Indiana Law Review

Volume 45 **2012** **Number 3**

TRIBUTES

A TRIBUTE TO RANDALL T. SHEPARD
JUSTICE, INDIANA SUPREME COURT, 1985-2012
CHIEF JUSTICE OF INDIANA, 1987-2012

CHIEF JUSTICE BRENT E. DICKSON*

"We want to be a court known widely for the clarity of its thinking, the common sense and humanity of its decisions, and the excellence of its writing. We want to be a court so well regarded that judges in other states, when considering the toughest legal issues of our time, will be led to look at each other and ask: 'I wonder what Indiana has done about this?'"[1] With remarks including these words, Randall Shepard accepted the responsibilities of Chief Justice of Indiana in 1987. After more than twenty years in the position, he reflected, "Just as the adjudicating judge long ago ceased being the passive non-manager of litigation, today's judge must take interest and responsibility for building better systems of justice."[2] He has similarly noted the growing role of state supreme courts in "remaking the American court system," and urged that "the judiciary must do what lies within us to help our fellow citizens in fostering a decent, safe, and prosperous society, by building a system of justice that befits a great nation."[3] These themes have been the hallmarks of Randall Shepard's tenure on the court and his historic string of five consecutive five-year terms as Chief Justice.[4]

In his judicial opinions, the Chief Justice's contributions to Indiana jurisprudence have been significant, particularly in areas of state constitutional law,[5] local government law,[6] shareholder rights,[7] choice of law,[8] criminal law and

---

* Chief Justice, Indiana Supreme Court. B.A., 1964, Purdue University; J.D., 1968, Indiana University Robert H. McKinney School of Law. Justice Dickson and Chief Justice Shepard have served together on the Indiana Supreme Court for over twenty-six years.

1. Randall T. Shepard, Remarks on Assuming the Office of Chief Justice at the Old Vanderburgh County Courthouse (Mar. 4, 1987) (on file with author).

2. Randall T. Shepard, *The Changing Nature of Judicial Leadership*, 42 IND. L. REV. 767, 772 (2009) [hereinafter Shepard, *Changing Nature*].

3. Randall T. Shepard, *The New Role of State Supreme Courts as Engines of Court Reform*, 81 N.Y.U. L. REV. 1535, 1536, 1552 (2006) [hereinafter Shepard, *The New Role of State Supreme Courts*].

4. Shepard is the longest-serving Chief Justice of Indiana, far exceeding the thirteen years served by Chief Justice Richard M. Givan from 1974 to 1987, the next longest-serving Chief Justice. As to his total length of service as a member of the supreme court, only Isaac N. Blackford (36 years) and Roger O. DeBruler (28 years) have served longer.

5. *See, e.g.*, Branham v. Varble, 952 N.E.2d 744 (Ind. 2011); Alpha Psi Chapter of Pi Kappa

procedure,[9] commercial law,[10] family law,[11] and torts.[12] His majority opinion in *Price v. State*[13] found particular favor with Indiana University law professor Patrick Baude, who compared it approvingly with *Marbury v. Madison*.[14] During his enormously productive twenty-six plus years on the court, Shepard has written 917 majority opinions which have been cited more than 729 times by appellate courts of other states.

Chief Justice Shepard has also been a prolific author of articles published in bar journals and historical periodicals, plus fifty articles published in various law reviews throughout the nation.[15] One of these, his 1996 publication on campaign

---

Phi Fraternity, Inc. v. Auditor of Monroe Cnty., 849 N.E.2d 1131 (Ind. 2006); Malinski v. State, 794 N.E.2d 1071 (Ind. 2003); Songer v. Civitas Bank, 771 N.E.2d 61 (Ind. 2002); State v. Moss-Dwyer, 686 N.E.2d 109 (Ind. 1997); Bayh v. Ind. State Bldg. & Constr. Trades Council, 674 N.E.2d 176 (Ind. 1996); Citizens Nat'l Bank of Evansville v. Foster, 668 N.E.2d 1236 (Ind. 1996); Ind. Gaming Comm'n v. Moseley, 643 N.E.2d 296 (Ind. 1994); Conner v. State, 626 N.E.2d 803 (Ind. 1993); *In re* Zumbrun, 626 N.E.2d 452 (Ind. 1993); Price v. State, 622 N.E.2d 954 (Ind. 1993); Campbell v. Criterion Grp., 605 N.E.2d 150 (Ind. 1992); Bayh v. Sonnenburg, 573 N.E.2d 398 (Ind. 1991).

6. *See, e.g.*, Town of Avon v. W. Cent. Conservancy Dist., 957 N.E.2d 598 (Ind. 2011); Cooper Indus., LLC v. City of South Bend, 899 N.E.2d 1274 (Ind. 2009); City of Carmel v. Certain Sw. Clay Twp. Annexation Territory Landowners, 868 N.E.2d 793 (Ind. 2007); Tippecanoe Cnty. v. Ind. Mfr.'s Ass'n, 784 N.E.2d 463 (Ind. 2003); Dep't of Local Gov't Fin. v. Griffin, 784 N.E.2d 448 (Ind. 2003); Farley Neighborhood Ass'n v. Town of Speedway, 765 N.E.2d 1226 (Ind. 2002); City of Fort Wayne v. Certain Sw. Annexation Area Landowners, 764 N.E.2d 221 (Ind. 2002); Bradley v. City of New Castle, 764 N.E.2d 212 (Ind. 2002); Tax Certificate Invs., Inc. v. Smethers, 714 N.E.2d 131 (Ind. 1999); Rogers v. City of Elkhart, 688 N.E.2d 1238 (Ind. 1997); Noblesville Redev. Comm'n v. Noblesville Assocs. Ltd. P'ship, 674 N.E.2d 558 (Ind. 1996); City of Hobart v. Chidester, 596 N.E.2d 1374 (Ind. 1992); Town of Beverly Shores v. Bagnall, 590 N.E.2d 1059 (Ind. 1992); Natural Res. Comm'n v. Porter Cnty. Drainage Bd., 576 N.E.2d 587 (Ind. 1991); Schloss v. City of Indianapolis, 553 N.E.2d 1204 (Ind. 1990); City of Crown Point v. Lake Cnty., 510 N.E.2d 684 (Ind. 1987); City of Crown Point v. Knesek, 499 N.E.2d 261 (Ind. 1986).

7. *See, e.g.*, *In re* ITT Derivative Litig., 932 N.E.2d 664 (Ind. 2010); *In re* Guidant S'holders Derivative Litig., 841 N.E.2d 571 (Ind. 2006).

8. *See, e.g.*, Hubbard Mfg. Co. v. Greeson, 515 N.E.2d 1071 (Ind. 1987).

9. *See, e.g.*, Lannan v. State, 600 N.E.2d 1334 (Ind. 1992); Cooper v. State, 540 N.E.2d 1216 (Ind. 1989); White v. State, 497 N.E.2d 893 (Ind. 1986).

10. *See, e.g.*, Insul-Mark Midwest, Inc. v. Modern Materials, Inc., 612 N.E.2d 550 (Ind. 1993).

11. *See, e.g.*, Voigt v. Voigt, 670 N.E.2d 1271 (Ind. 1996); Straub v. B.M.T. *ex rel* Todd, 645 N.E.2d 597 (Ind. 1994); *In re* Lawrance, 579 N.E.2d 32 (Ind. 1991).

12. *See, e.g.*, Doe v. Methodist Hosp., 690 N.E.2d 681 (Ind. 1997); Burrell v. Meads, 569 N.E.2d 637 (Ind. 1991); Peavler v. Bd. of Comm'rs, 528 N.E.2d 40 (Ind. 1988).

13. 622 N.E.2d 954, 956 (Ind. 1993).

14. 5 U.S. (1 Cranch) 137 (1803); Patrick Baude, *Has the Indiana Constitution Found Its Epic?*, 69 IND. L.J. 849 (1994).

15. Randall T. Shepard, *Changing the Constitutional Jurisdiction of the Indiana Supreme*

Court: Letting a Court of Last Resort Act Like One, 63 IND. L.J. 669 (1988); Randall T. Shepard, *Land Use Regulation in the Rehnquist Court: The Fifth Amendment and Judicial Intervention*, 38 CATH. U. L. REV. 847 (1989); Randall T. Shepard, *Second Wind for the Indiana Bill of Rights*, 22 IND. L. REV. 575 (1989); Randall T. Shepard, *Indiana Law, the Supreme Court, and a New Decade*, 24 IND. L. REV. 499 (1991); Randall T. Shepard, *A Bill of Rights for the Whole Nation*, 26 VAL. U. L. REV. 27 (1991); Randall T. Shepard, *Indiana Law and the Idea of Progress*, 25 IND. L. REV. 943 (1992); Randall T. Shepard, *Classrooms, Clinics, and Client Counseling*, 18 OHIO N.U. L. REV. 751 (1992); Randall T. Shepard, *Lawyer-Bashing and the Challenge of a Sensible Response*, 27 IND. L. REV. 699 (1994); Randall T. Shepard, *On Lawyers and Writing: Pass the Constitutional Mustard, Please*, 28 IND. L. REV. 811 (1995); Randall T. Shepard, *Campaign Speech: Restraint and Liberty in Judicial Ethics*, 9 GEO. J. LEGAL ETHICS 1059 (1996) [hereinafter Shepard, *Campaign Speech*]; Randall T. Shepard, *Why Law Review Survey Issues Are a Good Idea*, 29 IND. L. REV. 765 (1996); Randall T. Shepard, *The Maturing Nature of State Constitution Jurisprudence*, 30 VAL. U. L. REV. 421 (1996); Randall T. Shepard, *The Importance of Legal History for Modern Lawyering*, 30 IND. L. REV. 1 (1997); Randall T. Shepard, *Reflections on a Decade at the Indiana Supreme Court, 1987-1997*, 30 IND. L. REV. 921 (1997); Randall T. Shepard, *State High Courts as Central Figures in the Future of the American Legal System*, 72 NOTRE DAME L. REV. 1009 (1997); Randall T. Shepard, *What Judges Can Do About Legal Professionalism*, 32 WAKE FOREST L. REV. 621 (1997); Randall T. Shepard, *The Renaissance in State Constitutional Law: There Are a Few Dangers, But What's the Alternative?*, 61 ALB. L. REV. 1529 (1998); Randall T. Shepard, *From Students to Lawyers: Joint Ventures in Legal Learning for the Academy, Bench, and Bar*, 31 IND. L. REV. 445 (1998); Randall T. Shepard, *Moving the Rock: The Constant Need to Re-Invent the Profession Using the Nation's Judiciary as Leaders*, 32 IND. L. REV. 591 (1999); Randall T. Shepard, *Why Changing the Supreme Court's Mandatory Jurisdiction Is Critical to Lawyers and Clients*, 33 IND. L. REV. 1101 (2000); Randall T. Shepard, *What the Profession Expects of Law Schools*, 34 IND. L. REV. 7 (2000); Randall T. Shepard, *Judicial Professionalism and the Relations Between Judges and Lawyers*, 14 NOTRE DAME J.L. ETHICS & PUB. POL'Y 223 (2000); Randall T. Shepard, *Building Indiana's Legal Profession*, 34 IND. L. REV. 529 (2001); Randall T. Shepard, *Judicial Independence and the Problem of Elections: "We Have Met the Enemy and He Is Us.,"* 20 QUINNIPIAC L. REV. 753 (2001); Randall T. Shepard, *Telephone Justice, Pandering, and Judges Who Speak Out of School*, 29 FORDHAM URB. L.J. 811 (2002); Randall T. Shepard, *Takings Law: Do We Really Want More Judicial Intervention in State Land Use Regulation?*, 1 GEO. J.L. & PUB. POL'Y 99 (2002); Randall T. Shepard, *The Special Professional Challenges of Appellate Judging*, 35 IND. L. REV. 381 (2002); Randall T. Shepard, *Making Good Law Requires More Lawyers*, 35 IND. L. REV. 1111 (2002); Randall T. Shepard, *The Personal and Professional Meaning of Lawyer Satisfaction*, 37 VAL. U. L. REV. 161 (2003); Randall T. Shepard, *Norman Lefstein—Splendid Dean, Legitimate Hoosier*, 36 IND. L. REV. 1 (2003); Randall T. Shepard, *Why the Courts Matter in Building a Strong Economy*, 36 IND. L. REV. 913 (2003); Randall T. Shepard, *Plu Ça Change: Indiana Judges and Salaries*, 37 IND. L. REV. 885 (2004); Randall T. Shepard, *On Licensing Lawyers: Why Uniformity Is Good and Nationalization Is Bad*, 60 N.Y.U. ANN. SURV. AM. L. 453 (2004); Randall T. Shepard, *Is Making State Constitutional Law Through Certified Questions a Good Idea or a Bad Idea?*, 38 VAL. U. L. REV. 327 (2004); Randall T. Shepard, *What Can Dissents Teach Us?*, 68 ALB. L. REV. 337 (2005); Randall T. Shepard, *Jury Trials Aren't What They Used to Be*, 38 IND. L. REV. 859 (2005); Randall T. Shepard, *In a Federal Case, Is the State Constitution Something Important or Just Another Piece of Paper?*, 46 WM. & MARY L. REV. 1437 (2005);

speech and judicial ethics,[16] was cited by both the concurring opinion of Justice Kennedy and the dissenting opinion of Justice Ginsburg in the United States Supreme Court's landmark decision on judicial campaign speech.[17]

Notwithstanding his extremely significant contributions to the law and jurisprudence of our state and nation, perhaps Randy Shepard's most lasting and noteworthy achievements have been in the institutional and programmatic changes resulting not only from his own passion and innovation, but particularly as a result of his own remarkable management style that motivates, encourages, empowers, and inspires others to assume responsibility and to achieve. Several current and former agency staff members aptly describe his remarkable management effectiveness as being characterized by "dignity without arrogance," "quick to listen, slow to talk," "makes people feel they're working with him, not for him," "the ability to make people feel special," and "inspires others to be more than they are." One person described it this way: "His is not a high energy, high charisma, whirlwind sort of leadership [but] the force of his quiet leadership has the power to move mountains, but so subtly that you often don't even realize it is happening." The result, of course, has been enormous creativity, productivity, satisfaction, and loyalty achieved in the staff and supervisors responsible for the court's numerous administrative functions.[18]

Randall T. Shepard, *Indiana's Constitution as a Document of Special Aspirations*, 69 ALB. L. REV. 529 (2006); Randall T. Shepard, *Indiana's Place in American Court Reform: Rarely First, Occasionally Last, Frequently Early*, 39 IND. L. REV. 723 (2006); Shepard, *The New Role of State Supreme Courts*, *supra* note 3; Randall T. Shepard, *The "L" in "CLE" Stands for "Legal,"* 40 VAL. U. L. REV. 311 (2006); Randall T. Shepard, *Introduction: The Hundred-Year Run of Roscoe Pound*, 82 IND. L.J. 1153 (2007); Randall T. Shepard, *Access to Justice for People Who Do Not Speak English*, 40 IND. L. REV. 643 (2007); Randall T. Shepard, *Defining Community in a Society Based on Rights*, 2 ALB. GOV'T L. REV. 354 (2009); Shepard, *Changing Nature*, *supra* note 2; Randall T. Shepard, *Robust Appellate Review of Sentences: Just How British is Indiana?*, 93 MARQ. L. REV. 671 (2009); Randall T. Shepard, *Judith Kaye as a Chief Among Chiefs*, 84 N.Y.U. L. REV. 671 (2009); Randall T. Shepard, *Four Big, Dumb Trends Affecting State Courts*, 43 IND. L. REV. 533 (2010); Randall T. Shepard, *The Judiciary's Role in Economic Prosperity*, 44 IND. L. REV. 987 (2011); Randall T. Shepard, *State Constitutional Remedies and Judicial Exit Strategies*, 45 NEW ENG. L. REV. 879 (2011); Randall T. Shepard, *Elements of Modern Court Reform*, 45 IND. L. REV. (forthcoming Aug. 2012).

16. Shepard, *Campaign Speech*, *supra* note 15.

17. Republican Party of Minn. v. White, 536 U.S. 765, 793 (Kennedy, J., concurring), 816-17 (Ginsburg, J., dissenting) (2002).

18. Although vastly incomplete, the following are a representative sampling of some of the many advances and accomplishments during Chief Justice Shepard's twenty-five years of leadership: creation of Indiana's Interest on Lawyer Trust Accounts (IOLTA) program to benefit the administration of Indiana's pro-bono legal services initiative; promulgation of child support and parenting time guidelines; dramatic improvement in the adequacy and reliability of criminal defense representation for indigent persons; development of a system for assessing and improving workload equality among Indiana's trial judges; creation, development, and substantial deployment of an advanced electronic case management system for Indiana trial courts; establishment of the

On a more personal note, a crucial component of Randy Shepard's leadership is seen in the way he has treated others, especially his judicial colleagues during both our public proceedings and the justices' decision conferences—where he has effectively influenced others not by demand, insistence, or attempt to exercise authority, but rather by openness, institutional awareness, quiet strength of conviction, and a heart for Hoosier institutions and citizens. During his twenty-five years of presiding at oral arguments and leading our weekly court conferences, the Chief Justice has been a model of courtesy, respect, collegiality, consideration, and decency. Always kind and welcoming to lawyers presenting their cases, the Chief Justice has often been lenient in enforcing oral argument time limits. In the court conferences that follow, the Chief moved through the agenda with gentle efficiency but invariably allowed each justice to fully express his or her views. When expressing his own, it has not been unusual for the Chief Justice to introduce them by saying, "You have stated a perfectly respectable point of view, but it's just not one with which I agree."[19] His comments at conference often reflected his own common sense wisdom about the judicial function and its limits: "We can't do something about everything,"[20] "Let's not

Indiana Conference on Legal Education Opportunity (Indiana CLEO) to assist disadvantaged students to pursue and succeed in law school; adoption of court rules promoting Indiana's use of alternative dispute resolution techniques; providing Spanish language training for court and clerk personnel throughout Indiana; adoption of a foreign language interpreter certification program; providing access-to-justice assistance for non-English-speaking persons and also those desiring to use the legal system without lawyers; enhancement of Indiana's jury procedures including the comprehensive revision of civil trial jury instructions to replace "legalese" with ordinary, everyday English; restoration of the supreme court's historic courtroom, conference room, and clerk's offices; allowing television coverage of appellate case presentations, including the now routine webcasting of all supreme court oral arguments; development and implementation of a fourth generation offender risk assessment system utilizing evidence-based tools for more effective treatment of offenders and resulting reduced recidivism and prison cost; adoption of the bar admission requirement of graduation from an ABA-accredited law school; addition of the Multistate Bar Exam and Performance Exam as components to Indiana's bar admission regimen; creation of the award-winning "Courts in the Classroom" program to assist civics education; creation of the annual "Law School for Journalists" program; adoption of rules to address the issues of public access and privacy of court records; development of a robust program of special grants for local court improvement projects, law school lecture series, and judicial education scholarships for individual judges; creation and operation of the Indiana Graduate Program for Judges; providing assistance to local judges in high-profile cases; initiating programs for family courts, drug courts, and a variety of other problem-solving courts; creation of a foreclosure assistance program; and development and implementation of numerous technology-based improvements in a variety of areas of judicial administration.

19. Reoccurring comment by Chief Justice Shepard as recalled by former Supreme Court Administrator Doug Cressler, now Chief Deputy Clerk for the United States Court of Appeals for the Tenth Circuit.

20. Comment by Chief Justice Shepard at the court's conference on September 1, 2011.

doctrine it up,"[21] and "A little studied ambiguity here is probably the best the system will be able to achieve."[22]

Chief Justice Shepard has also been a powerful influence in encouraging and improving the quality of trial court judicial service and administration. From the personal touch of knowing and addressing each trial judge by name at the court's annual series of district meetings, to his encouragement of trial judge participation in state-wide judicial improvement programs and judicial education, to his development of a series of grants for local court improvement projects, and much more, the Chief Justice's keen interest in the work of trial court judges has been apparent and appreciated. With great respect for local trial judges, and, in characteristic Shepard leadership style, several trial court administration measures[23] were successfully implemented not by general state-wide uniform supreme court regulation but rather by assigning specific achievement goals to groups of courts in the same geographical area to develop methods to achieve the objectives in ways most suitable to local needs and bar culture.

In the discharge of his constitutional responsibilities to submit regular reports to the Indiana General Assembly,[24] Chief Justice Shepard has given a series of twenty-five remarkable annual "State of the Judiciary" addresses to joint sessions of the Indiana Senate and House of Representatives, usually also attended by the governor and lieutenant governor. Reflecting his oratorical skills, natural eloquence, and political acumen, his uplifting addresses not only have reported achievements of Indiana's trial and appellate judiciary, usually replete with specific expressions of appreciation for legislative cooperation, but they also have advocated new advancements and envisioned a brighter future. These masterful addresses have been consistently well received by legislators of both parties, often drawing warm praise and support. He has also maintained a warm, cordial, and approachable relationship with legislators and all of Indiana's governors and elected executive officers, regardless of political affiliation. To the applause of many, when Governor Mitch Daniels organized the Indiana Commission on Local Government Reform in 2007, he named Chief Justice Shepard and former Governor Joseph E. Kernan to co-chair the commission.

Significant tangible benefits to the judiciary have resulted from Shepard's inter-governmental effectiveness. The Indiana Constitution was amended to modernize the jurisdiction of the supreme court, enabling it to serve primarily as a court of discretionary review and thereby address legal issues on a greater variety of matters affecting Indiana citizens.[25] Another transformational achievement was legislation revising the method of compensation of Indiana

---

21. Comment by Chief Justice Shepard at the court's conference on November 8, 2007.

22. Comment by Chief Justice Shepard at the court's conference on May 11, 2011.

23. Among these were programs to provide robust pro-bono representation, to improve equalization of case workloads among judges, and to eliminate the possibility of forum-shopping in the filing of criminal cases.

24. IND. CONST. art. 7, § 3.

25. *See* H.R.J. Res. 1, 110th Leg., 2d Reg. Sess. (Ind. 1998) (eliminating cases of life imprisonment and imprisonment greater than fifty years from the court's mandatory jurisdiction).

judges to provide ongoing fair and equitable treatment, which resulted in a major restoration of judicial morale and productivity.[26]

Any reflection on the contributions of Chief Justice Randall Shepard would be incomplete without recognizing his significant national influence and leadership. His efforts with the American Bar Association included service as Chair of the ABA Appellate Judges Conference, 1996-1997, and Chair of the ABA Council of the Section for Legal Education and Admissions to the Bar, 1998-1999. He served as President of the Conference of Chief Justices and Chairman of the National Center for State Courts, 2005-2006. In 2006, United States Chief Justice John Roberts appointed Shepard to the U.S. Judicial Conference Advisory Committee on Civil Rules. In addition, he has done considerable state and national teaching, including serving as an adjunct professor at Indiana University Robert H. McKinney School of law, Yale Law School, and the University of Evansville, and as a Lecturer at the Opperman Institute of Judicial Administration at New York University School of Law.

As a profound visionary with great passion for Indiana judicial excellence; an abiding concern and respect for the citizens whose lives may be affected by the judiciary; a thorough understanding and appreciation for the role and responsibilities of the judiciary as one of the three co-equal branches of government; a deeply embedded personal civility, decency, and integrity; an enormous intellect; and an uncanny gift for inspiring, encouraging, and motivating colleagues and staff to undertake and achieve substantial accomplishments, Chief Justice Randall Shepard has dramatically exceeded his goal of building Indiana's judiciary into "a system of justice that befits a great nation."[27] Under his leadership, the Indiana Supreme Court has become nationally recognized for the quality of its judicial opinions and an abundance of innovative advancements in judicial administration. His beloved state and nation are rightfully grateful for his twenty-five spectacular years as Chief Justice of Indiana and his over twenty-six years of service as a member of the Indiana Supreme Court.[28]

---

26. IND. CODE § 33-38-5-8.1 (2011).

27. Shepard, *The New Role of State Supreme Courts*, *supra* note 3, at 1552.

28. For information regarding Randall Shepard's personal background, see Kevin W. Betz et al., *Randall T. Shepard*, *in* JUSTICES OF THE INDIANA SUPREME COURT 395, 396 (Linda C. Gugin & James E. St. Clair eds., 2010).

RANDALL T. SHEPARD
A TRIBUTE

JAMES P. WHITE*

I am pleased to contribute to this issue of the *Indiana Law Review* honoring retiring Chief Justice Randall T. Shepard who has served for twenty-five years as chief justice of the Indiana Supreme Court and as a member of that court for twenty-seven years. I have been privileged to have been associated with Chief Justice Shepard in many capacities for these past twenty-seven years.

Chief Justice Shepard is a unique individual in the firmament of the American legal profession. Many have spoken and written about his leadership in the advancement of American jurisprudence. He is a leader among the justices of America's highest courts. He is also a leader in American legal education. It is in this context I should like to reflect on his influence and contributions to American legal education.

In preparing this tribute I have reflected upon aspects of Chief Justice Shepard's personality and character and how they relate to legal education, which formed the basis of his professional career as a jurist. He has epitomized the axiom that a legal calling includes not only service to one's clients but to society as a whole.

SHEPARD AND LAW SCHOOLS

In an article in the *Indiana Law Review* entitled "What the Profession Expects of Law Schools," Chief Justice Shepard stated:

> The able organizers of this symposium printed my title in the program, and that is how I learned about it. I am glad they did so. Thinking through the title, "What the Legal Profession Expects of Law Schools," has been an interesting experience.
>
> For one thing, the demands of law practice are such that most members of the profession do not actually expect anything at all from their law schools once they have graduated. Lawyers receive most of what they expect from their school during the three years they spend as students. By and large, they take their diploma and seldom give the matter another thought.
>
> On the other hand, the organs and institutions of the profession, such as the bar associations and the courts, and those individual lawyers who pay close attention to legal education and admissions to the bar, actually do form and articulate discrete expectations with respect to the schools. These expectations, of course, are shifting and often conflicting.
>
> In thinking about both the modest expectations of the great bulk of lawyers and the specific expectations of the organizations, I have come to rest on five enumerated demands. I list them here, in no particular order, and specifically disclaim any authority or presumption to speak for

* Professor of Law Emeritus, Indiana University Robert H. McKinney School of Law. A.B., University of Iowa; J.D., LL.M., George Washington University Law School.

the profession.[1]

He listed the following expectations:

1. Honor my Degree
2. Train Good Lawyers
3. Provide Useful Scholarship
4. Contribute Toward Ethical Conduct; and
5. Honor the Practitioners[2]

In addressing the topic of "Honor my Degree" he stated:

> [T]his chance to move modestly upwards in life along with one's school plays some role in leading graduates to contribute to the success of the alma mater. The new building campaign, the new professorship, or the new lecture series, to name a few building blocks of thriving schools, are all things that likely enhance the value of the ticket which the graduate has received from the school.
>
> This saga in which America's lawyers find greater fulfillment through the world of higher education has run for more than a century. Lawyers and bar associations spent several decades attempting to transfer the training of lawyers from the Nineteenth Century system of reading the law to the present system of university education, believing in part that a trade that required a post-graduate degree in a seat of higher learning would always be thought a more noble or important one. This burnishing of the trade and its schools began ever so modestly, for when AALS established its first membership requirements in 1901, it elected to admit only schools that limited admission to high school graduates.
>
> Our trade has come a long way since then, and lawyers expect we can go further yet. Thus, one of their expectations is that their schools will manage their own affairs so that people will say, "Ah!", when we tell them where we went to law school.[3]

In his work with American law schools, particularly his year as chair of the ABA Section of Legal Education and Admissions to the Bar, he stressed these expectations of the Bench and Bar.

SHEPARD AND OPPORTUNITIES FOR ENTERING THE LEGAL PROFESSION

In 1997, with the support of the Indiana General Assembly and Governor O'Bannon, Indiana approved funding for an Indiana Conference for Legal Education Opportunity. Under Chief Justice Shepard's leadership, Indiana was the first state to adopt a state CLEO program to assist minority, low income or

---

1. Randall T. Shepard, *What the Profession Expects of Law Schools*, 34 IND. L. REV. 7, 7 (2000).
2. *Id.* at 7-13.
3. *Id.* at 8-9.

educationally disadvantaged college graduates to enter law school, graduate, and enter the legal profession.

In his 1997 Annual Report of the Judiciary to the General Assembly of Indiana, Chief Justice Shepard stated:

> I want to talk with you about another area where I want Indiana to be first. All of you know that the demographics of America are changing, with more women in the workforce and more citizens from minority groups. The legal profession is changing, and the people in it are changing. I recently received a notice for the 25th reunion of my law school class. My class was unusual in that it contained about ten percent women and about five percent African-Americans. These were considered groundbreaking numbers. Today, of course, the number of women in law school has risen to [forty-three] percent.
>
> The number of minority students in law schools has been growing, but it is still far too small and the number of minority lawyers serving people even in the urban areas of our state is pitifully small[,] [i]n Lake County, for [example].[4]

He has continued his support for this program by stating:

> The legal profession benefits when we recruit minority applicants and help them succeed. Our job is to ensure all citizens have faith and confidence in our system—faith and confidence that they are truly equal in the eyes of the law. When there is diversity at all levels of the legal system, our entire system better serves society.[5]

In 1998 Chief Justice Shepard was awarded the 1998 Mark of Distinction Award from the National Association for Law Placement (NALP). This award recognizes people or organizations whose programs or practices "effectively and ethically meet the needs of participants in the process of legal employment."[6]

In its citation NALP stated:

> NALP decided to honor Chief Justice Randall T. Shepard of the Indiana State Supreme Court for his visionary leadership in the creation of the first state-sponsored Conference on Legal Education Opportunity (CLEO) Program for minority and disadvantaged students. This innovative residential program provides students with an intensive training and "preview" of legal education that is effective in preparing them for the rigorous challenges of law school. In addition, successful graduates of the program are provided with a significant financial award

---

4. Randall T. Shepard, Chief Justice, Ind. Supreme Court, Annual State of the Judiciary Address to the Indiana General Assembly: Trying Something New (Jan. 30, 1997).

