offering such plans report great success so far.[142] One such third-party administrator is ELAP Services, a small benefits consulting firm, based in Chester Springs, Pennsylvania.[143] According to an ELAP spokesperson, when ELAP clients are sued by hospitals in pursuit of balance bills, the company "fights back with lawyers and several arguments. How can hospitals justifiably charge employers and their workers so much more than they accept from Medicare . . . ? How can hospitals bill $30 for a gauze pad? How can employee-patients consent to prices they will never see until after they've been discharged?"[144] ELAP reports that "[e]ventually, 'overwhelmingly, the providers just accept the payment' and leave patients alone."[145] Whether this approach will be upheld in court in hospitals' balance-billing claims against patients is unknown.

ELAP was named as a defendant in a federal district court case in Georgia, which was decided in 2012 in ELAP's and the employer's favor—but the issue in that case was limited to whether the administration of the direct payment plan was consistent with the Employee Retirement Income Security Act (ERISA).[146] The court held that it was, but as the provider was acting in the place of the insured pursuant to an assignment of the insured's rights to benefits under the plan the case did not involve balance billing.[147] That is, no state law claims, the balance-billing claim would be a contract claim, were part of the proceeding.[148]

139. *See id.* ("[ELAP] estimate[s] costs . . . based on the hospital's financial reports filed with Medicare. Then it add[s] a cushion so the hospital [can] make a modest profit.").

140. *See id.* (discussing how ELAP fights back when clients are balance billed).

141. *See id.*

142. *See id.*

143. *See id.*

144. *Id.*

145. *See id.* (quoting Steve Kelly, Chief Exec. Officer, ELAP).

146. *See*, Floyd Med. Ctr. v. Warehouse Home Furnishings Distribs. Inc., No. 4:11–CV–15(CDL), 2012 WL 1438470, at *5 (M.D. Ga. Apr. 25, 2012) (addressing whether charges were "reasonable" and "customary" and whether they were covered).

147. *See id.* at *5–6.

148. *See id.* (stating that plaintiff did not "assert[ ] any type of state law claim that was independent of the federal ERISA claim").

III. WHAT COURTS CAN AND SHOULD DO NOW

Courts can use the common law of contracts to rein in abusive balance-billing practices. As this Part discusses, the common law of contracts is based on the premise that courts will enforce agreements that are entered into knowingly and freely by individuals.[149] Contracts entered into between healthcare providers and patients are not knowingly and freely entered into with respect to price, and therefore the courts should not

149. *See, e.g.*, Mark Klock, *Unconscionability and Price Discrimination*, 69 TENN. L. REV. 317 (2002).

> The consideration can be as nominal as a peppercorn for the agreement to be legally enforceable. Courts do not inquire into the distribution of benefits between the parties. This legal fact is deeply rooted in a strong faith in the efficiency of free markets. Individuals do not voluntarily enter into agreements that they expect to make them worse off than before the agreement. If the agreement was made voluntarily, everyone is presumed to have been made better off by the agreement. This presumption can be justified by economic thought which, given a few simple axioms, demonstrates that markets will channel resources to their most valued use and maximize society's wealth when all market participants are permitted to freely make their own decisions. Government intervention cannot improve the allocation of resources and can even impede it.

Id. at 343–44 (footnotes omitted); *see also, e.g.*, JOHN D. CALAMARI & JOSEPH M. PERILLO, THE LAW OF CONTRACTS §§ 1.3–1.4, at 4–10 (4th ed. 1998) (discussing freedom of contract and philosophical foundation of contract law respectively). *Ellsworth Dobbs, Inc. v. Johnson*, for example, involved a real estate broker who found a buyer for the seller. *See* 236 A.2d 843, 846 (N.J. 1967). The seller and buyer entered into a contract for sale, but the contract was never performed due to breach by the buyer. The broker brought suit against the seller, alleging that, based on the express terms of the listing agreement, the commission was earned upon execution of the contract between buyer and seller. The court ruled that any contractual provision in the listing agreement that required the seller to pay the commission even though the buyer of the land was unable to arrange financing and therefore breached the contract of sale, was "so contrary to the common understanding of men, and also so contrary to common fairness, as to require a court to condemn it as unconscionable." *Id.* at 857. In so ruling, the court applied the following reasoning, which is equally applicable to hospital admission contracts:

> Courts and legislatures have grown increasingly sensitive to imposition, conscious or otherwise, on members of the public by persons with whom they deal, who through experience, specialization, licensure, economic strength or position, or membership in associations created for their mutual benefit and education, have acquired such expertise or monopolistic or practical control in the business transaction involved as to give them an undue advantage. Grossly unfair contractual obligations resulting from the use of such expertise or control by the one possessing it, which result in assumption by the other contracting party of a burden which is at odds with the common understanding of the ordinary and untrained member of the public, are considered unconscionable and therefore unenforceable.
>
> . . . .
>
> The perimeter of public policy is an ever increasing one. Although courts continue to recognize that persons should not be unnecessarily restricted in their freedom to contract, there is an increasing willingness to invalidate unconscionable contractual provisions which clearly tend to injure the public in some way.

Id. at 856, 857 (citation omitted).

enforce the ridiculous prices alleged to be due pursuant to these contracts.[150]

A. *Common Law of Contracts*

1. *Law of Voluntary Agreements: Freedom of Contract*

A contract may be defined as a promise the courts will enforce.[151] The concept of freedom of contract, which plays a central role in the law of contracts, reflects the idea that individuals should be free to enter into contracts on any terms they wish.[152] There are, of course, limits to the

---

150. *See supra* notes 43–48 and accompanying text; *see also, e.g.*, Phoenix Baptist Hosp. & Med. Ctr., Inc., v. Aiken, 877 P.2d 1345 (Ariz. Ct. App. 1994) (denying summary judgment where husband signed admission agreement for wife that purported to make husband as signer personally liable for services provided to his wife, noting that husband may not have understood agreement or felt he had no choice but to sign); Tunkl v. Regents Univ. of Cal., 383 P.2d 441, 447 (Cal. 1963) (rejecting contract with exculpatory clause, noting that patients are in "no position to reject the proffered agreement, to bargain with the hospital, or in lieu of agreement to find another hospital"); Wheeler v. St. Joseph Hosp., 133 Cal. Rptr. 775, 783–86 (Ct. App. 1976) (rejecting agreement to arbitrate, noting that hospital admission agreement "possess[es] all of the characteristics of a contract of adhesion"—contracts offered on a take-it-or-leave-it basis with no realistic opportunity to bargain and such that the goods or services cannot be acquired without agreeing to the terms offered—and that "admission to a hospital is an anxious, stressful, and frequently a traumatic experience" as result patient cannot reasonably be expected to read the printed agreement in detail much less to fully comprehend its terms); St. John's Episcopal Hosp. v. McAdoo, 405 N.Y.S. 2d 935, 937 (Civ. Ct. 1978) (holding that in emergency admission hospital should not be permitted to enforce contractual obligation entered into under such tension-laden circumstances).

151. *See* RESTATEMENT (SECOND) OF CONTRACTS § 1 (1981) ("A contract is a promise or a set of promises for the breach of which the law gives a remedy, or the performance of which the law in some way recognizes as a duty.").

152. *See* CALAMARI & PERILLO, *supra* note 149, § 1.3, at 5.

> The crux is that as England changed from a relatively primitive backwater to a commercial center with a capitalistic ethic, the law changed with it. As freedom became a rallying cry for political reforms, freedom of contract was the ideological principle for development of the law of contract. In Maine's classic phrase, it was widely believed that "the movement of the progressive societies has hitherto been a movement from *Status to Contract*." Williston adds: "Economic writers adopted the same line of thought. Adam Smith, Ricardo, Bentham and John Stuart Mill successively insisted on freedom of bargaining as the fundamental and indispensable requisite of progress; and imposed their theories on the educated thought of their times with a thoroughness not common in economic speculation."
>
> In the twentieth century the tide turned away from the nineteenth century tendency toward unrestricted freedom of contract. Today, while the parties' power to contract as they please for lawful purposes remains a basic principle of our legal system, it is hemmed in by increasing legislative restrictions. . . .
>
> Apart from legislative restrictions on freedom of contract, it seems likely that in the future there will be greater restrictions imposed by courts in the exercise of their function of developing the common law. There has been increasing recognition in legal literature that the bar-

freedom of contract doctrine, which reflect other policy concerns of our legal system.[153] For example, one cannot enter into an enforceable contract to sell illegal drugs.[154] However, within the broad parameters established by the legal system, freedom of contract is a fundamental principle of contract law. Moreover, the principle of freedom of contract is consistent with the idea of a free-market economy, and one of the most important functions served by contract law is to facilitate the operation of the free market.[155] That is, buyers and sellers may use contract law to create agreements on whatever terms they see fit.[156] In addition, the principle of freedom of contract counsels that the role of the courts is to enforce the agreement created by the parties.[157] Courts are not to create contracts for the parties, nor are courts required to approve the agreement the parties created, except to ensure that it falls within the broad parameters of contract law principles.[158] The parties must create their own contracts because it is only through the aggregate interactions (contracts) of market participants that the market can determine the appropriate allocation of resources.[159]

However, the doctrine of freedom of contract does not mean that the courts do not play a role in the contract formation process; on the contrary courts play a very important role by both establishing and enforcing the rules that must be followed to create a contract.[160] For example, courts apply these rules to ensure that only agreements knowingly and freely entered into by the parties are enforced.[161] Courts require that

---

> gaining process has become more limited in modern society. In purchasing a new automobile, for example, the individual may be able to dicker over price, model, color and certain other factors, but, in order to consummate the contract to purchase, the individual usually must sign the standard form prepared by the manufacturer (although the contract is with an independent dealer). The individual has no real choice and must take that form or leave it. Such contracts, called contracts of "adhesion," constitute a serious challenge to much of contract theory.
>
> Most of contract law is premised upon a model consisting of two alert individuals, mindful of their self-interest, hammering out an agreement by a process of hard bargaining. The process of entering into a contract of adhesion, however ". . . is not one of haggle or cooperative process but rather of a fly and flypaper." Courts, legislators and scholars have become increasingly aware of this divergence between the theory and practice of contract formation, and new techniques are evolving for coping with the challenges stemming from this divergence.

Id. § 1.3, at 5–6 (second alteration in original) (footnotes omitted).

153. *See id.*

154. S*ee id.*

155. *See id.*

156. *See id.*

157. *See id.*

158. *See id.*

159. *See id.*

160. *See id.*

161. *See, e.g.*, *supra* note 150.

contracts be knowingly and freely entered into because, to agree, both parties must know what they are agreeing to and voluntarily agree.[162]

It is not possible to agree to nothing; similarly it is not possible to agree to allow the other party to charge any amount he wishes.[163] Such terms are not contracts or even agreements—they are the opposite of an agreement; they are simply an exercise of power by the stronger party against the weaker party.[164] The doctrine of freedom of contract allows parties to agree to whatever terms they wish, but it does require that the parties agree.[165] Freedom of contract does not allow for the enforcement of the law of the jungle in the guise of a contract. If one party purports to agree to allow the other party to charge any amount he wishes as the price of the contract, such an agreement is not likely a contract, because it was not likely knowingly and freely entered into.[166] Who would knowingly and freely agree to such a term? No one. Yet, as discussed *infra*, this is precisely what hospitals claim.

Traditionally courts have required the parties to specifically agree to all necessary terms in order to create a contract.[167] However, modern courts are more willing to supply missing terms for the parties if the court is convinced that the parties in fact intended to enter into a contract even if the words or writings of the parties do not clearly establish all necessary terms.[168] When the court fills in this term, it provides a term that is consistent with the reasonable expectations of the parties.[169] For example, if

---

162. *See supra* note 150.

163. *See* RESTATEMENT (SECOND) OF CONTRACTS § 77 (1981) (defining *illusory promise* as one that furnishes no basis for a contract because of lack of consideration); CALAMARI & PERILLO, *supra* note 149, §§ 2.1, 2.13, at 25, 73 (discussing mutual assent and indefiniteness respectively).

164. *See, e.g.*, *supra* note 152.

165. *See supra* notes 150 & 164.

166. *See supra* note 150.

167. *See supra* note 163.

168. *See* CALAMARI & PERILLO, *supra* note 149, § 2–13, at 45–46 (2d ed. 1977).

> There is an important distinction between cases in which the parties have purported to agree on a contractual provision and have done so in a vague and indefinite manner and cases in which they have remained silent as to a material term or have discussed the term but did not purport to agree upon it. When parties are silent as to a material term, often the reasonable conclusion is that they intended that the term be supplied by implication. Thus, if *A* and *B* agree that *A* will perform a service for *B* and no mention is made of price, it will be implied that the parties intended that a reasonable price should be paid and received. The same is true if goods are involved. It will be assumed that the parties contracted in terms of a reasonable price which will ordinarily be the market price. Where there is no market price the reasonable price may be determined by actual cost plus a reasonable profit or other means of valuation.

Id. (emphasis added) (footnotes omitted); *see also* RESTATEMENT (SECOND) OF CONTRACTS § 204 (1981) (discussing how to supply omitted essential terms).

169. *See* RESTATEMENT (SECOND) OF CONTRACTS § 204 ("When the parties to a bargain sufficiently defined to be a contract have not agreed with respect to a term which is essential to a determination of their rights and duties, a term which is

the court is called upon to fill in the price term of the contract, it will seek to determine a fair and reasonable price, because this is what reasonable contracting parties would intend.[170]

2. *Objective Intent and the Essential Requirements of Contract*

While a contract is a promise the courts will enforce and while all contracts are promises, not all promises are contracts. The common law of contracts uses the concept of objective intent and from it has derived the essential requirements of contract in order to determine which promises are contracts and which are not.[171] Objective intent is very different than subjective intent.[172] Objective intent makes no attempt to pick the brain of the person whose intent is in question.[173] Rather, to determine objective intent, we do not focus at all on the person whose intent we are trying to determine; we instead focus on a hypothetical reasonable person in the position of the other party—the party interacting with the person whose intent we are trying to determine.[174] The law uses the hypothetical reasonable person to determine the objective intent of the other person based on the other person's objective manifestations.[175] Objective manifestations are the things done and said—in this context by the person who's intent we are trying to determine—that may be perceived by others.[176] Thus, for contract law purposes, a person's intent is

---

reasonable in the circumstances is supplied by the court."); CALAMARI & PERILLO, *supra* note 149, § 2.9.

170. *See* RESTATEMENT (SECOND) OF CONTRACTS § 204 cmt. d (citing U.C.C. § 2-305).

171. *See, e.g.*, CALAMARI & PERILLO *supra* note 149, § 2–2, at 24–25 (2d ed. 1977) (discussing objective and subjective intention).

> A party's intention will be held to be what a reasonable man in the position of the other party would conclude his manifestations to mean. By testing the meaning to be given to a party's words from the point of view of the reasonable man in the second party's position, the subjective element of this party's particular knowledge is incorporated into the objective test. In other words, the test considers what the second party knows or should know about the intention of the first party.
>
> The objective theory is strongly supported by those who place the basis of contract law upon the promisee's justified reliance upon a promise or upon the needs of society and trade. An objective test is believed to protect "the fundamental principle of the security of business transactions." Even those who espouse intention as the basis of contract obligations are generally willing to hold a promisor to the reasonable meaning of his words, basing such liability on a theory of negligence, but are inclined to wish that the objective theory be held on a short leash, and to allow subjective intention a high degree of relevance in the resolution of many contractual issues.

Id. (footnotes omitted).

172. *See id.*

173. *See id.*

174. *See id.*

175. *See id.*

176. *See id.*

deemed to be that which is consistent with the perceptions of a hypothetical reasonable person in the other party's position.[177]

Using an objective standard to determine intent for contracts provides certainty, which is especially important given the role that contract law plays in facilitating business transactions. Under the doctrine of freedom of contract, whether the parties have entered into a legally enforceable agreement, a contract is and should be based on the intent of the parties involved. However, the intent we are concerned with is their objective intent. As a result, the essential requirements of contract reflect what we would expect hypothetical reasonable people to do if they intended to enter into an enforceable exchange of promises. Specifically, the essential requirements of contract are (1) offer and acceptance, (2) consideration, (3) legal abject, and (4) capacity. In particular, the specific requirements for the offer and acceptance reflect the objective intent idea.[178]

3. *Offer and Acceptance: Knowingly and Freely Agreeing*

In order for an offer to be made, three requirements must be satisfied. The first is intent to contract.[179] That is, the person making the offer, the offeror, must indicate a willingness to exchange one thing for another.[180] The second requirement is that the terms of the offer must be reasonably definite and certain.[181] Not only does this requirement indicate that the parties have fully developed their agreement, it also, as a practical matter, is necessary for the court to determine if a breach has occurred and, if so, the appropriate remedy.[182] The third requirement is that the offer must be communicated to the person who is trying to accept it.[183]

For example, if one party, a new car dealer, says to another, "I would like to sell this car," and the other party, after test driving the car and fully examining it, says, "I would like to buy it," no contract has yet been formed, because the parties have not specified the terms of the contract with reasonable certainty. That is, we do not know the price the seller is willing to accept or the price the buyer is willing to pay. Moreover, without this information, it seems unlikely that the parties have actually entered into a contract; rather, they seem to still be in the process of

177. *See id.*

178. *See e.g.*, *id.* § 2.1, at 26 (5th ed. 2003) (noting that essential prerequisite to formation of informal contract is agreement—mutual manifestation of assent to same terms).

179. *See id.*

180. *See id.* § 2.5, at 31 (citing RESTATEMENT (SECOND) OF CONTRACTS § 2) (specifying that parties must make commitment to do or refrain from doing some specified thing in future).

181. *See id.* § 2.13, at 75 (noting performances to be rendered by both parties must be reasonably certain).

182. *See id.*

183. *See id.* § 2.15, at 78 (adding that generally offeree must know of offer to accept).

negotiation.[184] However, a problem arises where the parties act as though they have a contract, even though their words or writings have not spelled out all of the terms.[185]

In the above example, this situation would be illustrated if the seller had delivered the car to the buyer, and the buyer accepted it notwithstanding the fact that they had not agreed on a specific price.[186] Given the actions of the parties, it is clear that they both thought that they had a contract; why else would the car have been delivered and accepted?[187] However, if the parties cannot agree on a specific price, for example, the seller sends a bill to the buyer, which is higher than the buyer is willing to pay, then a court may be called upon to determine the price.[188] In doing so, the court will seek to determine a reasonable price, one to which reasonable contracting parties would agree.[189] While reasonable people may differ somewhat concerning what constitutes a fair price, the MSRP or manufacturer's suggested retail price of the car is clearly not a reasonable price because it is set to be discounted and not paid.[190] The same is true of list prices in healthcare except that in healthcare the list price is set much higher as a percentage of reasonable value than in the case of an automobile.[191]

4. *Types of Contracts*

There are various types of contracts, including express, implied-in-fact, and implied-in-law.[192] All real contracts (as discussed below, implied-in-law contracts are not real contracts) must satisfy the same essential requirements discussed above. The difference between an express contract and an implied-in-fact contract is the way in which the parties create the contract.[193]

An *express contract* is created by the use of words, either written or oral.[194] For example, a typical type of express contract is an Agreement of Sale for real estate. This contract is formed by words, the words written in the Agreement of Sale and agreed to by the parties as evidenced by their signature at the end of the agreement. An *implied-in-law contract* is created

184. *See supra* notes 170–83 and accompanying text.

185. *See, e.g.*, CALAMARI & PERILLO, *supra* note 149, § 2.18, at 82 (5th ed. 2003).

186. *See, e.g.*, U.C.C. § 2-207(3) (2013) (describing "[c]onduct by both parties which recognizes the existence of a contract").

187. *See id.* § 2-207(1)–(2).

188. *See supra* notes 170–83 and accompanying text.

189. *See* CALAMARI & PERILLO, *supra* note 149, § 2.9, at 55 (5th ed. 2003) (discussing that courts fill gaps with terms that "comport[ ] with community standards or fairness and policy").

190. *See id.*

191. *See supra* notes 46–51 and accompanying text.

192. *See* CALAMARI & PERILLO, *supra* note 149, § 1.11, at 21 (4th ed. 1990) (discussing express, implied, and quasi contracts).

193. *See generally id.*

194. *See id.*

by the actions of the parties. For example, assume that Joe often visits Jane's Candy Store on his lunch break and usually buys a candy bar for one dollar. One day during his lunch break, Joe visits Jane's candy store and selects a candy-bar but observes that Jane is busy with a customer and has several other customers waiting. Joe catches Jane's attention shows her the candy bar and walks out of the store.[195] Based on these facts, Jane and Joe have entered into a contract pursuant to which Joe is legally obligated to pay for the candy bar.[196] The contract is based on the actions of the parties.[197] Specifically, Jane has made an offer by setting up her store, inviting the public to shop there, and placing the candy out for sale. Joe has accepted by taking the candy bar, acknowledging it to Jane and leaving the store. The requirement of consideration is satisfied because the candy is being exchanged for the money. There is nothing illegal about buying and selling candy, and as long as both parties have capacity to contract, all of the essential requirements for a contract have been satisfied and enforceable contract has been formed. In this case, the price, like the other terms of the contract, is established by the conduct of the parties.[198] Specifically, their prior dealings establish one dollar as the price.[199]

Of course, in a free market, Jane is free to change her prices at any time. However, in order for a changed price to become part of the contract, the other party must agree to it.[200] For example, if Jane had decided to raise her prices the night before Joe came into the store, would Joe be bound to pay the new price? The answer depends on whether the parties intended (objectively) the new price to apply.[201] In this case, if Jane had clearly marked the new price on the candy, then the new price would become part of the contract.[202] In other words, if the new price were clearly marked, a reasonable person in Joe's position would only have taken the candy bar if he were willing to pay the posted price.[203] On the other hand, if Jane had failed to take actions necessary to bring the new price to the attention of a reasonable person in the buyer's position, then the new price would not apply to the contract.[204]

An implied-in-law contract, also called a *quasi-contract,* is not a real contract and therefore does not depend upon these essential requirements for its enforceability.[205] The requirements for a quasi-contract are that a benefit has been given by one party to the other, the party receiving

195. *See* RESTATEMENT (SECOND) OF CONTRACTS § 4 illus. 2 (1981).
196. *See id.*
197. *See id.*
198. *See id.*
199. *See id.*
200. *See supra* notes 151–83 and accompanying text.
201. *See supra* note 171.
202. *See id.*
203. *See id.*
204. *See id.*
205. *See* CALAMARI & PERILLO, *supra* note 149, § 1.11, at 21.

the benefit has kept it—that is, refused to return it—and the court finds that allowing the party to keep the benefit without payment is unfair.[206] When a court decides that a quasi-contract exists, it will require the benefitted party to pay the amount necessary to remedy the unfairness. A quasi-contract is not a real contract because it is not based on the agreement of the parties; rather it is based on the equitable policy of furthering justice by preventing an unjust enrichment.[207] Moreover, since there is no agreement between the parties, the court has to provide the appropriate amount to be paid.[208] In most real contract cases, the court enforces the amount that the parties have agreed upon.

A fair question at this point is *what does all of this have to do with contracting for healthcare?* The principles just discussed apply to contracting in general and thus apply to contracting for healthcare. Moreover, some commentators get confused regarding the distinction between implied-in-law contracts and implied-in-fact contracts, especially in the healthcare context.[209] This may be due in part to a very common law school example involving healthcare services.

The law school example involves a doctor, licensed and in good professional standing under the applicable state law, happening upon an injured motorist on the side of the highway who has been in an accident.[210] The injured person is unconscious and in need of medical attention.[211] The doctor provides medical attention and then sends a bill to the patient.[212] May the doctor recover for the services he provided?[213] There is certainly no real contract between the doctor and the patient, since there was never any agreement between them due to the fact that the patient was unconscious. However, under the doctrine of implied-in-law contract, the doctor may recover the fair value of his services from the patient.[214] It is reasonable to assume that most patients in this situation would want the

206. *See id.* at 22 (describing quasi-contract as "the body of law to which we look for the reallocation of gains and losses between the parties").

207. *See id.*

208. *See id.*

209. *See id.*

210. *See id.* at 21.

211. *See, e.g.*, Barak D. Richman, Mark A. Hall & Kevin A. Schulman, *Overbilling and Informed Financial Consent—A Contractual Solution*, 367 NEW ENG. J. MED. 396 (2012) (recounting this example and mislabeling it as implied-in-fact contract).

212. *See id.* ("A staple law-school hypothetical illustrates the usual function of implied contracts: a physician encounters an unconscious stranger in the street who requires immediate medical attention. The physician promptly gives the stranger the requisite emergency care and later submits a bill for her services. Is she entitled to payment?").

213. *Id.*

214. *Id.*; *see also* CALAMARI & PERILLO, *supra* note 149, § 1.11, at 21 (providing similar example).

doctor's services, and therefore the patient will receive an unjust enrichment unless required to pay the doctor.[215]

When it comes to the question of how much the doctor should be paid, obviously this cannot be based on the agreement of the parties, so the court must fill in the amount.[216] In setting the price, the court uses the fair market value of the services because it is reasonable.[217] While this scenario may be relevant to an emergency department situation, it is not directly relevant to the balance-billing context. As discussed in the next Section, in typical balance-billing contracts, there is no question that the parties have entered into an agreement for medical services, pursuant to which the patient expects to pay for the medical services received.[218] The problem in the balance-billing context is that the parties have not specifically agreed to a price; hospitals typically require patients to sign some agreement pursuant to which the patient purports to agree to pay for the services received at the hospital's chargemaster, usual, or customary prices.[219] The problem, as discussed above, is that chargemaster prices are not usual or customary; they are exorbitant, grossly unfair, and are set to be discounted rather than paid.[220] In other words, in the balance-billing context, it is as though the parties have entered into a contract for medical services but have not specified a price.[221]

5. *Good Faith*

The common law recognizes an obligation of good faith in every contract.[222] Good faith is generally defined as honesty in-fact and the observance of reasonable commercial standards of fair dealing.[223] As a result, all parties to a contract must act in good faith in the performance and enforcement of every term of a contract. The obligation to act in good faith strengthens this Article's conclusion. That is, to pay the hospital based on its chargemaster or similar rates in cases where a patient has agreed by signing a "Statement of Financial Responsibility" or some other open-ended financial agreement at the time of admission to the hospital, the common law requires that this provision be performed and enforced in good faith. Good faith, as noted, requires observance of reasonable commercial standards of fair dealing. As a result, it would be inconsistent

---

215. *See id.*

216. *See, e.g.*, Temple Univ. Hosp., Inc. v. Healthcare Mgmt. Alts., Inc., 832 A. 2d 501, 508–11 (Pa. Super. Ct. 2003) (rejecting hospital's billed charges as proper measure of reasonable value under theory of unjust enrichment and instead used market value based on amount most patients actually paid for services).

217. *See id.*

218. *See infra* notes 227–29 and accompanying text.

219. *See supra* notes 1–17 and accompanying text.

220. *See supra* notes 43–48 and accompanying text.

221. *See supra* notes 178–90 and accompanying text.

222. *See* RESTATEMENT (SECOND) OF CONTRACTS § 205 (1981) (imposing duty of good faith on parties to contract).

223. *See id.*

with the common law of contracts to allow the hospital to recover based on its chargemaster rates. Rather, the observance of reasonable commercial standards of fair dealing would require that the hospital be permitted to recover no more than a reasonable price for its services.[224]

B. *Contracting for Healthcare*

1. *Promise to Pay Regular, List, Customary, or Chargemaster Rates*

In the balance-billing context, however, hospitals argue that the parties have established a formula to arrive at the ultimate price that the patient will pay for the services provided based on the hospital's chargemaster, and therefore the court must use this formula to determine the amount the patient owes.[225] But have the hospital and the patient really agreed on a price? The answer is clearly no, as the supposed chargemaster based formula is illusory; all aspects of it remain completely within the control of the hospital.[226] The problem of inexact price information at the time of contracting is not unique to the sale of healthcare, but healthcare is the only area in which the parties purport to use a blank check as payment.[227] For example, many professionals such as lawyers base their charges on an hourly rate, and it is often not possible to know at the time of engagement how many hours a matter will take. However, unlike the healthcare situation, the hourly rate is agreed to at the time of engagement and cannot be changed unilaterally by the lawyer. In contrast, hospitals often do not provide any price information, a copy of their chargemaster, or any other specific information to the patient at the time of contracting, and worse—hospitals retain the right to change their chargemaster rates at any time.[228]

This situation with the hospital more closely resembles that of an auto mechanic and a customer, when the customer takes his car to the mechanic because it is having some unspecified problem and the mechanic refuses to give an exact price to repair the car because the mechanic does not yet know how much repair will be necessary. If the car owner and the mechanic agree that the mechanic will work on the car and then determine the price, an express contract has been entered into, even though no specific price was agreed upon. However, the law does not grant the mechanic the right to charge whatever price they will; rather, the

224. *See supra* notes 1–2 and accompanying text.

225. *See e.g.*, Allen v. Clarian Health Partners, Inc., 980 N.E.2d 306, 310–11 (Ind. 2012) (stating chargemaster rates should be used to set contract prices).

226. *See, e.g.*, Rosenthal, *supra* note 37 (quoting Professor Glenn Melnick) (noting that hospitals may set chargemaster rates at any level they want and set them to "maximize revenue"); *see also* Nation, *Chargemaster Insanity*, *supra* note 12, at *1–18 (same).

227. *See* Nation, *Determining Fair and Reasonable Value*, *supra* note 7, at 426–32 (describing patient as giving hospital ability to unilaterally determine prices later).

228. *See* Anderson, *supra* note 3, at 786 (noting that hospitals may change their chargemaster rates at any time).

obligation of good faith discussed in Part IV(A)(5) requires that the mechanic exercise discretion to set the price in good faith;[229] that is, the mechanic must set a commercially reasonable price.[230] The mechanic cannot charge any more than the fair market value for the work the mechanic has performed. If the mechanic and customer cannot agree as to what a reasonable price is, then the court will set the price.[231]

Similarly, in the case of contracting for healthcare, the court must step in and provide the price for the parties since it is clear—based on both the conduct and words of the patient and the hospital—that they intend to have a contract. The court should reject the hospital's claim that it should be free to calculate the price, based on its elusive and ever-changing chargemaster rates, and should, consistent with the common law principles of contract, imply a term into the contract that requires the patient to pay a reasonable price for the services received.[232]

2. *Patients Required to Pay Prior to Treatment*

Where the hospital requires an upfront payment of all, or a certain percentage of the overall amount that the patient will be liable for, the parties have effectively established a specific price for the services that will be provided by the hospital. In this case, usually there is still a good argument that under contract law principles, the patient should be liable for no more than the reasonable value of the services received. This argument focuses on the common law requirement that contracts be knowingly and freely entered into.[233] Chargemaster prices are simply unreasonable; no reasonable person would knowingly and freely agree to pay such exorbitant rates.[234] That is, if the hospital had informed the patient that they were being asked to pay at least 2.5 times the amount that in-network patients pay and that the hospital gladly accepts this lower amount as full payment for the same exact services, no reasonable patient who had a choice would agree to pay the exorbitant chargemaster rates.[235] As a result, a good argument also remains that an agreement to pay chargemaster prices is unconscionable.[236] That is, the patient either did not make an informed choice to pay these prices or the patient had no

229. *See supra* notes 224–26 and accompanying text.

230. *See supra* notes 169–74 and accompanying text (discussing implied contracts that rely on reasonable terms).

231. *See* RESTATEMENT (SECOND) OF CONTRACTS § 204 (explaining courts infer terms such as price that are reasonable under circumstances).

232. *See infra* notes 240–62 and accompanying text.

233. *See supra* notes 150–92 and accompanying text.

234. *See supra* notes 43–48 and accompanying text.

235. *See* Bai & Anderson, *supra* note 10, at 923 (noting that in 2012 on average U.S. hospital charges were 3.4 times Medicare-allowable cost); *see also supra* notes 43–48 and accompanying text.

236. *See* Nation, *Obscene Contracts*, *supra* note 3 (arguing that contracts with hospitals pursuant to which patients purportedly agreed to pay chargemaster rates are unconscionable).

choice but to agree. If the patient did not have the ability to acquire the services elsewhere, the contract is procedurally unconscionable, and because the chargemaster rates are grossly unfair, the contract is also substantively unconscionable.[237] Thus, the contract should not be enforced.[238] Other common law theories might also be applicable in this context, including fraud in the inducement or undue influence. Ultimately, what gives all of these arguments strength is the truth of the basic facts underlying the transaction. Specifically, chargemaster prices are exorbitant and unfair and no sane person properly informed would agree to pay them.

3. *Determining the Reasonable Price of Healthcare*

I have written previously and in detail about how courts should determine the reasonable price for healthcare.[239] It is sufficient here to note that I strongly support using, when possible, a price set by the free market as a basis from which to establish the reasonable price. When this is not possible, it may be necessary to alternatively use the Medicare price as a basis.

4. *Rejecting Usual, Customary, and Necessary*

How do we determine the fair and reasonable price for healthcare? Hospitals would prefer to frame the question in its traditional manner: what is the usual and customary charge for the necessary medical services provided? Hospitals, of course, answer this question with reference to their chargemasters. They support this argument with a deceptive statement that "all patients are billed chargemaster rates," and therefore those rates are their usual and customary rates.[240] However, while all patients may be *billed* chargemaster rates, all patients are not expected to and do not in fact *pay* chargemaster rates.[241]

Moreover, the hospital knows, accepts, and plans on treating the majority of its patient's pain on less—*much* less—than chargemaster rates.[242] Chargemaster rates are a fiction, they are not set to be paid; they are set to be discounted.[243] Statements by hospital administrators to the contrary are disingenuous. Once the relevance of chargemaster rates are rejected, then the question can be phrased properly: what is the value established by the market of the services provided by the hospital? In other words, what is the fair market value of the services the hospital has provided? Thus, the focus is on what is actually paid and accepted for the services,

237. *See id.* at 128–32 (discussing cases in which courts did not enforce excessive hospital bills).

238. *See id.* at 132–34.

239. *See generally* Nation, *Determining Fair and Reasonable Value*, *supra* note 7.

240. *See id.* at 430.

241. *See id.*

242. *See id.*

243. *See id.*

and not on some phantom chargemaster-based billed charges that very few ever pay and that no one should be expected to pay.[244]

5. *Prices Based on Fair Market Value*

The fair and reasonable value of medical expenses must be based on the usual amount actually paid to the provider, not by the amount billed by the provider.[245] A hospital invoice of itemized billed charges at chargemaster rates is, when it comes to measuring fair value, a complete fiction and should not be used by courts or others to establish the fair and reasonable value of medical services.[246] Hospitals engage in price discrimination. They charge different amounts to different patients for the same exact services, depending upon the identity of the party paying the bill.[247] Government insurers pay the least, private insurers pay more, and self-pay patients, including individuals with private insurance that receive balance-bills, pay the most.[248] As a result of the perceived unfairness of this price discrimination, some commentators have called for an all-payer system.[249]

Under an all-payer system, various methods may be used to arrive at a particular price for a good or service.[250] For example, the government may set prices or each hospital may set its own price. However, regardless of how prices are set, once they are set, the price must be posted for public view and must apply to all patients without discrimination.[251] Although I recognize that it is unfair to allow hospitals to charge any patient chargemaster rates, I have argued against all-payer systems, because a part of what appears to be price discrimination is really market-driven discounting designed to purchase value from specific buyers.[252] Because of this, imposing an all-payer system would be disruptive to the market and create inefficiency.[253]

Hospitals charge lower rates to insurers because insurers provide certain benefits to hospitals.[254] These benefits include an increased volume of business through access to patients who are insured by the insurance company, assurance of quick and full payment of discounted charges from the insurance company, as well as marketing and advertising benefits that result from being listed as "in-network" by the insurance company.[255]

244. *See id.*

245. *See id.* at 457–65 (arguing that amount should be "determined by the lowest amount the hospital/provider accepts as full payment from government or private insurers").

246. *See id.* at 446–57 (detailing hospital pricing and relationship to insurers).

247. *See id.*

248. *See id.*

249. *See id.*

250. *See id.*

251. *See id.*

252. *See id.*

253. *See id.*

254. *See id.*

255. *See id.*

These benefits are valuable to hospitals and result in a portion, but only a portion, of the discount from chargemaster rates that private insurers receive. It is likely that the most important of these benefits to hospitals is the increased volume of business that results from entering into a contract with a large insurer.[256] After all, any patient who pays with a credit card provides an assurance of payment similar to that provided by the insurance company. Hospitals reason that because individuals do not bring the extra benefits such as an increased volume of business, which insurance companies bring to hospitals, individuals should pay more than the amount paid by insurers.[257]

The best way for courts and others to determine the fair and reasonable value of medical services is to start with the average amount the hospital would be paid by private insurers and then add to this an amount between 10% and 15% to account for the value of the benefits private insurers provide to hospitals.[258] It is important to note that the system I recommend, while it may be described as a form of price-fixing, is different than a government-controlled, all-payer system. The formula I suggest to determine a fair and reasonable price is based on a price freely set by the market.[259] No individual market participant, provider or insurer, may control the base.[260] Encouraging a freer and more transparent market for the sale of healthcare is the only approach that will result in appropriate pricing while simultaneously encouraging the development of the best healthcare in the world. More price-fixing will only make the problem worse.[261]

IV. CONCLUSION

The chargemaster rates set by hospitals are exorbitant because they are set to be discounted and not paid. As a result, it is grossly unfair to enforce these rates against any patient; this is why balance billing is such a pernicious problem. There are many reasons for the development of exorbitant chargemaster rates, but the most important reason they continue to cause problems is a lack of price transparency in contracting for healthcare. Regulation, as suggested in this Article, to provide price transparency and make sure that patients knowingly and freely agree to the price term in contracts for healthcare would help to eliminate the chargemaster/balance-billing problem.

---

256. *See id.*

257. *See id.*

258. *See id.* at 460–65 (suggesting method to determine "fair and reasonable value of medical services"); *see also* Anderson, *supra* note 3 (developing another method).

259. *See* Nation, *Determining Fair and Reasonable Value*, *supra* note 7 (explaining how method takes supply and demand principles into account).

260. *See id.*

261. *See id.*

However, courts can and should begin now to rein in the balance-billing problem. As this Article explains, the common law of contracts provides the necessary tools for courts to use to enforce fair and reasonable prices in contracts for healthcare. When an individual contracts with a hospital for healthcare, courts should use these common law tools to ensure that the patient is not liable to pay any more than a fair and reasonable price for the services received. Moreover, a fair and reasonable price should be based on the average amount the hospital receives from private insurers for the services provided to the patient.

MOVING BEYOND BITCOIN TO AN ENDOGENOUS THEORY OF DECENTRALIZED LEDGER TECHNOLOGY REGULATION: AN INITIAL PROPOSAL

CARLA L. REYES*

INTRODUCTION

THE world is captivated by the emergence of decentralized ledger technologies[1] such as the blockchain[2] and their increasingly widespread use to facilitate everything from decentralized payments[3] to a decentral-

* J.D., LL.M. in International & Comparative Law, Duke University School of Law; M.P.P., Duke University Terry Sanford School of Public Policy. I am grateful for feedback received from my colleagues at Perkins Coie LLP, who work in the trenches with me on these issues. I would also like to thank my husband, Mario C. Reyes, for his unwavering support.

1. This Article uses the terms *decentralized ledger technology* or *decentralized ledger technologies* to refer broadly to distributed network technology that (1) enables users to upload programs and to leave the programs to self-execute; (2) maintains a permanent and public record (ledger) of the current and past states of every program; (3) is decentralized; (4) uses public key cryptography for authentication; and (5) uses economic incentives to ensure that the network maintains the technology. *See* Vitalik Buterin, *Visions, Part 1: The Value of Blockchain Technology*, ETHEREUM BLOG (Apr. 13, 2015), https://blog.ethereum.org/2015/04/13/visions-part-1-the-value-of-blockchain-technology/ [https://perma.cc/F3RB-E4TA]. The term *decentralized ledger technology* as used in this Article is therefore broad enough to encompass the blockchain that underlies the bitcoin payments application, which is currently receiving incredible attention, but is not so narrow as to exclude other forms of the technology or other practical applications other than payments. The term therefore intends to capture other distributed technology—such as Ripple and Ethereum—and is not intended to reference any one application of the technology.

2. The "blockchain" is the decentralized ledger technology that powers the payments application known as bitcoin. At a high level, the blockchain "combin[es] peer-to-peer networks, cryptographic algorithms, distributed data storage, and a decentralized consensus mechanisms [sic]" to "provide[ ] a way for people to agree on a particular state of affairs and record that agreement in a secure and verifiable manner." Aaron Wright & Primavera De Filippi, *Decentralized Blockchain Technology and the Rise of* Lex Cryptographia 4–5, 5 & n.15 (Mar. 12, 2015) (unpublished manuscript), *available at* https://www.intgovforum.org/cms/wks2015/uploads/proposal_background_paper/SSRN-id2580664.pdf [https://perma.cc/K7HM-4GG7].