5. Randall T. Shepard, *Hello from CLEO!*, IND. JUD. BRANCH: DIVISION STATE CT. ADMIN., http://www.in.gov/judiciary/cleo/2674.htm (last visited June 7, 2012).

6. *NALP Report: Randall T. Shepard Receives Mark of Distinction Award*, 29 SYLLABUS, no. 2, 1998, at 6, 6.

> to defray the cost of their legal education in each of their three years of law school. While assisting students in doing their academic best, this program will ultimately promote diversity in law schools and the legal profession.[7]

Following Indiana's lead, a number of states subsequently adopted their own CLEO program.

During his chairmanship of the American Bar Association's Section of Legal Education and Admissions to the Bar, Chief Justice Shepard placed particular emphasis on Standard 211 of the ABA Standards for the Approval of Law Schools which provides:

> Standards 211. NON-DISCRIMINATION AND EQUALITY OF OPPORTUNITY
>
> (a) A law school shall foster and maintain equality of opportunity in legal education, including employment of faculty and staff, without discrimination or segregation on the basis of race, color, religion, national origin, gender, sexual orientation, age or disability.
>
> (b) A law school shall not use admission policies or take other action to preclude admission of applicants or retention of students on the basis of race, color, religion, national origin, gender, sexual orientation, age or disability.[8]

Thus, in many ways Chief Justice Shepard has been a leader in expanding opportunities for entry to law school and the legal profession. He has been honored for his efforts by many organizations including the National Black Students Association and the Indiana Civil Rights Commission.

SHEPARD THE EDUCATOR

Throughout his career on the court, Chief Justice Shepard has been a frequent lecturer and visiting professor at America's law schools. He has written more than sixty-five law review articles in thirty different law journals. He has frequently taught at both Indiana University's Robert H. McKinney School of Law and at its Maurer School of Law. He has taught complex litigation at his alma mater, Yale Law School, and regularly teaches at New York University's law school new appellate judges seminar.

He served with distinction as Chair of the American Bar Association's Section of Legal Education and Admissions to the Bar, the entity responsible for accreditation of America's law schools.

Chief Justice Shepard's understanding of legal education is illustrated by an article he wrote for the National Law Journal at the beginning of his chairmanship

---

7. *Id.*

8. ABA Section of Legal Educ. & Admissions to the Bar, STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS 211 (2012).

of the Section of Legal Education and Admissions to the Bar.

> As the practice of law becomes more international, the Section of Legal Education and Admissions to the Bar is focusing increasingly on how to educate lawyers for the new world, how to reform both legal education and bar admissions, and how to build bridges with both the practicing bar and the universities that house America's law schools.
>
> The section is especially interested in opportunities for American students to conduct part of their education abroad, and for foreign lawyers to learn about U.S. law by visiting the United States. U.S. law schools have been expanding their overseas programs rapidly in recent years. There now are 150 programs operating in foreign countries on six continents.
>
> These programs provide excellent experiences for both students and faculty, but can also present management and quality-control problems for their sponsoring institutions. In December, the section will sponsor a Workshop on Overseas Programs to help these schools deal with these issues.
>
> At the same time, the number of foreign law graduates seeking education in the United States is growing. American schools have responded by creating [LL.M.] programs specifically aimed at this interest. In October, the section will conduct a workshop to help schools design better law masters programs.
>
> The section will also be expanding its consultations with other ABA sections and with foreign law societies about the admission of foreign lawyers to U.S. practice. In addition, the section will continue its work in central and eastern Europe and in Africa to help universities and law societies in these regions build stronger systems of legal education and bar admissions.
>
> The internationalization of law practice is only one of the many changes taking place within the profession. Both law schools and bar-admitting authorities must help shape the course of these myriad changes.
>
> The section's Curriculum Committee will work this year to identify innovative ideas for teaching and research, with plans to make the results of that effort available for use as models. The Professionalism Committee will be examining the movement to stage conclaves on legal education in various states, bringing together the bench, the bar and the academy to examine the continuing task of educating lawyers.
>
> Moreover, the section will be deeply involved in assessing the validity and reliability of the standards that the ABA uses to accredit law schools. This year, the section will examine the standards governing educational programs and faculty as part of its commitment to conduct a complete re-examination of every ABA standard over a period of [thirty-six] months.
>
> Reforming legal education is not a mission solely for lawyers and law schools; it must include those responsible for overseeing the

> institutions of higher education in which most law schools sit—namely presidents and provosts. In February, for the first time in five years, the section will convene a meeting of presidents, provosts and deans interested in issues for universities that educate lawyers.[9]

Chief Justice Shepard was nominated to chair the United States Conference of Chief Justices the same year he was scheduled to chair the ABA Section of Legal Education. He declined the Conference of Chief Justices nomination because of his commitment to the Section and legal education. Happily, the Chief Justices Conference tapped him following his Section chairmanship.

During the period of his Section leadership he provided wise and effective direction during a period of turmoil for the Section and for American legal education. He was a calming influence in the Section's interfacing with federal agencies and universities.

In July 2000, as part of the American Bar Association's London meeting, the Section of Legal Education and Admissions to the Bar sponsored a Joint Program with the University of London on the topic of "*Legal Education in the United Kingdom and the United States in the New Millennium*."

At that program, Chief Justice Shepard moderated a program entitled "*Advanced Legal Education, Academic and Continuing*." It was appropriate that Chief Justice Shepard moderated this session of the symposium and particularly the portion dealing with the role of ethics in legal education. He has been a forceful proponent of ethics as part of the legal education experience.

SHEPARD THE LEADER

In a tribute to New York Chief Judge Judith Kaye, Chief Justice Shepard wrote:

> Yet I propose that Judith Kaye be declared great on a quite different basis. It is one that receives far less attention than jurisprudence: building a better system of justice. Achieving important systemic change usually requires that judicial leaders, sometimes at the top of the pyramid and sometimes in the rank and file, decide to place themselves at risk in the arenas where change can be made. It requires a determination that those of us on the bench and in the bar can do more justice in individual cases if the court system is better organized.
>
> Recharting the course of a court system has never been an easy matter. The judiciary in form and inclination is both beholden to what has been done in the past and typically comfortable with it. Moreover, judges value the relative independence that they possess, an aspect of judicial work that confers benefits on litigants but that also leads to intransigence.[10]

---

9. Randall T. Shepard, *Globalization Leads Changes Affecting Practice of Law*, NAT'L L.J., Aug. 3, 1998, at C9.

10. Randall T. Shepard, *Judith Kaye as a Chief Among Chiefs*, 84 N.Y.U. L. REV. 671, 672

Justice Shepard's remarks about Judge Kaye can equally be applied to him. Under his leadership the Indiana court system has become a benchmark for court administrative reform.

Under his leadership, Indiana passed Proposition 2 delineating the responsibilities of the appellate courts in Indiana. He was instrumental in the establishment of the Indiana Pro Bono Commission and a statewide pro bono program funded by IOLTA funds.

He has served as president of the Conference of Chief Justices and the National Center for State Courts, chair of the Appellate Judges Conference of the American Bar Association and chair of the American Bar Association Section of Legal Education and Admissions to the Bar. In 2006, Chief Justice John Roberts appointed Justice Shepard to serve on the U.S. Judicial Conference Advisory Committee of Civil Rules, the only state court judge to receive such an appointment.

Chief Justice Shepard has been in the forefront of support for diversity in the legal profession. He has never wavered in his commitment to diversity. He has the ability to appreciate and to understand difficulties experienced by persons of color. He has always been committed to enlarging access to the legal profession so that it is more representative of American society, and the law school student body has changed not only to have more women and persons of color, but also persons from a variety of backgrounds, experiences and ethnic groups. He is forthright in addressing important issues relating to access to the profession.

He often is called upon for new public service such as co-chairing the Kernan-Shepard Indiana Commission on Local Government Reform or serving on the Butler University Presidential Search Committee.

Perhaps the greatest tribute to Randall Shepard can be stated thus—he has a quiet manner but is clear and genuine in his aspiration. He embodies the intellectual curiosity of an academic with the skills, wisdom and practicality of a jurist. His enthusiasm is genuine, sincere and effective. He possesses low-key but irresistible powers of persuasion. He knows and understands legal education and the legal profession. He has an understanding of what are the current successes and problems and what and how legal education and the profession should, will, and must change.

I am proud to have Randall Shepard as a friend, colleague, advisor and helper. He is a person of great character and strength, a person of conviction and principle. His contribution to the betterment of legal education and the profession are many. They will not diminish with his judicial retirement. He will continue to give of himself to the betterment of all.

(2009) (citation omitted).

Indiana Law Review

Volume 45 **2012** **Number 3**

ARTICLES

HONOR AMONGST THIEVES: ORGANIZED CRIME AND THE ILLICIT ANTIQUITIES TRADE

KIMBERLY L. ALDERMAN*

TABLE OF CONTENTS

ABSTRACT

Government agencies, non-profits, scholars, and advocacy groups alike assert that organized crime dominates the illicit antiquities trade. The illicit antiquities trade has been linked to money laundering, extortion, the drug and arms trades, terrorism and insurgency, and even slavery. This Article considers the connection between organized crime and the illicit antiquities trade, examines known criminal subcultures and evidence of their involvement in the trade, and

* J.D., Howard University School of Law; B.A., University of Maryland at College Park; M.A., ARCA International Art Crime Program.

analyzes lateral cooperation between loosely organized criminal groups. Finally, the Article poses the broader question of whether this lateral cooperation suggests that the antiquities trade as a whole operates as an organized criminal industry.

INTRODUCTION

Government agencies,[1] non-profits,[2] scholars,[3] and advocacy groups[4] alike assert that organized crime fuels the illicit antiquities trade. The illicit antiquities trade has been linked to organized criminal activities including money laundering,[5] extortion,[6] the drug and arms trades,[7] terrorism and insurgency,[8] and

---

1. *See The FBI Art Theft Program and Its Impact on Collecting: A Report from FBI Special Agent Robert Wittman and the Editor*, AM. SOC'Y ARMS COLLECTORS BULL. 52 (91st ed. 2005) [hereinafter *FBI Report*] ("[I]llicit trafficking of cultural property is often associated with various other criminal activities, such as organized crime, money laundering, extortion, and fraud.").

2. *See Art Crime Facts*, ASS'N FOR RES. INTO CRIMES AGAINST ART, http://www.artcrime.info/facts.htm (last visited June 14, 2012) [hereinafter ARCA Website] ("Since [1961] most art crime is perpetrated by, or on behalf of, Organized Crime, thereby fueling their other activities, including the drug and arms trades and terrorism . . . Art crime funds all organized crime enterprises, including terrorism.").

3. *See* Edgar Tijhuis, *Trafficking in Stolen Art, Antiques and Cultural Property*, *in* ORGANIZED CRIME: FROM TRAFFICKING TO TERRORISM 223, 223 (Frank G. Shanty ed., 2008) ("One of the most diverse and international types of organized crime is the trade in stolen art and antiquities.").

4. *See* Blythe Bowman, *Plunder of Antiquities: A Crime of Global Proportions*, SAVING ANTIQUITIES FOR EVERYONE (2008), http://www.savingantiquities.org/crime.php ("The drastic increase in the current market values of antiquities has precipitated an infiltration and monopoly of the black market by organized criminal syndicates. Plunder of ancient objects has become a thriving industry for these groups.").

5. *See FBI Report*, *supra* note 1, at 52; *see also* Patrick Hunt, *International Law and the Ethics of Antiquities Trafficking*, 11 STAN. J. INT'L REL. 80, 81 (2010) ("[T]here is growing concern that antiquities smuggling is often a money laundering operation for other criminal activity like narcotics, arms, and slavery.").

6. *See FBI Report*, *supra* note 1, at 52.

7. *See id.*; *see also* Hunt, *supra* note 5, at 81; ARCA Website, *supra* note 2; Gregory Elich, *Spoils of War: The Antiquities Trade and the Looting of Iraq*, CENTRE FOR RES. ON GLOBALISATION (Jan. 3, 2004), http://www.globalresearch.ca/articles/ELI401A.html ("[T]he trade in antiquities is closely intertwined with that of the drug trade.").

While the same trafficking channels are often used to move both drugs and antiquities, actual examples of this "intertwining" are rare. *See* NEIL BRODIE ET AL., STEALING HISTORY: THE ILLICIT TRADE IN CULTURAL MATERIAL 18 (2000), *available at* http://www.mcdonald.cam.ac.uk/projects/iarc/research/illicit_trade.pdf (describing such examples, one of which was, "In January 1999, Spanish police broke up a smuggling ring that had been planning to trade stolen art and antiquities for cocaine.").

8. Elena Becatoros, *Smuggled Antiquities Funding Iraq Extremists, U.S. Says*, ASSOCIATED

slavery.[9] Art crime experts have charged that organized crime has occupied a dominant role in the illicit antiquities trade since the early 1960s.[10] Despite these assertions, the connection between organized crime and the illicit antiquities trade is rarely explained or substantiated. Some have charged that the alleged connections are specious at best.[11]

This Article analyzes the repeated assertion that organized crime and the illicit antiquities trade are inextricably linked. In Part I, the Article examines definitions of organized crime, both vernacular and criminological, and suggests a definition for use in examining the connection between organized crime and the illicit antiquities trade. In Part II, this Article examines known criminal subcultures and evidence of their involvement in the trade, applying the suggested definition of organized crime to known excavators, middlemen, retailers, and owners of illicit antiquities to determine whether the posited connection to organized crime can be substantiated.

In Part III, this Article analyzes lateral cooperation between smaller, loosely organized criminal groups, posing the question of whether the antiquities trade as a whole is an organized criminal industry. The Article concludes that there is a connection between organized crime and the illicit antiquities trade if one employs a broader, criminological definition of organized crime. The conclusion further suggests a uniform definition of organized crime for use in discussions about the illicit antiquities trade, and disclosure of that definition in order to facilitate meaningful discourse on organized crime and the illicit antiquities

---

PRESS (Mar. 19, 2008), http://news.nationalgeographic.com/news/2008/03/080319-AP-iraq-insurg.html ("Bogdanos, a New York assistant district attorney . . . said the link between extremist groups and antiquities smuggling in Iraq was 'undeniable.'"); *see also* Hunt, *supra* note 5, at 81.

9. Hunt, *supra* note 5, at 81.

10. ARCA Website, *supra* note 2. Organized criminal groups have been charged with perpetuating up to 80% of art crime (illicit antiquities crime is considered by many to be a type of art crime). *Interview with Noah Charney*, VENTIQUATTRO (Apr. 14, 2007), *available at* http:/www.noahcharney.com/bio.htm (stating that international organized crime syndicates "have a division for stealing cars, another for drug trafficking, another for art crime, functioning like a huge multinational corporation," but concedes that no good statistics exist worldwide).

11. *See* Becatoros, *supra* note 8 ("Antonia Kimbell, an art trade manager at The Art Loss Register . . . said she had yet to see concrete evidence connecting the trade in illegal antiquities and insurgent financing."); *see also* Jonathan Lopez, *Stealing and Dealing*, ART & ANTIQUES, Summer 2010, at 42, 42 (asking Robert K. Wittman, founder of the FBI Art Crime team, "From time to time I've been told that most of the world's biggest art heists are carried out by organized crime syndicates, who then use the stolen works as collateral to finance drug deals, arms trafficking or even acts of terrorism," to which Wittman responds, "There are a lot of people who make things up to get their names in the paper."); Kimberly L. Alderman, *Does the Illicit Antiquities Trade Fund Terrorism?*, CULTURAL PROP. & ARCHAEOLOGY L. NEWS (Oct. 26, 2010), http://culturalpropertylaw.wordpress.com/2010/10/26/does-the-illicit-antiquities-trade-fund-terrorism; Peter Tompa, *Wikileaks, Tariq Aziz, and Archaeological Amnesia*, CULTURAL PROP. OBSERVER BLOG (Oct. 26, 2010, 3:39 AM), http://culturalpropertyobserver.blogspot.com/2010/10/wikileaks-tariq-aziz-and-archaeological.html.

trade.

I. Defining Organized Crime

There is significant dispute over whether organized crime is truly to blame for the flourishing illicit antiquities trade.[12] There are multiple meanings of the term "organized crime,"[13] and this impedes meaningful discourse on the subject. "Organized crime" (and especially "organized criminal syndicates"[14]) evokes imagery of Cosa Nostra, the Japanese Yakuza, and the Irish Mob. However, criminological definitions are broader than the vernacular use of the term. This contradiction causes misunderstandings in discussions about organized crime's role in the illicit antiquities trade.

The Federal Bureau of Investigation (FBI) defines organized crime as "any group having some manner of a formalized structure and whose primary objective is to obtain money through illegal activities."[15] This definition does not require that the activities be criminal, merely illegal; violation of domestic or international civil law would qualify as potential activities for organized criminal groups under the FBI's definition.

Compare the definition of organized crime in the United Nations Convention Against Transnational Organized Crime: "[A] structured group of three or more persons . . . acting in concert with the aim of committing one or more serious crimes or offences . . . in order to obtain, directly or indirectly, a financial or other material benefit."[16] The Convention defines a serious crime as one with a maximum punishment of at least four years incarceration.[17]

The UN definition of organized crime is broader than the FBI definition in the respect that it includes antiquities crime motivated by non-monetary benefits such as ideology, whereas the FBI definition does not. Further, the FBI definition requires a level of formal organization that would exclude many of the groups currently operating within the illicit antiquities trade. Under the UN

---

12. *See infra* note 13; s*ee also* Mark Durney, *Art Theft and Organized Crime?*, CULTURE DEV. (June 17, 2010), http://www.cultureindevelopment.nl/News/Discussing_Looting/365/Art_Theft_and_Organized_Crime%3F.

13. *See* Bojan Dobovšek, *Organised Crime—Can We Unify the Definition?*, POLICING CENT. & E. EUR. (1996), *available at* http://www.ncjrs.gov/policing/org323.htm ("There is no generally accepted definition of organised crime yet, due mostly to the quick development and changing of the forms in which organised crime appears."); *see also* William P. Jennings, Jr., *A Note on the Economics of Organized Crime*, 10 E. ECON. J. 315, 315 (1984) ("[T]here is no apparent consensus among researchers regarding the appropriate definition of the organized crime . . . .").

14. *See* Bowman, *supra* note 4.

15. *Glossary of Terms*, FED. BUREAU INVESTIGATION, http://www.fbi.gov/about-us/investigate/organizedcrime/glossary (last visited June 14, 2012) [hereinafter *FBI Glossary*].

16. United Nations Convention Against Transnational Organized Crime, G.A. Res. 55/25, art. 2(a), U.N. Doc. A/RES/55/25 (Nov. 15, 2000), *available at* http://www.unodc.org/documents/treaties/UNTOC/Publications/TOC%20Convention/TOCebook-e.pdf.

17. *Id.* at art. 2(b).

definition, smaller, informally organized groups that incidentally participate in the illicit antiquities trade would be considered organized criminal actors. The FBI definition is broader only in the respect that it includes civil offenses as well as criminal ones.

The level of organization in the illicit antiquities trade varies. Syndicates range from "smaller organized gangs to the Sicilian and Russian mafias."[18] Depending on the breadth of the definition used, legal actors may engage in activities that amount to organized crime.[19] For the purposes of this Article, and to create a broad definition of organized crime, a fusion of the FBI and UN definitions will be used:

> *A structured group of three of more persons, acting in concert to violate the law, whose primary objective is to obtain a financial or other material benefit.*

It will be in the context of this definition that the connection between organized crime and the illicit antiquities trade will be examined.

II. CRIMINAL GROUPS IN THE ILLICIT ANTIQUITIES TRADE

There is no clear distinction between the licit and the illicit antiquities markets. Archaeological materials are traded on the same market, regardless of whether they have been tainted[20] by an illegal act. Ricardo J. Elia of Boston University has explained, "People think that there is an illicit and a legitimate market. In fact, it is the same."[21] It is worth noting that not all participants in the illegal trade are illegal actors; legal actors often participate in illegal transactions, either knowingly or unknowingly.[22]

This makes identifying distinct criminal groups in the context of the illicit antiquities trade difficult. Motivation for illegal acts ranges from profit, to ideology, to neuroses. Some groups are organized for the purpose of engaging

---

18. Noah Charney, *Stolen Masterpieces: How Organized Crime Targets the International Art World,* CRIME REP. (Mar. 11, 2010, 5:00 AM), http://www.thecrimereport.org/archive/stolen-masterpieces-how-organized-crime-targets-the-international-art-world.

19. A.J.G. TIJHUIS, TRANSNATIONAL CRIME AND THE INTERFACE BETWEEN LEGAL AND ILLEGAL ACTORS 15-40 (2006).

20. *See* Kimberly L. Alderman, *Ethical Trade in Cultural Property: Ethics and Law in the Antiquity Auction Industry*, 14 ILSA J. INT'L & COMP. L. 549, 560 (2008) [hereinafter Alderman, *Ethical Trade*] ("A tainted item does not carry with it the appearance of having been illegally obtained or exported, so physical characteristics alone rarely reveal an object to be the fruit of past illegal acts.").

21. Mark J. Petr, *Trading Places: Illicit Antiquities, Foreign Cultural Patrimony Laws, and the U.S. National Stolen Property Act After* United States v. Schultz, 28 HASTINGS INT'L & COMP. L. REV. 503, 504 (2005) (quoting Ricardo J. Elia, *cited in* Barry Meier & Martin Gottlieb, *Loot, Along the Antique Trail: An Illicit Journey Out of Egypt, Only a Few Questions Asked*, N.Y. TIMES, Feb. 23, 2004, *available at* http://nytimes.com/2004/02/23/international/23ANTI.html).

22. *See* TIJHUIS, *supra* note 19, at 31.

in the illegal trade, while others participate only incidentally. Organizations range in size from transnational criminal syndicates to family-run operations.

Despite the oft-repeated assertion that organized crime controls the illicit trade, there has been little scholarly treatment of the claim.[23] Following are the groups that have been identified in scholarly or popular literature that (1) have a known pattern of participation in the illegal antiquities trade, and (2) meet the definition of organized crime suggested in Part I of this Article (a structured group of three or more people, acting in concert to violate the law, in order to obtain a financial or other material benefit).

A. Criminal Excavators

There is wide variety in the people and groups that illegally excavate archaeological materials. Most are motivated by poverty, but in some areas, there is substantial profit to be made in plundering.[24] Generally, illegal excavators receive less than one percent of the final market value of their discoveries, but the proceeds are a boon to them nonetheless.[25]

As demonstrated in this subsection, the majority of criminal excavators work in organized criminal groups as defined in the broad definition in Part I of this Article. They work in crews of three or more, often with a foreman or sponsor, and collaborate with a local middleman to pawn off their finds.[26] The objective of these operations is almost always financial benefit. One criminal excavator explained, "[T]here's no alternative for me or for my men. We work to put food on the table for our families. . . . We are all unemployed . . . ."[27]

1. Subsistence Diggers.—Subsistence diggers (usually called subsistence looters[28]) are motivated by poverty to excavate archaeological materials in their local region in violation of national law.[29] Source nations (those from which archaeological materials most often originate) tend to be less prosperous than

---

23. *See generally* VITO ANDREA IANNIZZOTTO, BENI CULTURALI E CRIMINALITÀ ORGANIZZATA (2001) (an exception to this general rule, exploring the connection between culture and organized crime, available only in Italian).

24. *See* Lisa J. Borodkin, *The Economics of Antiquities Looting and A Proposed Legal Alternative*, 95 COLUM. L. REV. 377, 378, 406 (1995).

25. *See, e.g.*, *id.* at 378 & n.10 (describing instances in Turkey and the United States where finders received very small amounts compared to the ultimate selling price of a marble statute and a Mimbres pot, respectively).

26. *See id.* at 378.

27. *See* Cristina Ruiz, *My Life as a Tombarolo*, ART NEWSPAPER No. 112, Mar. 2001, at 36, *available at* http://www.museum-security.org/tombarolo.htm.

28. This Article does not use the more common term "subsistence looter" because it implies a value judgment not helpful to the analysis herein.

29. John Alan Cohan, *An Examination of Archaeological Ethics and the Repatriation Movement Respecting Cultural Property (Part Two)*, 28 ENVIRONS ENVTL. L. & POL'Y J. 1, 8-9 (2004) (describing subsistence looters as "trapped in the abyss of poverty"); Jonathan S. Moore, *Enforcing Foreign Ownership Claims in the Antiquities Market*, 97 YALE L.J. 466, 486 (1988).

market nations (those in which archaeological materials are most often sold).[30] This economic disparity motivates locals to harvest untapped archaeological resources, often in agricultural off-seasons or when yields are poor.[31] David Matsuda, a scholar who has rigorously studied subsistence digging, has estimated that ninety-seven to ninety-nine percent of illicit excavations in Central America are performed by localized subsistence diggers.[32]

In Belize, subsistence diggers are called "huecheros."[33] Huecheros sell illicitly excavated artifacts to a wholesaler, who can move them to dealers.[34] They call their finds "semilla," meaning seed—planted by their ancestors to dig up and sell for money or actual corn seed.[35] Huecheros demonstrate how not all subsistence diggers consider themselves to be doing something illicit or criminal. Matsuda has explained, "Many of Latin America's indigenous peoples see themselves as the legitimate heirs to both seeds and artifacts, which are conceived of as ancestors' gifts, given to humanity by real or mythological patrons to be harvested, or excavated, as it were, by later generations."[36] Matsuda interviewed 400 subsistence diggers and concluded that most of them were "refugees from civil violence and economic despair."[37]

Subsistence digging is also widespread in Iraq and Afghanistan.[38] One Afghani villager explained, "[The villagers] tried to earn as much money as possible with their finds, to put food on the table for their children. You don't do this for fun. . . . If you're not desperate, you never do this."[39] Subsistence digging is not easy work; in the Middle East, men work in teams to dig holes straight down, and then tunnels radiating out on the horizontal plane in a star

---

30. Alderman, *Ethical Trade*, *supra* note 20, at 551 n.6.

31. *See, e.g.*, Aaron M. Boyce, *A Proposal to Combat the Illegal Trafficking of Pre-Columbian Artifacts*, 3 HISP. L.J. 91, 92 & n.3 (1997). David Matsuda has estimated that up to 500,000 people conduct seasonal subsistence digging in El Salvador and Guatemala, 30,000 to 50,000 in Belize, and 250,000 in the northern state of Jalisco in Mexico. David Matsuda, *The Ethics of Archaeology, Subsistence Digging, and Artifact Looting in Latin America: Point, Muted Counterpoint*, 7 INT'L J. CULTURAL PROP. 87, 91 (1998).

32. Matsuda, *supra* note 31, at 91.

33. *Id.* at 88.

34. *See id.*

35. *Id.* at 87-88.

36. *Id.* at 88.

37. *Id.* at 91 (explaining that "decades of prolonged civil violence forced subsistence farmers into unfamiliar ecosystems . . . . It is in these outback areas—with no seed crops, poor soil, and unstable weather patterns—that the abundance of uncharted archaeological ruins became a viable socioeconomic alternative to starvation").

38. Peter Brems & Wim Van den Eynde, *Blood Antiques* (LinkTV broadcast Oct. 8, 2009), *available at* http://www.journeyman.tv/?lid=59906&tmpl=transcript; John Russell, *Looters in Iraq* (2004), *available at* http://www.earlham.edu/~vanbma/20th%20century/images/1967daytwentyfour06.htm (showing photograph of looters at Isin waving at the photographer, who is overhead in a helicopter).

39. Brems & Van den Eynde, *supra* note 38.

shape—called an underground gallery.[40] In Tepe Zargaran, a site plundered in the 1990s,[41] the galleries ran at a depth of twenty feet.[42]

Some sites in southern Iraq have been so extensively plundered that they look like "moonscapes."[43] Archaeologist Maguire Gibson observed from a helicopter approximately 200 looters working in broad daylight at the site of Adab, and 200 to 300 at Isin.[44] In a 2004 photograph by John Russell, dozens of looters at Isin appear to be waving at him as he flies overhead in a helicopter.[45]

2. Tombaroli.—In Italy, illegal excavators are called "tombaroli,"[46] and they may or may not be impoverished.[47] Tombaroli work in crews to search the Italian countryside, usually under cover of night, for ancient tombs and burial chambers.[48] Crews operate under the supervision or authority of a foreman.[49] They sometimes use "mechanical digger[s] to break into the roofs of tombs."[50] Then, several men dig inside the chamber while a lookout watches for police, sometimes using infrared technology.[51] In an anonymous interview with a journalist, one tombarolo, who had his own team, reported that each tomb yielded between $3300 and $5700 worth of artifacts, of which the tombarolo took sixty to seventy percent, splitting the rest among his men.[52]

Burial tombs in Italy are often located on what is now farmland; experienced tombaroli use subtle signs such as changes in vegetation patterns to identify likely locations for these tombs.[53] Tombaroli may give the nearby farmers a cut

---

40. *Id.*

41. *Afghanistan's Ancient Treasures a Worrying Modern-Day Trade*, NEWS INT'L (Mar. 20, 2007), http://www.rawa.org/temp/runews/2007/03/20/afghanistan-s-ancient-treasures-a-worrying-modern-day-trade.html.