3. Although bitcoin, the decentralized virtual currency that may be exchanged through the blockchain, is currently receiving the most widespread attention, bitcoin itself is simply one payments-related application of the underlying technology. It is worth noting here that this Article adopts the term *decentralized virtual currency* to refer to decentralized payments applications because that term is now commonplace in the industry. However, the term *decentralized virtual currency* itself was somewhat controversial just several years ago. *Decentralized virtual currency* only became widely adopted as a term after the U.S. Department of Treasury's Financial Crimes Enforcement Network (FinCEN) released its guidance on virtual currencies on March 18, 2013. In that document, FinCEN defined *decentralized*

ized Internet of Things.[4] Software developers,[5] start-up companies,[6] and venture capitalists[7] all jubilantly declare the efficiency-maximizing,[8] cost-reducing,[9] and accessibility-enhancing[10] characteristics of decentralized cryptographic technologies. Academics predict that the blockchain and

---

virtual currency as "virtual currency (1) that has no central repository and no single administrator, and (2) that persons may obtain by their own computing or manufacturing effort." U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2013-G001, APPLICATION OF FINCEN'S REGULATIONS TO PERSONS ADMINISTERING, EXCHANGING, OR USING VIRTUAL CURRENCIES 5 (2013) [hereinafter FINCEN, VIRTUAL CURRENCY GUIDANCE], https://www.fincen.gov/statutes_regs/guidance/pdf/FIN-2013-G001.pdf [https://perma.cc/49PB-SU55]. Without assuming its accuracy, this Article adopts FinCEN's definition of decentralized virtual currency.

4. *See generally* VEENA PURESWARAN & PAUL BRODY, IBM INST. FOR BUS. VALUE, DEVICE DEMOCRACY: SAVING THE FUTURE OF THE INTERNET OF THINGS (2015), http://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03620usen/GBE03620USEN.pdf [https://perma.cc/HKG9-NPXR].

5. For example, the Open Mustard Seed Framework, being developed by ID3, combines blockchain technology with other decentralized technology with the intent "to provide a powerful new self-deploying and self-administering infrastructure layer for the Internet, which gives individuals control over their identities and their data, and which enables the formation and self-governance of Decentralized Autonomous Organizations, Authorities, and Enterprises for the creation and exchange of 'digital assets.'" *Open Mustard Seed (OMS) Framework*, ID3, https://idcubed.org/open-platform/platform/ [https://perma.cc/R7G9-FA63] (last visited Feb. 28, 2016).

6. *See, e.g.*, CHANGETIP, www.changetip.com [https://perma.cc/XLZ7-MB6D] (last visited Feb. 28, 2016) (providing "a love button for the Internet" through which users can engage in micro-tipping using bitcoin); COINBASE, www.coinbase.com [https://perma.cc/EF2Y-YYPP?type=image] (last visited Feb. 28, 2016) (providing Bitcoin wallet, sales, purchases, and exchange services); COINBEYOND, www.coinbeyond.com [https://perma.cc/GPX3-EU3W] (last visited Feb. 28, 2016) (providing bitcoin merchant services).

7. *See* Marc Andreessen, *Why Bitcoin Matters*, N.Y. TIMES, Jan. 21, 2014, http://dealbook.nytimes.com/2014/01/21/why-bitcoin-matters/?_r=0 [https://perma.cc/5MY8-378P].

> Bitcoin is the first practical solution to a longstanding problem in computer science called the Byzantine Generals Problem. . . .
>
> . . . .
>
> The practical consequence of solving this problem is that Bitcoin gives us, for the first time, a way for one Internet user to transfer a unique piece of digital property to another Internet user, such that the transfer is guaranteed to be safe and secure, everyone knows that the transfer has taken place, and nobody can challenge the legitimacy of the transfer. The consequences of this breakthrough are hard to overstate.

Id.

8. *See* BRIAN KELLY, THE BITCOIN BIG BANG: HOW ALTERNATIVE CURRENCIES ARE ABOUT TO CHANGE THE WORLD 163 (2015) ("Decentralization places the economic power into the hands of the citizens and removes many of the regulations that prevent capitalism from functioning efficiently.").

9. *See* J. ANTHONY MALONE, BITCOIN AND OTHER VIRTUAL CURRENCIES FOR THE 21ST CENTURY 47 (2014) ("Because there is no third party intermediary, Bitcoin transactions are substantially cheaper and quicker than traditional payment methods.").

10. *See id.* ("It has been estimated that half of the adults worldwide are unbanked due to barriers such as high costs, physical distance, and lack of infrastruc-

similar technologies will revolutionize the way people order their affairs and conduct transactions through the evolution of smart contracts,[11] decentralized autonomous organizations,[12] distributed property registries,[13] and distributed and secure data stores.[14] Although similarly captivated by these developing technologies, governments,[15] individual regulators,[16] and various policy makers[17] remain less optimistic that the good contributions of the technology will outweigh the way bad actors use the technology for illicit purposes.

ture and documentation. Bitcoin can exist wherever there is Internet or cell phone access and the requisite hardware." (footnote omitted)).

11. *See* Wright & De Filippi, *supra* note 2, at 10–12.

12. *See* William Mougayar, *What Does It Take to Succeed as a Decentralized Autonomous Organization?*, COINDESK (Feb. 21, 2015), http://www.coindesk.com/succeed-as-decentralized-autonomous-organization/ [https://perma.cc/3NJZ-34BY].

13. *See generally* Joshua A.T. Fairfield, *BitProperty*, 88 S. CAL. L. REV. 805 (2015) [hereinafter Fairfield, *BitProperty*].

14. *See* Wright & De Filippi, *supra* note 2, at 12–13.

15. For example, China's central bank completely prohibits financial institutions from handling fiat currency derived from bitcoin transactions. *See* Bloomberg News, *China Bans Financial Companies from Bitcoin Transactions*, BLOOMBERG (Dec. 5, 2013), http://www.bloomberg.com/news/articles/2013-12-05/china-s-pboc-bans-financial-companies-from-bitcoin-transactions [https://perma.cc/V5WD-6BQE]. Bolivia similarly bans the use of bitcoin and other decentralized virtual currencies. *See* Pete Rizzo, *Bolivia's Central Bank Bans Bitcoin*, COINDESK (June 19, 2014), http://www.coindesk.com/bolivias-central-bank-bans-bitcoin-digital-currencies/ [https://perma.cc/6DZY-PV3X]. Russia has considered whether to pass a similar ban on cryptocurrency. *See* Allen Scott, *Russia's Bitcoin Ban Expected in August; Expert Recommends Businesses 'Get Out,'* COINTELEGRAPH (Apr. 2, 2015), http://cointelegraph.com/news/113857/russias-bitcoin-ban-expected-in-august-expert-recommends-businesses-get-out [https://perma.cc/YUU2-HA3K]. Other examples of government skepticism of decentralized virtual currency abound.

16. *See, e.g.*, Jennifer Shasky Calvery, Dir., FinCEN, Prepared Remarks at the Association of Certified Anti-Money Laundering Specialists (ACAMS) 19th Annual International AML and Financial Crime Conference 2 (Mar. 18, 2014), transcript *available at* http://www.fincen.gov/news_room/speech/pdf/20140318.pdf [https://perma.cc/Y5E3-4ZYD] ("Indeed, the idea that illicit actors might exploit the vulnerabilities of virtual currency to launder money is not merely theoretical. We have seen both centralized and decentralized virtual currencies exploited by illicit actors.").

17. *See, e.g.*, CONF. OF STATE BANK SUPERVISORS, POLICY ON STATE VIRTUAL CURRENCY REGULATION 1 (Dec. 16, 2014), *available at* http://www.csbs.org/regulatory/ep/Documents/CSBS%20Policy%20on%20State%20Virtual%20Currency%20Regulation%20—%20Dec.%2016%202014.pdf [https://perma.cc/7SE4-FKVG] ("State regulators have determined that certain virtual currency activities raise concerns in the areas of consumer protection, marketplace stability, and law enforcement.").

Under the weight of various historical indicators[18] and in the wake of significant recent events,[19] regulators adopted an increasingly aggressive approach to enforcing existing regulations against the drastically new, different, and emerging technology.[20] The resulting barriers to entry[21] and climate of legal stigma[22] are stifling the nascent decentralized technology industry and preventing further innovation. In response, the decentralized virtual currency industry and other businesses interested in exploring the potential uses of decentralized technologies in commerce call for self-regulation.[23] Current literature, for its part, suggests a variety of regulatory models with each approach varying in light of the characteristics of the underlying technology that the commentator suggesting the model considers to threaten the most potential harm.[24]

History intimates that the self-regulatory approach is unlikely to sufficiently resolve the market failures that will ultimately allow illicit and fraudulent uses of decentralized technologies to occur. Meanwhile, the regulatory approaches suggested in the literature each impose a new regulatory barrier to entry even while trying to alleviate the inefficiencies of the

18. *See generally* Lawrence Trautman, *Virtual Currencies; Bitcoin & What Now After Liberty Reserve, Silk Road, and Mt. Gox?*, 20 RICH. J.L. & TECH. 13 (2014) (providing detailed historical review of illicit and criminal uses of bitcoin).

19. *See, e.g.*, Application and Affidavit for Seizure Warrant, In the Matter of the Seizure of the Contents of One Dwolla Account, No. 13–1162 SKG (D. Md. May 14, 2013) [hereinafter Application for Seizure Warrant], *available at* http://cdn.arstechnica.net/wp-content/uploads/2013/05/Mt-Gox-Dwolla-Warrant-5-14-13.pdf [http://cdn.arstechnica.net/wp-content/uploads/2013/05/Mt-Gox-Dwolla-Warrant-5-14-13.pdf]; Notice of Finding That Liberty Reserve S.A. Is a Financial Institution of Primary Money Laundering Concern, 78 Fed. Reg. 34169-01 (June 6, 2013) [hereinafter FinCEN, Notice of Finding].

20. *See, e.g.*, FINCEN, VIRTUAL CURRENCY GUIDANCE, *supra* note 3; Kashmir Hill, *Every Important Person in Bitcoin Just Got Subpoenaed by New York's Financial Regulator*, FORBES (Aug. 12, 2013), http://www.forbes.com/sites/kashmirhill/2013/08/12/every-important-person-in-bitcoin-just-got-subpoenaed-by-new-yorks-financial-regulator/.

21. Prior to the announcement of the Virtual Currency Guidance, state regulators remained silent about the potential for their existing laws to regulate decentralized virtual currency. The federal action seemed to spark a wave of activity at the state level as well. In fact, the state regulatory barriers to entry are now so high that they prevent even established players, such as Coinbase Inc., from entering all markets. *See* Jamie Redman, *Coinbase to Stop Service in Wyoming Due to 'Impractical' Regulations*, COINTELEGRAPH (June 4, 2015), http://cointelegraph.com/news/114470/coinbase-to-stop-service-in-wyoming-due-to-impractical-regulations [https://perma.cc/7ZKG-A9GM].

22. *See* Robin Sidel, *Bitcoin Evangelist Is Bound but Not Out*, WALL ST. J., June 30, 2014, http://www.wsj.com/articles/bitcoin-evangelist-is-bound-but-not-out-1404185518 [https://perma.cc/LA7X-UGZT].

23. *See, e.g.*, *Standards*, DIGITAL ASSET TRANSFER AUTHORITY, http://info.datauthority.org/wordpress/standards/ [https://perma.cc/FG2M-UPS2] (last visited Feb. 28, 2016).

24. *See, e.g.*, Omri Marian, *A Conceptual Framework for the Regulation of Cryptocurrencies*, 8 U. CHI. L. REV. DIALOGUE 53 (2015) (suggesting target anonymity to curb criminal uses); Wright & De Filippi, *supra* note 2 (suggesting do-nothing approach because encouraging innovation is tantamount).

current landscape. This Article suggests a different approach, one designed to encourage organic regulation that both addresses potential market and governance failures and takes into account the unique nature of the technology at issue.

This Article lays the foundation for adopting an endogenous theory of decentralized technology regulation. Drawing on theories of endogenous economic regulation, endogenous development, and functional financial regulation, this Article proposes that decentralized technologies, including the blockchain technology underlying decentralized payment systems such as bitcoin, are robust enough to support a theory of endogenous, technology-assisted regulation. Specifically, when this Article proposes an endogenous theory of regulation, it suggests that regulators undertake the dual task of enacting a law or regulation via statute and then implementing that statute through code by engaging in an iterative and cooperative process with the technologies' core developers and with consensus from the network, so that regulation is endogenously incorporated into the decentralized ledger technology and the applications running on top of the technology.

This Article makes the case for such an approach in five parts. Part I briefly introduces the decentralized ledger technology ecosystem, setting the stage for the rest of the Article by focusing on the protocol layer before addressing applications that might run on the protocol, such as bitcoin payments. Part II examines the current regulatory landscape, describing the "law lag" experienced by the decentralized ledger technology ecosystem and evaluating whether the choice between ex ante and ex post regulation is contributing to the lag. In light of the technology, its present uses, and the current regulatory landscape facing those uses, Part III proposes a set of criteria for constructing a regulatory framework for decentralized ledger technology and evaluates the alternative regulatory approaches that have been considered to date in light of those criteria. Ultimately, Part III identifies a regulatory lacuna common to both the present landscape and the alternative proposals presented in the literature. Part IV synthesizes an endogenous theory of decentralized ledger technology regulation from concepts found in economic regulation, international development, comparative law, and financial regulation literature.

In particular, Part IV argues that the regulatory approach with the potential to fulfill a majority of the criteria to a high degree is an endogenous, or functional, approach that simultaneously governs from within and without, and sidesteps the ex ante/ex post regulatory dichotomy by building compliance into the protocol as regulation-through-code. Linking the endogenous regulatory approach to the academic discourse regarding Lawrence Lessig's concept of code-as-law,[25] this Article proposes regulation that is endogenous at two levels: (1) as an iterative and cooperative process requiring participation of both regulators and industry actors approaching regulation design from a functional perspective and (2) as regulation implemented literally from within the decentralized ledger system as part of its operative code. Recognizing certain practical and theoretical challenges to implementing this endogenous theory, this Article concludes by identifying areas for further research and suggesting ways that successful implementation could disrupt core aspects of regulatory theory.

I. THE DECENTRALIZED LEDGER TECHNOLOGY ECOSYSTEM: A BRIEF INTRODUCTION

Commentators often compare the technical development and stages of public adoption of bitcoin to the development of the Internet.[26] The reality, however, is that the technical development and public adoption of bitcoin is better compared to the development and adoption of Internet Explorer or Apple TV (both of which are applications that run on the underlying Internet technology), while the development and adoption of the blockchain is akin to that of the Internet. Bitcoin is simply one application, a payments application, of the blockchain technology on which it runs. Further, the blockchain is just one of a variety of similar technologies often referred to as decentralized public ledgers or trustless public ledgers.[27] Because failure to appreciate these distinctions constitutes a core element in the regulatory difficulty facing entrepreneurs integrating decentralized ledger technology into their products and services, a brief introduction to the varied technology that makes up the decentralized ledger ecosystem, including decentralized public ledgers such as the blockchain, decentralized payments applications such as bitcoin, and other decentralized applications of the technology is warranted.[28] As the

25. *See* Lawrence Lessig, *Open Code and Open Societies: Values of Internet Governance*, 74 CHI.-KENT L. REV. 1405, 1408 (1999) [hereinafter Lessig, *Open Code and Open Societies*].

> The code of cyberspace—whether the Internet, or a net within the Internet—defines that space. It constitutes that space. And as with any constitution, it builds within itself a set of values and possibilities that governs life there. . . . And the design of code is something that people are doing. Engineers make the choices about how the world will be. Engineers in this sense are governors.

Id.

26. *See, e.g.*, Daniel Folkinshteyn, Mark Lennon & Tim Reilly, *A Tale of Twin Tech: Bitcoin and the WWW*, 10 J. STRATEGIC & INT'L STUD. 82 (2015); Clifton B. Parker, *The Promise and Peril of Bitcoin*, STAN. GRADUATE SCH. BUS. (Feb. 25, 2014), http://www.gsb.stanford.edu/news/headlines/promise-peril-bitcoin [https://perma.cc/DX2S-JNLY].

27. *See generally* Fairfield, *BitProperty*, *supra* note 13.

28. An exhaustive explanation of each of these elements is beyond the scope of this Article, and, as will be evident by the discussion that follows, this effort has

ultimate goal of this Article is to flesh out the initial contours of a regulatory approach that operates as organically and fluidly as the technology itself, this Section begins with an examination of decentralized ledgers and then describes various applications of that technology, such as bitcoin. This Section sets the tone for the regulatory approach ultimately proposed (which focuses on regulating the decentralized ledger technology, rather than the applications of that technology) and develops a common language for use throughout the rest of the Article.

A. *Decentralized Ledger Technologies, Including the Blockchain*

Decentralized ledger technologies "combin[e] peer-to-peer networks, cryptographic algorithms, distributed data storage, and [ ] decentralized consensus mechanisms" to enable "people to agree on a particular state of affairs and record that agreement in a secure and verifiable manner."[29] In other words, decentralized ledger technologies create "online lists, maintained by no one and available to everyone, [and] are maintained by a consensus protocol."[30] In the case of bitcoin, the decentralized public ledger is referred to as the blockchain.[31] The blockchain is "a chronological database of transactions recorded by a network of computers,"[32] which is encrypted and broken into smaller sets of aggregated transactions called "blocks."[33] A block is often described as "a container data structure,"[34] that groups transactions, marks them with a timestamp, and connects them to the previous block in the blockchain.[35] A new block of aggre-

already been eloquently undertaken by others. Instead, this Section explains the components of the decentralized ledger technology ecosystem that are important to understand before attempting to undertake the regulatory enterprise.

29. Wright & De Filippi, *supra* note 2, at 4, 5.

30. Joshua Fairfield, *Smart Contracts, Bitcoin Bots, and Consumer Protection*, 71 WASH. & LEE L. REV. ONLINE 35, 36 (2014) [hereinafter Fairfield, *Bitcoin Bots*].

31. *See* MALONE, *supra* note 9, at 35; Paul H. Farmer, Jr., Note & Comment, *Speculative Tech: The Bitcoin Legal Quagmire & the Need for Legal Innovation*, 9 J. BUS. & TECH. L. 85, 88–89 (2014) ("The Bitcoin peer-to-peer network that allows for miners to generate Bitcoins also serves as a public ledger for all Bitcoin transactions. A timestamp server records the time of creation of each Bitcoin and any other Bitcoin transaction within the network. The full record of transactions is called a block chain, a sequence of records composing a virtual ledger." (footnotes omitted)).

32. Wright & De Filippi, *supra* note 2, at 6 (citation omitted); *see also* PAUL VIGNA & MICHAEL J. CASEY, THE AGE OF CRYPTOCURRENCY: HOW BITCOIN AND DIGITAL MONEY ARE CHALLENGING THE GLOBAL ECONOMIC ORDER 124 (2015) ("The blockchain doesn't live on a single computer or server but . . . is shared around that community of computer owners, or *nodes*.").

33. Wright & De Filippi, *supra* note 2, at 6 (quoting *Blockchain*, BITCOIN FOUND. WIKI, https://en.bitcoin.it/wiki/Block_chain (last visited Mar. 1, 2015)) (internal quotation marks omitted); *see also* MALONE, *supra* note 9, at 35.

34. ANDREAS M. ANTONOPOULOS, MASTERING BITCOIN 160 (2015).

35. *See id.* at xix; Wright & De Filippi, *supra* note 2, at 7.

gated transactions will only be added to the ledger after "the computers on the network reach consensus as to the validity of the transaction."[36]

In the case of bitcoin, the method for reaching consensus is referred to as *mining*, a process of solving complex mathematical problems to validate the block.[37] Mining is an example of a proof-of-work consensus model.[38] A proof-of-work consensus model "require[s] the client requesting the service prove that some work has been done" in order to process the request.[39] Other decentralized ledger technologies employ different consensus models. For example, the Ripple protocol, a shared, public, distributed database,[40] validates transactions by creating a candidate list of transactions that is distributed to and voted on by a subset of trusted nodes, called the "unique node list."[41] The candidate set of transactions is validated and becomes part of the permanent, authoritative ledger once the "voting of server nodes reaches a consensus of 80% . . . ."[42] In a third

36. Wright & De Filippi, *supra* note 2, at 7; *see also* Fairfield, *BitProperty*, *supra* note 13, at 814 ("The Bitcoin protocol creates a ledger out of a series of groups of transactions, termed simply 'blocks,' which as a whole form a log of all transfers, termed the 'block chain.' The block chain is not maintained by any single entity, but instead relies on a mathematically innovative consensus model." (footnote omitted)).

37. *See* ANTONOPOULOS, *supra* note 34, at xx (describing miner as "[a] network node that finds valid proof of work for new blocks, by repeated hashing"); MALONE, *supra* note 9, at 36.

38. *See* PEDRO FRANCO, UNDERSTANDING BITCOIN: CRYPTOGRAPHY, ENGINEERING, AND ECONOMICS 101 (2015) ("To secure the blockchain—the distributed transaction database—Bitcoin requires proof-of-work to be performed on blocks of transactions following the Solution-Verification protocol.").

39. *Id.* at 100; *see also* ANTONOPOULOS, *supra* note 34, at xx (defining proof-of-work as "[a] piece of data that requires significant computation to find. In bitcoin, miners must find a numeric solution to the SHA256 algorithm that meets a network-wide target, the difficulty target."); Wright & De Filippi, *supra* note 2, at 7 n.29 (citing Joseph Bonneau et al., *SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies*, 36 SECURITY & PRIVACY (IEEE, San Fran., Cal.), May 18–20, 2015, *available at* www.jbonneau.com/doc/BMCNKF15-IEEESP-bitcoin .pdf.) ("The Proof of Work consensus mechanism requires that certain computers on the network (colloquially referred to as a [sic] 'miners') solve computationally-intensive mathematical puzzles, while others verify that the solution to that puzzle does not correspond to a previous transaction.").

40. *See* Bryant Gehring, *How Ripple Works*, RIPPLE (Oct. 16, 2014), https://ripple.com/knowledge_center/how-ripple-works/ [https://perma.cc/6LRU-E9JZ].

41. *See* Adrian Blundell-Wignall, *The Bitcoin Question: Currency Versus Trust-Less Transfer Technology* 15 (OECD Working Papers on Fin., Ins. & Private Pensions, No. 37, 2014), *available at* http://www.oecd.org/daf/fin/financial-markets/The-Bitcoin-Question-2014.pdf [https://perma.cc/N4D5-ZRZ4]; *see also* DAVID SCHWARTZ, NOAH YOUNGS & ARTHUR BRITTO, RIPPLE LABS, INC., THE RIPPLE PROTOCOL CONSENSUS ALGORITHM 3 (2014), *available at* https://ripple.com/files/ripple_consensus_whitepaper.pdf [https://perma.cc/X4WL-TQLA].

42. Blundell-Wignall, *supra* note 41, at 15; *see also* SCHWARTZ ET AL., *supra* note 41, at 4. At least one commentator has referred to this approach as a "Byzantine agreement system;" however, the creators of the Ripple protocol do not themselves adopt that terminology. *See generally* DAVID MAZIÈRES, STELLAR DEV. FOUND., THE STELLAR CONSENSUS PROTOCOL: A FEDERATED MODEL FOR INTERNET-LEVEL CONSENSUS (2015) (draft), *available at* https://www.stellar.org/papers/stellar-consensus-

consensus model, "voting rights depend on the amount of resources (e.g., a virtual currency) held by every computer connected to the network."[43] Researchers are presently pursuing the development of other consensus models as well.[44]

Regardless of the consensus method used, these decentralized public ledgers all share key qualities that make the technology revolutionary: they are distributed, run on peer-to-peer networks, and offer a public, permanent record management system "that does not require trust in other parties or in a central list authority, and is robustly resistant to falsification," employing known technology.[45] The technologies are also generally open-source and non-proprietary, meaning that no single person or entity controls the decentralized public ledgers.[46] Rather, the technology was developed, and is maintained and updated "by a worldwide collaborative community of volunteer[ ]" programmers.[47] The community can make improvements to the underlying code and alter the function of the blockchain protocol by proposing the code change and obtaining network consensus.[48]

B. *Decentralized Applications, Including Bitcoin and Smart Contracts*

There are a variety of applications for decentralized public ledger technologies such as the blockchain. Currently, the most well-known of those applications is decentralized virtual currency. Bitcoin is the decentralized virtual currency that runs on the blockchain.[49] "Bitcoins are computer files, similar to an mp3 or a text file and can be destroyed or lost just

---

protocol.pdf [https://perma.cc/J2AG-ZDMF]. A similar, but not identical, approach is adopted by the Stellar Consensus Protocol, which it terms a "Federated Byzantine Agreement" consensus model. *See id.*

43. Wright & De Filippi, *supra* note 2, at 7 n.30; *see also* Nicolas Houy, *It Will Cost You Nothing to "Kill" a Proof-of-Stake Crypto-Currency*, 34 ECON. BULL. 1038, 1040 (2014), *available at* http://www.accessecon.com/Pubs/EB/2014/Volume34/EB-14-V34-I2-P96.pdf [https://perma.cc/P6PH-4WZZ] (describing proof-of-stake as consensus mechanism in which "the expected reward for inserting transactions in the blockchain does not depend on the computational power of miners but on the amount of crypto-currency they already own").

44. *See, e.g.*, IDDO BENTOV, CHARLES LEE, ALEX MIZRAHI & MENI ROSENFELD, PROOF OF ACTIVITY: EXTENDING BITCOIN'S PROOF OF WORK VIA PROOF OF STAKE (2014), *available at* http://eprint.iacr.org/2014/452.pdf [https://perma.cc/9FAM-HDJA].

45. Fairfield, *BitProperty*, *supra* note 13, at 808.

46. *See* MALONE, *supra* note 9, at 34.

47. *Id.*

48. *See* Sarah Jeong, The Bitcoin Protocol as Law, and the Politics of a Stateless Currency 31 (May 8, 2013) (unpublished manuscript), *available at* http://www.modernmoneynetwork.org/sites/default/files/biblio/Bitcoin%20Protocol%20as%20Law.pdf [https://perma.cc/43S9-V75G] ("There is evidence that community consensus can result in the resolution of crisis.").

49. *See* MALONE, *supra* note 9, at 35.

like cash."[50] Just as an mp3 file can be stored either locally on the owner's computer or remotely through a cloud service, bitcoins can be stored either on the owner's computer or with an online service, called a wallet provider or an exchange.[51] When bitcoin owners refer to storing their bitcoins in a "wallet," the reference is to the public-key encryption technique used to secure bitcoins and bitcoin transactions.[52]

Specifically, a wallet is made of two mathematically related keys: a private key and a public key.[53] The public key is the outward facing destination address of the wallet, like a bank account number or an email address.[54] The private key functions as a PIN to a bank account or a password to an email address.[55] To execute a transaction, bitcoin owners use their private key to authorize the transfer of bitcoin to the public address representing the recipient's wallet.[56] The total amount of bitcoins is capped at 21 million,[57] and as a scarce resource with no government-affixed price, the value of an individual bitcoin is set by market forces.[58] Other decentralized virtual currencies run on other protocols layered on top of the blockchain[59] or run as payments applications on other protocols.[60]

---

50. Nikolei M. Kaplanov, Student Article, *Nerdy Money: Bitcoin, the Private Digital Currency, and the Case Against Its Regulation*, 25 LOY. CONSUMER L. REV. 111, 116 (2012).

51. *See id.*

52. *See id.* at 117.

53. *See id.*

54. *See id.*

55. *See id.* For a more complete description of how wallets are created, bitcoins are stored, and transactions are signed using the public-private key encryption pair, see Sarah Gruber, Note, *Trust, Identity, and Disclosure: Are Bitcoin Exchanges the Next Virtual Havens for Money Laundering and Tax Evasion?*, 32 QUINNIPIAC L. REV. 135, 141–45 (2013).

56. *See* Danton Bryans, Note, *Bitcoin and Money Laundering: Mining for an Effective Solution*, 89 IND. L.J. 441, 446 (2014).

57. *See id.*

58. *See id.* at 445; *see also* Reuben Grinberg, *Bitcoin: An Innovative Alternative Digital Currency*, 4 HASTINGS SCI. & TECH. L.J. 159, 160 (2012) ("Bitcoin is a digital, decentralized, partially anonymous currency, not backed by any government or other legal entity, and not redeemable for gold or other commodity.").

59. An example of this was Mastercoin, which was "an entirely new network of currencies, commodities and securities on top of Bitcoin itself." Vitalik Buterin, *Mastercoin: A Second-Generation Protocol on the Bitcoin Blockchain*, BITCOIN MAG. (Nov. 4, 2013), https://bitcoinmagazine.com/7961/mastercoin-a-second-generation-protocol-on-the-bitcoin-blockchain/ [http://perma.cc/53D8-MFKD]. Mastercoin sought to use the blockchain "as a protocol layer, on top of which new currency layers with new rules can be built without changing the foundation." *Id.* (quoting J.R. WILLETT, THE SECOND BITCOIN WHITEPAPER (2012) (draft), *available at* https://sites.google.com/site/2ndbtcwpaper/2ndBitcoinWhitepaper.pdf?attredirects=0) (internal quotation marks omitted).

60. The Ripple Protocol, for example, allows users to transfer "Balances," which represent fiat currency deposited with a gateway. Ripple describes it as follows:

Smart contracts represent another application of decentralized public ledger technology that is beginning to garner attention. Smart contracts can be thought of as self-executing transactions, or as "automated programs that transfer digital assets within the block-chain upon certain triggering conditions . . . ."[61] IBM has demoed a working prototype of a washing machine that orders its own detergent using the smart contracting applications of distributed ledger technology.[62] In practice, entrepreneurs rely on decentralized ledger technology, including the smart contracting features of that technology, to create decentralized marketplaces,[63] decentralized crowdfunding applications,[64] and distributed securities.[65] The more theoretical potential applications for distributed ledger technology include independently wealthy software,[66] zero-member LLCs,[67] and decentralized autonomous organizations,[68] among others.

Despite the potentially broad applications of decentralized ledger technology, regulatory efforts predominately center on payments applica-

> USD balances traded on the Ripple protocol are redeemable at the specific "gateway" from which the currency was issued. A gateway is the place where fiat money enters and exits the Ripple protocol. In practice, this can look very similar to traditional banks. However, a gateway can be any business that provides access to the Ripple protocol. Gateways can be banks, money service businesses, marketplaces, or any other financial institution.

Gehring, *supra* note 40.

61. Fairfield, *Bitcoin Bots*, *supra* note 30, at 38. Smart contracts have also been defined as "cryptographic 'boxes' that contain value and only unlock it if certain conditions are met." Wright & De Filippi, *supra* note 2, at 10 n.46 (quoting Vitalik Buterin, *A Next Generation Smart Contract & Decentralized Application Platform*, GITHUB (last edited Jan. 5, 2015), https://github.com/ethereum/wiki/wiki/White-Paper) (internal quotation marks omitted).

62. *See* Wright & De Filippi, *supra* note 2, at 15 n.71; *see also* PURESWARAN & BRODY, *supra* note 4.

63. Two examples of this include OpenBazaar and Bitmarkets. *See* Joon Ian Wong, *Bitmarkets Launches Decentralised Bitcoin Marketplace with Tor Support*, COINDESK (Dec. 8, 2014), http://www.coindesk.com/bitmarkets-launches-decentralised-bitcoin-marketplace-tor-support/ [https://perma.cc/5UQC-4U92].

64. An example includes BlockTrust, an "all-inclusive funding platform for Blockchain projects with project certification and ClearingHouse notarization services, along with multi-sig Bitcoin escrow." Diana Ngo, *BlockTrust Launches Crowdfunding Platform for Blockchain-Based Projects*, COINTELEGRAPH (Dec. 19, 2014), http://cointelegraph.com/news/113146/blocktrust-launches-crowdfunding-platform-for-blockchain-based-projects [https://perma.cc/8KUW-FLU6].

65. *See* Cade Metz, *Overstock Files to Offer Stock That Works Like Bitcoin*, WIRED (Apr. 27, 2015), http://www.wired.com/2015/04/overstock-files-offer-stock-works-like-bitcoin/ [https://perma.cc/B2H4-CK8M].

66. *See* Shawn Bayern, *Of Bitcoins, Independently Wealthy Software, and the Zero-Member LLC*, 108 NW. U. L. REV. 1485, 1492–95 (2014).

67. *See id.* at 1496–98.

68. For a description of the possible use case for a democratic autonomous organization, see generally ETHEREUM FRONTIER, https://www.ethereum.org/ [https://perma.cc/GG6Q-36WA] (last visited Feb. 28, 2016); *see also* Wright & De Filippi, *supra* note 2, at 15–17.

tions of decentralized ledger technologies, including bitcoin. Unfortunately, "[t]he danger is that this could imperil innovations built on top of the [blockchain] that do not fit a financial services model."[69] Indeed, such regulation has already made it more difficult for companies integrating decentralized ledger technologies into their products and services to obtain and maintain banking relationships that enable the company to function;[70] enter certain markets within the United States;[71] and develop alternative use cases for the technology without fear of inadvertently violating financial services laws.[72] To understand why the current regulatory landscape is marked by such dedication to taming bitcoin to the exclusion of all other uses of the technology, it is important to know the regulatory history of virtual currency generally.

II. THIS PRESENT DARKNESS: DECENTRALIZED LEDGER TECHNOLOGY REGULATION IN A CENTRALIZED AGE

At its core, the regulatory enterprise involves creating appropriate incentives to coax desired behavior out of market actors to address externalities otherwise naturally caused by the pursuit of maximizing private interests.[73] The ever-present dilemma of regulators involves whether to offer the incentive before (ex ante) or after (ex post) the activity at issue takes place.[74] In the context of the decentralized ledger technology ecosystem, determining which incentives to present and when to present them is further complicated by the "law lag," a term often used in law and technology literature to refer to the circumstances in which "existing legal provisions are inadequate to deal with a social, cultural or commercial context created by rapid advances in information and communication technology . . . ."[75] This Section evaluates the present state of decentral-

69. Fairfield, *BitProperty*, *supra* note 13, at 830 (citation omitted).

70. *See* Evander Smart, *Banks Handcuff Bitcoin Businesses in Opening Bank Accounts*, CRYPTOCOINSNEWS (Apr. 13, 2015), https://www.cryptocoinsnews.com/banks-handcuff-bitcoin-businesses-opening-bank-accounts/ [https://perma.cc/H7ZN-27YR].

71. *See* Redman, *supra* note 21.

72. *See* U.S. DEP'T OF TREASURY, FINCEN., ADMIN. RULING, FIN-2015-R001, APPLICATION OF FINCEN'S REGULATIONS TO PERSONS ISSUING PHYSICAL OR DIGITAL NEGOTIABLE CERTIFICATES OF OWNERSHIP OF PRECIOUS METAL (2015) [hereinafter FINCEN, DIGITAL CERTIFICATES], *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2015-R001.pdf [https://perma.cc/7JZ5-WGUL].

73. *See generally* Brian Galle, *In Praise of Ex Ante Regulation*, 68 VAND. L. REV. 1715 (2015).

74. *See generally* Donald Wittman, *Prior Regulation Versus Post Liability: The Choice Between Input and Output Monitoring*, 6 J. LEGAL STUD. 193 (1977) (exploring choice between regulating ex ante and punishing ex post).

75. Jeremy Pitt & Ada Diaconescu, *The Algorithmic Governance of Common-Pool Resources*, *in* FROM BITCOIN TO BURNING MAN AND BEYOND: THE QUEST FOR IDENTITY AND AUTONOMY IN A DIGITAL SOCIETY 130, 137–38 (John H. Clippinger & David Bollier eds., 2014), *available at* https://drive.google.com/file/d/0B2aVIexl9TLZSkNfTGx0U29WcHc/view [https://perma.cc/ZG27-M26S].

ized technology regulation in the United States. Despite the robustness of decentralized public ledger technology and its myriad potential uses, regulatory activity to date has focused on the decentralized virtual currency applications of such technology, with most prominent attention directed at bitcoin. This focus is partially explained by the fact that decentralized virtual currencies are the first and most widely adopted use of the underlying technology, receiving extensive media coverage and venture-capital attention. However, a full understanding of the present regulatory focus on decentralized virtual currency can only be achieved by looking at the recent regulatory history. To that end, this Section offers a brief overview of the regulatory developments in the decentralized ledger technology ecosystem to date.

This Section begins in the same spot in which regulators found themselves when bitcoin and the blockchain were introduced in 2009: addressing anti-money laundering and other criminal activity in the centralized technology context, including centralized virtual currencies. This Section then examines how regulators have attempted to apply the same regulatory lessons and policy priorities prevalent in the centralized context to decentralized ledger technology regulation. At each step, this Section considers the extent to which efforts are proceeding ex ante or ex post and queries whether preferences for one over the other vary in light of industry developments at the time. This Section concludes by examining the impact of the current regulatory landscape on the decentralized ledger technology industry and attempts to identify the regulatory gaps that leave actors in the industry facing regulatory uncertainty and railing against their very real experience with law lag.

A. *The Backdrop: Regulating Centralized Technology*

The largest virtual currency-related legal development before the open-sourced bitcoin protocol was released in 2009 was the prosecution of E-Gold, Ltd. (E-Gold). E-Gold was a centralized virtual currency allegedly backed by physical gold reserves.[76] E-Gold users could register for an account using an email address without verification of any identifying information.[77] As a result, E-Gold accounts existed under names such as "Mickey Mouse," "Anonymous Man," "bud wieser," and "No Name."[78] The E-Gold currency and the E-Gold platform became popular among criminals and money launderers, prompting various law enforcement ac-

76. *See* Catherine Martin Christopher, *Whack-A-Mole: Why Prosecuting Digital Currency Exchanges Won't Stop Online Money Laundering*, 18 LEWIS & CLARK L. REV. 1, 24 (2014); Stephen T. Middlebrook & Sarah Jane Hughes, *Regulating Cryptocurrencies in the United States: Current Issues and Future Directions*, 40 WM. MITCHELL L. REV. 813, 823 (2014).

77. *See* Christopher, *supra* note 76, at 24.

78. *Id.* (internal quotation marks omitted); *see also* Middlebrook & Hughes, *supra* note 76, at 824 (internal quotation marks omitted).

tions between 2005 and 2008.[79] Regulatory and enforcement activity after E-Gold remained publically quiet until 2013, but it is clear that law enforcement did not remain inactive. For example, the FBI authored an intelligence assessment, dated April 24, 2012, assessing the challenges faced in deterring illicit activity undertaken through the use of bitcoin.[80] The assessment evidences an initial shift in enforcement, from sole focus on centralized virtual currencies to a growing awareness of the unique regulatory and enforcement challenges posed by decentralized virtual currencies.

Then, nearly one year later, the federal government publicly unleashed the full weight of the quiet regulatory consideration it had been directing at decentralized virtual currencies. In March 2013, the Financial Crimes Enforcement Network of the United States Department of the Treasury (FinCEN) issued guidance on the application of the Bank Secrecy Act and its implementing regulations to virtual currencies (the Virtual Currency Guidance).[81] The Virtual Currency Guidance outlines the applicability of the existing federal anti-money laundering (AML) regime to convertible virtual currencies, including decentralized virtual currencies, and concludes that administrators and exchangers of such currencies are subject to the AML requirements to the extent that they transmit decentralized virtual currency or legal tender from one user to another, or from one location to another.[82] Shortly after FinCEN issued the Virtual Currency Guidance, the federal government announced two significant enforcement actions. First, on May 14, 2013, the Department of Homeland Security (DHS) seized a Dwolla account belonging to Mt. Gox, a leading Japan-based bitcoin exchange.[83] Second, a mere fourteen days later, FinCEN exercised its powers under Section 311 of the U.S. PATRIOT Act by designating the Costa Rican company Liberty Reserve a financial insti-

---

79. *See* Christopher, *supra* note 76, at 24. For details on the charges and guilty pleas entered, see United States v. e-Gold, Ltd., 550 F. Supp. 2d 82, 90 (D.D.C. 2008).

80. *See* FBI, INTELLIGENCE ASSESSMENT, (U) BITCOIN VIRTUAL CURRENCY: UNIQUE FEATURES PRESENT DISTINCT CHALLENGES FOR DETERRING ILLICIT ACTIVITY (2012), *available at* http://wired.com/images_blogs/threatlevel/2012/05/Bitcoin-FBI.pdf [https://perma.cc/3Q9C-9T3P].

81. *See* FINCEN, VIRTUAL CURRENCY GUIDANCE, *supra* note 3.

82. *See id.* at 4. For detailed descriptions of the Virtual Currency Guidance, see generally J. Dax Hansen, Carla L. Reyes & Joseph P. Cutler, *New FinCEN Guidance Changes Regulatory Landscape for Virtual Currencies and Some Prepaid Programs*, CYBERSPACE LAW. (Thomson/Legalworks, New York, N.Y.), Apr. 2013, at 12; Stephen T. Middlebrook & Sarah Jane Hughes, *Virtual Uncertainty: Developments in the Law of Electronic Payments and Financial Services*, 69 BUS. LAW. 263, 264–69 (2013).

83. *See* Application for Seizure Warrant, *supra* note 19. DHS alleged that Mt. Gox's U.S. subsidiary, Mutum Sigillum, LLC, operated an unlicensed money transmitting business in violation of 18 U.S.C. § 1960 (2012). *See id.*

tution of primary money laundering concern[84] and proposing the imposition of special measures against Liberty Reserve.[85]

The Virtual Currency Guidance and the Mt. Gox seizure paved the way for further regulation of decentralized technology. Key themes from each of these actions dating back to the beginning of the E-Gold prosecution in 2005 persist. In particular, the regulatory and law enforcement agencies active in the decentralized ledger technology industry continue to focus on centralized actors: companies with traditional centralized structures that just happen to offer products and services linked to a decentralized ledger technology. Further, like E-Gold, Mt. Gox, and Liberty Reserve, each of which processed very high volumes of transactions, regulators continue to target large-scale operations that offer the potential for high-impact regulatory or enforcement measures. Finally, and perhaps more importantly, the focus remains entirely on virtual currency applications of decentralized ledger technology—the decentralized versions of E-Gold and Liberty Reserve. Notably, this period of regulatory activity evidenced an intent to apply traditional AML and terrorist financing controls, a form of ex ante regulation in the payments space,[86] to virtual currencies, including decentralized virtual currencies. This intent was made most clear by the Virtual Currency Guidance and Liberty Reserve special measures, and continues to permeate the current regulatory landscape, as evidenced by the regulatory activity from 2013 to the present.