42. Brems & Van den Eynde, *supra* note 38.

43. Elich, *supra* note 7, para. 27; Russell, *supra* note 38 (showing photograph of Iraqi sites).

44. Elich, *supra* note 7, para. 27.

45. Russell, *supra* note 38 (illustrating Russell's 2004 photo, "Looters in Iraq").

46. Meaning, literally, "tomb raiders" or "grave robbers." Sue J. Park, *The Cultural Property Regime in Italy: An Industrialized Source Nation's Difficulties in Retaining and Recovering Its Antiquities*, 23 U. PA. J. INT'L ECON. L. 931, 931 n.1 (2002).

47. *Compare* Ruiz, *supra* note 27 (recounting the story of a tombarolo who explained, "I'd like to have an honest job, to spend my nights in bed with my wife . . . , but there's no alternative for me or for my men. We work to put food on the table for our families."), *with* Ariel David, *Modern-Day 'Tomb Raiders' Feel the Heat*, ASSOCIATED PRESS, July 6, 2007, http://www.msnbc.msn.com/id/19585643/ (telling the story of Pietro Casasanta, who "worked undisturbed in daylight with a bulldozer, posing as a construction worker . . . to dig into ancient Roman villas and unearth statues, pottery and other artifacts which he then sold for millions of dollars on the illegal antiquities market").

48. Park, *supra* note 46, at 931.

49. Ruiz, *supra* note 27.

50. BRODIE ET AL., *supra* note 7, at 27.

51. Ruiz, *supra* note 27.

52. *Id.*

53. *Id.*

of the profits for illegal excavations, although farmers are becoming less cooperative due to fear of state seizure.[54] One tombarolo explains the tombarolo's relationship with the public and each other, "We're not like the Mafia; there is no violence, but there is a code of honour."[55]

Tombaroli occasionally sell directly to local collectors, if they have the right connections and find an object they know a particular collector will like.[56] Some more seasoned tombaroli also make money by "selling" tombs, pointing less experienced tombaroli to unexcavated sites, and taking a flat fee or a cut for the tip off.[57] Like most criminal excavators, however, tombaroli usually sell to wholesalers.[58]

3. Terrorists.—There is little in the way of concrete evidence that terrorist groups participate in the illicit antiquities trade in any capacity other than incidentally. The occasional episodes brought to light by the media serve as indicators that some connection exists, although the extent of the connection remains contested.[59]

In 2005, the German secret service released that Mohammed Atta, the hijacker of the first plane to strike the World Trade Center on September 11, 2001, had attempted to sell Afghan antiques to a Göttingen professor in 1999.[60] Atta told the professor that he needed the money to purchase an aircraft.[61] This is the most cited example of the connection between terrorism and the illicit antiquities trade.

In a recent interview, Amir Ganor, the Director of the Robbery Prevention Unit of the Israel Antiquities Authority, stated that some of the antiquities looters in Israel are "connected to terror," and that terrorists "use the money [earned] from the antiquities to buy weapons."[62] He recounted a story where his unit caught a team of terrorists trying to sell an illegally excavated sarcophagus for "a lot of money."[63]

The 2009 documentary "Blood Antiques" demonstrates how antiquities from Iraq and Afghanistan are transported to Europe (Brussels in particular) in an

---

54. *Id.*

55. *Id.*

56. *Id.*; *see also* Park, *supra* note 46, at 931.

57. Ruiz, *supra* note 27.

58. Christine L. Green, Comment, *Antiquities Trafficking in Modern Times: How Italian Skullduggery Will Affect United States Museums*, 14 VILL. SPORTS & ENT. L.J. 35, 53 (2007).

59. *See* Dobovšek, *supra* note 13.

60. Cristina Ruiz, *9/11 Hijacker Attempted to Sell Afghan Loot*, ART NEWSPAPER, Feb. 2010, http://theartnewspaper.com/articles/9-11-hijacker-attempted-to-sell-Afghan-loot%20/20188; *Todespilot Atta Wollte Terroranschläge vom 11. September 2001 Durch Illegalen Kunsthandel Finanzieren*, DER SPIEGEL, July 16, 2005, http://www.spiegel.de/spiegel/vorab/0,1518,365376,00.html [hereinafter DER SPIEGEL].

61. DER SPIEGEL, *supra* note 60.

62. Eti Bonn-Muller, *Amir Ganor, Director, Robbery Prevention Division,* ARCHAEOLOGY, http://www.archaeology.org/israel_antiquities_authority/ganor.html (last visited June 15, 2012).

63. *Id.*

organized fashion.[64] The documentary argues that because terrorist organizations "run the show" in the "unstable countries" where the antiquities originate, "a large part" of the estimated €5 billion[65] spent on stolen antiquities annually finances terrorism.[66] The documentary offers anecdotal evidence that the Taliban harvests antiquities in the areas under their control, using the proceeds "to finance their war."[67]

New York assistant district attorney and U.S. Marine Reserve Colonel Matthew Bogdanos has attested that "the link between terrorism and antiquities smuggling [is] undeniable."[68] He argues that in late 2004, the Sunnis and al-Qaida in Iraq started using illicit antiquities as a source of funding, and the Shiite militias later followed suit.[69] However, he has said he is not able to release the hard data or intelligence that forms the factual basis for his assertions due to its sensitive nature.[70]

---

64. Brems & Van den Eynde, *supra* note 38.

65. The documentary does not explain where this number comes from, but it is a commonly used figure. According to the FBI Art Theft Program, "Art and cultural property crime—which includes theft, fraud, looting, and trafficking across state and international lines—is a looming criminal enterprise with estimated losses running as high as $6 billion annually." *Art Theft*, FED. BUREAU INVESTIGATION, http://www.fbi.gov/about-us/investigate/vc_majorthefts/arttheft/ arttheft (last visited June 15, 2012). However, INTERPOL cautions, "[I]t is very difficult to gain an exact idea of how many items of cultural property are stolen throughout the world and it is unlikely that there will ever be any accurate statistics." *Frequently Asked Questions*, INTERPOL, http://www.interpol.int/Crime-areas/Works-of-art/Frequently-asked-questions (last visited June 15, 2012).

66. Brems & Van den Eynde, *supra* note 38. This echoes the sentiment expressed by art collector Arthur Brand, who commented, "The obvious link is when a hundred thousand people in the United States buy antiquities coming out of the [sic] Iraq, Iran, Lebanon and Pakistan. . . . Nine out of [ten] artifacts that come out of the Middle East are controlled by Hezbollah and Islamic Jihad." Laura de la Torre, *Terrorists Raise Cash by Selling Antiquities*, GOV'T SECURITY NEWS, Feb. 20, 2006, http://www.savingantiquities.org/up-content/pdf/GSNarticle.pdf. The author conceded that "Brand's numbers cannot be confirmed." *Id.*

67. One Afghani trafficker reported, "The Taliban are also in this business in the areas under their control. They send pieces abroad. We've heard they've found something in Nimruz." Brums & Van den Eynde, *supra* note 38. An Afghani villager explained, "We used to graze our cattle here until the Taliban came. The Taliban started to dig here. They worked for [forty] days and nights. They dug a tunnel, but we don't know what they found. Everyone said something different. Some said they'd found a mummy. There were lots of [rumors], but the Taliban never let anyone near." *Id.*

68. *See Antiquity Smuggling Finances Terror*, ASSOCIATED PRESS, Mar. 19, 2008, http://www.heritagewatchinternational.org/km/antiquity-smuggling-finances-terror.html [hereinafter *Antiquity Smuggling*]; ARCA Website, *supra* note 2.

69. *Antiquity Smuggling*, *supra* note 68.

70. de la Torre, *supra* note 66; *see also* Matthew Bogdanos, *Thieves of Baghdad: Combatting Global Traffic in Stolen Antiquities*, 31 FORDHAM INT'L L.J. 725, 730 (2008) (alleging that antiquities are the new "cash crop" for insurgents, then explaining, "We do not have hard numbers—the traffic in art for arms is still too recent a phenomenon, and some of the investigations

After 9/11, Bogdanos investigated the looting of the Iraq museum alongside agents from Immigration and Customs Enforcement (ICE).[71] He has described a December 2005 raid of a terrorist bunker in northwest Iraq that yielded automatic weapons, ammunition stockpiles, infrared goggles, and uniforms alongside "30 vases, cylinder seals[72] and statuettes that had been stolen from the Iraq Museum."[73]

ICE is the primary agency in the United States responsible for policing the illicit antiquities trade, and its representatives appear less convinced of the link between terrorism and antiquities. ICE Senior Special Agent James McAndrew acknowledged only the possibility of that link.[74] Antonia Kimbell at the Art Loss Register also awaits hard evidence of the connection, explaining, "We haven't come across a direct link."[75]

4. Mafia and Gangs.—While the claim that organized crime controls the illicit antiquities trade evokes imagery from a James Bond story,[76] there is no known evidence that international criminal syndicates are involved in the trade.[77] However, reports of local mafias becoming involved in the illicit excavation and trafficking occasionally emerge, most often out of Eastern Europe. Although mafias and gangs are listed under criminal excavators in this Article, various local mafias have been reported to serve in three roles—as illegal excavators, wholesalers, and traffickers.

Bulgaria's mafia, for instance, has been increasingly involved in illegal excavation and trade in recent years. In 2008, Bulgarian prime minister Sergey Stanishev alleged that Bulgaria's mafia had a key role in trafficking "everything from sex slaves to antiquities."[78] One news report explained: "Tens of

---

remain classified because of the connection to terrorists.").

71. *See* de la Torre, *supra* note 66.

72. Cylinder seals are particularly valuable, retailing for up to $500,000 with collectors. *See* de la Torre, *supra* note 66; Elich, *supra* note 7, para. 5 (describing seals as "[e]steemed by collectors" and "highly marketable"). Although excavators or wholesalers in source nations usually receive only a small fraction of the final retail value of an object, dealing in very valuable objects can still prove profitable. *Id.* para. 14.

73. de la Torre, *supra* note 66; *see also* Becatoros, *supra* note 8.

74. "I think it's definitely possible without a doubt and probably more than likely probable," McAndrew has explained. de la Torre, *supra* note 66.

75. Becatoros, *supra* note 8.

76. Consider SPECTRE (SPecial Executive for Counter-intelligence, Terrorism, Revenge and Extortion), the fictional global terrorist organization from the James Bond stories. *See generally* IAN FLEMING, THUNDERBALL (1961).

77. Some art crime experts will protest because there is evidence that international criminal syndicates are involved in the trade of stolen art. However, as to the illicit *antiquities* trade in particular, the evidence of their involvement is limited to the type described in Part II of this Article.

78. Harry de Quetteville, *Bulgaria Acts to Smash Mafia's EU Network*, TELEGRAPH, May 6, 2008, http://www.telegraph.co.uk/news/worldnews/europe/bulgaria/1932962/Bulgaria-acts-to-smash-mafias-EU-network.html.

thousands of tomb raiders are systematically stripping Bulgaria. In some parts of the country, whole villages have taken up tomb-raiding and many of the digs are organised by the local mafia."[79]

Bulgaria's Unit for Combating Organized Crime (the country's anti-mafia police) has a department specifically assigned to address antiquities smuggling.[80] This department has conducted targeted raids to break up what are described as "antique smuggling rings."[81] Volodya Velkov, head of the Unit, has reported that there are "200 to 300 illegal treasure hunting groups in Bulgaria, digging on orders from customers and having top of the line equipment such as 3D radars at the cost of USD [$50,000] each."[82] Velkov believes there are five existing channels for smuggling illicit antiquities from Bulgaria to the United States, with the main traffic moving through Austria.[83]

This echoed a similar story that emerged from the Ukraine in 2002, when mafia groups reportedly added looting to existing drug, prostitution, and protection rackets.[84] Mafia families were reported to directly employ archaeologists "after making them an offer that they can't refuse."[85] Some were running night digs on sites that were being legitimately excavated during the day.[86] Reports indicated that "archaeologists employed by the mafia gangs move[d] in on promising historical sites with diggers, floodlights, and armed guards to keep out concerned academics, rival gangs, and Ukraine's under-funded police force."[87]

The use of local mafias for illegal excavation is not limited to Eastern Europe. Russian officials have identified over forty Russian organized crime groups as dealing in stolen art and cultural property.[88] In Central America, it is estimated that one to three percent of criminal excavators are full-time artifact looters or smugglers working within "the command and control structure of Mafia-like consortiums" and interfacing with international markets.[89]

Beyond formally organized national mafias and gangs, there are also

---

79. Malcolm Moore, *Tomb Raiders Strip Bulgaria of Its Treasures*, DAILY TELEGRAPH, Aug. 29, 2007, *available at* http://novinite.com/view_news.php?id=84723.

80. *Bulgaria with 50 000 Illegal Antique Collections,* SOFIA NEWS AGENCY, Nov. 6, 2010, http://novinite.com/view_news.php?id=121892 [hereinafter *Bulgaria Collections*].

81. *Bulgaria Anti-Mafia Police Break up Antique Smuggling Ring*, SOFIA NEWS AGENCY, Apr. 30, 2010, http://mafiatoday.com/general-breaking-news/bulgaria-anti-mafia-police-break-up-antique-smuggling-ring; *Bulgaria Collections*, *supra* note 80.

82. *Bulgaria Collections*, *supra* note 80.

83. *Id.*

84. Murdo MacLeod, *Archaeology Is New Target for Ukraine's Mafia Gangs*, SCOTSMAN, Oct. 13, 2002, http://www.scotsman.com/news/international/archaeology_is_new_target_for_Ukraine_s_mafia_gangs_1_1377655.

85. *Id.*

86. *Id.*

87. *Id.*

88. *See FBI Report*, *supra* note 1.

89. *See* Matsuda, *supra* note 31, at 91.

incidents of smaller gangs operating opportunistically in the illicit trade. In the days following the U.S. entry into Baghdad in 2003, Iraq's National Museum was left with few guards, one of whom explained, "Gangs of several dozen came. Some had guns. They threatened to kill us if we did not open up."[90] The guards unlocked the doors and the mob pushed their way through, although some still chose to smash and enter through a glass window.[91]

Many members of the mob quickly seized furniture and electronics.[92] Others, however, were well-dressed, "European-looking" men talking on cell phones or walkie talkies while perusing the available antiquities.[93] Professionals stormed in with equipment to lift heavier pieces, equipment to break into vaults, and glasscutters of a type not found in Iraq.[94] Some experts believe that unscrupulous art dealers had placed orders in advance, anticipating that the limited guard on the National Museum would inevitably fail in the crux of war.[95]

Even if professional thieves were involved in the plundering of Iraq's National Museum, some artifacts were later recovered from residents who took them to prevent them falling into the hands of looters.[96] Others were recovered during raids by occupation troops and searches by customs officials.[97] Overall, however, there has been a low recovery rate of artifacts looted from the National Museum, and experts believe this supports the idea that professional thieves were involved.[98]

A similar scene played out in the northern city of Mosul after the fall of the Iraqi government. The first thieves who entered the museum there appeared to be "acting on behalf of Western dealers."[99] They took only genuine pieces, ignoring replicas, and when they went through the basement library, they took only the rarest manuscripts.[100]

The use of gangs with professional thieves to plunder collections of valuable archaeological material is not new. In 1993, for instance, an "organized gang of 300 bandits" used hand grenades to blow open the Angkor Wat conservation compound.[101] They used a rocket launcher to blast through the main warehouse door before looting ten priceless statues inside.[102]

Under the broad definition of organized crime from Part I of this Article, even loosely structured mobs and gangs would constitute organized criminal

---

90. Elich, *supra* note 7, para. 2.
91. *Id.*
92. *Id.* para. 3.
93. *Id.*
94. *Id.*
95. *Id.*
96. *Id.* para. 11.
97. *Id.*
98. *Id.*
99. *Id.* para. 24.
100. *Id.*
101. *Id.* para. 16.
102. *Id.*

groups because they are comprised of three or more people working in concert to violate the law and seeking a material benefit. This would be particularly true of the professionals who work with people on the outside to select antiquities that will prove most valuable and purportedly take orders from dealers and collectors in anticipation of frenzied pillaging.

5. *Methamphetamine-Fueled Diggers*.—There is a well-documented connection between methamphetamines and looting in the United States, in particular the Southwest.[103] Garry Cantley, an archaeologist with the Bureau of Indian Affairs, has acknowledged the looting-meth connection, and others even call it an epidemic.[104] Meth addicts illegally excavate Native American sites, including burial grounds, in order to find pottery shards and arrowheads to sell to wholesalers to finance their drug habit.[105]

Monetary rewards are often secondary to the "twiggers"[106] true incentive, however; methamphetamine, like many stimulants, creates a compulsion for repetitive, focused behavior.[107] Meth addicts quench this thirst by artifact hunting, which is tedious, repetitive work. "It makes them the perfect, tireless looting workforce."[108] Accordingly, meth addicts are the exception to the general rule that criminal excavators dig with the motivation of obtaining a financial benefit.

Although meth addicts do not necessarily organize with one another to hunt for artifacts, they are hooked into a larger system in which they can sell the artifacts to wholesalers or collectors, or trade them directly for drugs.[109]

B. Middlemen

Those who illegally excavate most often sell to "middlemen"[110] (sometimes

---

103. Nicholas Phillips, *Tweakers 'N Diggers: Looters Are Pillaging Native American Burial Grounds to Finance Their Meth Habits*, RIVERFRONT TIMES NEWS, Mar. 24, 2010, http://www.riverfronttimes.com/2010-03-24/news/tweakers-n-diggers-looters-pillaging-native-american-burial-grounds-finance-meth-habits.

104. Samir S. Patel, *Drugs, Guns and Dirt*, 62 ARCHAEOLOGY, Mar.-Apr. 2009, *available at* http://www.archaeology.org/0903/etc/drugs.html.

105. *Id.*; Phillips, *supra* note 103.

106. Patel, *supra* note 104. A combination of the slang for methamphetamine addict, "tweaker," and digger. *Id.*

107. *Id.* ("The energizing and obsessive effects make it fun, almost pleasurable, for tweakers to do the tedious work of artifact hunting. They have the steam to wander sites and dig holes for hours, the focus to scan the ground closely, and the compulsive need to find more and more. According to those who have spoken to twiggers directly, the ability to sell artifacts seems almost secondary to the addictive thrill of discovery.").

108. *Id.*

109. *Id.*; Phillips, *supra* note 103.

110. Bonn-Muller, *supra* note 62.

called "wholesalers"[111]). Middlemen include both wholesalers and traffickers,[112] who are both essential to the illicit antiquities trade, although neither excavates archaeological material themselves or sells to the ultimate owners. Middlemen are, quite literally, in the middle of the distribution ladder for illicit antiquities. As a rule, criminal excavators sell plundered objects to a middleman, who arranges for the artifacts to be smuggled into a market nation, where the objects can then be sold to collectors or museums.[113]

Middlemen must necessarily work with both criminal excavators and retailers in order to accomplish their goals within the illicit antiquities trade. Whether these actors are participants in organized crime under the broad definition from Part I of this Article depends on whether a wholesaler buying objects from criminal excavators is viewed as working "in concert" with them. To some extent, the two are working toward different ends and have competing interests. Yet none of the criminological definitions of organized crime require a common goal—merely that the participants each have a goal, and work in concert with one another, in at least a loose structure. Middlemen satisfy these qualifications, and can thus be considered organized criminal actors.

1. Wholesalers.—Wholesalers are the contacts for criminal excavators. Wholesalers make much larger financial gains than do the individuals who dig up the archaeological materials.[114] In addition to finding traffickers or dealers for

---

111. Green, *supra* note 58, at 53.

112. In literature on the illicit antiquities trade, the term "middleman" is generally synonymous with "wholesaler," and falls into a different category from trafficker. These general distinctions are challenged when wholesalers employ crews of diggers or themselves engage in trafficking. Regardless of the business structure employed, wholesalers and traffickers do share the common characteristic of being in the middle of the distribution ladder for illicit antiquities. Accordingly, for the purposes of this Article and to avoid confusion, "middlemen" refers to wholesalers and traffickers generally, and the person to whom criminal excavators sell their goods is a "wholesaler."

113. BRODIE ET AL., *supra* note 7, at 29.

114. *See* Kimberly L. Alderman, *Ethical Issues in Cultural Property Law Pertaining to Indigenous Peoples,* 45 IDAHO L. REV. 515, 519-20 (2009) [hereinafter Alderman, *Ethical Issues*] ("[T]he enormous revenues that are ultimately generated in the galleries or on the auction blocks do not filter back to the indigenous descendants of creator cultures. An ancient cylinder seal that fetches $100 in a Baghdad souk may later sell for five figures in a New York gallery."); *see also* BRODIE ET AL., *supra* note 7, at 13 (stating that the original finder receives less than two percent of the final price); *id.* at 14 (indicating the disproportionate allocation of profits is not limited to source nations: "[A] landowner in the United States accepted $2,000 for a late Cretaceous Ankylosaur which was subsequently sold for $440,000."); Juliana V. Campagna, *War or Peace: It Is Time for the United States to Ratify the 1954 Hague Convention for the Protection of Cultural Property in the Event of Armed Conflicts*, 17 FLA. J. INT'L L. 271, 286 (2005); Asif Efrat, *A Theory of Internationally Regulated Goods*, 32 FORDHAM INT'L L.J. 1466, 1479 (2009); Elich, *supra* note 7, para. 13 (describing how the criminal excavators of the Morgantina acroliths were paid $1000 in 1979, and the acroliths were sold by renowned dealer Robin Symes a year later for over $1 million); Ruiz, *supra* note 27 (wherein a tombarolo explains, "I estimate that [the middleman I work with] sells the stuff for [ten] times what he paid me. Let me put it this way: he drives a Mercedes,

the illegally excavated artifacts, wholesalers may also arrange for false bills of provenance to aid traffickers in smuggling artifacts out of the source nation or into an intermediate or market nation.[115]

Wholesalers occasionally directly employ teams of diggers to perform illegal excavations.[116] This type of wholesaler pays his diggers salaries and supplies them with the necessary tools and equipment.[117] This benefits workers who might dig all night and not find anything of value and who cannot independently afford expensive equipment like metal detectors.[118] The diggers bring the loot back to the wholesaler in the morning, and the wholesaler contacts local dealers who can move the merchandise.[119]

In Italy, the wholesaler is called a *ricettatore*, meaning "receiver."[120] Ricettatori in Italy are connected to traffickers who can move the antiquities to market nations.[121] Tombaroli are paid fixed rates for certain types of objects, so they are not necessarily loyal to any particular ricettatore.[122] One tombarolo reported that it is not to his benefit to haggle too much with a ricettatore, because the ricettatore could always anonymously report the tombarolo to police.[123]

While tombaroli are most often the unemployed working class, ricettatori are well-educated and "part of the establishment."[124] If an object is too big for the tombarolo to transport from the site, a ricettatore will arrange for immediate pick up.[125] On rare occasions, a tombarolo will work as a sort of middleman himself, connecting a less experienced tombarolo with the right ricettatore, and taking a ten percent cut for his efforts.[126]

2. Traffickers.—Wholesalers either traffic illicit antiquities themselves or outsource to someone who specializes in moving illicit objects. The shared characteristic of traffickers is that they smuggle illicit antiquities—across national and international borders[127] or from remote to urban areas within the source nation.[128] Objects removed from source nations are often funneled

---

I drive a Fiat Panda.").

115. Chauncey D. Steele IV, *The Morgantina Treasure: Italy's Quest for Repatriation of Looted Artifacts*, 23 SUFFOLK TRANSNAT'L L. REV. 667, 681 (2000).

116. Bonn-Muller, *supra* note 62.

117. *Id.*

118. *Id.*

119. *Id.*

120. Ruiz, *supra* note 27.

121. *Id.*

122. *Id.*

123. *Id.*

124. *Id.*

125. *Id.*

126. *Id.*

127. *See* Steele, *supra* note 115, at 681.

128. Aaron Kyle Briggs, *Consequences of the Met-Italy Accord for the International Restitution of Cultural Property*, 7 CHI. J. INT'L L. 623, 625 (2007) (noting too that archaeological objects are illicit when they either were illegally excavated or are being illegally exported).

through intermediary countries with lax import laws, and then legally imported into market nations.[129]

In Italy, professional smugglers are called *clandestini*.[130] Like traffickers in other countries, clandestini organize the route across national borders, ascertain who to bribe and arrange for the payments, and they may even identify buyers before sending the artifacts on their way.[131] Many illicit antiquities moving out of Italy leave Rome in a shipping container carrying legal goods such as car parts, food, or marble.[132] One technique used to hide the antiquities is to place them inside a hollowed-out marble slab, and then to seal the slab with stucco.[133]

Another technique used to smuggle illicit antiquities is to lie on the bill of laden. In 2006, for instance, officials in Pakistan reported seizing a shipping container in Karachi packed with 625 Buddhist sculptures worth over €1 million, bound for the United Arab Emirates.[134] The bill of laden listed the shipment as "furniture."[135]

The sheer amount of international cargo makes it impossible for customs agents to inspect the contents of every shipping container. The agents do open a tiny percentage, but they are looking for drugs, not stolen antiquities.[136] For those objects transported by air, traffickers use private companies such as FedEx, UPS, and DHL,[137] which are monitored even less stringently.[138]

This allows traffickers to develop paths to move illicit goods out of each source nation, and to use those paths habitually. In Iraq, for instance, antiquities recovered from the harvesting of insufficiently protected museums "are driven overland either west to Jordan or north to Syria."[139] The objects are then sent to either Beirut, Dubai, or Geneva, where papers can be obtained.[140] Once the objects have papers, they can then "surface," so dealers can sell them to private collectors, sometimes through auction houses.[141]

In Afghanistan, traffickers fly illicit antiquities out of the airports at Kabul or Kandahar with the help of corrupt customs officers.[142] The antiquities pass

---

129. Alderman, *Ethical Issues*, *supra* note 114, at 520; *see also* BRODIE ET AL., *supra* note 7, at 31-33.

130. Steele, *supra* note 115, at 680.

131. *Id.*

132. Ruiz, *supra* note 27, at 37.

133. *Id.*

134. Adil Najam, *Smuggling Historical Artifacts to Sharjah*, ALL THINGS PAKISTAN (Nov. 19, 2006), http://pakistaniat.com/2006/11/19/smuggling-historical-artifacts-to-sharjah.

135. Imitiaz Gul, *China to Build Nuke Plants in Pak*, TELEGRAPH, Nov. 19, 2006, http://www.telegraphindia.com/1061119/asp/foreign/story_7021846.asp.

136. Ruiz, *supra* note 27, at 38.

137. de la Torre, *supra* note 66.

138. *See* Brems & Van den Eynde, *supra* note 38.

139. *Antiquity Smuggling*, *supra* note 68.

140. *Id.*

141. *Id.*

142. Brems & Van den Eynde, *supra* note 38.

through either Bangkok, Singapore, or Dubai, and then on to Brussels.[143] After the objects slip through Belgium customs, they are then transported to warehouses near the Zavel[144] (Sablon Square[145]). Dealers come into the warehouses on Saturdays to shop for fresh merchandise.[146]

The paths of illicit antiquities and drugs most often cross at the trafficking stage, because traffickers sometimes use the same routes and means to smuggle all kinds of illegal goods.[147]

C. Retailers

Retailers, consisting of dealers and auction houses, sell illegally excavated or exported antiquities to the public. Selling illicit antiquities to collectors or museums is often legal under the laws of the market nations where sales usually take place.[148] This raises an interesting question when considering whether retailers are organized criminal actors in the illicit antiquities trade. Even under the broader definition of organized criminal groups, the participants must act in concert *to violate the law*. If a dealer or auction house is selling an object they know to have been illegally exported from a source nation, but their home nation does not have a bilateral agreement to enforce the export laws of that source nation, then the transaction is legal.

Accordingly, retailers would not fall under the umbrella of organized criminal groups unless they specifically facilitate illegal transactions. Nonetheless, they play a critical role in the illicit antiquities trade, and their contribution must be acknowledged.

1. Dealers.—Dealers serve as intermediaries between wholesalers and purchasers of antiquities (collectors and museums). Antiquities that have been illegally excavated or exported often pass through several dealers before being sold to collectors or museums.[149] Some dealers own antique shops[150] while others work directly with specific collectors or curators to locate desired objects.[151]

The typical dealer works within a market nation, selling finds to wealthy collectors and museums. However, retailers with antique shops in source nations are just as deserving of the title "antiquities dealers," as they serve as a public interface for the illicit antiquities trade.

---

143. *Id.*

144. *Id.*

145. A square in Brussels famous for its antiques market. *See* Myriam Jansen-Verbeke & Robert Govers, *Brussels: A Multilayered Capital City*, *in* CITY TOURISM: NATIONAL CAPITAL PERSPECTIVES 142, 152 (Robert Maitland & Brent W. Ritchie eds., 2009) (describing the Zavlon as an "antique cluster").

146. Brems & Van den Eynde, *supra* note 38.

147. *See* BRODIE ET AL., *supra* note 7, at 16.

148. *See infra* note 170-71 and accompanying text.

149. Elich, *supra* note 7, para. 12.

150. *See, e.g.*, Bonn-Muller, *supra* note 62.

151. *See id.*

Antique shops are mediums through which some dealers move their wares.[152] There are antique shops both inside and outside of source nations. The benefit of selling illicit artifacts in source nations is that they do not have to go through a trafficker to get smuggled out of the country; tourists do the smuggling themselves.[153] The drawback is that the objects will not sell for as high a price as they would in a market nation.