B. *The Current Landscape: Applying Centralized Themes to Decentralized Technology*

Although state regulators had not yet voiced an opinion on decentralized virtual currency regulation, shortly after the federal Mt. Gox and Liberty Reserve actions, the California Department of Business Oversight (DBO) sent letters to industry participants in order to gather information to help the DBO decide how to regulate decentralized virtual currencies and related activities.[87] The New York Department of Financial Services (NYDFS) quickly followed suit, issuing subpoenas to twenty-two companies

84. *See* FinCEN, Notice of Finding, *supra* note 19.

85. *See* FinCEN, Imposition of Special Measure Against Liberty Reserve S.A. as a Financial Institution of Primary Money Laundering Concern, 78 Fed. Reg. 34008 (proposed June 6, 2013) (to be codified at 31 C.F.R. pt. 1010); *see also* Jean-Jacques Cabou, J. Dax Hansen, Carla L. Reyes, Ashley A. Locke & Kieth Miller, *Legal Landscape for E-Currency Is Getting More Complicated*, LAW 360 (June 3, 2013), http://www.law360.com/articles/446907/legal-landscape-for-e-currency-is-getting-more-complicated [https://perma.cc/MN2R-6ECQ].

86. *See* Terri Bradford, Fumiko Hayashi, Christian Hung, Simonetta Rosati, Richard J. Sullivan, Zhu Wang & Stuart E. Weiner, *Nonbanks and Risk in Retail Payments: EU and U.S., in* MANAGING INFORMATION RISK AND THE ECONOMICS OF SECURITY 17, 35 (M. Eric Johnson ed., 2009).

87. *See, e.g.*, Letter from Paul T. Crayton, Senior Counsel, Cal. Dep't of Fin. Inst., to BitCoin Found. (May 30, 2013), *available at* http://www.scribd.com/doc/149335233/ca-state-cease-and-desist-may-30 [https://perma.cc/U5PK-G2BY].

in the decentralized technology ecosystem in August 2013.[88] The NYDFS issued the subpoenas as part of "an inquiry into the appropriate regulatory guidelines that it should put in place for virtual currencies."[89] Not to be outdone, the federal government also continued to apply regulatory pressure to the decentralized technology ecosystem. In October 2013, federal law enforcement agents shut down the Silk Road website and arrested its founder and administrators, who were charged with a wide range of crimes, including money laundering, drug trafficking, and computer hacking.[90]

The negative publicity surrounding bitcoin from the Silk Road events prompted the U.S. Senate to hold public hearings on virtual currencies in November 2013.[91] In her testimony during the Senate hearings, Jennifer Shasky Calvery, the Director of FinCEN, noted the "attributes that make virtual currency vulnerable to illicit use" but also recognized the "innovation virtual currencies provide."[92] Early 2014 brought additional federal criminal prosecutorial activity,[93] and the first state prosecutions related to decentralized virtual currency.[94] Furthermore, states continued to evaluate regulation: in January 2014, the NYDFS held a series of hearings on virtual currencies,[95] and U.S. state financial services regulators issued warnings on the risks of virtual currencies, either individually or in collab-

88. *See* Greg Farrell, *N.Y. Subpoenas Bitcoin Firms in Probe on Criminal Risk*, BLOOMBERG (Aug. 12, 2013), http://www.bloomberg.com/news/2013-08-12/n-y-regulator-subpoenas-firms-over-bitcoin-crime-risks.html [https://perma.cc/28L5-C38K].

89. N.Y. Dep't of Fin. Servs., Notice of Inquiry on Virtual Currencies (Aug. 12, 2013), *available at* http://www.dfs.ny.gov/about/letters/ltr130812_virtual_currency.pdf [https://perma.cc/BZ8J-EFMC].

90. *See* Kim Zetter, *How the Feds Took Down the Silk Road Drug Wonderland*, WIRED (Nov. 18, 2013), http://www.wired.com/2013/11/silk-road/ [https://perma.cc/UDN5-CFQ9].

91. *See Beyond Silk Road: Potential Risks, Threats, and Promises of Virtual Currencies Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 113th Cong. (2013), video *available at* http://www.hsgac.senate.gov/hearings/beyond-silk-road-potential-risks-threats-and-promises-of-virtual-currencies [https://perma.cc/3WHN-YN7V].

92. *Id.* (statement of Jennifer Shasky Calvery, Dir., FinCEN).

93. In January 2014, Charlie Shrem, a bitcoin exchange CEO, was arrested and charged with illegal money transmission and money laundering under federal laws. *See* Sidel, *supra* note 22.

94. In February 2014, two Florida-based users of a bitcoin local trading site were arrested and charged with illegal money transmission and money laundering under state laws. *See* United States v. Faiella, 39 F. Supp. 3d 544 (S.D.N.Y. 2014); Florida v. Reid, No. 14-002935 (Fla. Miami-Dade Cnty. Ct., Sept. 16, 2015); Florida v. Espinoza, No. 14-002923 (Fla. Miami-Dade Cnty. Ct., filed Feb. 7, 2014).

95. For more information, see generally N.Y. State Dep't of Fin. Servs., NYDFS Outlines Additional Details on Witnesses and Panels for Virtual Currency Hearing on January 28 and 29 in New York City (Jan. 23, 2014), *available at* http://www.dfs.ny.gov/about/panels_witnesses_virtual_currency_hearing.pdf [https://perma.cc/SP2Z-MNQH].

oration with other state regulators.[96] Other federal regulatory agencies, including the Internal Revenue Service[97] and the Federal Reserve, also waded into the decentralized technology morass.[98] Finally, FinCEN issued

96. *See generally* CONF. OF STATE BANK SUPERVISORS, MODEL STATE CONSUMER AND INVESTOR GUIDANCE ON VIRTUAL CURRENCY (Apr. 23, 2014), http://www.csbs.org/legislative/testimony/Documents/ModelConsumerGuidance—Virtual%20 Currencies.pdf [https://perma.cc/K63Y-TEU4]. To date, ten states have issued consumer guidance based on the CSBS model. *See* Press Release, Susan Bell, Comm'r, Alaska Dep't of Commerce, Cmty. & Econ. Dev., No. 14-014, Virtual Currency Investor Advisory Alert: What's in Your e-Wallet? (Apr. 30, 2014), *available at* https://www.commerce.alaska.gov/web/portals/3/pub/pr-14-014-virtua-currency-investor-advisory-alert-043014.pdf [https://perma.cc/S89T-MY4W]; Media Release, Jan Lynn Owen, Comm'r, Cal. Dep't of Bus. Oversight, California Department of Business Oversight Issues Advisory on Virtual Currencies (Apr. 30, 2014), *available at* http://www.dbo.ca.gov/Press/press_releases/VC_Advisory_PR_043014.pdf [https://perma.cc/M9QR-MJ36]; Advisory, Idaho Dep't of Fin., What's in Your e-Wallet? (Apr. 29, 2014), *available at* http://finance.idaho.gov/Securities/Documents/Virtual_Currency_Advisory.pdf [https://perma.cc/4VRL-EWQS]; Investor Alert, Ind. Sec'y of State Secs. Div., What's in Your e-Wallet? (May 12, 2014), *available at* http://www.in.gov/sos/files/Virtual_Currency_IN.pdf [https://perma.cc/M9ZP-KRBE]; Advisory Notice, Md. Comm'r of Fin. Regulation, Virtual Currencies: Risks for Buying, Selling, Transacting, and Investing, No. 14-01 (Apr. 24, 2014), *available at* http://www.dllr.state.md.us/finance/advisories/advisoryvirtual.pdf [https://perma.cc/U8XZ-3GCG]; News Release, Mo. Sec'y of State, Investor Alert: Kander Cautions Missouri Investors on Bitcoin (Apr. 24, 2014), *available at* http://www.sos.mo.gov/news.asp?id=1383 [https://perma.cc/AU9Y-RFLY]; Press Release, Nev. Dep't of Bus. & Indus., Nevada Financial Institutions Division Issues Consumer and Investor Guidance on Virtual Currency (Apr. 25, 2014), *available at* http://www.fid.state.nv.us/GuidanceVirtualCurrency.pdf [https://perma.cc/NGQ8-5XP6]; Press Release, N.M. Regulation & Licensing Dep't, State Regulations Alert Consumers About Risks Involved with Investing in Virtual Currencies (Apr. 30, 2014), *available at* http://www.rld.state.nm.us/uploads/PressRelease/28b4bcdf4 f8d46899c723b2b76f16 b5b/Virtual_Currency_Warning.pdf [https://perma.cc/8PYK-J6KE]; Press Release, Pa. Dep't of Banking & Secs., Pennsylvania Banking and Securities Department Urges Consumers, Businesses and Investors to Be Cautious with New Virtual Currencies (May 19, 2014), *available at* http://www.prnewswire.com/news-releases/pennsylvania-banking-and-securities-department-urges-consumers-businesses-and-investors-to-be-cautious-with-new-virtual-currencies-259822731.html [https://perma.cc/PH8M-XWFT]; Press Release, George Althoff, Commc'ns Dir., Wis. Dep't of Fin. Insts., State Agency Warns Consumers to Be Cautious with Virtual Currencies (Apr. 30, 2014), *available at* https://www.wdfi.org/newsroom/press/2014/DFIVirtualCurrencyAdvisory.pdf [https://perma.cc/5LKC-C6PS].

97. In May 2014, the Internal Revenue Service (IRS) of the Department of Treasury issued a notice classifying bitcoin and similar virtual currencies as property. *See* I.R.S. Notice 21-16, 2014-21 C.B. 938 (May 19, 2014), *available at* https://www.irs.gov/pub/irs-irbs/irb14-21.pdf [perma.cc/3FDT-56DT].

98. The Federal Reserve, for its part, used 2014 to evolve its thinking regarding bitcoin and other decentralized virtual currencies. On February 27, 2014, Federal Reserve Chair Janet Yellen, in testimony before the Senate Banking Committee, opined that the Federal Reserve had no role in regulating bitcoin because it is developing outside of the banks that the Federal Reserve has the authority to regulate. *See Semiannual Monetary Policy Report to the Cong.: Before the H. Comm. on Fin. Servs.*, 113th Cong. (2014) (testimony of Janet Yellen, Chair, Bd. of Governors of the Fed. Reserve Sys.), *available at* http://www.federalreserve.gov/newsevents/testimony/yellen20140211a.htm [https://perma.cc/RD87-VKDG]. At its

a series of administrative letter rulings in 2014 intended to clarify the application of the Virtual Currency Guidance to various business models used in the decentralized technology ecosystem.[99] By the end of 2014, FinCEN had addressed the status of miners,[100] software development,[101] investment activity,[102] virtual currency trading platforms,[103] and a virtual currency payment system.[104]

The year 2015 brought similar developments. At the state level, the NYDFS finalized its BitLicense regime,[105] several other states took a wait-

May 9, 2014, quarterly meeting, however, the Federal Reserve concluded that bitcoin did not presently pose a threat to the system, but that "regulation is advisable" to address consumer protection and illicit use. *See* Fed. Advisory Council & Bd. of Governors, Record of Meeting 11 (May 9, 2014), *available at* http://www.federalreserve.gov/aboutthefed/fac-20140513.pdf. The Federal Reserve also noted that "one area of focus should be [b]itcoin's circumvention of currency controls." *Id.* at 10.

99. *See generally* Joseph Cutler, J. Dax Hansen & Carla L. Reyes, *Admin Rulings: FinCEN Clarifies Money Transmitter Exemptions for Virtual Currency and Payment-Related Services*, JD SUPRA BUS. ADVISOR (June 10, 2014), http://www.jdsupra.com/legalnews/admin-rulings-fincen-clarifies-money-tr-92506/ [https://perma.cc/XF7C-4UBC]; Joseph Cutler, J. Dax Hansen & Carla L. Reyes, *New FinCEN Guidance on Virtual Currency Transactions*, JD SUPRA BUS. ADVISOR (Feb. 12, 2014), http://www.jdsupra.com/legalnews/new-fincen-guidance-on-virtual-currency-14922/ [https://perma.cc/L7GC-RFE7]; Carla L. Reyes, *FinCEN Issues New Rulings Covering Virtual Currency Exchanges and Payment Processors*, JD SUPRA BUS. ADVISOR (Oct. 30, 2014), http://www.jdsupra.com/legalnews/fincen-issues-new-rulings-covering-virtu-44222/ [https://perma.cc/HT3U-YB7F].

100. *See* U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2014-R007, APPLICATION OF MONEY SERVICES BUSINESS REGULATIONS TO THE RENTAL OF COMPUTER SYSTEMS FOR MINING VIRTUAL CURRENCY (Apr. 29, 2014), *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R007.pdf [https://perma.cc/J6CG-U2DD]; U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2014-R001, APPLICATION OF FINCEN'S REGULATIONS TO VIRTUAL CURRENCY MINING OPERATIONS (Jan. 30, 2014), *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R001.pdf [https://perma.cc/4XNA-7R3L].

101. *See* U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2014-R002, APPLICATION OF FINCEN'S REGULATIONS TO VIRTUAL CURRENCY SOFTWARE DEVELOPMENT AND CERTAIN INVESTMENT ACTIVITY (Jan. 30, 2014), *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R002.pdf [https://perma.cc/AB7K-22AW].

102. *See id.*

103. *See* U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2014-R011, REQUEST FOR ADMINISTRATIVE RULING ON THE APPLICATION OF FINCEN'S REGULATIONS TO A VIRTUAL CURRENCY TRADING PLATFORM (Oct. 27, 2014), *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R011.pdf [https://perma.cc/KPH2-TMSA].

104. *See* U.S. DEP'T OF TREASURY, FINCEN, ADMIN. RULING, FIN-2014-R012, REQUEST FOR ADMINISTRATIVE RULING ON THE APPLICATION OF FINCEN'S REGULATIONS TO A VIRTUAL CURRENCY PAYMENT SYSTEM (Oct. 27, 2014), *available at* http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R012.pdf [https://perma.cc/XC62-KEMF].

105. *See* N.Y. COMP. CODES R. & REGS. tit. 23, §§ 200.1–200.22 (2015), *available at* http://www.dfs.ny.gov/legal/regulations/adoptions/dfsp200t.pdf [https://perma.cc/L58E-QPQP].

and-see approach,[106] and state legislatures,[107] the Uniform Law Commission,[108] and the Conference of State Banking Supervisors[109] undertook legislative drafting efforts with a best practices approach in mind. FinCEN issued yet another administrative ruling, this time examining the ways in which digitally issued certificates representing ownership in commodities (such as gold) intersect with money transmission regulation.[110] The Commodity Futures Trading Commission (CFTC) concluded that bitcoin is a commodity[111] and took enforcement action against two bitcoin-related businesses for various violations of the Commodity Exchange Act.[112] The CFTC also issued an order of temporary registration as a swap exchange facility to a third bitcoin-related business, indicating that the CFTC is actively considering and acting upon applications.[113] The Security and Exchange Commission (SEC) refrained from conclusively ruling how it will

106. *See, e.g.*, *Sellers of Checks*, WIS. DEP'T FIN. INSTS., https://www.wdfi.org/fi/lfs/soc/ [https://perma.cc/P4RW-FE3M] (last visited Mar. 1, 2016) ("The division is unwilling, at this time, to license companies to transmit virtual currency.").

107. *See, e.g.*, *Bitcoin Week in Review 06.27.15 - 07.10.15*, PERKINS COIE (July 10, 2015), http://www.virtualcurrencyreport.com/2015/07/bitcoin-week-in-review-06-27-15-07-10-15/ [https://perma.cc/URB8-UAQX] (discussing California's proposed bill AB-1326).

108. *See generally* NAT'L CONF. OF COMM'RS ON UNIF. STATE LAWS, REGULATION OF VIRTUAL CURRENCIES ACT (2015) (draft), *available at* http://www.uniformlaws.org/shared/docs/regulation%20of%20virtual%20currencies/2015oct_RVCA_Mtg%20Draft.pdf [https://perma.cc/MH24-SHJ9]; *Regulation of Virtual Currencies*, UNIF. L. COMM'N, http://www.uniformlawcommission.com/Committee.aspx?title=Regulation%20of%20Virtual%20Currencies [https://perma.cc/3VLA-ZHTS] (last visited Mar. 1, 2016).

109. *See generally* CONF. OF STATE BANK SUPERVISORS, STATE REGULATORY REQUIREMENTS FOR VIRTUAL CURRENCY ACTIVITIES: CSBS MODEL REGULATORY FRAMEWORK (2015), *available at* https://www.csbs.org/regulatory/ep/Documents/CSBS-Model-Regulatory-Framework(September%2015%202015).pdf [https://perma.cc/AV3R-ZZA8].

110. *See* FINCEN, DIGITAL CERTIFICATES, *supra* note 72. The recent CFTC Complaint against MintCo LLC indicates that the CFTC may also consider itself to have jurisdiction over such digital commodity certificates, identified by the Virtual Currency Guidance as a form of convertible virtual currency. *See* U.S. Commodity Futures Trading Comm'n v. MintCo LLC, 15-cv-61960-BB (S.D. Fla. filed Sept. 16, 2015), *available at* http://www.cftc.gov/idc/groups/public/ @lrenforcementactions/documents/legalpleading/enfmintcocomplaint091615.pdf [https://perma.cc/Z39L-4W3Q].

111. *See* Coinflip, Inc., CFTC, Docket No. 15-29, at 3 (Sept. 17, 2015) (order), *available at* http://www.cftc.gov/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfcoinfliprorder09172015.pdf [https://perma.cc/WA4W-ZCUJ].

112. *See id.*; TeraExchange LLC, CFTC Docket No. 15-33 (Sept. 24, 2015) (order), *available at* http://www.cftc.gov/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfteraexchangeorder92415.pdf [https://perma.cc/PW6U-ZN3C].

113. *See* LedgerX LLC, CFTC Staff Letter, [Current Decisions] Comm. Fut. L. Rep. (CCH) ¶33,537 (Sept. 8, 2015), *available at* http://www.cftc.gov/idc/groups/public/@lrlettergeneral/documents/letter/15-49.pdf [https://perma.cc/PD9Y-FDM9].

treat bitcoin, despite efforts from entrepreneurs to force the SEC's hand by filing shelf registrations for offerings that integrate blockchain technology.[114] Meanwhile, prosecutions for criminal activity perpetrated to some degree through the use of bitcoin moved forward,[115] and several consent orders were entered against decentralized technology companies for failure to properly comply with the Bank Secrecy Act regulations.[116]

Although federal regulatory agencies, including FinCEN, have carried out their intent to impose ex ante anti-money laundering regulations on decentralized virtual currency systems, industry developments have revealed that compliance with those regulations through the mechanics of the decentralized ledger technology is not always intuitive. As a result, where practical difficulties or other technology mismatches arise, regulators also used ex post criminal prosecutions to heighten compliance incentives. Even as federal authorities expanded the tools in their regulatory toolbox, they also expanded the policy priorities away from a singular focus on money-laundering, terrorist financing, and identity verification towards a more comprehensive set of payments-related issues, including privacy and security, tax compliance, and the potential for use of unfair and deceptive businesses practices in the industry. State regulatory activity also added new policy concerns to the mix, with a primary focus on consumer protection. Although other agencies are clearly apprised of the possibility that decentralized ledger technology may touch on areas within their regulatory purview, they have remained silent.

---

114. *See* Romain Dillet, *Winklevoss Twins Eat, Sleep, Breathe Bitcoin*, TECHCRUNCH (May 6, 2015), http://techcrunch.com/2015/05/06/winklevoss-twins-eat-sleep-breathe-bitcoin/ [https://perma.cc/22DT-2FAA] (discussing Winklevoss's efforts to build exchange traded fund to allow trading of bitcoins on NASDAQ OMX); Cade Metz, *Overstock Will Issue a Private Bond Powered by Bitcoin Tech*, WIRED (June 5, 2015) http://www.wired.com/2015/06/overstock-will-issue-private-bond-powered-bitcoin-tech/ [https://perma.cc/6G6E-KLF3] (discussing Overstock .com's intention to offer $25 million private bond using blockchain). Although the *Shavers* case involved bitcoin, the ruling did not determine whether or how bitcoin would be regulated by the SEC and focused instead on the criminal Ponzi scheme Shavers operated. The fact that Shavers used bitcoin to operate the Ponzi scheme did not change the underlying nature of the criminal activity. *See* SEC v. Shavers, No. 4:13–CV–416, 2013 WL 4028182 (E.D. Tex. Aug. 6, 2013). For arguments that bitcoin should be considered an investment contract or a security, see Derek A. Dion, Note, *I'll Gladly Trade You Two Bits on Tuesday for a Byte Today: Bitcoin, Regulating Fraud in the E-Conomy of Hacker-Cash*, 2013 J.L. TECH. & POL'Y 165, 193–94; Farmer, *supra* note 31, at 98–104.