Antique shops in source nations target tourists, who are often willing to slip a token of their visit in their suitcases.[154] These shops are numerous; there are forty-five in Israel, for instance, most located in Jerusalem.[155] Illegally excavated beads are for sale "by the buckets full" in Bangkok antique shops.[156] Bangkok is reputed as a clearinghouse for illegal antiquities from all over Indochina, including Myanmar, Vietnam, Laos, Bengal, and Cambodia.[157]

In mainland Europe, Brussels serves as a hub for purchasing antiquities because of lax regulation of the trade.[158] Fifty thousand passengers and 2000 tons of freight pass through the international airport in Brussels daily,[159] and screening for illegal antiquities is "basically an impossible task."[160] There is no law in Belgium requiring dealers to prove that an archaeological object came in through legitimate channels.[161] Illegally excavated antiquities are on display in the front windows of antique shops in Brussels.[162] Dealers are often fully aware that they are selling illegally excavated objects.[163]

Most illicit antiquities dealers use the traditional channels of antique shops

---

152. *See, e.g.*, *id.*

153. *See* Thomas Kean, *Asia Fights to Stem Loss of Cultural Treasures*, MYANMAR TIMES, Dec. 22, 2008, http://www.heritagewatchinternational.org/km/asia-fights-to-stem-loss-of-cultural-treasures.html.

154. Bonn-Muller, *supra* note 62.

155. *Id.*

156. Elich, *supra* note 7, para. 16.

157. Kean, *supra* note 153 (explaining how there is a movement toward using reproductions instead, in no small part because the antiquities are simply running out—most of Myanmar's treasures, for instance, have already left the country).

158. Brems & Van den Eynde, *supra* note 38 (noting a Belgian government official explaining, "[illegal antiquities are] not on the top of anyone's list").

159. *Id.*

160. *Id.* (statement by Belgian customs officer Pol Meuleneire).

161. *Id.*

162. *Id.*

163. *Id.* (wherein a dealer says, "It's obvious that if a piece comes from Afghanistan, it has been stolen." When asked, "Every piece?" she responds, "Yes, of course."). *But cf.* Susan Montoya Bryan, *Fed Crackdown Puts Tribal Artifact Dealers on Edge*, ASSOCIATED PRESS, Aug. 19, 2009, http://seattletimes.newsorce.com/html/nationworld/2009691125_apusartifactsmarket.html (describing Walter Knox, a dealer in a gallery in Scottsdale, AZ, who says every week he has to run someone out of his gallery for trying to sell him stolen pots. "I post my rules so people know I'm not going to deal with anything shady." As a retired police officer, Knox may be the exception rather than the rule.).

and building their clientele. However, those who use less traditional means to hawk their wares are using eBay to reach a worldwide audience.[164] Industrious dealers may also employ "loot-to-order" services, as they did in the case of the twelfth-century Cambodian temple of Banteay Chmar.[165] The temple was covered in "breathtaking" bas reliefs, but has since been gutted by thieves, some of whom worked for dealers who would allow collectors to order their sculptures directly off the temple walls.[166] Local archaeologists describe looting in the region as being "on a virtually industrial scale."[167]

2. Auction Houses.—Auction houses and museums are both legal actors frequently credited with perpetuating the illegal antiquities trade by way of quasi-legal or illegal actions.[168] While auction houses may not be organized criminal actors per se, the significance of their involvement in the illicit antiquities trade as a whole cannot be overlooked.

An estimated "[t]hirty to forty percent of all the world's antiquities pass through auction houses, making [them] the most public portion of the antiquities market."[169] Yet, "studies repeatedly show that about ninety percent of antiquities auctioned through the houses are illegally excavated or exported."[170] This is because even if an object has been illegally excavated or exported, it is frequently still legal to sell antiquities in a market nation.[171]

Ricardo Elia of Boston University completed a study in which he examined 1881 Apulian vases sold through the Sotheby's saleroom from 1960 to 1997, concluding that "*not one* had a published provenance."[172] This finding suggested "every single Apulian vase sold at Sotheby's over a 37-year period might have been illegally excavated or at least illegally exported from Italy."[173]

---

164. *See, e.g.*, Patrick E. Tyler, *Museum Asks eBay to Block Some Sales*, N.Y. TIMES, Oct. 30, 2004, at B7, *available at* http://www.nytimes.com/2004/10/30/arts/30trea.html; *see also* Dalya Alberge, *Illicit Artefacts Sold as eBay Turns Blind Eye*, THE TIMES, Dec. 18, 2006, at 7. *But see* Charles Stanish, *Forging Ahead: Or, How I Learned to Stop Worrying and Love eBay*, 62 ARCHAEOLOGY, May/June 2009, *available at* http://www.archaeology.org/0905/etc/insider.html (explaining that online selling has hurt the antiquities trade because of mass production of fakes).

165. *See* BRODIE ET AL., *supra* note 7, at 20; Elich, *supra* note 7, para. 16.

166. Elich, *supra* note 7, para. 16.

167. *Id.*

168. *See generally* Alderman, *Ethical Trade*, *supra* note 20.

169. *Id.* at 561 n.67.

170. *Id.* at 560 (citing PETER WATSON, SOTHEBY'S: THE INSIDE STORY 304 (1997)); *see also* Ricardo J. Elia, *Digging Up Dirt: Antiquities Case Unearths Corruption*, WALL ST. J., June 19, 2002, http://online.wsj.com/article/0,,SB1024442830326084200-search,00.html; Elich, *supra* note 7, para. 12 (describing a study that estimated between sixty-five and ninety percent of antiquities sold at auction in London lacked established provenance, "a fairly reliable indication of illicit origin").

171. *See* BRODIE ET AL., *supra* note 7, at 32-33 (containing a section on "the international loophole"); Alderman, *Ethical Trade*, *supra* note 20, at 559-60; Bonn-Muller, *supra* note 62.

172. BRODIE ET AL., *supra* note 7, at 27; *see also* Elich, *supra* note 7, para. 17.

173. BRODIE ET AL., *supra* note 7, at 27.

Aside from quasi-legal sales, it is not uncommon for plainly illicit artifacts to show up in the auction catalogue, only to be pulled just before the auction.[174] An Attic jar and a Greek jug listed for £2,000 and £4,000 respectively were pulled from an October 2010 auction at Bonhams after the Italian authorities filed a claim that the objects had originated with infamous trafficker Giacomo Medici.[175] This was despite a warning received in April 2010, when Bonhams had to pull another object after listing it because it was revealed to be Medici loot.[176]

The antiquities at stake are not always of inconsequential value; the Sevso treasure was pulled from the Sotheby's auction block in 1990 after the documentation of provenance was determined false.[177] The hoard is now estimated to be worth more than $100 million.[178]

In the past, auction houses have been more directly involved in the illicit trade. The investigation into infamous dealer Giacomo Medici yielded documents showing that "in individual cases, regarding more valuable pieces, Sotheby's personnel had either been aware that objects sold on their premises had been illegally exported from Italy, or had themselves had a hand in the arrangements."[179] The carabinieri (Italian police) found that Sotheby had allowed Medici to purchase back some of his own objects in order to acquire for them an apparently legitimate provenance.[180]

---

174. The Italian government has not released the photographs of the extant Medici objects to the public (or the auction houses), so the houses defend that they lack the requisite knowledge to know what objects are illicit in origin. *See generally* Kimberly L. Alderman, *Is Italy "Asking for It" By Refusing to Release the Medici Photographs? Three Items at Christie's Raise Questions*, CULTURAL PROP. & ARCHAEOLOGY L. NEWS (June 6, 2010), http://culturalpropertylaw.net/2010/06/06/is-italy-asking-for-it-by-refusing-to-release-the-medici-photographs-three-items-at-christies-raise-questions.

175. Mark Hughes, *Lots of Trouble on New Bond Street*, INDEP., Oct. 27, 2010, at 20, *available at* http://www.independent.co.uk/news/uk/crime/bonhams-lots-of-trouble-on-new-bond-street-2117393.html (reporting that "[d]ays before the auction the house received an email from an eminent academic alerting them to the questionable provenance of the lots, but it pressed ahead with the sale.").

176. *Id.*

177. *See Sesvo Treasure*, BBC (Mar. 23, 2007), http://www.bbc.co.uk/insideout/east/series11/week9_castle_ashby.shtml.

178. *See id.* (estimating the value of the collection as between £50-100 million); *see also* Peter Landesman, *The Curse of the Sesvo Silver*, ATL. MONTHLY, Nov. 2001, at 66, *available at* http://www.theatlantic.com/magazine/archive/2001/11/the-curse-of-the-sevso-silver/2331 (estimating a worth of up to $200 million); Alan Riding, *14 Roman Treasures, on View and Debated*, N.Y. TIMES, Oct. 25, 2006, at E1, *available at* http://www.nytimes.com/2006/10/25/arts/design/25sevs.html?ref=arts&pagewanted=all (estimating the value at more than $187 million).

179. BRODIE ET AL., supra note 7, at 27.

180. *Id.* at 27; *see also id.* at 29 ("Looted antiquities then acquire a patina of legitimacy when ultimately they are sold, without provenance, by dealers and auction houses.").

In recent years, the auction houses have pledged via codes of conduct not to knowingly sell illegally exported artifacts and have established compliance departments to oversee implementation of the codes.[181] While this is a step in the right direction, it does not necessarily prevent the listing of illicit antiquities. The auction houses rely on the provenance of objects as described by their owners, who are unlikely to be forthcoming about illicit origin.[182]

D. Antiquities Owners

The two major groups of people who own antiquities are collectors and museums. While collectors and museums may not constitute organized criminal groups, even under a broad definition, no discussion about the organization of the illicit trade in antiquities is complete without consideration of their role in the trade.

Collectors and museums are most often located in market nations, where it is sometimes legal to purchase objects that have been the subject of an illicit act.[183] Even when it is illegal, they are often isolated from criminal liability. The crimes with which unscrupulous purchasers can be charged have a scienter element, requiring the state make a showing that it *knew* the objects were illicitly excavated and exported. Because collectors and museums are the "end of the line," they are less likely to take part in a lateral structure of organized criminal activity.[184] Instead of committing what they know to be criminal acts, retail buyers of antiquities are more likely to not ask too many questions or turn a blind eye to questions about an object's provenance.

1. Collectors.—Collectors of antiquities—licit or illicit—tend to be wealthy, "high society" people.[185] Maguire Gibson of the Oriental Institute of Chicago has explained, "[Collectors] are very, very well-connected people . . . . They are able to get a meeting with whoever [sic] they like, when they like . . . . They are very affluent people, too. One of the leading lights is a former State Department man . . . ."[186] The affluence of collectors allows them to pay premium prices for objects that generated only small profits for the local people who unearthed or plundered them. Their collective influence permits them to lobby for less restrictive import regulations, and to individually avoid prosecution for collecting

181. *Id.* at 28-29 (describing the code of conduct and compliance department established by Sotheby's in the wake of the Peter Watson investigation); *see generally* Alderman, *Ethical Trade*, *supra* note 20.

182. Alderman, *Ethical Trade*, *supra* note 20, at 553.

183. *See supra* note 171.

184. *See infra* Part III (discussing the concept of lateral and vertical criminal structures in more detail).

185. Brems & Van den Eynde, *supra* note 38 (reporting, "This is a world of it[s] own, one that doesn't take to being investigated. It's a high society world. If [you] really start digging in the art world, heads of people in high places would roll. There's also the aristocracy, some of whom also have collections, of [these] so-called 'blood antiquities.'").

186. Elich, *supra* note 7, para. 20.

illicit antiquities.

Collectors do not always rely on dealers to provide them with choice objects. Occasionally, the more industrious will themselves go out in search of antiquities. For example, when the Phnom Penh media reported on the discovery and looting of a site in Cambodia, collectors traveled to the nearby village and approached locals, looking to buy direct.[187]

The archaeological community tends to criticize antiquities collectors. Archaeologist Ricardo Elia, for instance, has stated, "Collectors are the real looters . . . Without their money and their demand, there would be no market."[188] Some collectors may never take part in a technically illegal transaction, since in market nations it is often legal to buy objects that have been illegally excavated or exported from a source nation. Nonetheless, collectors are an indispensable component of the antiquities trade and, consequently, the illicit antiquities trade.

Collectors do not necessarily organize with one another in a conventional model of organized crime. However, when they purchase objects that organized criminal groups have moved, they participate in a broader criminal system that has the primary purpose of trading illicit goods for monetary gain.

2. Museums.—Like collectors, museums are not likely to fit into any definition of organized criminal group—even a broad one. However, they too have historically participated in a system that moves illicitly excavated or exported antiquities from source to market nations.

In *The Medici Conspiracy: The Illicit Journey of Looted Antiquities, From Italy's Tomb Raiders to the World's Greatest Museums*, authors Peter Watson and Cecilia Todeschini describe a vast network of antiquities trafficking from Mediterranean tombs to museums in market nations.[189] This network operated for decades, with museum personnel turning a blind eye to issues about provenance or, in some instances, knowingly acquiring illicit objects.[190]

In a classic example, the Metropolitan Museum of Art in New York purchased the Lydian hoard in 1969, knowing that the 363 objects were illegally excavated and exported from Turkey.[191] Although the $1.5 million acquisition violated Turkish law, it was legal under U.S. law, so the purchase moved

---

187. *Id.* para. 16.

188. *See* Hunt, *supra* note 5, at 81 (alteration in original). Consider that "the late Archaeological Commissioner of Belize stated in a personal communication with [David Matsuda] that the intensity of subsistence digging [in Belize] is inextricably linked not to auction house prices but to the success or failure of subsistence agriculture." Matsuda, *supra* note 31, at 94.

189. *See generally* PETER WATSON & CECILIA TODESCHINI, THE MEDICI CONSPIRACY: THE ILLICIT JOURNEY OF LOOTED ANTIQUITIES FROM ITALY'S TOMB RAIDERS TO THE WORLD'S GREATEST MUSEUMS (2006).

190. *See generally id.*

191. BRODIE ET AL., *supra* note 7, at 10 ("The Metropolitan knew at the time of its purchase that the material had been looted and exported illegally—a junior curator had been to Turkey and visited the looted site, and had managed to identify the matching parts of a pair of sphinxes held by the museum—and did not mount a display until 1984, when the material was exhibited without provenance under the misleading title of the East Greek Treasure."); Briggs, *supra* note 128, at 628.

forward.[192] In 1993, after a prolonged lawsuit, the Metropolitan was forced to return the objects to Turkey without compensation.[193]

Since the enactment of the 1970 UNESCO Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, museums in market nations have been subject to more regulation from laws and codes of ethics.[194] Neil Brodie explained, "Museum professionals welcome clear guidelines from professional organisations, especially for dealing with issues like the illicit trade, which involves negotiating legal and ethical minefields."[195]

There has been increasing pressure on museums to tighten acquisition and loan policies, and many museums will no longer purchase objects without documentation establishing a legitimate ownership history.[196] Still, museums may receive donations of unprovenanced antiquities, for acquisition or loan, which are later revealed to be illicit.[197] At present, museums are also more likely than they were in the past to return objects on ethical grounds even when there is no legal basis to demand repatriation.[198] This shows an increased focus on ethical behavior on the part of museums and a move away from literalism concerning the law.

III. Lateral Cooperation Among Criminal Organizations: The Illicit Antiquities Trade as an Organized Criminal Megagroup

A pattern of interconnectedness emerges when examining discrete groups within the illicit antiquities trade. These groups cannot and do not function independently from one another. They are interdependent by necessity; without the wholesalers, there would be insufficient demand to keep criminal excavation profitable; without traffickers, unscrupulous dealers would not have access to illicitly excavated or exported antiquities. The manner in which the criminal groups collaborate with one another is distinctly less categorical than the organization of the groups themselves.

The interconnectedness among the criminal groups reveals a broader, organized system that comprises the illicit antiquities trade. If any of the criminal groups ceased to function—the excavators, the wholesalers, the

192. *See* BRODIE ET AL., *supra* note 7, at 10.

193. *See id.*

194. *See id.* at 43.

195. *See id.*

196. Briggs, *supra* note 128, at 628.

197. *See* BRODIE ET AL., *supra* note 7, at 47-48.

198. BRODIE ET AL., *supra* note 7, at 27 ("[I]n the spring of 1999 . . . [Medici] came to an agreement with the Italian *carabinieri* to return three fragments of a bowl made by the well-known ancient potter Euphronios. The other fragments of this bowl were in the J Paul Getty Museum in Los Angeles, and . . . the museum voluntarily returned what it possessed to Italy."); Kimberly Alderman, *Yale Agrees to Return Machu Picchu Artifacts to Peru: Ethics-Based Repatriation Efforts Gain Steam*, CULTURAL HERITAGE & ARTS REV., Fall/Winter 2010, at 3.

traffickers, or the retailers—the illicit antiquities trade would end. Therefore, the illicit antiquities trade serves as a biosphere of sorts for smaller, discrete criminal groups.

The argument can therefore be made that the illicit antiquities trade *itself* constitutes organized crime. This does not necessarily validate the claim that organized crime "controls" the illicit antiquities trade. That claim relies on the notion of a transnational criminal enterprise larger than the antiquities market itself.

As discussed in Part II of this Article, there is no legitimate distinction between the licit and illicit antiquities trades; all antiquities are bought and sold on the same market, regardless of legality.[199] The antiquities market has been called a "grey market,"[200] because there is no distinct "black market" or "white market" for antiquities. It is commonly accepted that the vast majority of the antiquities sold at auction were illegally excavated or exported.[201] For example, it is estimated that only five percent of the over 13,000 Apulian red-figure vases known to exist were excavated legally.[202] Therefore, there is no illicit division of the trade, because most participants deal in at least some illicitly excavated or exported material, even if unknowingly.

The auction industry is the most public portion of the antiquities market.[203] The sales that take place pursuant to auctions are, to as great a degree as possible, public. If ninety percent of the antiquities that go through these public sales were illicitly excavated or exported,[204] then it is fair to approximate that at least this percentage of undocumented, private deals are for materials with an illicit origin. The infrastructure for the antiquities market has developed amidst more pressure from the illicit trade than from the licit trade, due to sheer volume. Accordingly, many facets of the market infrastructure serve no legitimate purpose for licit traders, but do serve to benefit those who trade in illicit antiquities.

Whether or not the trade in licit antiquities could function independently of the illicit trade, it does not. The illicit trade depends upon the licit trade in order to promote an appearance of legitimacy. Collectors and dealers can present themselves as legitimate businesspeople because, at least in theory, they *could* deal exclusively in legally obtained antiquities. Dealers and collectors are perceived as "people who can speak seven languages, have university degrees and are highly intelligent."[205] They are just as much participants in and members of this organized criminal megagroup as are the traffickers and the crews who dig antiquities out of the ground under cover of night. Without wealthy dealers and

---

199. *See supra* Part II.

200. *See* Blythe A. Bowman, *Transnational Crimes Against Culture: Looting at Archaeological Sites and the "Grey" Market in Antiquities*, 24 J. CONTEMP. CRIM. JUST. 225, 225 (2008).

201. Elich, *supra* note 7, para. 12; *supra* Part II.C.

202. Elich, *supra* note 7, para. 17.

203. Alderman, *Ethical Trade*, *supra* note 20, at 561 & n.67.

204. *See supra* note 170 and accompanying text.

205. de la Torre, *supra* note 68, at 10 (quoting Arthur Brand, art collector).

collectors, the efforts of all the other actors would be for naught. Collectors tend to have political influence,[206] and they tend to oppose reform of the trade because they are benefiting from its largely unregulated nature.

Meanwhile, the licit trade in antiquities uses the illicit trade for its infrastructure. The sheer volume of unprovenanced antiquities justify the highly developed infrastructure for their trade. If the market were restricted to documented legal goods (three to ten percent of the current trade, depending on who you ask), it would not be possible for as many dealers, auction houses, and museums to survive.

Accordingly, the relationship between the licit and illicit antiquities trades echo the interconnectedness and interdependence of the criminal groups discussed in Part II of this Article. These criminal groups depend on one another to keep the antiquities trade functioning; pull any rung of the ladder out—the excavators, the middlemen, the traffickers, or the retailers—and the system will not function. Similarly, the licit and illicit antiquities trades so significantly depend on one another that they are not readily distinguishable. Again, "People think that there is an illicit and a legitimate market. . . . In fact, it is the same."[207]

If the illicit antiquities trade operates as an organized criminal industry, and the licit antiquities trade is indivisible from the illicit trade, it could be said that the antiquities trade as a whole functions as an organized criminal megagroup. Any actor who participates in the antiquities trade is contributing to the larger system that enables and perpetuates organized crime. This applies even to scrupulous participants who insist on provenance and have never handled a tainted object.

Auction houses continue to trade in antiquities because there are at least *some* licit objects on the market. There is a presumption of legitimacy, which is why proof of acceptable provenance is not required, merely suggested.[208] Licit transactions serve as a smokescreen and set the stage for the trade in objects of illicit origin. In this way, above-board transactions are indispensable to the illicit trade in antiquities.

The discrete organized criminal groups identified in Part III of this Article operate, for the most part, within national boundaries, cooperating with organized criminal groups in other jurisdictions to move illicit goods. This cooperation points toward an organized criminal megagroup that is transnational in character. Although antiquities crime crosses national borders, law enforcement does not, making effective prosecution and prevention particularly difficult.

---

206. *See supra* notes 185-86 and accompanying text.

207. Barry Meier & Martin Gottlieb, *LOOT: Along the Antiquities Trail; An Illicit Journey Out of Egypt, Only a Few Questions Asked*, N.Y. TIMES, Feb. 23, 2004, http://www.nytimes.com/2004/02/23/world/loot-along-antiquities-trail-illicit-journey-egypt-only-few-questions-asked.html?pagewanted=all&src-pm (internal quotation marks and citations omitted).

208. This is the general, not absolute, rule. Some auction houses require paperwork establishing provenance for certain locations or types of objects. *See* Alderman, *Ethical Trade*, *supra* note 20, at 570.

CONCLUSION

Multiple meanings of the term "organized crime" have impeded meaningful discourse in discussions about the illicit trade in antiquities. Accordingly, the author suggests that scholars and experts strive toward a unified definition of the term as it pertains to the trade. A broader definition, akin to the one propounded in Part I of this Article, would be helpful: *a structured group of three or more people, acting in concert to violate the law, in order to obtain a financial or other material benefit*. This broader definition encompasses most known participants in the illicit antiquities trade, ranging from tombaroli to the Ukrainian mafia, to unscrupulous collectors and dealers. It also takes into account violations of criminal or civil law, inside or outside of nations in which transactions take place.

This definition is criminological in nature and designed to facilitate meaningful discourse through uniformity and transparency. Compare the vernacular concept of "organized crime," used to increase public concern about the illicit trade.[209] Regardless of the definition employed, increased clarity is needed in discussions about the connection between organized crime and the illicit antiquities trade. As an alternative to a unified definition, discussants should account for variation in meaning of the term "organized crime."

Examining the actors in the illicit antiquities trade through the lens of this broad definition of organized crime reveals the interconnected relationship between those actors. Lateral cooperation of organized criminal groups in the illicit antiquities trade reveals that the trade itself is an organized criminal megagroup of sorts, transnational in character. This is the only framework through which the claim that organized crime controls the illicit antiquities trade is justified.

209. Putting the illicit antiquities trade on par with the arms trade, terrorism, and human trafficking justifies the study of the trade, and enforcement of regulations aimed at thwarting it. The groups and agencies who allege most strenuously that organized crime controls the illicit antiquities trade tend to be the same ones that depend upon public concern for funding or support.

FROM THE THIEF IN THE NIGHT TO THE GUEST WHO STAYED TOO LONG: THE EVOLUTION OF BURGLARY IN THE SHADOW OF THE COMMON LAW

HELEN A. ANDERSON*

ABSTRACT

Burglary began evolving from the common law crime almost as soon as Lord Coke defined it in 1641 as breaking and entering a dwelling of another in the night with the intent to commit a crime therein. But sometime between the Model Penal Code in 1962 and today, burglary lost its core *actus reus*, "entry." In the majority of jurisdictions, burglary can now be accomplished by simply remaining in a building or vehicle with the intent to commit a crime. Not only does such an offense cover a wide range of situations, but it allows burglary to be attached to almost any crime that occurred indoors, and justify a significant additional penalty—even death. Burglary thus functions as a "location aggravator" for other crimes. Paradoxically, it may be the shadow of the common law crime that has obscured the breadth and significance of these changes. Burglary's long tradition and pedigree give an illusion of solidity to the charge, even when it no longer necessarily describes real criminal conduct beyond the target offense.

This is the first survey of burglary in the United States since the Model Penal Code. It begins with a summary of burglary's history from the common law definition through the first two centuries of the republic, then explains the Model Penal Code proposal for burglary—as well as the Model Code authors' misgivings about the offense. The Article then looks in detail at what happened in the states after the Model Penal Code—how the common law elements continued to erode until we ended up with today's very thin crime. The Article shows what this has meant: a serious crime with significant penalties that can be invoked in a range of situations, e.g., shoplifting, hold-up of a business, or murder by a houseguest. It concludes that burglary's evolution has finally gone too far, and no longer necessarily describes a distinct offense. It is only the memory of the common law offense that keeps courts and lawmakers from recognizing how empty the crime has become.

INTRODUCTION

What do the following scenarios have in common? An eleven-year-old girl enters a store during business hours and eats a chocolate Easter egg without paying. A man enters an open gas station, robs it and shoots two employees dead. A man invited into a home to socialize turns on his host and kills her. Answer: all can be prosecuted as burglaries.[1] Such prosecutions would not have

---

* Associate Professor, University of Washington School of Law. I wish to thank Miriam Korngold for her excellent research assistance. I also wish to thank Mary Fan and Elizabeth Porter for their very helpful comments on an earlier draft.

1. *See* Davis v. State, 737 So. 2d 480 (Ala. 1999) (holding that evidence of a struggle during

been possible in the time of Blackstone or Coke, when the common law definition of burglary (a capital offense) was breaking and entering the dwelling house of another in the night.[2]

Few crimes have changed as much over the years as burglary. From its origins as the Anglo-Saxon crime of "hamsoken,"[3] or forcible housebreaking, this offense has evolved into a flexible modern one of entering, or merely remaining, some place with intent to commit a crime. Not only is the offense generally simpler, covering much more conduct, but burglary now also functions increasingly as a way to add to the punishment for the target offenses, those intended by the defendant.[4] Some of the most significant changes have occurred during the past fifty years. Criticized heavily in the mid-twentieth century by scholars and law reformers, the crime of burglary nevertheless survived their challenges and emerged leaner, and meaner, stripped of its more restrictive common law requirements and effective as a "location aggravator"[5]—even to the point of justifying the death penalty for certain murders. The modern crime is a far cry from our idea of the common law offense—a forcible night time intrusion into the home.

Yet paradoxically, it may be that the idea of the common law offense has allowed burglary to survive and change. Burglary's long tradition and pedigree gives an illusion of solidity to the charge, even when it no longer necessarily describes real criminal conduct beyond the target offense. And courts return to the common law treatises occasionally to resolve burglary conundrums—as if the common law offense were still relevant. Like an impoverished aristocrat, burglary can rely on its name and reputation to keep courts and lawmakers from realizing just how little remains of its former estate.

This Article explores the history of burglary law in the United States, with particular emphasis on the last fifty years, when burglary evolved so far as to lose the central conduct of its "actus reus," entry. In the majority of jurisdictions, burglary can now be accomplished by simply remaining in a building or vehicle

---

a murder can support finding of unlawfully remaining on premises and therefore support burglary charge); Bowling v. Commonwealth, 942 S.W.2d 293 (Ky. 1997), *overruled on other grounds by* McQueen v. Commonwealth, 339 S.W.3d 441 (Ky. 2011) (upholding burglary conviction where defendant entered two different gas stations, and shot employees at both); *In re* T.J.E., 426 N.W.2d 23 (S.D. 1988), *superseded by statute as stated in* State v. Miranda, 776 N.W.2d 77 (S.D. 2009), discussed *infra* notes 127-28 and accompanying text.

2. *See infra* notes 17-25 and accompanying text.

3. "[Burglary's] origins lie in the ancient Anglo-Saxon crime of hamsocn or hamsoken, which was an attack upon, or forcible entry into, a man's house." Theodore E. Lauer, *Burglary in Wyoming*, 32 LAND & WATER L. REV. 721, 721 (1997).

4. *See infra* Part II.

5. As used here, the term "location aggravator" refers to the use of burglary charges to add significant penalties to other completed crimes based on where those crimes took place. *See infra* Part IV. Thus, for example, burglary might be charged along with assault and robbery where the victims are assaulted and robbed in a building. *See infra* Part IV.

with the intent to commit a crime.[6] Not only does such an offense cover a wide range of situations, but it allows burglary to be attached to almost any crime that occurs indoors. This significant transformation has generated little discussion.

This Article provides the first comprehensive study of the law of burglary in the United States since the Model Penal Code. Two law review articles surveyed state burglary statutes in 1951[7] and the Model Penal Code revised comments looked at many of the new state criminal codes in 1980.[8] This Article takes these surveys as comparison points for a review of state and federal burglary laws today. This survey shows that since the Model Penal Code, burglary has continued to shed common law requirements, and increasingly is charged together with other serious crimes to function as a location aggravator. The elements of "night time," "breaking," "dwelling," and even "entry" are no longer required under many statutes. The survey shows burglary's continuing evolution from a six-element common law crime to (in its most extreme form) a streamlined modern offense of being in the wrong place with the wrong intent. While there is still a core of conduct that is popularly understood and prosecuted as burglary, at the margins the reach of burglary statutes has extended considerably.

Academic interest in burglary was high during the Model Penal Code era, as scholars pointed out the illogical aspects of the law. Indeed, the law of burglary frustrates those of us who want a rational criminal code.[9] No one wonders why murder and theft are crimes, but unlike with other offenses that began as common law crimes, the purpose of burglary is unclear. Is it to punish attempted crime? To protect personal security in the home? Or to protect property? All of these justifications have been offered.[10] More importantly, from a rational standpoint, the crime of burglary seems unnecessary. The offense is really a combination of offenses: criminal trespass plus the attempt to commit another offense, or criminal trespass plus a completed offense. In fact, it is often charged together with the attempted or completed offense.[11]

This lack of a clear rationale almost led the Modal Penal Code authors to propose abolishing the crime altogether.[12] Out of deference to the prevailing view and tradition, however, they included it in the 1962 Model Code.[13] Their proposal rejected some of the ways in which the crime had already been broadened, and the authors cautioned against trends to further extend the crime.

6. *See infra* note 113.

7. *See* Note, *A Rationale of the Law of Burglary*, 51 COLUM. L. REV. 1009 (1951) [hereinafter *A Rationale of the Law of Burglary*]; Minturn T. Wright III, Note, *Statutory Burglary—The Magic of Four Walls and a Roof*, 100 U. PA. L. REV. 411 (1951).

8. *See* MODEL PENAL CODE § 221.1 (Official Draft & Revised Comments 1980).

9. *See, e.g.*, WAYNE R. LAFAVE, CRIMINAL LAW 1079-80 (5th ed. 2010).

10. *Id*. at 1080.

11. *See id.*

12. *See infra* notes 81-89 and accompanying text.

13. *See infra* text accompanying note 90.

For the most part, states ignored this advice. Burglary evolved[14] and survived—despite the lack of clear rationale and in the face of significant critique. However, the Model Code authors' misgivings about the offense proved well-founded as its reach expanded.

Part I of this Article lays out Lord Coke's influential common law definition of burglary and then summarizes the evolution of burglary law in the states, providing the background for the Model Penal Code proposal in 1962. Part II looks at the Model Penal Code proposal for burglary, and explains the drafters' critique of burglary as it then existed in most American jurisdictions. Part III addresses developments since the Model Penal Code until the present day, examining how states have addressed each of the common law elements, particularly "entry." Part IV describes the several federal breaking and/or entering statutes. Part V examines burglary's increasingly significant role as a location aggravator—the way in which it is used to add punishment (even the death penalty) to offenses committed in particular locations. Part VI looks at how, despite all these changes, the common law crime continues to play a role in discussions of modern burglary, demonstrating that burglary still exists in the shadow of the common law.

I. HISTORICAL DEVELOPMENT OF BURGLARY

A study of the law of burglary in the United States presents significant challenges. Every state has its own statutory scheme, and the variation is enormous. Some states define a single crime of burglary. Others divide the crime into degrees. Still others have developed different statutes for the type of structure entered, the type of crime intended, or the status of the victim. Jurisdictions with determinate sentencing schemes may deal with aggravating factors through sentencing, rather than in the definition of the crime. There is little agreement among states as to the essential elements of burglary. A study in 1951 commented on the wide variation in state approaches.[15] Revised comments to the Model Penal Code made the same point in 1980.[16] More than

---

14. "Evolution" is a useful metaphor for the process of change over time to the offenses that fall under the heading of burglary. The term suggests adaptation and the gradual withering away of aspects (elements) that are no longer useful. But obviously this metaphor is not perfect. Burglary is not an organism; it exists only as an artifact of human lawmakers, and any changes occur in their (successive) minds and actions. *See* Kay L. Levine, *The External Evolution of Criminal Law*, 45 AM. CRIM. L. REV. 1039, 1047-48 (2008) (discussing the usefulness, with caveats, of considering change in criminal law through the lens of "external evolution"). At the same time, the metaphor of evolution reinforces the enduring link to the common law crime, and mirrors the way in which judges even now refer back to the common law when deciding questions about modern statutory versions of the crime. *See infra* Part V.

15. *See* Wright, *supra* note 7, at 415.

16. MODEL PENAL CODE § 221.1 cmt. 1, at 66 (Official Draft & Revised Comments 1980) ("It is also worth noting that a haphazardly defined burglary offense impedes scientific study of crime and its treatment by making statistical studies based on this categorization virtually

thirty years later, it is safe to say that there has been no subsequent trend toward uniformity.

Nevertheless, the current array of statutes has a common ancestor in the English common law.

A. Common Law Burglary

In his *Institutes of the Laws of England*, Lord Coke in 1641 defined burglary as:

> A Burglar (or the person that committeth burglary) is by the common law a felon, that in the night breaketh and entreth into a mansion house of another, of intent to kill some reasonable creature, or to commit some other felony within the same, whether his felonious intent be executed or not.[17]

This is the oft-cited[18] common law definition that still influences the law today. Burglary had been an offense long before Lord Coke's formulation, having its roots in the crime known as "housebreaking" or "hamsecken."[19] But Coke's six-element crime took hold.

Coke also defined each element of burglary. *Night* he defined as when "darkeness comes" and "you cannot discerne the countenance of a man."[20] Night "doth aggravate the offence, for the night is the time wherein man is to rest, and wherein beasts runne about seeking their prey."[21] *Breaking* he defined through examples: If a thief enters the house through an open door or window, there is no breaking; but if the thief breaks the glass of the window and uses a hook to

---

meaningless.").

17. SIR EDWARD COKE, THE THIRD PART OF THE INSTITUTES OF THE LAWS OF ENGLAND 63 (London, W. Clarke & Sons 1809) (1644).

18. Coke "through his Institutes turn[ed] a jumble of law into an astonishingly complete, reconciled, organized body of propositions which concealed all the 'inconsistencies and difficulties which were inherent in his position[.]'" GERHARD O. W. MUELLER, CRIME, LAW AND THE SCHOLARS 17 (1969). Coke's burglary definition is cited in Wright, *supra* note 7, at 411, and *A Rationale of the Law of Burglary*, *supra* note 7, at 1009. This is substantially the common law definition referred to, without attribution, in the commentaries to the Model Penal Code. MODEL PENAL CODE § 221.1 cmt. 1, at 61 (Official Draft & Revised Comments 1980).

19. Commonwealth v. Hope, 39 Mass. (22 Pick.) 1, 4-5 (quoting 1 Hale's P.C. 547). The gravamen of the offense was to break and enter a house, and a felonious intent was more or less assumed from the breaking and entering. *Id*. The crime was intended to protect the sanctity and security of the home, and was punishable by death. Lauer, *supra* note 3, at 724. The crime was mentioned as early as 942. *Id.* A thirteenth century reference to burglars defines them as "all those who feloniously in the time of peace break churches, or the houses of others, or the walls or gates of our cities or boroughs." BRITTON: AN ENGLISH TRANSLATION AND NOTES 36 (Francis Morgan Nichols ed. 1901), *cited in* Lauer, *supra* note 3, at 725 n.15.

20. COKE, *supra* note 17, at 63.

21. *Id.*

pull goods out of the house through the window, there has been a breaking (and an entry).[22] *Entry* could be made by any part of the body or a weapon or a tool.[23] A *mansion house* did not mean only grand homes, but included "every house for the dwelling and habitation of man."[24]

Coke treated "breaking and entering" together, almost as if they were one element. Separately or together, they caused great difficulty for the English courts. For example, was there a breaking if a man climbed down a chimney opening? Did it make a difference if he dislodged some bricks during his descent?[25] As one scholar wrote of the eighteenth and early nineteenth centuries in England:

> A person who assaulted a home-owner at the latter's threshold was held to be a burglar because his pistol passed over the line of the doorway. An offender who, in the act of unfastening a window, allowed his finger to pass over the sill, but who was apprehended at that point, was held to have entered the house. But the conviction of another offender was reversed because, in the act of prying open a shutter, no part of his body or any instrument entered the space between the shutter and the window.[26]

The mansion-house or dwelling requirement also led to fine distinctions and expansion. For example, the dwelling was expanded to "include all out-buildings within the curtilage of the dwelling provided they were enclosed by a common fence."[27] If a home was left in the care of domestic servants, it was still a dwelling, but not if left in the care of other employees.[28] The other elements of burglary lead to similar hair-splitting decisions.[29]

B. Burglary in America Before 1960

The colonists brought the common law with them from England, but they did not simply replicate the English system.[30] Crime became a matter for legislation. The colonies had their own penal codes—Massachusetts' colonial code provided that a burglar be branded on the forehead with the letter "B" for a first offense.[31] Throughout the nineteenth century "[t]he concept of the common-law crime was

---

22. *Id*. at 64.

23. *Id*.

24. *Id*.

25. *See* JOHN HOSTETTLER, A HISTORY OF CRIMINAL JUSTICE IN ENGLAND AND WALES 76 (2009) (discussing Sir Matthew Hale's analysis of such a case in 1736).

26. Wright, *supra* note 7, at 412 (citations omitted).

27. *Id.* at 413.

28. *Id*.

29. *See id*. at 413-14.

30. *See* MUELLER, *supra* note 18, at 9-15.

31. Lauer, *supra* note 3, at 730.

in retreat."[32] More and more states passed comprehensive penal codes, and there was a prevailing idea that judges should not make up crimes.[33]

As burglary became a creature of statute, it varied by jurisdiction. By the end of the nineteenth century, the elements of night time, entry, breaking, and the structure to be entered, varied considerably.[34] "Some states retained the common law elements of dwelling house and in the night time; others broadened burglary to include structures of nearly all kinds, and embraced entries made in both day and night."[35] These variations persisted through the first half of the twentieth century. By 1950, breaking and night time were on the wane, the mansion house or dwelling had been expanded to include many types of structures, but entry was still required, as shown by the following summary of the two 1951 surveys of burglary law.[36]

1. Night Time.—Already in 1951, night time was not an element in eleven jurisdictions.[37] In nine jurisdictions, however, it was the sole aggravating circumstance for first degree burglary,[38] and in thirty-two jurisdictions it was a requirement for the highest level of burglary.[39] Night time was usually defined as the hours between sunset and sunrise, or thirty minutes after sunset until thirty minutes before sunrise.[40] An alternative definition was Coke's "when a man's face could not be discerned."[41]

2. Breaking.—In 1951, breaking was an element of burglary in eighteen[42] or nineteen[43] jurisdictions. At that time, twelve jurisdictions did not require breaking.[44] But even where breaking was a statutory element, courts were willing, as under the common law, to stretch the requirement to include minimal force such as raising a partly open window, or pushing open an unlatched door.[45] Breaking might also be satisfied by "constructive breaking," gaining entry by ruse or deceit.[46] As noted later in the Model Penal Code commentaries, "the 'breaking' had become little more than symbolic."[47]

---

32. LAWRENCE M. FRIEDMAN, CRIME AND PUNISHMENT IN AMERICAN HISTORY 65 (1993).

33. *See id*.

34. *See* Lauer, *supra* note 3, at 728-30.

35. *Id.* at 731-32.

36. *See A Rationale of the Law of Burglary*, *supra* note 7; Wright, *supra* note 7.

37. Wright, *supra* note 7, at 417.

38. *Id.*

39. *A Rationale of the Law of Burglary*, *supra* note 7, at 1015.

40. Wright, *supra* note 7, at 417 & n.52.

41. Martha Grace Duncan, *Beauty in the Dark of Night: The Pleasures of Form in Criminal Law*, 59 EMORY L.J. 1203, 1242 (2010) (citation omitted).

42. *A Rationale of the Law of Burglary*, *supra* note 7, at 1014.

43. Wright, *supra* note 7, at 415.

44. *A Rationale of the Law of Burglary*, *supra* note 7, at 1013.

45. Wright, *supra* note 7, at 416 n.44.

46. *A Rationale of the Law of Burglary, supra* note 7, at 1012 n.19 (citing cases where defendants lied or otherwise tricked a person into admitting them onto premises).

47. MODEL PENAL CODE § 221.1 cmt. 3, at 69 (Official Draft & Revised Comments 1980).

3. Entry.—In 1951, "virtually all" jurisdictions required an entry for burglary, although the requirement was often applied in surprising ways.[48] Texas, for example, included the firing of a bullet into a house as an entry.[49] As under common law, the entry of any part of the body or a tool connected to the body might qualify.[50] Some jurisdictions had statutes that required breaking *or* entering, but a breaking without entry would most likely be treated as an attempted burglary.[51]

4. Dwelling.—The common law requirement that the site of burglary be a dwelling was early on broadened to include other types of buildings, especially outbuildings within the curtilage.[52] Well before the mid-twentieth century, many statutes included shops, storehouses, ships, churches, etc., as buildings that could be burglarized.[53] By 1951, many statutes contained long lists of structures, including vehicles and railroad cars.[54] Not all statutes were so broad, but none was restricted to dwellings.[55]

In thirty-one jurisdictions, burglary of a dwelling was a distinguishing aspect of the highest degree of burglary in 1951.[56] But, while the burglary of a dwelling, especially an occupied dwelling, continued to be treated severely, by 1951 the burglary laws had expanded to include all kinds of structures and vehicles.

5. Of Another.—The common law requirement that the place entered be that "of another" had almost disappeared from the statutory definitions of burglary by the mid-twentieth century.[57] However, it remained an implied element. Courts consistently ruled that "a man cannot commit burglary in his own house."[58] The right protected was possession. Possession meant a right of

---

48. *See* Wright, *supra* note 7, at 416.

49. *Id*. at 416 n.47. The statutory definition of entry no longer includes shooting into a building. *See* TEX. PENAL CODE ANN. § 30.02 (West 2011).

50. Wright, *supra* note 7, at 416.

51. *Id*. at 416 n.45.

52. *See* State v. Engel, 210 P.3d 1007, 1011 (Wash. 2009) (discussing the curtilage concept in determining the scope of a "fenced area" under the burglary statute).

53. Lauer, *supra* note 3, at 731.

54. One example comes from Nebraska: "dwelling house, kitchen, smokehouse, slaughterhouse, shop, office, storehouse, mill, pottery, factory, watercraft, schoolhouse, church or meetinghouse, barn, chicken house, stable, warehouse, malthouse, stillhouse, public building, or other private building, railroad car factory, station house, railroad car, public or private telephone pay station or booth." Wright, *supra* note 7, at 417 (quoting NEB. REV. STAT. § 28-532 (1948)).

55. *See id.* at 418. "In every jurisdiction, virtually all buildings are covered by one or more degrees of the crime." *A Rationale of the Law of Burglary*, *supra* note 7, at 1011.

56. *A Rationale of the Law of Burglary*, *supra* note 7, at 1011. According to the other 1951 article, the element of "dwelling" was an additional aggravating circumstance in fourteen states, and the sole aggravating circumstance in seven. Wright, *supra* note 7, at 418-19. Many of these statutes required that the dwelling be occupied or inhabited—in contrast to the common law. *See id*. at 419.

57. Wright, *supra* note 7, at 419.

58. *Id*.

occupancy, not necessarily ownership. Thus, for example, a lessor could burglarize a structure leased to another.[59]

6. Intent to Commit a Felony Therein.—Lord Coke's definition required the intent to commit a felony.[60] By 1951, most state statutes still required the "intent to commit a felony, or a felony or any larceny."[61] (The inclusion of "any larceny" or "any theft" allowed prosecutors to proceed without proving a felonious monetary value for the goods intended to be stolen.[62] Such an approach would be helpful to prosecutors where the defendant was apprehended before he took any property.) Fifteen state statutes then allowed the intent to commit any crime, and seven state statutes limited the intended crimes to a specified list.[63]

7. Grading of Burglary.—At common law, burglary was a single offense, punishable by death.[64] In the United States, the penalty varied by state, even in the eighteenth century.[65] As with the other common law offenses, over time, most jurisdictions developed different grades of burglary or increased penalties if different factors were present. In the mid-twentieth century, close to two-thirds of the states had a separate offense for "burglary with explosives," which was punished more severely than other burglary crimes.[66] Seventeen states provided a heavier penalty if the burglar was armed.[67] Seventeen required that the structure entered be "occupied" or "inhabited" to support first degree burglary charges.[68] Looking at aggravation slightly differently, another author reported that, in 1951, twelve states had what is essentially common law burglary as burglary in the first degree.[69]

Thus, during the 1950s, as the American Law Institute was drafting the

---

59. *Id*. at 420.

60. *See supra* text accompanying note 17.

61. Wright, *supra* note 7, at 420 (citation omitted). At that time, twenty-eight jurisdictions required the intent to commit a felony or any larceny. *A Rationale of the Law of Burglary, supra* note 7, at 1017.

62. *See* MODEL PENAL CODE § 221.1 cmt. 3(c), at 77 (Official Draft & Revised Comments 1980).

63. *A Rationale of the Law of Burglary, supra* note 7, at 1017.

64. *See* HOSTETTLER, *supra* note 25, at 76.

65. For example, in the early years of the Republic, Pennsylvania and New York abolished the death penalty for burglary and some other serious offenses. FRIEDMAN, *supra* note 32, at 73.

66. *A Rationale of the Law of Burglary*, *supra* note 7, at 1018 (stating that twenty-nine states had separate offense of burglary with explosives); *see also* Wright, *supra* note 7, at 430 n.148 (stating that all but seventeen state jurisdictions had a separate offense of burglary with explosives).

67. *A Rationale of the Law of Burglary*, *supra* note 7, at 1018.

68. *Id.* at 1019 n.72. To be occupied or inhabited, structure need not have a person present; such structures were much like the common law definition of a dwelling. *See id.*

69. Wright, *supra* note 7, at 421. According to this source, another group of states then used one or more of the following factors to elevate the crime: a person is present, the defendant is armed, the defendant assaults someone, the defendant brings confederates, or the defendant uses false keys. *Id.* at 422. One or more of the common law elements of "night time" or "dwelling" were also used to elevate the crime in some jurisdictions. *Id.* at 422-23.

Model Penal Code, the critical elements of burglary were entry of a structure with intent to commit a crime. The more serious grades of burglary often included the elements of an occupied dwelling, or weapons.

II. The Model Penal Code's Burglary Proposal

The Model Penal Code Final Draft appeared in 1962.[70] The final draft included general provisions about criminal liability and defenses, as well as proposals for specific offenses.[71] The comments to the model burglary offense noted the way in which the crime had already been significantly broadened from the common law definition, and criticized the harshness and irrationality of the offense.[72] The comments suggest that the authors considered eliminating burglary as a distinct offense, but that "[c]enturies of history and a deeply imbedded Anglo-American conception such as burglary, however, are not easily discarded."[73] Instead, the Model Code proposed an offense "limited . . . to the invasion of premises under circumstances especially likely to terrorize occupants."[74]

Section 221.1 defined the crime of burglary as follows:

> (1) <u>Burglary Defined.</u> A person is guilty of burglary if he enters a building or occupied structure, or separately secured or occupied portion thereof, with purpose to commit a crime therein, unless the premises are at the time open to the public or the actor is licensed or privileged to enter. It is an affirmative defense to prosecution for burglary that the building or structure was abandoned.
>
> (2) <u>Grading.</u> Burglary is a felony of the second degree if it is perpetrated in the dwelling of another at night, or if, in the course of committing the offense, the actor:
>
> (a) purposely, knowingly or recklessly inflicts or attempts to inflict bodily injury on anyone; or
>
> (b) is armed with explosives or a deadly weapon.
>
> Otherwise, burglary is a felony of the third degree. An act shall be deemed "in the course of committing" an offense if it occurs in an attempt to commit the offense or in flight after the attempt or commission.
>
> (3) <u>Multiple convictions.</u> A person may not be convicted both for burglary and for the offense which it was his purpose to commit after the

---

70. *See* MODEL PENAL CODE: COMPLETE STATUTORY TEXT (1985) (containing the completed official draft adopted in 1962).

71. *See generally id.*

72. *See* MODEL PENAL CODE § 221.1 cmts. 1-2, at 61-68 (Official Draft & Revised Comments 1980).

73. *Id.* § 221.1 cmt. 2, at 67.

74. *Id.*

> burglarious entry or for an attempt to commit that offense, unless the additional offense constitutes a felony of the first or second degree.[75]

In addition to this model burglary offense, the Model Code proposed the crime of criminal trespass, which did not require proof of intent to commit a crime.[76] At common law, trespass was not a crime, only a private wrong.[77]

The model burglary statute thus is broader than the common law crime because it encompasses entry of buildings, not just dwellings, and is not restricted to night time. The intended crime need not be a felony. However, the authors rejected other expansions of the crime already endorsed by various jurisdictions. For example, the proposed statute applies only to buildings and occupied structures, not vehicles or storage containers, and it requires an unprivileged entry.[78] However, entry of a dwelling at night results in a higher degree of the model offense.[79] The crime is also aggravated if the actor attempts bodily injury or is armed[80]—reflecting a concern with the risk to personal security. Thus, the Model Code increases the punishment for conduct that threatens human life or safety. Moreover, reflecting a concern for proportionality, the Model Code prohibits adding a burglary conviction to the conviction for the completed or attempted target offense.

The authors of the Model Penal Code believed that the crime had broadened to compensate for defects in the law of attempts, and that the expanded crime was no longer necessary in light of the Model Penal Code reform of attempt law.[81] According to the comments, common law attempt was difficult to prove because the actor must have come very close to achieving the criminal goal—sometimes requiring that the actor commit the "final act," and only being thwarted by circumstances beyond the actor's control.[82]

Burglary was a common law offense long before attempts were made generally punishable.[83] A possible illustration of burglary substituting for attempt is a seventeenth century English case in which the defendants were convicted and executed for burglary after unsuccessfully trying to kill a man by shooting him through a hole in the wall of their adjoining houses.[84] Today, such

75. *Id.* § 221.1.

76. *Id.* § 221.2.

77. *See, e.g.*, State v. Pierce, 417 A.2d 1085, 1087 (N.J. Super. Ct. Law Div. 1980).

78. *See* MODEL PENAL CODE § 221.1 cmt. 3. The authors sought to avoid the application of burglary to what would otherwise be considered theft or shoplifting from a commercial establishment open to the public. *See id.* They specifically rejected the expansion of entry to include "remaining," or even "remaining surreptitiously," drawn from the model criminal trespass statute. *See id.* § 221.2.

79. *Id.* § 221.1(2).

80. *Id.*

81. *Id.* § 221.1 cmt. 2.

82. *Id*.

83. *See* LAFAVE, *supra* note 9, § 11.4(e).

84. This case is recounted in HOSTETTLER, *supra* note 25, at 76. "Presumably it was shown

an act would probably be charged as both attempted murder and burglary. The modern—and Model Penal Code—approach is to allow conviction for attempt based on a "substantial step" toward the criminal goal.[85]

According to the Model Penal Code's comments, and other contemporary scholars, the crime of burglary had been expanded by courts and legislatures to reach conduct that threatened persons and property but could not otherwise be punished under the common law as an attempt.[86] The comments noted the irrational results of this expansion in many jurisdictions: stealing a car might be punished less severely than breaking into the car to take something from the glovebox; stealing a chicken might be petty larceny, but entering a henhouse to steal the chicken would be a serious offense.[87] The comments also noted the unfairness resulting from adding burglary sentences to the punishment for the completed target offense.[88] This "double punishment" was not possible with attempts; an ordinary attempt would merge with the completed offense so that a defendant could not be convicted of both the completed offense and an attempt.[89]

Despite these criticisms of burglary law, the Model Code retained the offense of burglary in part out of "deference to the momentum of historical tradition," and because the offense "reflects a considered judgment that especially severe sanctions are appropriate for criminal invasion of premises under circumstances likely to terrorize occupants."[90] The Model Code's approach reflected its "principled pragmatism."[91]

The Model Penal Code critique of burglary reflected contemporary concerns

---

that the tip of the pistol entered the injured man's room." *Id*.

85. MODEL PENAL CODE § 5.01(1)(c); *see* LAFAVE, *supra* note 9, § 11.4(e) (noting that the Model Penal Code's substantial step language has been adopted by the majority of states in the "modern recodifications").

86. *See* MODEL PENAL CODE § 221.1 cmt. 1; *see also A Rationale of the Law of Burglary*, *supra* note 7, at 1023 (noting that the law of attempt provides an "incidental justification for the law of burglary"); Susan Bundy Cocke, Note, *Reformation of Burglary*, 11 WM. & MARY L. REV. 211, 224 (1969) (arguing that burglary had evolved to compensate for defects in attempt law leading to "chaotic and anomalous theoretical and practical results"); Wright, *supra* note 7, at 433 (arguing that burglary is a type of attempt: "No other reasoning sufficiently explains the great expansion of burglary made by our statutes.").

87. MODEL PENAL CODE § 221.1 cmt. 1, at 63-65.

88. *See id.* at 65-66.

89. *Id*.

90. *Id.* § 221 introductory note.

91. Herbert L. Packer, *The Model Penal Code and Beyond*, 63 COLUM. L. REV. 594, 594 (1963).

> [T]he dominant tone of the Code is one of principled pragmatism. Perhaps the adjective and the noun should be reversed, because fidelity to principle is the solid base on which the Code is built. But its provisions reflect an awareness that the discernment of right principles is only the beginning of rational lawmaking and that the besetting sin of rationality is the temptation to press a principle to the outer limits of its logic.

Id.

with deterrence and rationalization of the criminal code. Burglary had been broadened far beyond housebreaking, to the point where it could function as a "generalized law of attempts" in many jurisdictions, protecting personal and property security.[92] But burglary was never solely aimed at attempted crimes. Its original concern was for the security of the home and the potential for violence and terror resulting from home invasion.[93] In any event, burglary has picked up multiple justifications along its journey to its present form.[94]

The Model Penal Code drafters were correct, however, in their observations that burglary unfairly added excessive penalties to otherwise ordinary thefts. They could not have foreseen how much further the crime would expand, and how its role as a location aggravator would grow.

---

92. "There exists persuasive argument that statutory burglary has been enlarged to such an extent that it has become, in reality, a generalized law of attempts, and there exists conclusive support for the proposition that burglary is no longer aimed at the protection of the habitation." Cocke, *supra* note 86, at 213 (citations omitted).

93.

> It is evident that the offense of burglary at common law was considered one aimed at the security of the habitation rather than against property. That is to say, it was the circumstance of midnight terror aimed toward a man or his family who were in rightful repose in the sanctuary of the home, that was punished, and not the fact that the intended felony was unsuccessful. Such attempted immunity extended to a man's dwelling or mansion house has been said to be attributable to the early common-law principle that a man's home is his castle. The jealousy with which the law guarded against any infringement of this ancient right of peaceful habitation is best illustrated by the severe penalties which at common law were assessed against a person convicted of burglary, even though the enterprise, except for the essential elements of breaking and entering a mansion house or dwelling house at night with intent to commit a felony therein, was unsuccessful.

Id. at 211 n.5 (quoting Annotation, *Burglary: Outbuildings or the Like as Part of a "Dwelling House,"* 43 A.L.R.2d 831, 834-35 (1955)).

94. *See* 13 AM. JUR. 2D *Burglary* § 3 (2009).

> The purpose of burglary statutes is to protect possessory rights with respect to structures and conveyances, to define prohibited space and to protect the integrity of the home. The historical principle underlying the law of burglary is protection of the right of habitation. Thus, burglary is actually an offense against the possession of property and not necessarily against the ownership thereof.
>
> However, it has also been stated that burglary laws are based primarily upon a recognition of the dangers to personal safety created by the usual burglary situation—the danger that the intruder will harm the occupants in attempting to perpetrate the intended crime or to escape, and the danger that the occupants will in anger or panic react violently to the invasion, thereby inviting more violence. Thus, burglary laws are also designed primarily to protect against the creation of a situation dangerous to personal safety caused by unauthorized entry.

Id. (citations omitted).

III. BURGLARY AFTER THE MODEL PENAL CODE

In 1980-85, revised commentaries to the Model Penal Code were published "to reflect and explore the far reaching legislative and judicial response to the Code."[95] By then, there had been "widespread revision and codification of the substantive criminal law of the United States";[96] at least twenty-nine revised codes by 1980, with more in process.[97] With respect to burglary laws, some aspects of the Model Penal Code were incorporated in many revisions, but there was great variation, as shown below.[98] Some aspects of the proposed code were soundly rejected, as reflected by, for example, the widespread adoption of "remaining" as an alternative to "entry" in defining the elements of burglary.

When states adopted aspects of the model crime, such as basing a higher degree on possession of a deadly weapon or infliction of bodily injury, it is difficult to know whether the Model Code influenced the legislatures, or vice versa. Regardless of the cause, there are some very clear trends in burglary statutes since the Model Penal Code. The element of "night time" has almost disappeared. "Breaking" remains a formal element in only twelve jurisdictions, and even in these, it is broadly interpreted. "Entry," once considered an essential aspect of burglary, is no longer required in a majority of jurisdictions—"remaining" may suffice. The structure entered or remained in need not always be that "of another," especially where domestic violence is involved. The majority of states no longer require the "intent to commit a felony," but more commonly require only the intent to commit a crime. Thus, since the Model Penal Code, the erosion of the common law elements has continued. The most significant change is the widespread elimination of entry as a requirement.

Developments in the law of burglary since the Model Penal Code are addressed in detail below with respect to each of the major common law elements.