115. *See* Joel Rosenblatt, *Second U.S. Agent Agrees to Plead Guilty to Bitcoin Theft*, BLOOMBERG (June 22, 2015), http://www.bloomberg.com/news/articles/2015-06-23/second-u-s-agent-agrees-to-plead-guilty-to-bitcoin-theft [https://perma.cc/K9U9-BEV9]; Sam Thielman, *Silk Road Operator Ross Ulbricht Sentenced to Life in Prison*, GUARDIAN (May 29, 2015), http://www.theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced [https://perma.cc/99MH-TPE8].

116. *See* U.S. DEP'T OF JUSTICE, SETTLEMENT AGREEMENT BETWEEN U.S. ATT'Y N. DIST. OF CAL. AND RIPPLE LABS, INC. (2015) [hereinafter DOJ, RIPPLE SETTLEMENT AGREEMENT], *available at* http://www.justice.gov/file/421626/download [https://perma.cc/DHF4-AE39].

The collective impact of this regulatory history is that new entrants into the decentralized ledger technology industry must tread carefully, for fear of triggering one or more known or unknown regulatory priorities by introducing a new and innovative application of the underlying technology. The 2015 FinCEN ruling and the SEC's silence on the Winkelvoss endeavor is a clear example of the regulatory lacuna and its effects. Even when given the opportunity to comment on regulation of the underlying technology, the SEC has not done so, and when presented with a novel application of the technology, FinCEN shoehorned it into the payments regulatory landscape. In other words, the lacuna in the present regulatory landscape is characterized by the lack of regulation addressing distributed ledger technology, and the result is that, when a new use emerges, the application may be subject to payments laws that are ill-suited to the issues presented by the technology.

C. *The Impact of the Current Landscape: Confusion, High Risk, and Disincentives to Innovation*

The policy priorities related to bitcoin and other decentralized virtual currencies are now relatively clear. The implementation of those policies through existing law remains a mystery in several areas, however. As a result, a common critique leveled at the regulatory efforts related to decentralized ledger technologies described above is that it evidences what appears to be development and implementation absent a master plan. To date, the United States has made great efforts to understand the unique characteristics and potential risks of virtual currency. Nonetheless, virtual currency hearings have not yielded any formal recommendations or guidance. As a result, regulatory bodies, courts and state legislatures have acted independently resulting in a regulatory mishmash of guidance, clarification, extension and ongoing discussion.[117]

The reality, however, is that the present regulatory approach to decentralized virtual currencies reflects mainstream approaches to financial regulation generally, using a combination of ex ante and ex post regulation to mitigate systemic risk in the financial system.[118] Leading literature on approaches to financial regulation offers this insight:

> Complete ex ante financial regulation, whereby regulators prevent every failure, is [ ] a futile goal. And even if it were feasible, it would not necessarily be desirable. Ex ante regulation can provide an incentive for regulatory arbitrage. Furthermore, any ex ante regulation that attempts to prevent all financial failures may end up being too chilling, thereby dampening economic growth. Ex post remedies will therefore always be needed to try

117. Kevin V. Tu & Michael W. Meredith, *Rethinking Virtual Currency Regulation in the Bitcoin Age*, 90 WASH. L. REV. 271, 304 (2015).

118. *See generally* Steven L. Schwarcz, *Systemic Risk*, 97 GEO. L.J. 193 (2008) [hereinafter Schwarcz, *Systemic Risk*].

> to prevent financial failures—when they inevitably occur—from spreading and becoming systemic.[119]

The regulatory history of decentralized virtual currency to date reads like a textbook example of this commentary: a heavy emphasis on ex ante attempts to prevent financial harm in the decentralized virtual currency industry, complemented by ex post prosecutions of harmful activity when failures nevertheless occurred.

Nontheless, the heavy emphasis on ex ante efforts to prevent decentralized virtual currency-related market failures when coupled with the innovative nature of the underlying technology has led to confusion and uncertainty, while the ex post efforts to punish and correct failures highlight the high risk faced by anyone connected to the decentralized ledger technology industry, whether working with a payments applications like virtual currency or not. Even a cursory review of the state of the industry supports this position. There presently exists no fully licensed, U.S.-based decentralized virtual currency exchange.[120] Any entity offering a product or service connected in any way to decentralized public ledger technology finds difficulty obtaining and maintaining banking relationships to enable the operation of its business.[121] Furthermore, the regulatory history is replete with evidence of the practical difficulties resulting from shoehorning decentralized ledger technologies, related applications, and the businesses that offer products and services related to them into a regulatory scheme first designed for centralized technologies. Compliance with the Funds

119. Steven L. Schwarcz, Keynote Address at Chapman Law Review Symposium: Ex Ante Versus Ex Post Approaches to Financial Regulation (Jan. 28, 2011), *in* 15 CHAP. L. REV. 257, 260 (2011) (footnotes omitted).

120. Coinbase launched its exchange to claims of being the first licensed bitcoin exchange in the United States. *See* Davey Alba, *Coinbase Opens First Licensed Bitcoin Exchange in the US*, WIRED (Jan. 26, 2015), http://www.wired.com/2015/01/coinbase-opens-first-licensed-bitcoin-exchange-u-s/ [https://perma.cc/3HME-UXLY]. However, in reality, Coinbase remains unable to obtain the requisite licenses in all states, and even pulled out of certain states where the licensure process proved too burdensome. *See* Redman, *supra* note 21. Similarly, itBit initially claimed that the New York trust charter it obtained from the New York Department of Financial Institutions enabled it to lawfully operate an exchange service in all fifty states. *See* Cade Metz, *NY Backs Bitcoin Exchange. But It May Not Fly in California*, WIRED (May 8, 2015), http://www.wired.com/2015/05/new-york-backs-bitcoin-exchange-may-not-fly-california [https://perma.cc/LG49-FA85]. However, the analyses of practitioners and academics alike have failed to prove out itBit's theory and instead openly call it into question. *See id.* (quoting Carol Van Cleef, Partner, Manatt, Phelps & Phillips, as saying that charter "is not necessarily going to be a blank pass to offer services in all states"); Houman Shadab, *What itBit's Banking Law Charter Really Means*, COINDESK (May 17, 2015), http://www.coindesk.com/in-itbit-we-trust/ [https://perma.cc/8FL5-PQ9R] ("Nonetheless, there seems to be some uncertainty about whether itBit—a banking law trust—automatically qualifies to do business without a money transmitter license in certain states.").

121. *See* Smart, *supra* note 70.

Transfer and Funds Travel Rules[122] and the valuation of virtual currency assets when applying for money transmission licenses under statutes that contemplate only fiat currency[123] offer two examples of such difficulties. If the regulatory approach to decentralized ledger technologies to date is a clear example of traditional financial regulation, and the result is confusion, uncertainty, and disincentive for innovation, the urgent question facing regulators, consumers, and entrepreneurs alike is whether an alternative approach that nevertheless remains consistent with leading legal regulatory theory is possible.

III. EVALUATION OF CURRENT ALTERNATIVE REGULATORY PROPOSALS ECHOES THE GAP IN THE CURRENT APPROACH

Recognizing the inadequacy of the current regulatory approach to the decentralized technology industry,[124] the current literature suggests several alternative proposals. The proposals can be grouped into three categories. First, various proposals attempt to apply existing law to bitcoin and other decentralized virtual currencies by shoehorning the decentralized payment applications into a specific type of asset or property category. Second, many commentators argue that federal financial services law should apply to all decentralized virtual currencies in order to address the money laundering risk, but that the remaining policy issues should be left to the states to address. Third, a variety of proposals call for various methods and levels of self-regulation.

Notably, most of the proposals in each category focus on building a regulatory approach to bitcoin and other decentralized virtual currencies, and do not address regulation of the underlying decentralized ledger technology. When the literature does turn its attention to the legal implications of decentralized ledger technology, it tends to skip the question of how to regulate the blockchain and moves straight to jurisprudential questions of how the blockchain might disrupt or alter known legal structures such as contract law,[125] property law,[126] and judicial decision making.[127]

---

122. *See* DOJ, RIPPLE SETTLEMENT AGREEMENT, *supra* note 116, at Attachment A, paras. 13–14.

123. *See* WASH. DEP'T OF FIN. INSTS., DIV. OF CONSUMER SERVS., INTERIM REGULATORY GUIDANCE ON VIRTUAL CURRENCY ACTIVITIES (Dec. 8, 2014), *available at* http://www.dfi.wa.gov/documents/money-transmitters/virtual-currency-interim-guidance.pdf [https://perma.cc/TL2S-LQNK] (explaining how minimum tangible net worth and permissible investment statutory requirements can be applied to virtual currency-related businesses).

124. Some of the literature is solely dedicated to proving out the extensiveness of this inadequacy. *See, e.g.*, Grinberg, *supra* note 58; Tu & Meredith, *supra* note 117.

125. *See generally* Fairfield, *Bitcoin Bots*, *supra* note 30.

126. *See generally* Fairfield, *BitProperty*, *supra* note 13.

127. *See generally* Michael Abramowicz, *Cryptocurrency-Based Law* (George Wash. Law Sch., Research Paper No. 2015-9, 2015), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2573788 [https://perma.cc/NHL7-G3MA].

The literature leaves a significant gap, which will have a substantial impact on both current entrepreneurial efforts to build regulatory-compliant businesses and the ability of decentralized ledger technology to actually impact the other jurisprudential questions raised by the literature. The blockchain and other similar technologies will never revolutionize property law, for example, if the regulatory environment is so hostile that no software developer can risk developing the product that will spark the revolution. To develop a regulatory approach robust enough to account for the many uniquely beneficial aspects of decentralized ledger technology, the various market and governance failures that have historically plagued such technologies, and the variety of forms taken by decentralized ledger technology, and to allow for continued innovation, this Section first develops certain criteria for evaluating alternative regulatory approaches. This Section then weighs existing alternative regulatory proposals against these criteria, finding that none of the current proposals meet all of the criteria to a high degree.[128] This Section concludes by attempting to define the boundaries of the regulatory lacuna left by both the current regulatory landscape and the academic literature, thereby setting the stage for developing a gap-filling alternative approach in Part IV.

A. *Criteria for Evaluating Alternative Regulatory Proposals*

To objectively consider whether any of the presently proposed alternative regulatory approaches adequately fill the lacunae in the current regulatory landscape, this Article proposes the use of the following seven standards: (1) minimize compliance risk; (2) minimize risk of illicit use; (3) minimize malfunctions and related problems; (4) minimize data security risk; (5) minimize systemic risk; (6) promote innovation and adaptability, and (7) maximize political feasibility. These criteria are drawn from the policy priorities reflected in the current state of decentralized ledger technology regulation and from financial regulatory theory.

Compliance risk refers to "the risk of loss associated with non-compliance with laws, rules, regulations, prescribed practices, or ethical standards."[129] This criterion aims to capture the policy priorities that underlie many of the ex ante regulatory measures presently applied to the decentralized ledger technology industry, including anti-money laundering and terrorist financing controls,[130] and minimum net worth and bond require-

128. This approach is patterned after a common analytical tool used in public policy literature. For an overview of the approach, see EUGENE BARDACH, A PRACTICAL GUIDE FOR POLICY ANALYSIS: THE EIGHTFOLD PATH TO MORE EFFECTIVE PROBLEM SOLVING 79–82 (4th ed. 2012). For examples of applications of this approach in the legal literature more broadly, see generally Carla L. Reyes, *Access to Counsel in Removal Proceedings: A Case Study for Exploring the Legal and Societal Imperative to Expand the Civil Right to Counsel*, 17 UDC/DCSL L. REV. 131 (2014); Schwarcz, *Systemic Risk*, *supra* note 118.

129. Bradford et al., *supra* note 86, at 35.

130. *See id.*

ments found in state money transmission laws. *Risk of illicit* use refers to "the risk of penalties if the failure to comply with required guidelines to curb illicit use . . . is discovered."[131] This criterion reflects ex post punishments for criminal behavior using decentralized ledger technology. *Malfunctions and related problems* refers to "malfunctions that are the result of unintentional circumstances or events . . . or intentional circumstances or events . . . ."[132] This criterion reflects the history of malfunctions resulting from unintentional circumstances in the decentralized ledger technology industry, including, for example, the fork of the blockchain in March 2013,[133] theft of consumer bitcoins held in trust by industry actors (such as the case of the Mt. Gox failure),[134] and the potential for an intentional 51% attack on the blockchain[135] or similar attacks on other decentralized ledger technologies.[136]

Data-security risk refers to "unauthorized modification, destruction, or disclosure of data . . . ."[137] Although this criterion may seem counterintuitive in the context of decentralized ledger technology, which sports encryption and cryptographic security features, the security of all applications that interact with decentralized ledger technologies are not created equal. This criterion is designed to ensure that any remaining data-security risk not intrinsically addressed by the underlying technology is addressed by consumer-facing uses of the technology. *Systemic risk* refers to

> [t]he risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either (X) the failure of a chain of markets or institutions or (Y) a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility.[138]

---

131. *Id.*

132. *Id.* at 34.

133. *See* Jeong, *supra* note 48, at 5.

134. *See* Yoshifumi Takemoto & Sophie Knight, *Mt. Gox Files for Bankruptcy, Hit with Lawsuit*, REUTERS (Feb. 28, 2014), http://www.reuters.com/article/2014/02/28/us-bitcoin-mtgox-bankruptcy-idUSBREA1R0FX20140228 [https://perma.cc/WYV3-FDUW].

135. *See* Daniel Cawrey, *Are 51% Attacks a Real Threat to Bitcoin?*, COINDESK (June 20, 2014), http://www.coindesk.com/51-attacks-real-threat-bitcoin/ [https://perma.cc/NQ9Z-ZT8J].

136. *See, e.g.*, Vitalik Buterin, *Long-Range Attacks: The Serious Problem with Adaptive Proof of Work*, ETHEREUM BLOG (May 15, 2014), https://blog.ethereum.org/2014/05/15/long-range-attacks-the-serious-problem-with-adaptive-proof-of-work/ [https://perma.cc/K7FB-TG73] (describing potential for long-range attack on Ethereum protocol).

137. Bradford et al., *supra* note 86, at 34.

138. Iman Anabtawi & Steven Schwarcz, *Regulating Systemic Risk: Towards an Analytical Framework*, 86 NOTRE DAME L. REV. 1349, 1353 (2011) (quoting Schwarcz, *Systemic Risk*, *supra* note 118, at 204).

Although this criterion relies on a definition from financial regulatory literature, here, it is not intended to be used only with reference to the payments applications of decentralized ledger technologies. Rather, as each decentralized ledger protocol represents a system, albeit not necessarily a financial system, systemic shocks may result in parallel effects. This criterion is intended to be used by way of extrapolation, drawing parallels between the complexities and interrelatedness of the financial system and those of decentralized ledger technologies.[139]

Promoting innovation and adaptability refers to the ability to reduce the lag between the law and the technology it regulates through mechanisms designed to promote innovation and enable relatively fast regulatory adaptation to those innovations. This criterion reflects both the recognition that decentralized ledger technology has quickly evolved and that it will likely continue to do so. It also captures the complexity of the technology, which like the financial system it frequently interacts with, traditionally "innovate[s] more quickly than regulators can adapt."[140] *Political feasibility* refers to the variety and number of actors that will support the regulatory approach.[141] Given the decentralized nature of the technology to be regulated, wide stakeholder buy-in will be crucial to the effectiveness of any alternative regulatory proposal.[142]

B. *Categorizing Decentralized Virtual Currencies as a Recognized Legal Asset*

The first group of alternative proposals presented by the academic literature focuses on the payments application of decentralized technologies and argues that bitcoin and related decentralized virtual currencies should be regulated as a specific type of asset, such as intangible prop-

139. Indeed, this Article has drawn such parallels throughout, including with reference to the ex ante/ex post regulatory dichotomy and with regard to elements of the endogenous approach proposed below. This is in no way intended to suggest that the payments applications of decentralized ledger technologies are the focus of the analysis presented here. In fact, the opposite is intended. However, the complexity of issues, the speed of innovation, the potential for system-wide impact (both good and bad), and the number of points-of-entry for consumers are all parallels that make financial regulatory literature a useful point of reference. Notably, the analysis in this Article does not end with financial regulatory literature. Rather, the diffuse nature of the inquiry, incorporating elements of technology law, financial law, comparative law, and international development law, reflects the depth of the technology and related points of regulatory inflection at issue.

140. Schwarcz, *Systemic Risk*, *supra* note 118, at 260.

141. *See* BARDACH, *supra* note 128, at 41–42 (referring to this criterion as *political acceptability* and defining it as "a combination of two conditions: too much opposition (which may be wide or intense or both) and/or too little support (which may be insufficiently broad or insufficiently intense or both)").

142. *See* Wright & De Filippi, *supra* note 2, at 57 (arguing that if decentralized ledger community does not buy-in to regulatory proposal, it could defeat it by refusing to adopt new rules).

erty,[143] money,[144] securities,[145] uncertificated securities,[146] or some other presently recognized form of legal asset.[147] One of the goals of these proposals appears to be reducing uncertainty as to which legal regimes apply to bitcoins. If bitcoins are categorically money, property, or some other form of recognized legal asset, the path to regulatory compliance becomes clearer: follow the traditional rules applicable to that asset. Various practical difficulties plague this approach. First, not all decentralized virtual currencies behave in precisely the same way and therefore may not all fit within the definition of the traditional legal asset at issue. Second, implementation issues may arise that leave the sufficiency of compliance efforts uncertain. For example, issues of adequate implementation have plagued even the most well-meaning industry actors with regard to compliance with federal money service business regulations and state money transmitter laws.

Two significant theoretical difficulties face these approaches as well. First, these approaches only consider the regulation of the payments application of decentralized ledger technologies, ignoring the underlying technology and its other potential use cases entirely. In other words, if the primary criteria for these approaches is to provide regulatory clarity, to the extent it does so at all, it only achieves that criteria for one application of

---

143. *See, e.g.*, Rhys Bollen, *The Legal Status of Online Currencies: Are Bitcoins the Future?*, 24 J. BANK. & FIN. L. & PRAC. 272, 279 (2013) (arguing that "Bitcoins are a form of intangible private property, a valuable digital artefact" and are therefore "analogous with other forms of intangible private property, such as digital music, shares, licenses, trademarks, copyright, goodwill, domain names, frequent flier points and brands"); Nelson DaCunha, *Virtual Property, Real Concerns*, 4 AKRON INTELL. PROP. J. 35, 41 (2010) (noting that if decentralized virtual currencies are form of intangible property, tort law offers one way to protect interests of both users and service providers).

144. *See* Bryans, *supra* note 56, at 441; Nicholas A. Plassaras, Comment, *Regulating Digital Currencies: Bringing Bitcoin Within the Reach of the IMF*, 14 CHI. J. INT'L L. 377, 403 (2013).

145. *See* J. Scott Colesanti, *Trotting Out the White Horse: How the S.E.C. Can Handle Bitcoin's Threat to American Investors*, 65 SYRACUSE L. REV. 1, 38 (2014); Ruoke Yang, *When Is Bitcoin a Security Under U.S. Securities Law?*, 18 J. TECH L. & POL'Y 99, 108 (2013).

146. *See generally* Jeanne L. Schroeder, Bitcoin and the Uniform Commercial Code (Benjamin N. Cardozo Sch. Law Research Paper No. 458, 2015), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2649441 [https://perma.cc/LS6G-EVF2].

147. *See, e.g.*, Houman B. Shadab, *Regulating Bitcoin and Block Chain Derivatives* (2014), *available at* http://www.cftc.gov/ucm/groups/public/@aboutcftc/documents/file/gmac_100914_bitcoin.pdf [https://perma.cc/F7PD-42PG] (arguing that bitcoins fall into definition of commodity under Commodity Exchange Act and that derivative contracts that reference bitcoins are subject to regulation by Commodity Futures Trading Commission); George K. Fogg, *Perkins Coie: The UCC and Bitcoins—Solution to Existing Fatal Flaw*, COINDESK (Jan. 29, 2015), http://www.coindesk.com/perkins-coie-bitcoin-can-learn-real-estate-law/ [https://perma.cc/72M7-XRR8] (arguing that bitcoins would presently be considered general intangibles under UCC, and presenting solution for changing that classification to one of investment property).

the technology, leaving the rest of the industry either in unchartered waters or, perhaps worse, attempting to comply with a regime that is ill-suited to the technology use at issue. Second, these approaches do nothing to resolve the underlying tension of a combined ex ante and ex post approach to regulating decentralized ledger technologies. In sum, although these approaches address the criteria of compliance risk and illicit use to a high degree, they do not promote innovation and adaptability, minimize systemic risk, or solicit broad stakeholder buy-in.