A. Night Time

The Model Penal Code burglary proposal did not have night time as an element of ordinary burglary but included it as an element of the higher degree offense.[99] This approach has not been taken up by the states. Today, only two

---

95. 1 MODEL PENAL CODE AND COMMENTARIES, pt. I, at xii (1985) [hereinafter MPC COMMENTARIES]. The revised commentaries to the model burglary statute were published in 1980. *See* MODEL PENAL CODE (Official Draft & Revised Comments 1980).

96. MPC COMMENTARIES, *supra* note 95, at xi.

97. *See id*.

98. The general provisions of the Model Penal Code were very influential, but its proposals for the definitions of particular crimes were less so. *See* Gerard E. Lynch, *Towards a Model Penal Code, Second (Federal?): The Challenge of the Special Part*, 2 BUFF. CRIM. L. REV. 297, 299 (1998).

99. MODEL PENAL CODE § 221.1 (Official Draft & Revised Comments 1980).

states retain night time as an element of burglary—although both jurisdictions also have lesser offenses for what would in most places be considered burglary.[100] Only seven other states use night time as a factor to elevate the degree of the offense.[101]

The use of night time as an element of burglary offenses has thus diminished considerably. Its lesser importance in a world of artificial light probably makes sense: In the urban environments where most people now live, it is rare to encounter a darkness so deep that "a man's face could not be discerned." In addition, with burglary now applicable to many more structures and places than the home, time of day seems less relevant. Historically, night time invasions of the home were seen as particularly threatening.[102] The special protection of the dwelling at night is seen in search and seizure law.[103] ("[I]t is difficult to imagine a more severe invasion of privacy than the nighttime intrusion into a private home. . . ."[104])

Burglary is no longer concerned only, or even primarily, with protecting the

---

100. *See* MASS. GEN. LAWS ANN. ch. 266, §§ 14, 18 (West 2012) (defining burglary as breaking and entering a dwelling house in the night time, but also defining an offense of breaking and entering in the daytime); VA. CODE ANN. §§ 18.2-89, -90 (2011) (defining burglary as breaking and entering in the night time with intent to commit a felony, but also defining an offense of breaking and entering a dwelling in the daytime with intent to commit certain felonies).

101. *See* CONN. GEN. STAT. ANN. § 53a-101 (West 2012); DEL. CODE ANN. tit. 11, § 826 (2012); N.D. CENT. CODE § 12.1-22-02 (2011); R.I. GEN. LAWS § 11-8-5 (2011) (defining offense of entering a building in the night time with intent to commit a crime). Other breaking and entering offenses do not require that they occur at night. *Id.* §§ 11-8-2, -2.1, -2.2; S.C. CODE ANN. §16-11-311(2011); S.D. CODIFIED LAWS § 22-32-1 (2011); W. VA. CODE ANN. § 61-3-11 (West 2012).

102. *See* Duncan, *supra* note 41, at 1242 (commenting on the role of darkness and night in criminal law).

103.

> At common law there was a strong hostility to nighttime searches of a dwelling house. *Commonwealth v. Hinds*, 145 Mass. 182, 13 N.E. 397 (1887). Nighttime searches were regarded with revulsion because of the indignity of rousing people from their beds. *Commonwealth v. DiStefano*, 22 Mass. App. Ct. 535, 541, 495 N.E.2d 328 (1986). The underlying rationale was that night time police intrusion posed a great threat to privacy, violated the sanctity of home, and endangered the police and slumbering citizens. 2 W.R. LaFave, Search and Seizure § 4.7(b), at 266 (2d ed. 1987). See, e.g., *Gooding v. United States*, 416 U.S. 430, 463, 94 S.Ct. 1780, 1797, 40 L.Ed.2d 250 (1974) (Marshall, J., dissenting) (nighttime searches involve greater intrusion than ordinary searches); *Monroe v. Pape*, 365 U.S. 167, 210, 81 S.Ct. 473, 496, 5 L.Ed.2d 492 (1961) (Frankfurter, J., dissenting in part) ("Searches of the dwelling house were the special object of this universal condemnation of official intrusion. Night-time search was [considered] the evil in its most obnoxious form"); *Jones v. United States*, 357 U.S. 493, 498, 78 S.Ct. 1253, 1257, 2 L.Ed.2d 1514 (1958) ("[I]t is difficult to imagine a more severe invasion of privacy than the nighttime intrusion into a private home").

Commonwealth v. Grimshaw, 595 N.E.2d 302, 304 (Mass. 1992) (alteration in original).

104. Jones v. United States, 357 U.S. 493, 498 (1958).

home, and night time is no longer what it was when the common law was developed. The fact that the element of night time is retained at all in nine jurisdictions is a testament to the enduring influence of the common law definition.

B. Breaking

The element of breaking was not included in the Model Penal Code proposal. As with night time, even the formal breaking requirement has significantly eroded in the last sixty years. As an element, breaking has long been interpreted liberally, and now it has been eliminated altogether in more than two-thirds of United States jurisdictions.

Twelve jurisdictions retain breaking as an element, although not for all degrees of the offense, and in most, it has been judicially interpreted to mean little more than unlawful entry.[105] Most jurisdictions permit "constructive breaking," meaning entry gained by artifice, trick, fraud or threat.[106] In Virginia and Massachusetts, breaking is required for daytime burglary, but entry alone is sufficient for night.[107] The use of force to gain entry is no longer an essential aspect of burglary.

---

105. *See* MD. CODE ANN. CRIM. LAW §§ 6-202, -203, -204 (West 2011) (includes breaking as an element, but breaking includes entry gained by artifice or fraud, Winder v. State, 765 A.2d 97, 124 (Md. 2001)); TEX. PENAL CODE ANN. §§ 30.02, -.03, -.04 (West 2011) (breaking is an element only of burglary of coin-operated machines or burglary of vehicles); W. VA. CODE ANN. §§ 61-3-11, -12 (West 2012) (breaking only required for some burglary crimes); *see also* Commonwealth v. Labare, 416 N.E.2d 534, 538 (Mass. App. Ct. 1981) (holding breaking includes "constructive breaking"); People v. Toole, 576 N.W.2d 441, 443 (Mich. Ct. App. 1998) (holding any force, however slight, used to open a door or window constitutes a breaking); Magee v. State, 966 So. 2d 173, 180 (Miss. Ct. App. 2007) (holding any effort, such as turning a door knob, constitutes a breaking); State v. McDowell, 522 N.W.2d 738, 744 (Neb. 1994) (holding that opening of a closed door is breaking); State v. Abdullah, 967 A.2d 469, 476 (R.I. 2009) (holding breaking requires force, however slight, to gain entry); Bright v. Commonwealth, 356 S.E.2d 443, 445 (Va. Ct. App. 1987) (opening of secured window is breaking).

106. *See, e.g.*, State v. Parker, 516 S.E.2d 106, 117 (N.C. 1999) (holding that for first degree burglary, "breaking may be actual or constructive"); Patton v. State, 973 P.2d 270, 287 (Okla. Crim. App. 1998) (holding breaking includes constructive breaking by fraud, trick or threat).

107. In Massachusetts, breaking is accomplished by any slight force, and includes constructive breaking. *See* 14A HOWARD J. ALPERIN, MASS. PRACTICE, SUMMARY OF BASIC LAW § 7.206 (4th ed. 2011). The law is similar in Virginia. *See, e.g.*, Johnson v. Commonwealth, 275 S.E.2d 592, 594 (Va. 1981). The breaking requirement cannot be completely ignored, however. *See, e.g.*, Finney v. Commonwealth, 671 S.E.2d 169, 173-74 (Va. 2009) (holding no evidence of breaking where no indication defendant applied even slight force to enter owner's shed); Broady v. Commonwealth, 429 S.E.2d 468, 473-74 (Va. Ct. App. 1993) (reversing burglary conviction where no evidence of actual or constructive breaking when defendant followed couple into their hotel room).

C. Entry

One mid-century researcher noted that if breaking were not a requirement, the practical result might be the elimination of the element of unlawful entry.[108] "The result is that, where the scope of the crime extends to stores, every shoplifter who enters with the requisite intent . . . is liable to far larger penalties than those for larceny."[109] His example was not that far-fetched,[110] and his analysis of the relationship between a decline in breaking and a decline in the requirement of entry proved prescient. The breaking requirement, however weak, was one way to ensure entry was unlawful.[111]

The Model Penal Code proposal required entry, but it attached no liability if "the premises are at the time open to the public or the actor is licensed or privileged to enter."[112] Thus, under the Model Code, entry must be unauthorized.

Since that time, the requirement of entry has become the minority approach. At least twenty-nine jurisdictions have modified the statutory entry requirement to include "remaining unlawfully" or "remaining."[113] Only some of these

---

108. *A Rationale of the Law of Burglary*, *supra* note 7, at 1014-15.

109. *Id.* at 1014. He advocated a requirement that if breaking is not an element, the entry be unpermitted or trespassery. *Id*.

110. Even one hundred years ago, there were courts that upheld burglary convictions where the entry was lawful. *See, e.g.*, Pinson v. State, 121 S.W. 751, 753-54 (Ark. 1909) (defendant entered saloon through main door during business hours with intent to steal whiskey); People v. Barry, 29 P. 1026, 1027 (Cal. 1892) (defendant entered grocery store during business hours with intent to steal food), *cited in* Wright, *supra* note 7, at 419 n.71. California has continued to adhere to this rule. *See* Comment, *Criminal Law—Development of the Law of Burglary in California*, 25 S. CAL. L. REV. 75, 88 (1951) (citing cases); *see also* Magness v. Superior Court, 126 Cal. Rptr. 3d 318, 325 (Ct. App.), *petition for review granted*, 260 P.3d 284 (Cal. 2011) (citing *Barry*, 29 P. 1026).

111. In fact, Coke treated breaking and entry together, almost as a single requirement. COKE, *supra* note 17, at 64.

112. MODEL PENAL CODE § 221.1 (Official Draft & Revised Comments 1980).

113. *See* ALA. CODE §§ 13A-7-5, -6 (2012); ALASKA STAT. § 11.46.310 (2011); ARIZ. REV. STAT. ANN. §§ 13-1506, -1507 (2012); ARK. CODE ANN. § 5-39-201 (2012); COLO. REV. STAT. §§ 18-4-202, -203 (2011); CONN. GEN. STAT. ANN. §§ 53a-101,-102, -103 (West 2012); DEL. CODE ANN. tit. 11, §§ 824, 825, 826 (West 2012); FLA. STAT. ANN. § 810.02 (West 2012); GA. CODE ANN. § 16-7-1 (2011); HAW. REV. STAT. §§ 708-810, -811 (West 2011); 720 ILL. COMP. STAT. ANN. 5/19-1 (West 2012); IOWA CODE ANN. § 713.1 (West 2012); KAN. STAT. ANN. § 21-5807 (West 2011); KY. REV. STAT. ANN. §§ 511.020, -.030, -.040 (West 2011); ME. REV. STAT. ANN. tit. 17-A, § 401 (2011); MO. REV. STAT. §§ 569.160, -170 (2011); MONT. CODE ANN. § 45-6-204 (2009); N.J. REV. STAT. § 2C:18-2 (2011); N.Y. PENAL LAW §§ 140.20, -.30 (McKinney 2010); N.D. CENT. CODE § 12.1-22-02 (2011); OHIO REV. CODE ANN. §§ 2911.11, -.12, -.13 (West 2011) ("No person by force, stealth, or deception, shall trespass . . . .") (Trespass is defined in part as to "knowingly enter or remain." *Id.* § 2911.21); OR. REV. STAT. ANN. § 164.215 (West 2011); S.D. CODIFIED LAWS §§ 22-32-1, -3, -8 (2011); TENN. CODE ANN. § 39-14-402 (2011); TEX. PENAL CODE ANN. § 30.02 (West 2011); UTAH CODE ANN. § 76-6-202 (West 2011); VT. STAT. ANN. tit.

jurisdictions require that the remaining be "surreptitious"[114] (language recommended by the Model Penal Code provision for criminal trespass[115]) or "concealed."[116] The extension of burglary statutes explicitly to situations where the defendant "remains unlawfully" or simply "remains" seems to have begun sometime between the 1951 studies (which reported that entry was a technical requirement in every jurisdiction, although many courts did not require a trespassory entry[117]) and the revised comments to the Model Penal Code, published in 1980. The revised comments to burglary noted with regret that New York and several other states had "recently" adopted the "enters or remains unlawfully" language.[118]

As a result, even where the initial entry was permitted and lawful, a burglary conviction may result if the defendant remains, after permission is withdrawn, with the intent to commit a crime.[119] Under this theory, a defendant may be convicted of burglary as well as robbery when the defendant enters an open business and robs it.[120] Furthermore, where the statute includes "remaining" as an alternative to entry, the criminal intent may be formed at any time while the defendant remains on the premises and need not have been formed at the time of entry.[121] In these jurisdictions, a consensual visit that turns ugly might be prosecuted as a burglary.[122]

Whether the statute requires that entering or remaining be trespassory also matters. If there is no such requirement, then shoplifting may be burglary. Thus, without the requirement of breaking or even unlawful entry, the character of burglary in many places has expanded considerably from the common law crime

---

13, § 1201 (2011); WASH. REV. CODE ANN. §§ 9A.52.020, -.025, -.030 (West 2012); WYO. STAT. ANN. § 6-3-301 (2011).

114. FLA. STAT. ANN. § 810.02; ME. REV. STAT. ANN. tit. 17A, § 401; N.J. REV. STAT. § 2C:18-2; N.D. CENT. CODE § 12.1-22-02; VT. STAT. ANN. tit. 13, § 1201.

115. MODEL PENAL CODE § 221.2 (Official Draft & Revised Comments 1980).

116. TENN. CODE ANN. § 39-14-402; TEX. PENAL CODE ANN. § 30.02.

117. *See A Rationale of the Law of Burglary*, *supra* note 7, at 1012; Wright, *supra* note 7, at 416.

118. MODEL PENAL CODE § 221.1, cmt. 3, at 69-71.

119. *See, e.g.*, People v. Leonard, 921 N.Y.S.2d 337, 340 (App. Div. 2011); State v. Morton, 768 N.E.2d 730, 737-38 (Ohio Ct. App. 2002).

120. *See* Bowling v. Commonwealth, 942 S.W.2d 293, 307 (Ky. 1997), *overruled on other grounds by* McQueen v. Commonwealth, 339 S.W.3d 441 (Ky. 2011) (upholding burglary conviction where defendant entered gas station, shot two employees and fled with contents of cash register). But if the defendant leaves immediately after permission to remain is revoked, there is no burglary. *See* Wilburn v. Commonwealth, 312 S.W.3d 321, 325 (Ky. 2010) (finding insufficient evidence of burglary where defendant entered open liquor store, attempted to rob clerk, but fled when clerk revoked permission to enter by firing a gun).

121. *See* State v. Garcia, 236 P.3d 853, 856 (Utah Ct. App. 2010) (upholding burglary conviction where defendant may have formed intent to assault victim after entering but while remaining).

122. *See Leonard*, 921 N.Y.S.2d at 340; *Morton*, 768 N.E.2d at 737-38.

of house-breaking.[123]

The extreme breadth of a statute that only requires "remaining" with the intent to commit a crime is illustrated by a series of South Dakota cases. The South Dakota statute enacted in 1976 defined third degree burglary as follows: "Any person who enters or remains in an unoccupied structure, with intent to commit any crime therein, is guilty of third degree burglary."[124] In one case, a defendant entered a laundromat while it was open to the public and then pried open coin boxes with a crow bar, and stole a case of soda pop.[125] Another burglary conviction was upheld where the defendant entered a convenience store during business hours and attempted to steal a frozen pizza.[126] But the court decided the expansion had gone too far when an eleven-year-old girl was prosecuted for eating a chocolate Easter egg in a department store without paying.[127] The court reversed the conviction and held that "remains" means "unlawful presence."[128] The court eventually overruled this holding in a case where it upheld the conviction of twenty counts of burglary for a delivery driver who, over an extended period of time, took cases of soda pop from a store during his regular deliveries.[129] In these South Dakota cases, burglary functions as a location enhancement for what might otherwise be petty theft.

D. Dwelling

As noted, burglary had extended well beyond the home by the nineteenth

---

123. In this weakening or outright elimination of the common law elements of breaking and entering, the history of modern burglary bears a striking resemblance to the history of another crime with roots in the common law: rape. The common law definition of rape had elements that parallel those of burglary's breaking and entering: force and penetration. *See* LAFAVE, *supra* note 9, § 17.1, at 892. As with burglary, these elements are retained in most modern statutes for the most serious grades of the offense, but not necessarily for lesser degrees. Moreover, the degree of force may be minimal, and permission for entry may be revoked. Lesser offenses have been defined for sexual conduct that does not rise to the level of rape, just as lesser breaking or entering offenses such as trespass have developed in most jurisdictions. Like common law burglary, common law rape was both a crime against property and a crime against personal security. Unlike burglary, of course, modern rape is clearly an offense against a person, rather than property. It can nevertheless be viewed, like burglary, as essentially a kind of invasion.

124. S.D. CODIFIED LAWS § 22-32-8 (1976) (amended 1989 and 2005).

125. State v. Blair, 273 N.W.2d 187, 187-88 (S.D. 1979) The conviction was upheld under the plain language of the statute. *Id.* at 188.

126. State v. Shult, 380 N.W.2d 352, 356 (S.D. 1986).

127. *In re* T.J.E., 426 N.W.2d 23, 25 (S.D. 1988), *superseded by statute as stated in* State v. Miranda, 776 N.W.2d 77 (S.D. 2009).

128. *Id.*

129. State v. Burdick, 712 N.W.2d 5, 10 (S.D. 2006). The legislature then amended the statute to add language based on the Model Penal Code: "unless the premises are, at the time, open to the public or the person is licensed or privileged to enter or remain." S.D. CODIFIED LAWS § 22-32-8 (2011).

century. The Model Penal Code accepted this enlarged scope of the crime, referring to the entry of a "building or occupied structure, or separately secured or occupied portion thereof,"[130] but it stopped short of including vehicles or storage containers.[131]

The great variety in statutory schemes makes state approaches difficult to compare. Some states provide a list of places that may be burglarized in a definition section. Others include the list in the section that defines the crime of burglary. Still others create separate provisions or crimes for each type of structure, and others develop the list in caselaw. Despite this variation, however, the substantive law of the states is quite similar; buildings, structures, vehicles, and containers for storing or securing goods can all be objects of burglary.

Today, many jurisdictions have lists[132] that usually include the buildings and structures already commonly mentioned in 1951, but with some very specific additions: outhouse,[133] cash register,[134] vending machine,[135] "vault, safe, cash register, coin vending machine, product dispenser, . . . coin telephone,"[136] cemetery,[137] tent,[138] "outside showcase or other outside enclosed counter."[139] Other jurisdictions have more general provisions that can be read broadly to

130. MODEL PENAL CODE § 221.1 (Official Draft & Revised Comments 1980).

131. "Restricting the offense to buildings and other occupied structures confines it to those intrusions that are typically the most alarming and dangerous." MODEL PENAL CODE § 221.1 cmt. 3(b), at 72 (Official Draft & Revised Comments 1980).

132. *See, e.g.*, ALA. CODE § 13A-7-1(2) (2012); ARIZ. REV. STAT. ANN. §§ 13-1501(12) ("any vending machine or any building, object, vehicle, railroad car"), -1506(A)(1) ("a fenced commercial or residential yard") (2012); ARK. CODE ANN. § 5-39-202 (2012); CAL. PENAL CODE § 459 (West 2012); D.C. CODE § 22-801(b) (2012); FLA. STAT. ANN. § 810.011 (West 2012) (containing special provisions for structures damaged during a state of emergency—presumably meant to cover hurricane damaged property); IDAHO CODE ANN. § 18-1401 (2011); KAN. STAT. ANN. § 21-5807 (West 2011); MICH. COMP. LAWS ANN. § 750.110 (West 2012); MISS. CODE ANN. § 97-17-33 (2011); NEV. REV. STAT. ANN. § 205.060 (West 2011); OKLA. STAT. tit. 21, § 1435 (2011); WASH. REV. CODE ANN. § 9A.04.110(5) (West 2012); W. VA. CODE ANN. § 61-3-12 (West 2012).

Some jurisdictions define separate offenses for different categories of structures or vehicles entered. *See, e.g.*, MICH. COMP. LAWS ANN. §§ 750.113, -.114, -.115, -.356b (West 2012); N.C. GEN. STAT. §§ 14-51, -53, -54, -56, -56.1, -56.2, -56.3 (2011); R.I. GEN. LAWS § 11-8-2 to -6 (2011); TEX. PENAL CODE ANN. §30.02 to -.04 (West 2011).

133. *See* CAL. PENAL CODE § 459.

134. *See* ARK. CODE ANN. § 5-39-202.

135. *See* ARIZ. REV. STAT. ANN. § 13-1501(13) (2012); ARK. CODE ANN. § 5-39-202.

136. COLO. REV. STAT. § 18-4-204(1) (2011) (third degree burglary).

137. *See* LA. REV. STAT. ANN. § 14:62 (2011).

138. *See* KAN. STAT. ANN. § 21-5807 (West 2011); MICH. COMP. LAWS ANN. § 750.110 (West 2012); OKLA. STAT. tit. 21, § 1438 (2011).

139. MICH. COMP. LAWS ANN. § 750.114 (West 2012).

cover a variety of structures: "structure,"[140] "building,"[141] and "structure, vehicle, watercraft or aircraft."[142] All jurisdictions also recognize burglary of a portion of a building, such as separate apartments in an apartment building.[143] A few jurisdictions extend burglary to either a "fenced area,"[144] "fenced commercial or residential yard,"[145] or even the "land or premises of another."[146]

For the highest degree of the offense, in fourteen states the site of the burglary must be a dwelling.[147] Three jurisdictions require that the highest degree of the offense take place in an "occupied structure."[148] Tennessee and Texas require the highest degree of burglary to occur in a "habitation,"[149] while Vermont requires an "occupied dwelling."[150] In nine states, commission of the offense in a dwelling is one of (or part of) two or more alternative ways of committing the highest degree of the crime.[151] These numbers do not seem to have changed much since the 1951 research.

E. Of Another

This common law requirement—that the structure entered be that "of another"—had already almost disappeared from the statutory definitions of

---

140. ME. REV. STAT. ANN. tit. 17-A, § 401 (2011); *see also* IND. CODE § 35-43-2-1 (2011) ("building or structure").

141. HAW. REV. STAT. §§ 708-810, -811 (West 2011).

142. KY. REV. STAT. ANN. § 511.010 (West 2011).

143. *See, e.g.*, OR. REV. STAT. ANN. § 164.205(1) (West 2011); 18 PA. CONS. STAT. ANN. § 3502 (West 2012) (a building "or separately secured or occupied portion thereof"); VT. STAT. ANN. tit. 13, § 1201 (2011) ("any portion of a building, structure or premises which differs from one or more other portions of such building, structure or premises with respect to license or privilege to enter, or to being open to the public").

144. WASH. REV. CODE ANN. § 9A.04.110(5) (West 2012).

145. ARIZ. REV. STAT. ANN. § 13-1506 (2012).

146. *See* OHIO REV. CODE ANN. § 2911.13 (West 2011); *see also* NEB. REV. STAT. § 28-507 (2011) ("any real estate or any improvements erected thereon").

147. ALA. CODE § 13A-7-5 (2012); ARK. CODE ANN. § 5-39-201 (2012); CAL. PENAL CODE § 460 (West 2012); DEL. CODE ANN. tit. 11, § 826 (2012); D.C. CODE § 22-801 (2012); MD. CODE ANN., CRIM. LAW § 6-202 (West 2011); MASS. GEN. LAWS ANN. ch. 266, § 14 (West 2012); MISS. CODE ANN. § 97-17-23 (2011); N.Y. PENAL LAW § 140.30 (McKinney 2010); N.C. GEN. STAT. § 14-51 (2011); OKLA. STAT. tit. 21, § 1431 (2011); R.I. GEN. LAWS § 11-8-2 (2011); S.C. CODE ANN. § 16-11-311 (2011); W. VA. CODE ANN. § 61-3-11 (West 2012).

148. *See* MONT. CODE ANN. § 45-6-204 (2009); OHIO REV. CODE ANN. § 2911.11 (West 2011); S.D. CODIFIED LAWS § 22-32-1 (2011).

149. TENN. CODE ANN. § 39-14-403 (2011); TEX. PENAL CODE ANN. § 30.02(d) (West 2011).

150. VT. STAT. ANN. tit. 13, § 1201 (2011).

151. *See* ALASKA STAT. § 11.46.300 (2011); CONN. GEN. STAT. ANN. § 53a-101 (West 2012); HAW. REV. STAT. § 708-810(1)(c) (West 2011); LA. REV. STAT. ANN. § 14:60 (2011); MINN. STAT. ANN. § 609.582 (West 2011); N.H. REV. STAT. ANN. § 635:1 (2012); N.D. CENT. CODE § 12.1-22-02 (2011); OR. REV. STAT. ANN. § 164.225 (West 2011); WIS. STAT. § 943.10(2)(e) (2011).

burglary by the mid-twentieth century.[152] Even the Model Penal Code contains no such language, although the drafters believed that the requirement of an unprivileged entry would obviate the need for it.[153]

In the last thirty years, however, even this implied element has been eroded in an additional way, as courts convict defendants of burglary for entering a home in which they have a possessory right but where there is a no-contact or protection order preventing them from entering the home, or even where there is no court order but the defendant has moved out.[154] In New Hampshire, for example, a defendant whose name was on the lease was convicted of burglary when he entered the home after being asked to move out and having his belongings removed.[155] In this way, courts and prosecutors are able to use burglary law as another weapon against domestic violence.[156]

Today, ten jurisdictions have statutes that require that the structure entered be "of another."[157] Four of those jurisdictions include the phrase in only one of several burglary statutes, giving the impression that the language is vestigial, a remnant from the common law.[158] But many jurisdictions have other words or phrases implying that the "entry" or "remaining" must be trespassory or unpermitted by the one in possession. The most common word is "unlawfully," as in to "enter or remain unlawfully."[159] About two-thirds of the jurisdictions using this phrase also have a statutory definition, most commonly incorporating

---

152. Wright, *supra* note 7, at 419.

153. MODEL PENAL CODE § 221.1 cmt. 1, at 64-65 (Official Draft & Revised Comments 1980). The comment notes disapprovingly that "[t]he language of some statutes appeared to be broad enough to make a burglar out of one who entered his own house or office with the purpose of committing a crime, whether it be to prepare a fraudulent income tax return or to commit an assault upon his wife." *Id*.

154. *See* Jeannie Suk, *Criminal Law Comes Home*, 116 YALE L.J. 2 (2006). Suk writes of a case where the court upheld a burglary conviction where the defendant had moved out of the marital home, but there was no court order excluding him from the house. *Id.* at 28.

155. State v. McMillan, 973 A.2d 287 (N.H. 2009).

156. *See, e.g.*, State v. Hagedorn, 679 N.W.2d 666 (Iowa 2004).

157. *See* ARK. CODE ANN. § 5-39-201 (2012); GA. CODE ANN. § 16-7-1 (2011); IND. CODE § 35-43-2-1 (2011); MD. CODE ANN., CRIM. LAW §§ 6-202, -203 (West 2011); MASS. GEN. LAWS ANN. ch. 265, § 18C (West 2012); MISS. CODE ANN. § 97-17-23 (2011); N.C. GEN. STAT. § 14-53 (2011); OKLA. STAT. tit. 21, § 1431 (2011); VA. CODE ANN. § 18.2-89 (2011); W. VA. CODE ANN. § 61-3-11 (West 2012).

158. *See* MASS. GEN. LAWS ch. 265, § 18C; MISS. CODE ANN. § 97-17-23; OKLA. STAT. tit. 21, § 1431; VA. CODE ANN. § 18.2-89.

159. *See* ALA. CODE §§ 13A-7-5 to -7 (2012); ALASKA STAT. § 11.46.310 (2011); ARIZ. REV. STAT. § 13-1507 (2012); ARK. CODE ANN. § 5-39-201; COLO. REV. STAT. §§ 18-4-202, -203 (2011); CONN. GEN. STAT. ANN. § 53a-101, -102 (West 2012); DEL. CODE ANN. tit. 11, § 826 (2012); HAW. REV. STAT. § 708-810 (West 2011); KY. REV. STAT. ANN. § 511.020 (West 2011); MO. REV. STAT. § 569.160 (2011); MONT. CODE ANN. § 45-6-204 (2009); N.Y. PENAL LAW § 140.00 (McKinney 2010); OR. REV. STAT. ANN. § 164.215 (West 2011); UTAH CODE ANN. § 76-6-202 (West 2011); WASH. REV. CODE ANN. § 9A.52.020 (West 2012).

language similar to that of the Model Penal Code: "unless the premises are at the time open to the public or the actor is licensed or privileged to enter."[160] An additional group of states incorporates similar language into the definition of the crime itself.[161] Other phrases include, "without authority,"[162] "unauthorized,"[163] "having no right, license or privilege,"[164] or "without consent."[165] Thus, generally there is an implied element that the place of the burglary be that of another, but it does not always apply.

F. Intent to Commit a Felony Therein

The Model Penal Code proposed "with purpose to commit a crime therein," in large part to obviate proof problems for prosecutors who, under the common law formulation, were required to show that the intended crime was not a misdemeanor.