C. *Applying Federal Financial Service Laws and Otherwise Deferring to State Regulation*

The second group of proposals argues that decentralized ledger technology service providers, and especially those offering a service related to decentralized virtual currency, should remain subject to existing "customer-identification program and AML compliance program requirements of Sections 326 and 352 of the USA PATRIOT Act, and with the economic sanctions regulations enforced by OFAC" and FinCEN regulations as appropriate, and that the remaining regulatory functions should be left to state governments.[148] Within this second group, the proposals for state approaches to regulation vary widely and include a recommendation to require exchanges to pay premiums to a third-party insurer of the exchange's choice[149] and an elective tax on anonymity,[150] among others.[151] The wide variety of specific recommendations within this sec-

148. Trautman, *supra* note 18, at 43 (quoting *The Present and Future Impact of Virtual Currency: Joint Hearing Before the Subcomms. on Econ. Pol'y & on Nat'l Sec. & Int'l Trade & Fin. of the S. Banking Comm.*, 113th Cong. 2–3 (2013), *available at* http://www.banking.senate.gov/public/index.cfm/hearings?ID=955322CC-D648-4A00-A41F-C23BE8FF4CA (statement of Sarah Jane Hughes, Univ. Scholar & Fellow in Commercial Law, Ind. Univ. Maurer Sch. of Law)); *see also* Middlebrook & Hughes, *supra* note 76, at 840 (noting that regulating cryptocurrencies by analogy to existing regulatory schemes may have higher likelihood of success because "[t]he law has a tendency to address new products and technologies by analogizing to existing regulatory schemes"); Joshua J. Doguet, Comment, *The Nature of the Form: Legal and Regulatory Issues Surrounding the Bitcoin Digital Currency System*, 73 LA. L. REV. 1119, 1147–49 (2013) (examining three regulatory approaches and arguing that compliance with existing anti-money laundering regulations offers existing opportunity to curb illicit uses of bitcoin payment system).

149. *See* David Groshoff, *Kickstarter My Heart: Extraordinary Popular Delusions and the Madness of Crowdfunding Constraints and Bitcoin Bubbles*, 5 WM. & MARY BUS. L. REV. 489, 554 (2014).

150. *See* Marian, *supra* note 24, at 54.

151. *See, e.g.*, Gruber, *supra* note 55, at 204–08 (arguing for compliance with existing BSA and tax law and for additional prohibition of mixing technologies and use of Tor network to enhance anonymity); Misha Tsukerman, Note, *The Block Is Hot: A Survey of the State of Bitcoin Regulation and Suggestions for the Future*, 30 BERKELEY TECH. L.J. 1127 (2015) (arguing for registry linking bitcoin wallet public keys to user's identity). Notably, both of these proposals focus on the anonymous nature of bitcoin transactions as the primary policy problem to be addressed. *See* Kaplanov, *supra* note 50, at 113–14 (arguing that bitcoin should be regulated as community currency).

ond group stems from the emphasis each proposal places on specific characteristics of decentralized virtual currency. For example, the third-party insurance solution emphasizes the price volatility that characterizes bitcoin and uses economic theory to suggest a way to minimize investor risk and protect consumers against the price volatility with minimal government intervention.[152] Meanwhile, the elective tax on anonymity seeks to reduce what is commonly perceived as a core benefit that decentralized virtual currencies offer to criminal masterminds: anonymity.[153]

Despite the variety of approaches, each of the proposals in this group attempts to bridge the gap between the policy concerns presently voiced by regulators and the frustration of many decentralized industry participants. In particular, many of the proposals start from the premise that the payments applications of decentralized technologies present a real risk of money laundering and facilitation of other criminal activity.[154] Further, the proposals recognize that users of decentralized payments systems face real risk of loss of funds and other assets.[155] These proposals, however, face difficulties similar to the first group of proposals: practical difficulties to implementation, a singular focus on bitcoin and other payments applications of decentralized ledger technology, and proposals for regulatory frameworks that emphasize one perceived characteristic of decentralized virtual currencies at the expense of others. In other words, these proposals, like both the first group of proposals and current regulation, minimize compliance risk and illicit use at the expense of innovation and adaptability. Further, these proposals lack structured consideration of issues related to minimizing malfunctions, data security, or systemic risks. Finally these approaches will likely not garner broad stakeholder support, especially among industry stakeholders. As a result this group of proposals generally only upholds two of the criteria to a high degree.

D. *Proposals for Self-Regulation*

The third and most prominent group of proposals calls for various levels of industry self-regulation.[156] One such proposal argues for a three-tiered self-regulatory approach. The three tiers would include: (1) the code itself acting as law to restrain activity; (2) contractual obligations self-

152. *See* Groshoff, *supra* note 149, at 554.

153. *See* Marian, *supra* note 24, at 53.

154. *See, e.g.*, *id.*

155. *See, e.g.*, Middlebrook & Hughes, *supra* note 76, at 823; Christopher, *supra* note 76, at 24.

156. *See, e.g.*, Primavera De Filippi, *Bitcoin: A Regulatory Nightmare to a Libertarian Dream*, 3 INTERNET POL'Y REV. 1, 10 (2014) [hereinafter De Filippi, *Regulatory Nightmare*] ("[R]egulation of the protocol will most probably arise 'organically' as bitcoin adoption increases. Accordingly, before turning to regulation, it might be wise to first look at whether the solutions forthcoming from the market could actually provide a satisfactory answer to these aforementioned problems." (citation omitted) (citing Henrik Karlstrøm, *Do Libertarians Dream of Electronic Coins? The Material Embeddedness of Bitcoin*, 15 SCANDINAVIAN J. SOC. THEORY 23 (2014))).

imposed through a service provider's terms of service, privacy policy, and other consumer facing documents; and (3) private lawsuits to hold service providers "liable for all losses due to their negligence, recklessness, or disregard for" users' rights.[157] Other commentators suggest that a new body of law, which they term "*Lex Cryptographia*," will emerge as "a set of rules administered through self-executing smart contracts and decentralized (and potentially autonomous) organizations."[158] Recognizing *Lex Cryptographia* as a form of self-regulation in the vein of its predecessors, *Lex Mercatoria* and *Lex Informatica*, Wright and De Filippi argue that "[o]ne of the key consequences of the blockchain could be a rapid expansion of what Lawrence Lessig referred to as 'architecture'—the code, hardware, and structures that constrain how we behave—or at a minimum a redefinition of how laws and regulations are designed, implemented, and enforced."[159]

Although this Article contends that promoting innovation and adaptability will be a key aspect of regulating decentralized ledger technology, these self-regulatory proposals elevate this criterion over all others. Doing so discounts the actual market failures that have occurred since the introduction of decentralized ledger technologies, including massive loss of consumer funds caused by the failure of large industry players, such as Mt. Gox.[160] The self-regulatory proposals do not, however, limit their scope to the decentralized virtual currency applications of the underlying decentralized ledger technology. Rather, it is the very recognition of the vast and complex potential use cases for the underlying technology that causes these writers to counsel self-regulation. Proposals for self-regulation of decentralized ledger technologies face several challenges. First, the threat of ex post incentives for compliance has proved important to the regulatory history of decentralized virtual currencies to date. Without the power to impose ex post regulation, whether through enforcement activity or some other form of ex post incentive, it is unclear how proposals for self-regulation intend to address the compliance, illicit use, and malfunction risk criteria. Second, *Lex Mercatoria* and *Lex Informatica* largely provide rules for the private sphere, while many of the core policy concerns facing the decentralized ledger technology industry implicate matters of public law as well. As a result, a self-regulatory approach, including a *Lex Cryptographia* that follows the pattern of its predecessors and focuses on rules for the private sphere, will be ill-suited to address systemic risk.

---

157. DaCunha, *supra* note 143, at 45, 58, 71.

158. Wright & De Filippi, *supra* note 2, at 48.

159. *Id.* at 50 (footnote omitted) (citing LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE VERSION 2.0, at 24 (2d ed. 2006) [hereinafter LESSIG, CODE AND OTHER LAWS]).

160. *See* Takemoto & Knight, *supra* note 134.

E. *Crystalizing the Problem and Defining the Boundaries of the Regulatory Lacuna*

Taken as a whole, the present alternative regulatory proposals focus on the same policy priorities as the current regulatory approach: focusing overwhelmingly on the payments applications of decentralized ledger technologies, curbing illicit uses of such payments applications, reducing the perceived extreme level of anonymity afforded to use of such payments applications, protecting consumers from financial loss, and predominately focusing on ex ante measures followed by ex post supplemental enforcement actions as necessary. Furthermore, both the current regulatory landscape and the alternative proposals focus their efforts on the centralized actors in the ecosystem. Although this approach arguably worked well with regard to regulating the Internet,[161] it appears to be experiencing higher levels of inefficiency and ineffectiveness in the context of decentralized ledger technologies. As regulation of the Internet experienced in the 1990s, so regulation of decentralized ledger technologies is experiencing now: a lacuna in the regulatory approach that results in law lag. Regulation that focuses on the centralized actors in a decentralized ecosystem simply will not be able to keep pace. There is some level of urgency related to addressing the law lag, as the level of decentralization in the ecosystem is only expected to grow.[162]

The above discussion reveals four core elements of the gap between law and decentralized ledger technology that is causing the law lag. First, the complete failure to regulate at the decentralized ledger level rather than the payments application level occupies a large place in the regulatory gap. This failure implicates two of the criteria: promoting innovation and adaptability and minimizing systemic risk. Second, the current and alternative approaches are overly reactive to the past bitcoin market failures. The heavy emphasis on anti-money laundering and curbing other illicit uses are directly reactive to the Mt. Gox failure, and the E-Gold, Liberty Reserve, and Silk Road enforcement actions. Similar to the pattern of overreacting to past crises that characterizes most existing financial regulation,[163] this reactive approach implicates the criteria of compliance, illicit use, malfunction, and data security risks. Third, the present and proposed alternative regulatory approaches are grounded in the characteristics of bitcoin and decentralized virtual currencies as they appear at present. This artificially ties effective regulation to a specific moment in time and a specific market context. Again, this flaw echoes that of existing financial regulation more generally[164] and implicates the innovation and

161. *See* Wright & De Filippi, *supra* note 2, at 55.

162. *See id.* at 56–58; *see also* Abramowicz, *supra* note 127; Fairfield, *Bitcoin Bots*, *supra* note 30.

163. *See* Anabtawi & Schwarcz, *supra* note 138, at 1351.

164. *See id.* at 1369 ("It is our view, however, that additional measures are, and will continue to be, needed to protect the financial system from unforeseen economic shocks. In part, this is because the measures that have been adopted re-

adaptability and systemic risk criteria. Finally, it remains unclear whether the mishmash of ex ante and ex post regulation is purposeful or simply the result of various regulators applying existing laws to an emerging technology. When viewed in this light, the regulatory gap crystalizes and may be problematized in the following manner: what regulatory approach (ex ante, ex post, a combination of both, or neither) will treat decentralized ledger technology holistically, including taking a balanced approach to rectifying past crises, without tying regulation to specific implementations or applications of the technology, while maximizing the seven criteria to a high degree?

IV. Filling the Gap with an Endogenous Theory of Decentralized Technology Regulation: An Initial Proposal

Current regulation of decentralized virtual currencies and the existing literature assume that the form of regulation is limited to the following dichotomy: "[S]elf-regulation (i.e. through market-based mechanisms, or with the support of a private regulatory body like the Bitcoin Foundation) or by means of state regulation."[165] This Article, however, has demonstrated that neither choice effectively incentivizes decentralized technology ecosystem participants to prevent the market and governance failures that primarily concern regulators. This Article therefore challenges the conventional dichotomy and proposes a third alternative: an endogenous model of regulation that simultaneously governs from within and without, and sidesteps the ex ante/ex post regulatory choice by building compliance into the protocol and thereby eliminating the need for incentives. In so doing, this Article investigates the impact that decentralized technologies will have on the regulatory exercise itself.

A. *An Endogenous Theory of Regulation for Decentralized Ledger Technologies*

The concept of endogenous regulation can be found in various branches of academic literature, including economics, development, comparative law, and financial regulation. This Article suggests an approach to endogenous regulation that relies on insights from each of these areas in order to meet the criteria for decentralized ledger technology regulation identified above in Part III(A) and fill the regulatory lacuna identified in Parts II and III above. Synthesizing the various commentaries regarding endogenous regulation leads to an approach that is iterative, cooperative, focused on the functional purposes for enacting regulation, and implemented from within the market requiring regulation.

---

spond to practices specific to the recent global financial crisis, rather than addressing its fundamental causes.").

165. De Filippi, *Regulatory Nightmare*, *supra* note 156, at 10.

Generally speaking, the term *endogenous* refers to the concept of "[h]aving an internal cause or origin."[166] For economists, "[t]he theory of endogenous policy describes how self-interested agents influence the choices made regarding government policies."[167] In particular economic theory considers the endogenous evolution of regulation as one that is "incremental, and continuous, producing small effects at each individual step but cumulatively resulting in ever deeper levels of cooperation."[168] Economic development literature similarly conceptualizes an endogenous approach as an iterative and cooperative one, in which "the distinctiveness of individual cultures and societies should be at the center of determining the goals to be pursued."[169] This approach links culture to development and argues that because the other system possesses a unique cultural identity, it should be "entitled to freely choose [its] own method[ ] for achieving full political and economic independence."[170] The contribution of endogenous development theory relevant to the present regulatory inquiry is the premise that when promoting development in another system, whether through the imposition of new regulations[171] or otherwise, it is

166. *Endogenous*, OXFORD DICTIONARY, http://www.oxforddictionaries.com/us/definition/american_english/endogenous [https://perma.cc/DZ7D-A9RC] (last visited Mar. 1, 2016).

167. ISIDORO ADOLFO MAZZA, *An Endogenous Policy Model of Hierarchical Government*, *in* ESSAYS ON ENDOGENOUS ECONOMIC POLICY 11, 11 (2009).

168. Laurence R. Helfer, *Understanding Change in International Organizations: Globalization and Innovation in the ILO*, 59 VAND. L. REV. 649, 665 (2006); *see also* Stanley Reiter, *On Endogenous Economic Regulation*, 2 ECON. DESIGN 211, 211–14 (1996) (setting forth economic game theory underlying endogenous economic regulatory approach and describing it as iterative game process in which interactions between regulators and economic actors they regulate in one phase construct rules and regulations that govern next phase).

169. Ruth E. Gordon & Jon H. Sylvester, *Deconstructing Development*, 22 WIS. INT'L L.J. 1, 35 n.153 (2004).

170. Edmund H. Chiang, Note, *The UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions: A Look at the Convention and Its Potential Impact on the American Movie Industry*, 6 WASH. U. GLOBAL STUD. L. REV. 379, 392 n. 81 (2007) (citing Ivan Bernier, *A UNESCO International Convention on Cultural Diversity*, *in* FREE TRADE VERSUS CULTURAL DIVERSITY: WTO NEGOTIATIONS IN THE FIELD OF AUDIOVISUAL SERVICES 65, 72 (Cristoph Beat Graver, Michael Girsberger & Mira Nenova eds., 2004)).

171. Modernizing legal norms (by patterning them off of Western institutions) has long been considered a method for promoting economic development. For a brief history and description of the law and development movement, see Gordon & Sylvester, *supra* note 169, at 18–23; *see also* Amanda J. Perry, *International Economic Organizations and the Modern Law and Development Movement*, *in* MAKING DEVELOPMENT WORK: LEGISLATIVE REFORM FOR INSTITUTIONAL TRANSFORMATION AND GOOD GOVERNANCE 19 (Ann Seidman, Robert B. Seidman & Thomas W. Wälde eds., 1999); David M. Trubek, *The "Rule of Law" in Development Assistance: Past, Present, and Future*, *in* THE NEW LAW AND ECONOMIC DEVELOPMENT: A CRITICAL APPRAISAL 74 (David M. Trubek & Alvaro Santos eds., 2006).

necessary to understand "the ways in which local innovations and gains can be preserved as part of local economic and cultural power."[172]

The financial regulatory literature also explores an endogenous or functional approach to regulation.[173] The functional approach to financial regulation argues that in order to regulate "a dynamically changing financial system," which rapidly changes in unexpected ways, "it may be more effective—or at least instructive—to focus on the system's underlying, and thus less time-dependent, economic functions than to tie regulation to any specific financial architecture."[174] A functional approach is desirable, the theory argues, because it is well-suited to "analysis of a highly complex or unknown structure" and it "facilitates the analysis of a rapidly changing structure," both of which are qualities of the financial system.[175]

Comparative legal theory describes a similar approach to understanding regulation and, in particular, a methodology for determining the relevant functions of a system to be regulated, and the appropriate institution with which to achieve that regulation through the concept of the functional method.[176] The central idea of the comparative functional method is that of functional equivalence, which suggests that "similar functional needs [of society] can be fulfilled by different institutions . . . ."[177] Functional equivalence posits that social problems are generally universal, but the legal and institutional response to universal social problems need not be.[178] As a result, the functional method

172. Natsu Taylor Saito, *Decolonization, Development, and Denial*, 6 FLA. A & M U. L. REV. 1, 45 (2010) (quoting ARTURO ESCOBAR, ENCOUNTERING DEVELOPMENT: THE MAKING AND UNMAKING OF THE THIRD WORLD 98 (1995)) (internal quotation marks omitted).

173. *See generally* Steven L. Schwarcz, *Regulating Financial Change: A Functional Approach*, 100 MINN. L. REV. (forthcoming 2016) [hereinafter Schwarcz, *Functional Approach*], *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2469467 [perma.cc/V2DH-HUDM].

174. *Id.* at 5–6.

175. *Id.* at 6.

176. The functional method as used in the comparative legal enterprise is designed to help a comparative legal scholar understand the way that a foreign legal system approaches a particular problem through regulation. This Article suggests that this same approach can be used to consider how to approach a particular problem in a foreign system, with decentralized ledger technology being the foreign system. On the idea of code-as-law, which leads to the conception of decentralized ledger technology as a foreign legal system, see Lessig, *Open Code and Open Societies*, *supra* note 25.

177. Ralf Michaels, *The Functional Method of Comparative Law*, *in* THE OXFORD HANDBOOK OF COMPARATIVE LAW 339, 357 (Mathias Reimann & Reinhard Zimmermann eds., 2006); *see also* KONRAD ZWEIGERT & HEIN KÖTZ, INTRODUCTION TO COMPARATIVE LAW 34 (Tony Weir trans., Oxford Univ. Press 3d rev. ed. 1998) (1977).

178. *See* Esin Örücü, *Developing Comparative Law*, *in* COMPARATIVE LAW: A HANDBOOK 43, 51 (Esin Örücü & David Nelken eds., 2007). Stated another way, "if a society has a certain problem *a*, it must have a legal institution *y*, and different solutions to *a* are functionally equivalent." Michaels, *supra* note 177, at 371.

> aims at explaining the effects of legal institutions as functions . . . and it promises to look at non-legal responses to societal requisites, too. The functional method asks us to understand legal institutions not as doctrinal constructs but as societal responses to problems—not as isolated instances but in their relation to the whole legal system, and beyond, to the whole of society.[179]

The functional method forces comparative legal scholars to shed the assumption that the absence of a certain legal structure means that the society does not address the social problem the structure is designed to resolve.[180] Instead, comparative theorists look at the legal and social context of the foreign system as a whole, with the goal of uncovering a functional equivalent—another institution that, although different in structure, serves the same purpose.[181]

Comparative legal scholars have developed a methodology for uncovering functional equivalents in foreign legal systems.[182] The first step in functional equivalence analysis is to ask what social problem a certain legal institution seeks to resolve.[183] In framing this question, comparative theorists seek to eliminate any reference to concepts in their own legal system, but rather state the question in "purely functional terms."[184] The next step in the functional method is to search the foreign system for the legal institution responding to the social problem posed. Comparative theorists undertaking this research are obligated to explore all aspects of the foreign system and not expect the legal construct to appear in the same form and context as in the native legal system.[185]

---

179. Michaels, *supra* note 177, at 364.

180. *See id.*

181. *See id.*

182. Comparative legal scholarship has debated the relative importance of the functional method as opposed to other approaches to comparative law. *See generally* Jaakko Husa, *Farewell to Functionalism or Methodological Tolerance?*, 67 RABELS ZEITSCHRIFT 419 (2003) [hereinafter Husa, *Farewell to Functionalism*] (providing review of debate and arguing for "moderate version of functionalism"). To clarify, the functional method is one of several approaches to micro-comparison, including the study of legal transplants. *See* Michele Graziadei, *The Functionalist Heritage*, *in* COMPARATIVE LEGAL STUDIES: TRADITIONS AND TRANSITIONS 100, 100 (Pierre Legrand & Roderick Munday eds., 2003). The functional method and legal transplants, however, are of particular value to the rule of law development enterprise, and for that reason, are discussed in more detail than other approaches.

183. *See* Michaels, *supra* note 177, at 366.

184. ZWEIGERT & KÖTZ, *supra* note 177, at 34 (demonstrating, through example, that question should not be "What formal requirements are there for sales contacts in foreign law?" but rather, "How does foreign law protect parties from surprise, or from being held to an agreement not seriously intended?" (internal quotation marks omitted)).

185. *Id.* at 35 ("Even experienced comparatists sometimes look for the rule they want only in the particular place in the foreign system where their experience of their own system leads them to expect it: they are unconsciously looking at the problem with the eyes of their own system.").

In the event that a search of the foreign legal system fails to produce an institution designed to address the problem, the next step for comparative theorists is to ask why the legal system does not provide such an institution.[186] This step in functional methodology forces comparative theorists to search the entire foreign system, rather than narrowly focusing on the legal system.[187] In particular, comparative theorists investigate the foreign system's cultural and customary norms in search of an institution that addresses the problem.[188] By the end of the comparative endeavor, comparative theorists have either discovered a functionally equivalent institution, or have discovered the reason for the absence of one.[189]

As a result of the functional method, comparative theorists "know that law, just like culture, is not monolithic."[190] Instead of prescribing which laws work best, the comparative functional method "sets out to understand what makes law work."[191] By beginning the inquiry with the social function of law,[192] the comparative functional method creates a framework in which it is possible to avoid "the problem that one perceives the foreign systems mainly through the mind-set of one's own legal system."[193] In doing so, the functional method enables the comparative theorists to conceive of possible reasonable substitutes, rather than assume that one set of institutions is essential for meeting a certain set of regulatory criteria.

Ultimately, then, in undertaking the micro-comparison enterprise through the functional method, the first question is not, *How can this specific institution be successfully transplanted?* but rather, *Which institution can be transplanted that would successfully address the problem?* If decentralized ledger technology is considered a "foreign system," with protocol rules that govern its usage and a coding language all its own, the comparative functional method becomes a core element of an endogenous approach to decentralized ledger technology development. Essentially, when combined with foundational concepts from functional financial regulation, endogenous economic regulation, and endogenous development, the comparative functional method offers regulators a tool for engaging in a two-way regulation design process that requires the participation of core developers in the decentralized ledger technology ecosystem. By doing so, the construct of an endogenous or functional approach to regulating de-

186. *See id.*

187. *See id.*

188. *See id.*

189. *See id.*

190. Graziadei, *supra* note 182, at 115.

191. *Id.*

192. *See* Husa, *Farewell to Functionalism*, *supra* note 182, at 423; Jaakko Husa, *About the Methodology of Comparative Law – Some Comments Concerning the Wonderland. . .* 8 (Univ. of Maastricht Faculty of Law, Working Paper No. 2007/5, 2007) [hereinafter Husa, *Wonderland*] ("The point of departure for comparison ought to be . . . the socio-legal function.").

193. Husa, *Farewell to Functionalism*, *supra* note 182, at 423; *see also* Husa, *Wonderland*, *supra* note 192, at 8.

centralized ledger technology has the potential to meet a majority of the criteria set forth in Part III above to a high degree and therefore offers a potential avenue for filling the gap in the current regulatory landscape and academic literature. Because it is difficult to assess the fulfillment of the criteria in the abstract, this Article next offers one possible implementation and uses it to demonstrate the capacity for endogenous regulation to offer flexible, adaptable, and feasible regulation that minimizes a variety of risks to a high degree.

B. *Technology-Assisted Regulation: A Proposed Implementation of Endogenous Regulation in the Decentralized Ledger Technology Ecosystem*

Taking the various strands of endogenous and functional regulation together reveals a synthesized approach that is iterative, cooperative, focused on the functional purposes for enacting regulation, and implemented from within the market requiring regulation. These are the core elements of the endogenous theory of decentralized ledger technology proposed here. Taking the last element first, this Article proposes that regulation of decentralized ledger technology should build on the body of literature, most prominently led by Lawrence Lessig, that argues for the use of code-as-law.[194] The proposal here is not for allowing the industry to self-regulate through rules created by use of the blockchain or other similar technologies.[195] Nor is this a proposal for using smart contracts to create a code of law that enables the regulation of private actors and decentralized autonomous organizations.[196] Finally, this is not a proposal for layering law on top of technology architecture in order to form a com-

---

194. *See, e.g.*, LESSIG, CODE AND OTHER LAWS, *supra* note 159, at 24; Lawrence Lessig, *Foreword*, 52 STAN. L. REV. 987, 990 (2000) (explaining that phrase "*[c]ode is law*" was meant "[m]etaphorically, in that the code controls behavior as law might control behavior" (internal quotation marks omitted)). In Lessig's conception of code-as-law, the idea is that writers of code, private actors, constrain behavior as they write rules into the code. *See* Lessig, *Open Code and Open Societies*, *supra* note 25, at 1408.

> The code of cyberspace—whether the Internet, or a net within the Internet—defines that space. It constitutes that space. And as with any constitution, it builds within itself a set of values and possibilities that governs life there. . . . And the design of code is something that people are doing. Engineers make the choices about how the world will be. Engineers in this sense are governors.

Id.

The proposal here builds Lessig's conception of code-as-law into law-through-code, or technology-assisted regulation, something that Lessig doubted could be achieved in an open source environment. *See infra* note 219 and accompanying text.

195. *See generally* De Filippi, *Regulatory Nightmare*, *supra* note 156; David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1391–95 (1996).

196. In other words, this is not a proposal in the same vein as Wright & De Filippi's *Lex Cryptographia*. *See generally* Wright & De Filippi, *supra* note 2.

plete regulatory picture.[197] Rather, this is a proposal that regulators undertake the dual task of enacting a law or regulation via statute, and then implementing that statute through code, so that it is endogenously incorporated into the decentralized ledger technology or applications running on top of the technology. In other words, while several commentators have suggested that increasingly complex systems of smart contracts can be used to create new rules and constructs for organizations,[198] corporate entities,[199] property,[200] and government bodies,[201] this proposal contests that the first and primary target for regulation-through-code is the technology itself. In this way, the endogenous theory of regulation advanced here reflects Lessig's suggestion that

> [r]egulation in cyberspace is, or can be, different. If the regulator wants to induce a certain behavior, she need not threaten, or cajole, to inspire the change. She need only change the code—the software that defines the terms upon which the individual gains access to the system, or uses assets on the system.[202]

Achieving such implementation of regulation-through-code, which this Article will refer to as technology-assisted regulation, will organically require that regulators adopt the other elements of the endogenous approach. Namely, because the writing of code is limited by certain properties of the protocol, regulators will be unable to craft the requisite rules without cooperating with members of the decentralized ledger technology ecosystem. Further, only an iterative process of incorporation and feedback will lead both the drafters of statute and the writers of code to a place of thoughtful and meaningful regulation. Finally, cooperation between regulators and industry members in an iterative process will inevitably lead to discussions about the functional purpose of current regulation and how to translate those purposes into mechanisms that serve as a functional equivalent within the code. If such a level of cooperation could be achieved, the criteria of political feasibility would be met to a high degree, as buy-in would be required of regulators and industry members alike.

---

197. *See* LESSIG, CODE AND OTHER LAWS, *supra* note 159, at 123 (arguing that regulation is composed of various levels of constraints, including laws, market, social norms, and architecture (including code)); *see also* DaCunha, *supra* note 143, at 58, 64, 68 (arguing for three layers of bitcoin regulation); Andy Yee, *Internet Architecture and the Layers Principle: A Conceptual Framework for Regulating Bitcoin*, 3 INTERNET POL'Y REV. 1 (2014).

198. *See* Wright & De Filippi, *supra* note 2, at 50.

199. *See* Mihaela Ulieru, *Organic Governance Through the Logic of Holonic Systems*, *in* FROM BITCOIN TO BURNING MAN AND BEYOND: THE QUEST FOR IDENTITY AND AUTONOMY IN A DIGITAL SOCIETY, *supra* note 75, at 113, 128.

200. *See generally* Fairfield, *BitProperty*, *supra* note 13.

201. *See generally* Abramowicz, *supra* note 127.

202. Lawrence Lessig, *The Zones of Cyberspace*, 48 STAN. L. REV. 1403, 1408 (1996).

Writing regulation into the code is not only possible, but is organic to the system. It is possible that an interaction between the decentralized ledger technology industry and regulators could be coordinated through a centralized entity representing the community, such as the Bitcoin Foundation,[203] the Digital Asset Transfer Authority (DATA),[204] or a similar organization created specifically for this purpose. However, the chosen organization may dissolve while the need for regulation would remain constant. As a result, coordinating the cooperative, iterative endogenous regulation process through such a centralized entity, while familiar to regulators, may not be the most beneficial in the long run. In any event, such efforts need not be implemented through a centralized organization. "There is evidence that community consensus can result in the resolution of" issues related to decentralized public ledger maintenance.[205] As such, "if the [decentralized ledger technology] network came to a consensus about modifying the rules," the rules of the system could change.[206] In fact, it is likely that much decentralized public ledger maintenance is conducted by a small group of core developers and a handful of additional volunteers recruited by the core team.[207] Moreover, many of the prominent decentralized ledgers have adopted such a structure,[208] and evidence shows that "large successful open source projects typically have a benevolent dictator who makes decisions."[209] Furthermore, decentralized ledger technologies and, in particular, the payments applications that run on them, arguably already contain "a number of economic rules that serve as monetary policy enacted in code."[210] In fact, the very issues of who can conduct transactions, under what circumstances and when, are all determined by rules built into the protocol.[211]

By leveraging smart contracts and other features of decentralized ledger technologies, regulation can be implemented at a systemic, decentralized ledger technology level, as opposed to an application-specific level (e.g., focusing solely on payments). In doing so, this proposal fulfills two criteria to a high degree: minimizing systemic risk and promoting innovation and adaptability. Implementing regulation from within the code it-

203. *See* BITCOIN FOUNDATION, http://bitcoinfoundation.org/ [https://perma.cc/2SJZ-SQUS] (last visited Mar. 1, 2016).

204. *See* DATA ASSET TRANSFER AUTHORITY, http://info.datauthority.org/wordpress/ [https://perma.cc/QG8R-T7DR] (last visited Mar. 1, 2016).

205. Jeong, *supra* note 48, at 31.

206. *Id.* at 5.

207. *See* David S. Evans, *Economic Aspects of Bitcoin and Other Decentralized Public-Ledger Currency Platforms* 15 (Coase-Sandor Inst. for Law & Econ., Working Paper No. 685, 2014), *available at* http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=2349&context=law_and_economics [https://perma.cc/2T2D-X6SA].

208. *See id.* at 16.

209. *Id.* at 15 (citing CHARLES M. SCHWEIK & ROBERT C. ENGLISH, INTERNET SUCCESS: A STUDY OF OPEN-SOURCE SOFTWARE COMMONS (2012)).

210. Jeong, *supra* note 48, at 5.

211. *See* De Filippi, *Regulatory Nightmare*, *supra* note 156, at 1.

self promises an opportunity to create system-wide mechanisms for absorbing shocks such as market or institutional failure triggers. Further, like the financial system, "technology evolves in an unpredictable way"[212] and at a rapid pace. This is particularly true of decentralized ledger technologies, which only first emerged in 2009 and are already disrupting traditional markets and legal structures. Regulation at the decentralized ledger technology level will help avoid "[t]he danger [ ] that [blockchain] technologies as a whole may be cabined to one use case by the development of a specific legal regime, say, payments and money transmission services, and not left open to other potential uses."[213] Further, a functional approach to regulation, once built-in at a code-as-law level, and once supplemented by a formal decentralized consensus mechanism for updating regulation to reflect protocol and software changes in real time,[214] promises to keep pace with the changes in technology and decrease the length of the law lag that presently plagues the decentralized ledger technology industry. Fulfilling these criteria to a high degree is among the key contributions of this regulatory approach, as neither the current landscape nor current alternative proposals feature these characteristics.

A third key contribution of this approach regards minimizing compliance risk and the risk of illicit use.[215] Here, the heart of the regulatory exercise comes into motion: Will regulation seek to minimize such risks via ex ante or ex post measures? What if this dichotomy could be eliminated? If the system incorporates regulation-through-code, self-executing code will be regulatory-compliant, and the choice presented to individual actors will no longer be whether to comply or not, but will merely be whether or not to use the system.[216] When viewed in this light, the prospect of regulating decentralized ledger technology from within raises is-

212. ANTONIO VÁZQUEZ-BARQUERO, ENDOGENOUS DEVELOPMENT: NETWORKING, INNOVATION, INSTITUTIONS AND CITIES 77 (2002).

213. Fairfield, *BitProperty*, *supra* note 13, at 869.

214. I note here that the economic literature occasionally makes reference to a slow pace of endogenous changes. This appears to be a general reference to the time required for the endogenous regulatory effort to take root. I recognize this potential difficulty to the proposed approach and discuss it more fully in *infra* note 219 and accompanying text.

215. It is not presently clear that endogenous regulation will make any further contribution to minimizing data security risk, malfunctions, or related unintended problems other than that which is already present in the protocols. One issue that might be addressed through the functional method inquiry is the issue of privacy of financial information on a public ledger, especially as adoption becomes more widespread, or if mainstream banks adopt use of decentralized ledger technology. For now, however, these criteria are considered a wash—neither enhancing nor detracting from the suitability of the proposed endogenous regulatory approach.

216. See LESSIG, CODE AND OTHER LAWS, *supra* note 159, at 125 ("The code or software or architecture or protocols set these features, which are selected by code writers. They constrain some behavior by making other behavior possible or impossible. The code embeds certain values or makes certain values impossible.").

sues more important to regulatory theory than simply how to approach the regulation of decentralized ledger technology, or the subset of related applications that includes bitcoin. Rather, the possibility of endogenous regulation of decentralized ledger technology raises the possibility of disrupting the regulatory exercise as we know it. This is a particularly distinct possibility in the financial regulation realm for several reasons.

First, as this Article has noted throughout, regulation of decentralized ledger technology poses many of the same challenges as regulation of the financial system. Second, significant financial industry actors are moving to incorporate decentralized ledger technology into the mainstream financial system.[217] If decentralized ledger technology can make an endogenous theory of financial regulation or a functional approach to financial regulation[218] a practical reality and, if it can go a step further to enhance innovation and the ability of regulators to keep pace with such innovation, solving the riddle of how to regulate the blockchain and similar technologies will have far more significant consequences than alleviating the immediate regulatory quagmire facing the businesses in the decentralized ledger technology ecosystem.

C. *Challenges to Implementation and a Call for Further Research*

The above proposal lays out the foundation for pursuing endogenous regulation of decentralized ledger technologies, through a cooperative, iterative process that focuses on the functional purposes for enacting regulation, implemented from within the system requiring regulation. The proposal may, however, face various practical and theoretical challenges to adoption and implementation. First, it is not clear the extent to which endogenous regulation meets the criteria of political feasibility. It may be impossible to know the extent of stakeholder buy-in to the approach until after initial regulatory proposals are presented to the development team for integration into the code. This could pose a powerful practical obstacle to the endogenous regulatory enterprise. As Wright and Filippi explain:

> [T]he open nature of blockchain-based architecture means that most, if not all of the applications deployed on the blockchain could be reproduced and adjusted by anyone, in order to fulfill different functions and satisfy the needs of different groups and communities. As a result, dictating the manner in which software developers design a particular application protocol, or forcing software developers to introduce a particular feature into the code will only work to the extent that the user-base actually

217. *See* Yessi Bello Perez, *8 Banking Giants Embracing Bitcoin and Blockchain Tech*, COINDESK (July 27, 2015), http://www.coindesk.com/8-banking-giants-bitcoin-blockchain/ [https://perma.cc/Z4UK-GPS3].

218. *See generally* Schwarcz, *Functional Approach*, *supra* note 173.

agrees to switch to the new protocol. Failure to reach consensus amongst users means that software will remain in use.[219]

Another practical challenge to adopting endogenous regulation is the time that it may take to get the cooperative, iterative process up and running, and to put a consensus building mechanism into place that will allow regulators and the core development teams to gauge the political feasibility of any specific regulation being considered for inclusion as regulation-through-code. The economic and development literature both hint at a view that the endogenous approach is a slow method.[220] Such delays may temporarily increase the law lag, but not more so than if no new regulatory approach is taken.

A third challenge to adopting and implementing endogenous regulation of decentralized ledger technology is a theoretical one. Assuming the regulator determines to adopt this proposal for technology-assisted regulation and assuming the regulator uses the functional method to determine that the decentralized ledger technology already contains a mechanism approximating the same function as the real-world regulation intended to be imposed through code-as-law, "how should law regulate" under such circumstances?[221] Should the law change in response to the differences found in the code, "[o]r should the law try to change the features of cyberspace, to make them conform to the law? And if the latter, then what constraints should there be on the law's effort to change cyberspace's 'nature'?"[222]

Each of these challenges makes clear that further research into the practical, technical, and theoretical viability of wide-scale endogenous regulation of decentralized ledger technology is necessary. Further research into the implications of a successful implementation of this approach would also be useful. In particular, further exploration of the impact of technology-assisted regulation, or regulation-through-code, on the ex

219. Wright & De Filippi, *supra* note 2, at 57; *see also* Lawrence Lessig, *The Limits in Open Code: Regulatory Standards and the Future of the Net*, 14 BERKELEY TECH. L.J. 759, 764 (1999) ("Whether government can regulate code depends in part upon who controls that code. If the code is closed—controlled by private for-profit organizations—then government's power is assured. But if the code is open—outside of the control of any particular private for-profit organization—then the government's power is threatened."). While evidence suggests that the issue is no longer as black and white as Lessig would lead readers to believe, the point is well taken that the open source nature of decentralized ledger technologies may pose a practical challenge to implementing technology-assisted regulation.

220. *See* Helfer, *supra* note 168, at 665 (describing iterative process proposed here as "endogenous, incremental, and continuous, producing small effects at each individual step but cumulatively resulting in ever deeper levels of cooperation"); *see also* Reiter, *supra* note 168, at 211–14 (describing iterative process as one that occurs over time).

221. Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 505 (1999).

222. *Id.*

ante/ex post dichotomy is warranted in light of industry moves to integrate decentralized ledger technology into the mainstream financial system. Finally, examining whether comparative law's functional method offers some insight into the limits to be imposed upon regulators who elect to use technology-assisted regulation as a policy tool will be critical to ensuring that an endogenous approach to regulation remains cooperative, iterative, and properly focused on function and regulation from within.

V. CONCLUSION

This Article presents a framework for regulating decentralized technology in a holistic, organic, and functional way—an endogenous way. The current patchwork of regulations applied to businesses using decentralized ledger technology is compromised by its inability to adapt to the technology, its inefficient mechanisms for responding to market and governance failures, and its overwhelming tendency to quash innovation in the name of preventing crime and protecting consumers. Alternative regulatory methods proposed to date often fall victim to the same shortcomings. Most such alternative proposals center on regulating payments applications of decentralized technology, such as bitcoin, without regard to the collateral damage caused to other innovative uses of blockchain technology. Proposals that attempt to focus on the novel characteristics of the technology similarly focus on only a subset of novel characteristics, at the expense of the others. To the extent that the present literature discusses decentralized technology through a holistic lens, its inquiry centers on ways new applications (other than payments applications) of decentralized technology may evolve and disrupt the established order. Such literature often bypasses the most pressing current question facing the very innovators who hope to disrupt technology use as we know it: What regulatory approach to the decentralized ledger technology itself can keep pace with innovation while still addressing common market and governance failures? The endogenous theory of regulation proposed in this Article attempts to fill that lacuna.

Importantly, an endogenous approach to regulation will only be effective if each of the relevant stakeholders actively participates. The current regulatory nightmare facing decentralized virtual currencies and decentralized ledger technologies is a result of more than just the mismatch resulting from the application of antiquated laws to revolutionary technology. It also stems from a long history of boisterous, bad industry actors, high volumes of monetary losses suffered by consumers, and unsubstantiated myths regarding the level of anonymity and impunity afforded to criminals using the technology. Dispelling such myths and charting a new industry course will require a collaborative effort between industry actors and regulators. By working together and leveraging the consensus properties of decentralized technologies, industry actors and regulators are well-placed to make technology-assisted regulation, which this Article submits should be viewed as a practical application of architecture-as-law or code-as-law, a reality. Such endogenous regulation offers unique compliance incentives and stakeholder buy-in that should enable more efficient ex ante regulation while simultaneously reducing the need for expensive coercive enforcement action. Finally, if the endogenous regulation of decentralized ledger technology such as the blockchain proposed in this Article could be achieved, it might pave the way for an entirely new approach to financial regulation: technology-assisted regulation, or regulation-through-code, that bypasses the ex ante/ex post dichotomy and influences actions in real time.