Today, the largest group of states has taken the Model Penal Code's suggestion of "any crime,"[166] although about a quarter of that group uses different

---

160. MODEL PENAL CODE § 221.1(1) (Official Draft & Revised Comments 1980). *See, e.g.*, ALASKA STAT. § 11.46.350(a) (2011); ARIZ. REV. STAT. ANN. § 13-1501(2) (2012); ARK. CODE ANN. § 5-39-101 (2012); COLO. REV. STAT. § 18-4-201 (2011); CONN. GEN. STAT. ANN. § 53a-100 (West 2012); DEL. CODE ANN. tit. 11, § 829 (2012); MONT. CODE ANN. § 45-6-201 (2009); N.Y. PENAL LAW § 140.00(5) (McKinney 2012); OR. REV. STAT. ANN. § 164.205 (West 2011); UTAH CODE ANN. § 76-6-201(3) (West 2011); WASH. REV. CODE ANN. § 9A.52.010(5) (West 2012).

161. *See* N.J. REV. STAT. § 2C:18-2 (2011); N.D. CENT. CODE § 12.1-22-02 (2011); S.D. CODIFIED LAWS § 22-32-1 (2011); *see also* FLA. STAT. ANN. § 810.02 (West 2012). Florida limits the Model Penal Code exception, however, and allows conviction even where presence is lawful if the defendant "commit[s] or attempt[s] to commit a forcible felony." *Id.* § 810.02(1)(b)(2)(c).

162. 720 ILL. COMP. STAT. ANN. 5/19-1 (West 2012); KAN. STAT. ANN. § 21-5807 (West 2011); WYO. STAT. ANN. § 6-3-301 (2011).

163. LA. REV. STAT. ANN. § 14:60 (2011); N.M. STAT. ANN. § 30-14-8 (West 2012).

164. IOWA CODE ANN. § 713.1 (West 2012).

165. MINN. STAT. ANN. § 609.581 (West 2011); S.C. CODE ANN. § 16-11-311 (2011); R.I. GEN. LAWS § 11-8-2 (2011) ("without the consent of the owner or tenant"); TENN. CODE ANN. § 39-14-402 (2011) ("without the effective consent of the property owner"); TEX. PENAL CODE ANN. § 30.02 (West 2011) ("without the effective consent of the owner"); WIS. STAT. § 943.10 (2011) ("without the consent of the person in lawful possession").

166. *See* ALA. CODE § 13A-7-5 (2012); ALASKA STAT. § 11.46.310 (2011); COLO. REV. STAT. § 18-4-202 (2011); CONN. GEN. STAT. ANN. §§ 53a-101, -102, -103 (West 2012); DEL. CODE ANN. tit. 11, §§ 824, 825, 826 (2012); D.C. CODE § 22-801 (2012); FLA. STAT. ANN. § 810.02; KY. REV. STAT. ANN. §§ 511.020, -.030 (West 2011); ME. REV. STAT. ANN. tit. 17-A, § 401 (2011); MINN. STAT. ANN. § 609.582 (West 2011); MISS. CODE ANN. § 97-17-23 (2011); MO. REV. STAT. §§ 569.160, -170 (2011); MONT. CODE ANN. § 45-6-204 (2009); N.H. REV. STAT. ANN. § 635:1 (2012); N.J. REV. STAT. § 2C:18-2 (2011); N.Y. PENAL LAW §§ 140.20, -.30 (McKinney 2010); N.D. CENT. CODE § 12.1-22-02; OKLA. STAT. tit. 21, § 1431 (2011); OR. REV. STAT. ANN. § 164.215 (West 2011); 18 PA. CONS. STAT. ANN. § 3502 (West 2012); S.C. CODE ANN. §§ 16-11-311 to -313 (2011); S.D. CODIFIED LAWS §§ 22-32-1, -3, -8 (2011); W. VA. CODE ANN. § 61-3-11

intended offenses for different degrees of burglary.[167] The next largest group requires that the defendant intend to commit "any theft or any felony,"[168] although the precise wording of the statutes varies, and again, a number of these states have different intended crimes for different degrees or kinds of burglary.[169] A few states list particular intended crimes[170] or restrict the intended crimes to "crimes against a person or property therein."[171]

G. How Burglary Is Graded

The Model Penal Code drafters decided against including a first-degree burglary, and instead proposed that burglary be a third-degree felony unless "perpetrated in the dwelling of another at night," or if the actor "inflicts or attempts to inflict bodily injury" or "is armed with explosives or a deadly weapon."[172] Thus, the Model Code incorporated into its formulation of burglary some of the elements of common law burglary as aggravating facts, as well as some of the already commonly used aggravators such as injury or weapons. The comments explain that the drafters believed first-degree burglary was unnecessary because where additional serious crimes were committed, those crimes could be separately punished.[173] However, the second degree offense was proposed, in part, out of deference to common practice, but also because "[t]he sanctions of a second-degree felony normally could not be reached in such cases by prosecution for any additional felonies that may have been committed . . . ."[174]

(West 2012).

167. *See* ALA. CODE §§ 13A-7-5, -6 (2012); COLO. REV. STAT. §§ 18-4-202, -203 (2011); MISS. CODE ANN. §§ 97-17-23, -33 (2011); OKLA. STAT. tit. 21, §§ 1431, 1435 (2011); W. VA. CODE ANN. §§ 61-3-11, -12 (West 2012).

168. *See, e.g.*, ALA. CODE § 13A-7-6; ARIZ. REV. STAT. ANN. § 13-1507 (2012); ARK. CODE ANN. § 5-39-202 (2012); CAL. PENAL CODE § 459 (West 2012); GA. CODE ANN. § 16-7-1 (2011); IDAHO CODE ANN. § 18-1401 (2011); 720 ILL. COMP. STAT. ANN. 5/19-1 (West 2012); LA. REV. STAT. ANN. § 14:60 (2011); MICH. COMP. LAWS ANN. §§ 750.110, -.111 (West 2012); NEB. REV. STAT. § 28-507 (2011); N.M. STAT. ANN. § 30-16-3 (West 2012); N.C. GEN. STAT. §§ 14-53, -54, -54.1, -56 (2011); OHIO REV. CODE ANN. § 2911.12 (West 2011); OKLA. STAT. tit. 21, § 1435; VA. CODE ANN. § 18.2-89 (2011); W. VA. CODE ANN. § 61-3-12; WIS. STAT. § 943.10; WYO. STAT. ANN. § 6-3-301 (2011).

169. *See* ARK. CODE ANN. §§ 5-39-201, -202 (2012); OHIO REV. CODE ANN. §§ 2911.11, -.12, -.13 (West 2011); VA. CODE ANN. § 18.2-89.

170. *See, e.g.*, MD. CODE ANN., CRIM. LAW §§ 6-202, -203 (West 2011) ("theft or a crime of violence," "theft, a crime of violence, or arson"); R.I. GEN. LAWS § 11-8-3 (2011) ("murder, sexual assault, robbery, arson or larceny"); UTAH CODE ANN. § 76-6-202 (West 2011) (felony, theft, an assault on any person, lewdness, sexual battery, lewdness involving a child, voyeurism).

171. *See, e.g.*, HAW. REV. STAT. § 708-810 (West 2011); WASH. REV. CODE ANN. § 9A.52.020 (West 2012).

172. MODEL PENAL CODE § 221.1(2) (Official Draft & Revised Comments 1980).

173. *Id.* § 221.1 cmt. 4, at 78-79.

174. *Id.* § 221.1 cmt. 4, at 79. The drafters did not foresee the proliferation of weapon

The Model Penal Code proposal may have influenced states more than is generally recognized. Today, only eight states have a separate offense of burglary with explosives.[175] Although "night" is an aggravating factor in only seven states today,[176] in approximately thirty states whether the offense was committed in a dwelling or its equivalent is an element (or alternative element) of the most serious type of burglary.[177] At least twenty-eight states have as elements of a more severely punished burglary offense: (1) the fact that the defendant was armed with a firearm, deadly weapon, or dangerous instrument; and/or (2) that the defendant inflicted, or attempted to inflict bodily injury, assault or battery.[178] States vary considerably in the wording of these

---

enhancements in criminal codes.

175. *See* CAL. PENAL CODE § 464 (West 2012); IDAHO CODE ANN. § 18-1405 (2011); MD. CODE ANN., CRIM. LAW § 6-207 (West 2011); MICH. COMP. LAWS ANN. § 750.112 (West 2012); MISS. CODE ANN. § 97-17-37 (2011); NEV. REV. STAT. ANN. § 205.075 (West 2011); N.C. GEN. STAT. § 14-57 (2011); OKLA. STAT. tit. 21, § 1441 (2011). However, explosives are mentioned in fourteen states as a fact to elevate the crime. *See* ALA. CODE § 13A-7-5 (2012); COLO. REV. STAT. § 18-4-202 (2011); CONN. GEN. STAT. ANN. § 53a-101 (West 2012); DEL. CODE ANN. tit. 11, § 826 (2012); FLA. STAT. ANN. § 810.02 (West 2012); IOWA CODE ANN. § 713.3 (West 2012); KY. REV. STAT. ANN. § 511.020 (West 2011); MO. REV. STAT. § 569.160 (2011); MONT. CODE ANN. § 45-6-204 (2009); N.H. REV. STAT. ANN. § 635:1 (2012); N.J. REV. STAT. § 2C:18-2 (2011); N.Y. PENAL LAW §§ 140.25, -.30 (McKinney 2010); S.C. CODE ANN. § 16-11-311(2011); UTAH CODE ANN. § 76-6-203 (West 2011).

176. *See supra* note 101.

177. *See* ALA. CODE § 13A-7-5; ARIZ. REV. STAT. ANN. § 13-1508 (2012); ARK. CODE ANN. § 5-39-201 (2012); CAL. PENAL CODE § 460 (West 2012); COLO. REV. STAT. § 18-4-202; CONN. GEN. STAT. ANN. § 53a-101; DEL. CODE ANN. tit. 11, §§ 825, 826 (2012); D.C. CODE § 22-801(a) (2012); FLA. STAT. ANN. § 810.02; HAW. REV. STAT. § 708-810 (West 2011); IND. CODE § 35-43-2-1 (2011); KY. REV. STAT. ANN. § 511.030 (West 2011) (second degree); ME. REV. STAT. ANN. tit. 17A, § 401 (2011); MD. CODE ANN., CRIM. LAW § 6-202 (West 2011); MASS. GEN. LAWS ANN. ch. 266, §§ 14, 15 (West 2012); MICH. COMP. LAWS ANN. § 750.110a (West 2012); MINN. STAT. ANN. § 609.582 (West 2011); MISS. CODE ANN. § 97-17-23 (2011); N.H. REV. STAT. ANN. § 635:1; N.Y. PENAL LAW §§ 140.20, -.30 (McKinney 2010); N.C. GEN. STAT. § 14-51 (2011); N.D. CENT. CODE § 12.1-22-02 (2011); OHIO REV. CODE ANN. § 2911.11 (West 2011); OKLA. STAT. tit. 21, § 1431 (2011); OR. REV. STAT. ANN. § 164.225 (West 2011); S.C. CODE ANN. § 16-11-311; UTAH CODE ANN. § 76-6-202 (West 2011); VT. STAT. ANN. tit. 13, § 1201 (2011); WASH. REV. CODE ANN. § 9A.52.025 (West 2012); W. VA. CODE ANN. §§ 61-3-11, -12 (West 2012).

178. ALA. CODE § 13A-7-5; ALASKA STAT. § 11.46.300 (2011); COLO. REV. STAT. § 18-4-202; CONN. GEN. STAT. ANN. § 53a-101; DEL. CODE ANN. tit. 11, § 826; FLA. STAT. ANN. § 810.02; HAW. REV. STAT. § 708-810; IOWA CODE ANN. § 713.3 (West 2012); KY. REV. STAT. ANN. § 511.020; LA. REV. STAT. ANN. § 14:60 (2011); ME. REV. STAT. ANN. tit. 17-A, § 401 (2011); MINN. STAT. ANN. § 609.582; MO. REV. STAT. § 569.160; MONT. CODE ANN. § 45-6-204; N.H. REV. STAT. ANN. § 635:1; N.J. REV. STAT. § 2C:18-2; N.M. STAT. ANN. § 30-16-4 (West 2012); N.Y. PENAL LAW §§ 140.25, -.30; N.D. CENT. CODE § 12.1-22-02; OHIO REV. CODE ANN. § 2911.11; OR. REV. STAT. ANN. § 164.225; S.C. CODE ANN. § 16-11-311; S.D. CODIFIED LAWS § 22-32-1 (2011); TENN. CODE ANN. §§ 39-14-402, -403 (2011); UTAH CODE ANN. § 76-6-203;

requirements and whether one, two, or all of these aggravating elements are required.

Other slightly less common reasons to increase the severity of burglary are: the presence in the entered structure of a person who is not a criminal participant,[179] the defendant's prior burglary convictions,[180] and the nature of the crime intended.[181] Some states increase punishment for entering certain types of buildings, such as religious structures or day care centers.[182] A few states increase the penalty if the victim of the crime is over sixty years of age.[183] Two states have higher penalties for burglaries of pharmacies or other places that lawfully keep controlled substances.[184] Some states have separate offenses for entry of vehicles.[185]

States grade burglary offenses in a variety of ways. Some have very complicated schemes, with multiple offenses aimed at burglary of different structures, with different intents or with different victims.[186] These are generally

---

WASH. REV. CODE ANN. § 9A.52.020 (West 2012); WIS. STAT. § 943.10 (2011). Some include being armed with what seems to be a firearm or weapon. *See, e.g.*, MINN. STAT. ANN. § 609.582; N.Y. PENAL LAW § 140.30.

179. *See* CONN. GEN. STAT. ANN. § 53a-102 (West 2012); D.C. CODE § 22-801; FLA. STAT. ANN. § 810.02(3); IOWA CODE ANN. § 713.3; KAN. STAT. ANN. § 21-5807 (West 2011); MICH. COMP. LAWS ANN. § 750.110a; MINN. STAT. ANN. § 609.582; MISS. CODE ANN. § 97-17-23 (2011) (if terrorized); MO. REV. STAT. § 569.160; N.C. GEN. STAT. § 14-51; OHIO REV. CODE ANN. § 2911.11; OKLA. STAT. tit. 21, § 1431; 18 PA. CONS. STAT. ANN. § 3502 (West 2012); WIS. STAT. § 943.10.

180. *See* GA. CODE ANN. § 16-7-1(b) (2011); ME. REV. STAT. ANN. tit. 17-A, § 401; MASS. GEN. LAWS ANN. ch. 266, § 14, 15; NEV. REV. STAT. ANN. § 205.060 (West 2011) (defendant with prior conviction may not have suspended sentence or probation); R.I. GEN. LAWS §§ 11-8-2, -2.2 (2011); S.C. CODE ANN. § 16-11-311.

181. *See* MD. CODE ANN., CRIM. LAW § 6-203 (West 2011) (intent to steal firearm); MICH. COMP. LAWS ANN. § 750.110a ("felony, larceny or assault"); TEX. PENAL CODE ANN. § 30.02 (West 2011) ("a felony other than felony theft"); VA. CODE ANN. § 18.2-90 (2011) ("murder, rape, robbery or arson").

182. *See, e.g.*, 720 ILL. COMP. STAT. ANN. 5/19-1 (West 2012) (school, day care, place of worship); IND. CODE § 35-43-2-1 (structure for religious worship); MINN. STAT. ANN. § 609.582(2)(b) ("government building, religious establishment, historic property, or school building"); S.C. CODE ANN. § 16-11-380 (2011) (bank). Federal breaking and entering statutes are similarly tailored to the entry of particular structures. *See infra* Part IV.

183. *See* DEL. CODE ANN. tit. 11, § 826 (sixty-two years or older); MISS. CODE ANN. § 99-19-351 (2011) (sixty-five years or older); R.I. GEN. LAWS § 11-8-2.3 (2011) (sixty years or older).

184. *See* COLO. REV. STAT. §§ 18-4-202(3), - 204(2), -303(2) (2011); MINN. STAT. ANN. § 609.582(2)-(3).

185. *See, e.g.*, TEX. PENAL CODE ANN. § 30.04 (West 2011); UTAH CODE ANN. § 76-6-204 (West 2011).

186. *See, e.g.*, MASS. GEN. LAWS ANN. ch. 266 (West 2012); MICH. COMP. LAWS ANN. §§ 750.110 to -.115 (West 2012); N.C. GEN. STAT. §§ 14-51 to -57 (2011); R.I. GEN. LAWS §§ 11-8-1 to -9 (2011).

states that continued to add to their criminal codes in lieu of revising them. Most states have first degree or "aggravated" burglary, in addition to lower degrees or lesser versions. Only a few follow the Model Penal Code suggestion that burglary be no more than a second degree felony.[187] It is difficult to compare schemes between states due to the varying terminology and different approaches to ordering offenses.

Of course, any survey of the grading of burglary will only present part of the picture due to the prevalence of determinate sentencing. In states with determinate sentencing schemes like the Federal Sentencing Reform Act,[188] punishment will be graded by the sentencing laws in addition to the statutory scheme. These laws might have aggravating and mitigating circumstances based on the amount stolen, whether the defendant was armed, the status of the victim, the number of prior convictions, etc.[189] Sentencing law has the potential to usurp the traditional role of offense definition in grading offenses.

H. Conclusion

Since the publication of the Model Penal Code, the crime of burglary has continued to streamline itself, shedding the elements of night time, breaking, and felonious intent, while extending its reach to all kinds of locations. Most significantly, the element of entry has been replaced by the alternative of "remaining." Although state laws vary considerably, in its most streamlined forms, burglary now consists of being in a particular location with the intent to commit a crime. The most common factors that aggravate punishment include entry of a dwelling, infliction of injury, or being armed.

IV. FEDERAL BREAKING AND ENTERING OFFENSES

Neither of the 1951 surveys of American burglary law included federal crimes.[190] There is no general federal crime of burglary. There are, however, several very specific "entry" or "breaking and entry" offenses aimed at protecting particular federal interests. The federal bank robbery statute, for example, includes punishment for:

> Whoever enters or attempts to enter any bank, credit union, or any savings and loan association, or any building used in whole or in part as a bank, credit union, or as a savings and loan association, with intent to commit in such bank, credit union or in such savings and loan association, or building, or part thereof, so used, any felony affecting

---

187. *See, e.g.*, N.J. REV. STAT. § 2C:18-2 (2011); N.M. STAT. ANN. §§ 30-16-3, -4 (West 2012).

188. Pub. L. No. 98-473, 98 Stat. 1987 (1984).

189. *See, e.g.*, WASH. REV. CODE ANN. § 9.94A.535 (West 2012) (setting forth aggravating circumstances for sentencing purposes).

190. *See A Rationale of the Law of Burglary*, *supra* note 7, at 1009 n.3; Wright, *supra* note 7, at 415 n.36.

> such bank, credit union, or such savings and loan association and in violation of any statute of the United States, or any larceny[.][191]

This offense bears little resemblance to common law burglary or modern state burglary statutes. It applies to both entry and attempted entry, is restricted to certain financial institutions, and restricts the types of intended crimes. The provision was added in 1937 to the bank robbery statute "to cover the situation where a person enters a bank for the purpose of committing a crime, but is frustrated for some reason before completing the crime."[192] Like state burglary statutes, however, it appears to be used not only as a version of attempt, but also to aggravate a completed crime.[193] "Entry" now includes using a cash machine or drive-up window.[194]

A different offense for a post office is found at 18 U.S.C. § 2115:

> Whoever forcibly breaks into or attempts to break into any post office, or any building used in whole or in part as a post office, with intent to commit in such post office, or building or part thereof, so used, any larceny or other depredation, shall be fined under this title or imprisoned not more than five years, or both.[195]

Unlike § 2113, this offense requires forcible breaking. It has been referred to as "burglary of a United States Post Office"[196] as well as "post office robbery,"[197] and appears applicable to either situation.

Another federal statute, 18 U.S.C. § 2116, makes it a crime to enter "by violence" a post-office car, steamboat or vessel assigned to the use of the mail service.[198] Section 2117 makes it a crime to "break[] the seal or lock of any railroad car, vessel, aircraft, motortruck, wagon or other vehicle or of any pipeline system, containing interstate or foreign shipments of freight or express or other property, or enter[] any such vehicle or pipeline system with intent in either case to commit larceny therein."[199] The offense can be charged along with a completed larceny.[200]

Still another federal statute, 18 U.S.C. § 2118, defines the offenses of "robberies and burglaries involving controlled substances." It punishes:

---

191. 18 U.S.C. § 2113(a) (2006).

192. Prince v. United States, 352 U.S. 322, 406-07 (1957).

193. *See, e.g.*, United States v. Goudy, 792 F.2d 664, 678 (7th Cir. 1986) (upholding convictions for entering a bank, larceny, conspiracy to commit larceny, and additional offenses); United States v. Phillips, 609 F.2d 1271, 1272 (8th Cir. 1979) (convictions upheld for entering a bank, conspiracy to take money from a bank and for taking money from a bank).

194. *Goudy*, 792 F.2d at 675 n.11; *Phillips*, 609 F.2d at 1273.

195. 18 U.S.C. § 2115 (2006).

196. United States v. York, 171 F. App'x 655, 655 (9th Cir. 2006).

197. United States v. Jackson, 347 F.3d 598, 601 (6th Cir. 2003).

198. 18 U.S.C. § 2116 (2006).

199. *Id.* § 2117.

200. *See, e.g.*, United States v. Kiff, 377 F. Supp. 2d 586, 590 (E.D. La. 2005).

"Whoever, without authority, enters or attempts to enter, or remains in, the business premises or property of a person registered with the Drug Enforcement Administration . . . with the intent to steal any material or compound containing any quantity of a controlled substance . . . ."[201]

Enacted in 1992, this offense more resembles state burglary statutes than some of the other federal "entry" offenses in that it uses terms such as "without authority" and "or remains." However, unlike most state burglary offenses, it applies only to a narrow set of victims and intended offenses.

Two statutes come close to creating general federal burglary statutes for specific locations. The first, 18 U.S.C. § 2276, makes it a crime to "break[] or enter[] any vessel with intent to commit any felony," when the vessel is on the water and within the jurisdiction of the United States but "out of the jurisdiction of any particular State."[202] The second, 18 U.S.C. § 1991, punishes entering a train with the intent to commit "any crime or offense against any person or property thereon," and provides a higher punishment for entry of a train with intent to commit murder or robbery.[203] Although these statutes, dating from the first half of the twentieth century, are not used much now, they apparently can be used to "pile on" punishment for the completed crime.[204]

The Court has stated in another context that unless Congress uses common law terms in defining a crime, mere similarity between a statutory offense and a common law crime does not mean that common law concepts apply.[205] Although some of the federal offenses resemble burglary in that they criminalize entry with criminal intent, they do not appear to draw on the common law and there is no general federal burglary offense. Like state burglary offenses, however, they can function as "location aggravators," adding punishment to other completed or attempted crimes.

V. BURGLARY AS AGGRAVATION BY LOCATION

Modern day burglary, stripped of many of the common law requirements such as breaking or even entering, can function essentially to increase the punishment of the intended (and most likely completed) offense based on its location. This additional punishment might come through anti-merger statutes,

---

201. 18 U.S.C. § 2118(b) (2006).

202. *Id.* § 2276.

203. *Id.* § 1991.

204. *See, e.g.*, United States v. Garcia-Castillo,127 F. App'x 385, 386 (10th Cir. 2005) (defendant pled to "(1) conspiracy to violate the laws of the United States under 18 U.S.C. § 371 by committing a theft from an interstate shipment, in violation of 18 U.S.C. § 659; by entering a train to commit a crime, in violation of 18 U.S.C. § 1991; and by breaking or entering carrier facilities, in violation of 18 U.S.C. § 2117; (2) trespassing upon a railroad car with intent to commit a crime, in violation of 18 U.S.C. § 1991; (3) entering a railroad car with intent to commit a theft, in violation of 18 U.S.C. § 2117; and (4) assaulting a federal officer, in violation of 18 U.S.C. § 111(a)(1).").

205. Carter v. United States, 530 U.S. 255, 265-67 (2000) (interpreting 18 U.S.C. § 2113(b)).

sentencing laws or death penalty statutes. The concept of increasing punishment for an offense because of where it took place is not new to criminal law.[206] But because burglary is considered, and often functions as, a distinct offense, its role as a "location aggravator" is not fully recognized.

Burglary has taken on this role in addition to covering the traditional "unlawful entry to commit theft" offenses. This traditional concept of burglary still dominates the public understanding of the term,[207] and burglary offenses are still reported as "property crimes" in compilations of crime statistics.[208] But the role of burglary as an "aggravator" of other crimes is not much discussed.

Punishing a defendant for burglary as well as the underlying offense does not violate double jeopardy as long as there is a clear legislative intent to allow a burglary conviction separate from a conviction for the offense that was the object of the burglary.[209] The majority rule is that burglary and any underlying crimes

---

206. *See, e.g.*, WASH. REV. CODE ANN. § 69.50.435 (West 2012) (authorizing increased penalties for drug crimes committed in a school, on a school bus, on a public transit vehicle, in a public park, etc.).

207. Wikipedia reports, for example, "Burglary (also called breaking and entering and sometimes housebreaking) is a crime, the essence of which is illegal entry into a building for the purposes of committing an offense. Usually that offense will be theft, but most jurisdictions specify others which fall within the ambit of burglary." *Burglary*, WIKIPEDIA, http://en.wikipedia.org/wiki/Burglary (last visited June 10, 2012). An online thesaurus offers the following synonyms for burglar: "housebreaker, thief, robber, pilferer, filcher, cat burglar, sneak thief, picklock." *Burglar*, FREE DICTIONARY, http://www.thefreedictionary.com/burglar (last visited June 10, 2012). Many online dictionaries also include the old common law definition of the crime. *See, e.g.*, *Burglary*, MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/burglary (last visited June 10, 2012).

208. The Bureau of Justice Statistics uses a restricted definition of burglary for purposes of reporting burglary crimes:

> Burglary is defined as unlawful or forcible entry or attempted entry of a residence. This crime usually, but not always, involves theft. The illegal entry may be by force, such as breaking a window or slashing a screen, or may be without force by entering through an unlocked door or an open window. As long as the person entering has no legal right to be present in the structure a burglary has occurred. Furthermore, the structure need not be the house itself for a burglary to take place; illegal entry of a garage, shed, or any other structure on the premises also constitutes household burglary. If breaking and entering occurs in a hotel or vacation residence, it is still classified as a burglary for the household whose member or members were staying there at the time the entry occurred.

Burglary, BUREAU JUST. STAT., http://bjs.ojp.usdoj.gov/ index.cfm?ty=tp&tid=931 (last visited June 10, 2012). *But see* Commonwealth v. Pruitt, 951 A.2d 307, 315 (Pa. 2008) (holding burglary properly considered a crime of violence and therefore prior burglary convictions could be aggravating factors in penalty phase of capital trial).

209. *See* Blockburger v. United States, 284 U.S. 299 (1932) (holding punishment for multiple crimes in same proceeding does not violate double jeopardy where each crime requires proof of an element not required for conviction of the remaining crimes). "The applicable rule is that, where the same act or transaction constitutes a violation of two distinct statutory provisions, the test to be

do not merge.[210] Thus, in the majority of states that allow multiple punishments, burglary can provide a significant additional penalty.

A few cases, drawn from appeals involving burglary in 2010, are illustrative. In *Commonwealth v. Benson*,[211] the defendant was convicted of burglary, robbery, simple assault, reckless endangerment, unlawful restraint, theft and receipt of stolen property.[212] He received consecutive sentences on the burglary, robbery and theft offenses.[213] Testimony showed that he had entered behind an elderly woman as she returned home, then assaulted her and took her property.[214] In *Cooper v. State*,[215] the defendant was convicted of murder, burglary, armed robbery, and other offenses when he and his accomplices entered (most likely through an open door) the home of some persons they had been riding with, then assaulted the occupants with a gun and a bottle, ultimately taking property and strangling one person.[216] In *State v. Jacobs*,[217] the defendant was convicted of robbery, impersonating a law enforcement officer, first degree burglary and second-degree kidnapping.[218] The evidence showed that defendant and his accomplices pretended to be DEA agents conducting a raid, knocked on the victims' door, entered with guns drawn, bound the victims, and took their property.[219] In *Walker v. State*,[220] the defendant was convicted of burglary, robbery and criminal confinement when he kicked in the door of a home, entered with a drawn gun, bound the victims, and then took property.[221] In *People v. Leonard*,[222] the defendant was convicted of second degree kidnapping, third degree possession of a weapon, endangering the welfare of a child, and second degree burglary when he was allowed into his child's mother's home for a visit, but then used violence to hold the child and attempt to take her.[223]

---

applied to determine whether there are two offenses or only one, is whether each provision requires proof of a fact which the other does not." *Id.* at 304. Burglary will usually require either an entry or remaining, elements most likely not included in the target offense.

210. *See* State v. Pancake, 296 S.E.2d 37, 41-42 (W. Va. 1982) (collecting cases); *see also* OR. REV. STAT. ANN. § 161.067 (West 2011) (providing that burglary and the underlying offense do not merge); WASH. REV. CODE ANN. § 9A.52.50 (West 2012).

211. 10 A.3d 1268 (Pa. Super. Ct. 2010).

212. *Id.* at 1270.

213. *Id.*

214. *Id*. at 1269.

215. 700 S.E.2d 593 (Ga. 2010), *disapproved on other grounds by* Smith v. State, 290 Ga. 768 (2012).

216. *Id.* at 594.

217. 688 S.E.2d 726 (N.C. Ct. App. 2010).

218. *Id.* at 729.

219. *Id.*

220. 932 N.E.2d 733 (Ind. Ct. App. 2010).

221. *Id.* at 734-35.

222. 921 N.Y.S.2d 337 (App. Div. 2011).

223. *Id.* at 339-40.

In these cases, rather than functioning as a "generalized law of attempts,"[224] the burglary charges add a significant penalty to what are already very serious crimes. In effect, burglary operates as a location aggravator—increasing the penalty for offenses based on where they are committed. In the above instances, the offenses all took place in the victims' homes. But the expansion of burglary law would allow for an aggravating effect for offenses committed in any building or other structure specified by the statute of the particular jurisdiction. For example, in *Graham v. Florida*,[225] best known for its holding that a sentence of life without parole is unconstitutional when imposed on juvenile offenders for non-capital offenses, the defendant pleaded to armed burglary and attempted armed robbery.[226] The defendant's accomplice had let the defendant and another person into the restaurant where he worked shortly after closing.[227] One of the accomplices struck the manager on the head with a metal bar, and the offenders left without taking any money.[228] In a Kentucky case, the defendant was charged with murder, first degree robbery and first degree burglary when he entered a gas station, shot two employees and fled with money from the cash register.[229] In both cases, the burglary charge was possible simply because the robbery (like many robberies) occurred in a building.

Burglary plays a significant aggravation role in death penalty schemes. *Furman v. Georgia*,[230] the 1972 case in which the Supreme Court invalidated Georgia's death penalty, involved a felony murder where burglary was the underlying felony.[231] Surprised during the burglary of a home, Furman had tripped and not meant to shoot the victim. Although Furman's sentence was reversed,[232] and subsequent law would restrict capital punishment in felony-murder cases to those where the defendant was a major participant in the crime and acted with at least reckless indifference to human life,[233] burglary continues to be important in death penalty law because it is included in most death penalty schemes as an "aggravating" circumstance or "special circumstance" that will make the defendant eligible for the death penalty.

Modern capital sentencing statutes build upon the structure of the revised Georgia law upheld in *Gregg v. Georgia*.[234] These statutes typically narrow the class of murders eligible for the death penalty by requiring the jury to find

---

224. Cocke, *supra* note 86, at 213.

225. 130 S. Ct. 2011 (2010).

226. *Id.* at 2018, 2034.

227. *Id.* at 2018.

228. *Id.*

229. Bowling v. Commonwealth, 942 S.W.2d 293 (Ky. 1997), *overruled on other grounds by* McQueen v. Commonwealth, 339 S.W.2d 441 (Ky. 2011).

230. 408 U.S. 238 (1972).

231. *Id.* at 252 (Douglas, J., concurring).

232. *Id.* at 239-40.

233. *See* Tison v. Arizona, 481 U.S. 137, 168 (1987).

234. 428 U.S.153, 207 (1976); WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE § 26.1(b), at 1237 (5th ed. 2009).

beyond a reasonable doubt not only the elements of first-degree murder, but also the presence of an aggravating or special circumstance.[235] Only if the jury has found the defendant guilty of the crime and the aggravating circumstance[236] does the jury consider aggravating and mitigating factors for and against the death penalty. The aggravating circumstances, although intended to narrow the class of murders subject to consideration for capital punishment, have tended to proliferate under legislatures wishing to show they are tough on crime.[237] But from the beginning, most jurisdictions have included the contemporaneous commission of five felonies—arson, burglary, kidnapping, rape, or robbery—as aggravating circumstances.[238] This "contemporaneous felony" aggravating circumstance was suggested by the Model Penal Code.[239] Twenty-seven out of thirty-nine death penalty jurisdictions include contemporaneous burglary as an aggravating circumstance.[240]

Generally, the aggravating circumstance is worded along the lines of the Model Penal Code recommendation for the death penalty: "The murder was committed while the defendant was engaged or was an accomplice in the commission of, or an attempt to commit, or flight after committing or an attempt to commit . . . burglary . . . ."[241] In addition, some jurisdictions, such as California, allow the death penalty for felony-murder[242] where the underlying felony is a burglary, and then "double-count" the burglary by allowing the burglary to constitute the aggravating circumstance.[243]

An example of burglary as a death penalty aggravating circumstance is

---

235. *See, e.g.*, CAL. PENAL CODE § 190.2 (West 2012) (§ 190.2(4) held unconstitutional by People v. Sanders, 797 P.2d 561 (Ca. 1990)); TEX. PENAL CODE ANN. § 19.03 (West 2011).

236. "Aggravating factors may also be referred to as aggravating circumstances or eligibility factors. In California, the term 'special circumstance' is used to express this concept." Chelsea Creo Sharon, Note, *The "Most Deserving" of Death: The Narrowing Requirement and the Proliferation of Aggravating Factors in Capital Sentencing Statutes*, 46 HARV. C.R.-C.L. L. REV. 223, 223 n.3 (2011).

237. *See id*. at 232-35 (explaining how expansion of aggravating factors has undermined the Supreme Court's "narrowing requirement").

238. *See generally* David McCord, *Should Commission of a Contemporaneous Arson, Burglary, Kidnapping, Rape, or Robbery be Sufficient to Make a Murderer Eligible for a Death Sentence?—An Empirical and Normative Analysis*, 49 SANTA CLARA L. REV. 1 (2009).

239. *Id*. at 22 (quoting MODEL PENAL CODE § 210.6(3)(e) (Official Draft & Revised Comments 1980)).

240. *Id*. at 24.

241. *Id*. at 22 (quoting MODEL PENAL CODE § 210.6(3)(e)).

242. Indeed, the use of burglary to support felony-murder is an interesting topic in itself. *See, e.g.*, People v. Fuller, 150 Cal. Rptr. 515 (Ct. App. 1978) (upholding charge of felony-murder where defendant caused death during high-speed chase following burglary of a parked car).

243. *See* Steven F. Shatz, *The Eighth Amendment, the Death Penalty, and Ordinary Robbery-Burglary Murderers: A California Case Study*, 59 FLA. L. REV. 719, 730 (2007); *see also* McCord, *supra* note 238, at 23 n.100 (explaining four common statutory patterns for incorporating felonies—including burglary—as an aggravating circumstance in death penalty prosecutions).

provided in *McCray v. State*.[244] The Alabama Court of Criminal Appeals upheld a death penalty conviction where the murder was made death eligible by the contemporaneous commission of burglary.[245] The defendant argued that since he had entered the victim's home with permission, he could not be said to have entered or remained unlawfully.[246] But the court disagreed, reasoning that evidence of a struggle—shown by the profusion of blood loss and injuries—could support a finding that permission to enter had been revoked, thus supporting a finding of burglary.[247] Another example is the Kentucky gas station robbery and murder case discussed above,[248] where the burglary and robbery charges were used to qualify the defendant for the death penalty.

The use of burglary and other felonies to impose the death penalty has been justly criticized; one argument is that these aggravators apply in the majority of murders, and that they do not represent the "worst of the worst."[249] One author concludes that "ordinary" robbery-burglary murderers are "in every respect, the 'average' murderers."[250] The fact that burglary can be charged will usually mean only that the murder occurred in a building, perhaps a dwelling if the death penalty statute requires first-degree burglary and if that jurisdiction's first degree burglary requires entry of a dwelling. In most jurisdictions, the other elements of burglary (criminal intent, entering or remaining, or—for first degree burglary—armed with a deadly weapon or causing injury) will overlap in substance, if not in form, with the elements of murder. Thus, burglary functions as a location aggravator in the death penalty context, just as it can with non-capital crimes.

Of course, these reported opinions showing burglary's function as a location aggravator in capital and non-capital cases are just the tip of the iceberg, the visible examples. Beneath the water is the rest of the ice, composed mostly of the ninety percent to ninety-five percent of cases that are resolved through a guilty plea.[251] The real significance of burglary's additional role is that it provides prosecutors with yet another tool to induce these pleas.

---

244. No. CR-06-0360, 2010 WL 5130747 (Ala. Crim. App. Dec. 17, 2010).

245. *Id.* at *1.

246. *Id.* at *21.

247. *Id.*

248. Bowling v. Commonwealth, 942 S.W.2d 293 (Ky. 1997), *overruled on other grounds by* McQueen v. Commonwealth, 339 S.W.3d 441 (Ky. 2011).

249. Shatz, *supra* note 243, at 745; *see also* McCord, *supra* note 238, at 28 (discussing Illinois and Massachusetts state blue ribbon panels that recommended eliminating contemporaneous felonies as bases for death eligibility).

250. Shatz, *supra* note 243, at 770.

251. Simmons v. Kapture, 516 F.3d 450, 457 (6th Cir. 2008); U.S. SENTENCING COMM'N, SOURCEBOOK OF FEDERAL SENTENCING STATISTICS, at tbl.11 (2009), *available at* http://www.ussc.gov/Data_and_Statistics/Annual_Reports_and_Sourcebooks/2009/Table11.pdf.

VI. THE CONTINUING INFLUENCE OF COMMON LAW BURGLARY

Despite the significant evolution of burglary from Lord Coke's common law definition—the withering away of several common law elements, and burglary's increased role as a location aggravator of other serious crimes—the common law notion of burglary still exerts a powerful influence on the legal imagination. A search of state law cases for references to Lord Coke's definition of burglary turned up seven references from after the time of the Model Penal Code.[252] References to common law burglary in general are quite common, fifteen in 2010 alone.[253] The common law version of burglary is invoked occasionally as an analogy to support grading or creating other kinds of offenses. For example, securities fraud that decimated retirement accounts has been likened to home invasion,[254] and some have argued for a more serious type of identity theft to be called "identity burglary."[255] And, although courts recognize that modern statutory burglary is quite different from the common law crime, they nevertheless resort not infrequently to the common law to resolve disputes about the modern crime.[256]

For example, in Washington, the statutory definition of "building" for purposes of the burglary includes any "fenced area."[257] The Washington court referred to the common law notion of buildings within the home's "curtilage" being protected by burglary laws, and used this concept to limit the meaning of the term "fenced area" to the fenced curtilage.[258] The court did so despite the absence of any reference to the curtilage in the statute, asserting that such an interpretation was necessary to avoid "absurd results."[259] Similarly, an Arizona Court of Appeals case referred to the "historical purpose of sanctioning burglary at common law," "to punish the forcible invasion of a habitation and violation of

252. The most recent was *State v. Goldsmith*, 652 S.E.2d 336, 339 (N.C. Ct. App. 2007).

253. Search on Westlaw, 6/30/2011 using "allstates" database and query: "common law" "common-law" /5 (burglary burglar).

254. *See* Christine Hurt, *Of Breaches of the Peace, Home Invasions and Securities Fraud*, 44 AM. CRIM. L. REV. 1365, 1365-68 (2007). The author posited that personal financial ruin was a greater fear for many Americans than personal violence; that while in earlier times the home may have sheltered a person's most precious possessions, now the retirement account may be a person's "castle." *Id.*

255. Shane Pennington et al., *A Precise Model for Identity Theft Statutes*, 46 CRIM. LAW BULL. 137, 144-45 (2010) (arguing for gradations of identity theft, with the most serious version to be "identity burglary," accomplished either by invading a private space or a private computer). Another author urges us to prepare for "nanotechnology crime," and consider how some such offenses might be prosecuted as burglary. Susan W. Brenner, *Nanocrime?*, 2011 U. ILL. J.L., TECH. & POL'Y 39, 71-72 & nn.205-10.

256. The following examples are just a sampling of many case references to the common law of burglary. These examples were chosen in part for their recency.

257. State v. Engel, 210 P.3d 1007, 1009 (Wash. 2009).

258. *Id.* at 1011.

259. *Id*.

the heightened expectation of privacy and possessory rights of individuals in structures and conveyances," when it found that the yard the defendant entered did not meet the definition of a commercial yard under that state's statute.[260] Both courts referred to common law doctrine to answer a question about statutes that little resembled the common law crime.

Virginia courts referred to the common law concept of burglary as "primarily an offense against the security of the habitation," in determining that a vacation home is a "dwelling"[261] and in determining that entering a utility room from inside the garage is not breaking and entering a dwelling.[262] A California appellate court also referred to the common law crime in holding that an unenclosed balcony was not part of the "building" under the burglary statute.[263] That court noted, "The predominant factor underlying common law burglary was the desire to protect the security of the home, and the person within his home. Burglary was not an offense against property, real or personal, but an offense against the habitation . . . ."[264] The same court also extensively discussed a common law doctrine—burglary-by-instrument—in determining that inserting a burning pole into the victim's crawl space to start a fire was burglary.[265] A Mississippi court recently relied on the historical purpose of the common law offense to hold that first degree statutory burglary is a crime of violence.[266] A Nebraska court relied on common law principles to hold that the state must allege the crime intended by a burglary defendant.[267]

The common law crime is sometimes invoked as a reason to check the unlimited expansion of statutory burglary, as it was in the Washington and Arizona cases above. This notion of the common law offense as a limiting influence was also apparent in the recent back and forth between the courts and the legislatures of Florida and South Dakota. In Florida, the court in 2000 reversed a felony murder conviction where the underlying felony was burglary, although the state conceded that entry was consensual.[268] The state argued that, at some point, consent was withdrawn and the defendant remained on the premises with intent to commit murder.[269] In reversing, the court referred to the common law crime and the Model Penal Code, and held that "remaining" with

---

260. State v. Hinden, 233 P.3d 621, 624 (Ariz. Ct. App. 2010) (citations omitted).

261. Giles v. Commonwealth, 672 S.E.2d 879, 882 (Va. 2009) (citations omitted).

262. Lacey v. Commonwealth, 675 S.E.2d 846, 850-51 (Va. Ct. App. 2009).

263. People v. Yarbrough, 123 Cal. Rptr. 3d 376, 379 (Ct. App.), *review granted*, 253 P.3d 1152 (Cal. 2011). The court stated that entry onto the balcony could be charged as attempted burglary. *Id.*

264. *Id.* (quoting People v. Valencia, 46 P.3d 920 (Cal. 2002)).

265. People v. Glazier, 113 Cal. Rptr. 3d 108, 110-11, 114-15 (Ct. App. 2010).

266. Brown v. State, No. 2010-KA-00352-COA, 2011 WL 2449291, at *3, *6 (Miss Ct. App. June 21, 2011).

267. State v. Nero, 798 N.W.2d 597, 605 (Neb. 2011).

268. Delgado v. State, 776 So. 2d 233, 236 (Fla. 2000), *superseded by statute as stated in* Bradley v. State, 33 So. 2d 664 (Fla. 2010).

269. *Id.*

the intent to commit a crime could only be burglary if that remaining were surreptitious.[270] The legislature responded by amending the statute to clearly abrogate that decision.[271] Addressing a similar issue in *State v. Burdick*,[272] the South Dakota court reversed a line of cases which had held that "enters or remains in an unoccupied structure, with intent to commit any crime therein"[273] required that the entering or remaining be unauthorized.[274] The dissent noted that without such a requirement, the burglary statute "ensnares any offense committed indoors, no matter how petty," and referred to the common law offense and its traditional purpose as support.[275] In a mirror image of the Florida debate, the legislature amended the statute in accord with the dissent's interpretation.[276]

As the Florida and South Dakota examples illustrate, appeals to the common law offense as a kind of limiting principle are countered regularly by the recognition that modern statutes have little in common with the common law offense, and that it is the province of the legislature to define crimes, regardless of how they might deviate from common law.[277] Indeed, given how different most statutes are from the common law formulation, it is surprising that the common law offense is referenced at all in the caselaw.[278]

What this harkening back seems to show is the surprising influence of the common law over discussions of the statutory crime. Modern statutory burglary still operates in the shadow of Coke's definition, and ironically, it may be this shadow that obscures the radical changes that have occurred to the offense. The very name of the offense suggests something time-honored and traditional—a solid crime. To attempt yet another metaphor: The name cloaks the modern statutes in respectability—and prevents lawmakers and judges from seeing just how little lies beneath that cloak.

---

270. *Id.* at 240.

271. 2001-58 FLA. LAWS 1-2 (codified at FLA. STAT. §§ 810.015, -.02 (West 2002)).

272. 712 N.W.2d 5 (S.D. 2006).

273. S.D. CODIFIED LAWS § 22-32-8 (1976) (amended 1989 and 2005).

274. *Burdick*, 712 N.W.2d at 10.

275. *Id.* at 10-11 (Meierhenry, J., dissenting).

276. S.D. CODIFIED LAWS § 22-32-8 (2011).

277. *See, e.g.*, *Burdick*, 712 N.W.2d at 9-10.

278. The references to the common law as a limiting force may also reflect a more general discomfort some judges feel with the expansion of criminal law. Scholars have commented on "overcriminalization"—the proliferation of offenses, *see* Sanford H. Kadish, *The Crisis of Overcriminalization*, 374 ANNALS AM. ACAD. POL. & SOC. SCI. 157 (1967); William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505 (2001), and the increasing number of charges that can be attached to a single incident. Stuntz, *supra*, at 507 ("[F]ederal and state codes alike are filled with overlapping crimes, such that a single criminal incident typically violates a half dozen or more prohibitions. Lax double jeopardy doctrine generally permits the government to charge all these violations rather than selecting among them."). The ballooning of burglary is but another symptom of this general inflation. Appeals to common law rationalizations may in some sense reflect a wish to go back to a time when the possible charges were fewer and more tailored.

CONCLUSION

Modern burglary little resembles its common law ancestor, the offense of breaking and entering the dwelling of another at night with the intent to commit a felony therein. Changes that began early in our nation's history accelerated during the past fifty years as burglary in many jurisdictions shed common law elements such as breaking, entering, night time, and dwelling. In its broadest forms, burglary became the offense of being in the wrong place with the wrong intent. Its reach has extended far beyond the home. Burglary also took on a prominent role as a "location aggravator," a charge that could be added to the completed or attempted target offense and providing a significant additional penalty to crimes such as robbery, theft, or kidnapping, if they were committed in a place protected by the burglary statute. In many states, a murder committed in such a place may qualify for the death penalty on the basis of burglary.

This survey of burglary law in the United States is the first comprehensive look at burglary since the mid-twentieth century. It shows the enormous variation in state schemes, but it also shows some trends. In spite of strong critiques that burglary law was unfair, unnecessary and illogical, the crime of burglary has survived and evolved. Critics argued that burglary developed to make up for the difficulty in proving attempts at common law. Burglary can also be seen as an unnecessary combination crime; a combination of an attempted or completed crime and a criminal trespass. But legislatures for the most part resisted calls for reform, rejecting important aspects of the Model Penal Code burglary provision, and continuing to broaden the reach of the burglary statutes.

This history of burglary shows the path of evolution in criminal law; how calls for rational reform interact with pressure on judges to affirm convictions and pressure on legislatures to protect against more wrongdoing. It also shows the sometimes unexpected connections between elements of an offense: how night time became less important as burglary moved beyond dwellings, how eliminating the "breaking" requirement eroded the "entry" requirement, and how substituting "remaining" for "entry" then affected the timing of the wrongful intent. Just as pulling on a piece of yarn may start the unraveling of a sweater, it seems that removing one element of the offense leads to the weakening or elimination of others.

Burglary is still a somewhat illogical crime. When applied to housebreaking or surreptitious entry of premises to commit theft, it seems reasonable and supported by tradition. But when extended to cover what would otherwise be considered shoplifting or robbery, especially when entry was initially permitted, burglary appears less like an actual crime addressing a separate problem than a weapon for prosecutors to increase penalties, or extract pleas, based on where the crime occurred. This is especially true in the many states where burglary has lost its actus reus, "entering," and requires only remaining with criminal intent.

Burglary has traveled far from its origins. Yet the common law offense looms large even today in discussions of the statutory crime. Courts refer to the common law definition and rationales regularly when addressing questions about the modern statutes. It is very likely the influence of the common law crime in the legal imagination that has allowed burglary to survive as "burglary," even when it no longer necessarily describes a separate offense. Firmly lodged in our criminal codes, the current offense evolved far from its original form, but in the shadow of the common law.

"FASTEN YOUR SEAT BELT, ORVILLE!": EXPLORING THE RELATIONSHIP BETWEEN STATE-OF-THE-ART, TECHNOLOGICAL AND COMMERCIAL FEASIBILITY, AND THE *RESTATEMENT*'S REASONABLE ALTERNATIVE DESIGN REQUIREMENT

RICHARD C. AUSNESS*

INTRODUCTION

Of course Wilbur never told Orville to fasten his seat belt. Not only did the first Wright Flyer have no seat belts, it did not even have seats![1] Orville steered the Wright Flyer I from a prone position, not sitting up.[2] Nor was the Wright brothers' flying machine equipped with such useful safety features as brakes, landing gear, a radio, ailerons, or even much of a tail.[3] In fact, if the truth be told, the Wright brothers' airplane was a flying deathtrap. The good news for both modern pilots and their passengers is that aircraft technology, including safety technology, has progressed enormously since 1903. Aircraft manufacturers have now incorporated many forms of technology, such as radar, global positioning navigation systems, transponders, radios, anti-lock braking systems, and fire-resistant insulation in order to make their products safer. Unfortunately, progress has not been as impressive in other areas. Consider punch presses. Notwithstanding the fact that these machines cause hundreds of injuries in the workplace each year, many punch presses continue to employ Rube Goldberg-like devices such as "pullbacks" to protect punch press operators from serious injury.[4]

Why has safety technology improved so dramatically over the last hundred years for airplanes while punch press safety technology apparently has not? One explanation for this curious result is that product manufacturers effectively control the pace of technological development, including the development of safety-related technology by deciding how much to invest in research and development.[5] At the same time, the experience of aircraft and punch press

---

* Dorothy Salmon Professor of Law, University of Kentucky College of Law. B.A., 1966, J.D., 1968, University of Florida; LL.M., 1973, Yale University. The author would like to thank the University of Kentucky College of Law for its financial support for this Article in the form of a summer research grant.

1. The Wright brothers did not install seats in their aircraft until 1908. *See* T.K. DERRY & TREVOR I. WILLIAMS, A SHORT HISTORY OF TECHNOLOGY FROM THE EARLIEST TIMES TO A.D. 1900, at 401 fig. 199 (1961) (depicting Orville Wright lying prone on the Wright Flyer).

2. *See id.* (illustrating the Wright brothers' 1903 flying machine).

3. *See id.*

4. *See* Buzzell v. Bliss, 358 N.W.2d 695, 697 (Minn. Ct. App. 1984) (describing safety devices).

5. *See* Beshada v. Johns-Manville Prods. Corp., 447 A.2d 539, 548 (N.J. 1982); *see also* Omri Ben-Shahar, *Should Products Liability Be Based on Hindsight?*, 14 J.L. ECON. & ORG. 325, 340 (1998); Gary T. Schwartz, *Foreword: Understanding Products Liability*, 67 CALIF. L. REV.

manufacturers suggests that some manufacturers are willing to invest more aggressively in product safety technology than others. What can be done to encourage manufacturers who lag behind to devote more effort to the improvement of product safety? One solution may be for courts to adopt tort rules that impose liability on manufacturers who fail to develop safety technology for their products. This sort of logic led the New Jersey Supreme Court in *Beshada v. Johns-Manville Products Corp.*[6] to conclude that asbestos manufacturers would have a greater incentive to discover health risks associated with their products if knowledge of scientifically undiscoverable risks was imputed to them.[7] Other courts have endorsed this reasoning as well.[8]

However, this reasoning assumes that manufacturers can exercise almost complete control over the pace of technological development within their industry. This assumption may in fact be true when technological development is linear in nature. Linear technological development is largely independent of technological developments in other areas. However, technological development is often interdependent.[9] For example, developments in one technology may influence developments in another and vice versa, in what might be called a ping-pong model. Another form of development, which is analogous to the path of a pinball, occurs when technological progress flows from a number of different sources.[10] In both of these situations, manufacturers will not be able to significantly affect the pace of technological development outside of their industry and no amount of tort liability will change this.

A particular technological innovation, whether safety-related or not, must also be successful in the marketplace. However, in some cases, popular acceptance takes many years and in other cases, it never happens at all. There are a variety of reasons why a particular technological innovation may not succeed commercially, including cost, lack of demand, resistance by competitors or workers in other industries, social or cultural resistance and lack of a supportive infrastructure. Although manufacturers can try to encourage public acceptance of new products and designs by advertising or engaging in other

---

435, 484-85 (1979); Tyrus V. Dahl, Jr., Comment, *Strict Products Liability: The Irrelevance of Foreseeability and Related Negligence Concepts*, 14 TULSA L.J. 338, 353-54 (1978).

6. *Beshada*, 447 A.2d 539.

7. *Id.* at 545-48.

8. *See, e.g.*, Kisor v. Johns-Manville Corp., 783 F.2d 1337, 1341-42 (9th Cir. 1986); Johnson v. Raybestos-Manhattan, Inc., 740 P.2d 548, 549 (Haw.), *certifying questions sub nom. In re* Asbestos Cases, 829 F.2d 907 (9th Cir. 1987).

9. The idea for this comes from the BBC television series *Connections*, which aired in the United States in the late 1970s on many PBS stations. These highly informative and entertaining programs were written and hosted by science historian, James Burke. *Connections* (PBS television broadcast 1978). For background on *Connections*, see *Connections*, IMDB, http://www.imdb.com/title/tt0078588/ (last visited Feb. 2, 2012).

10. *See* JAMES BURKE, THE PINBALL EFFECT: HOW RENAISSANCE WATER GARDENS MADE THE CARBURETOR POSSIBLE—AND OTHER JOURNEYS THROUGH KNOWLEDGE 3-6 (1996) [hereinafter BURKE, PINBALL EFFECT].

promotional activities, there is no assurance that these efforts will succeed.

Is there a legal concept that will encourage manufacturers to make reasonable investments in product safety without punishing them for failing to achieve the impossible? One promising solution is the "state-of-the-art" concept. Although state-of-the-art has a number of different meanings,[11] in design defect cases, it should normally come into play only when a plaintiff tries to prove that a product is defectively designed because an alternative design would have prevented or reduced injury.[12] In such cases, the defendant should be allowed to defeat the plaintiff's claim by showing that his or her proposed alternative design was not within the state-of-the-art when the product was sold.

This Article begins by examining some of the case law involving the state-of-the-art concept and finds that it is principally concerned with technological feasibility. It also concludes that most cases do not treat state-of-the-art as conclusive on the design defect issue; rather, they merely consider it one of several factors that the fact finder may take into account when deciding whether a product's design is defective or not. Part II is concerned with technological development. This part examines two basic patterns of technological development and provides a number of historical examples for each. The first is a linear pattern, exemplified by violins and clipper ships. The second pattern includes development involving the interaction of two different technologies, as exemplified by the progress of shipbuilding and naval gunnery technology during the sixteenth century, as well as more complex interactions among a number of seemingly unrelated technologies. Examples of this include printing with movable type, railroads, and motor vehicles.

Part III discusses the concept of commercial feasibility. It identifies some of the conditions that often lead to prompt commercialization of new technology. These include sudden changes in the physical environment, depletion of natural resources, military competition among nations, popular dissatisfaction with the state of existing technology, as well as changing demographic and social conditions. At the same time, the Article points out that a particular technological innovation may not succeed commercially because of high cost, lack of demand, resistance by competitors or workers in other industries, social or cultural resistance, and lack of a supportive infrastructure.

Finally, the conclusion offers suggestions on how the state-of-the-art doctrine in design defect cases could be made more rational and coherent. First, state-of-the-art is not a useful concept when applied to the defendant's existing design. Instead, it should only be applied to evaluate a safer, alternative design proposed by the plaintiff. Second, the plaintiff must be required to prove that his or her proposed alternative design was technologically and commercially feasible at the time the product was sold. Third, the plaintiff's proposed alternative design can be hypothetical and does not have to be actually adopted by others in the industry. Fourth, the defendant should be allowed to dispute the plaintiff's claim by offering evidence that the proposed alternative design was not technologically

---

11. *See* DAVID G. OWEN, PRODUCTS LIABILITY LAW § 10.4 (2d ed. 2008).

12. *Id.* § 10.4, at 711-12.

or commercially feasible at the time the product was sold. Fifth, the defendant should be able to argue that the technology involved was interdependent and, therefore, it could not control the pace of its development. Sixth, technological and commercial feasibility should not be treated merely as factors for the jury to take into account; rather, they should be regarded as essential to the plaintiff's case. Consequently, a plaintiff who fails to prove that a proposed alternative design is technologically and commercially feasible should lose. Finally, even if the plaintiff proves that his or her proposed design is technologically and commercially feasible, the defendant should still be able to offer reasons, such as convenience, price, or consumer choice, to explain why its existing design should not make its product defective.

I. STATE-OF-THE-ART

A. Doctrinal Foundations

Courts have traditionally distinguished between three types of product defects: manufacturing defects, design defects, and inadequate warnings or instructions.[13] These categories are also recognized by legal commentators[14] and are embodied in the *Products Liability Restatement* as well.[15] A manufacturing defect exists when a product fails to conform to its intended design.[16] A design defect, on the other hand, occurs when a product-related risk exists which could be reduced or eliminated by an "alternative design" and the failure to do so makes "the product not reasonably safe."[17] A failure-to-warn claim is based on an assertion that the manufacturer has failed to provide "reasonable instructions or warnings" which cause the product to be "not reasonably safe."[18] While there is general agreement that the state-of-the-art concept is not relevant to cases involving manufacturing defects,[19] it is potentially applicable to both design

---

13. *See, e.g.*, Lantis v. Astec Indus., Inc., 648 F.2d 1118, 1120 (7th Cir. 1981); Piper v. Bear Med. Sys., Inc., 883 P.2d 407, 410-11 (Ariz. Ct. App. 1993); Hurley v. Heart Physicians, P.C., 898 A.2d 777, 783 (Conn. 2006); Cooper Tire & Rubber Co. v. Mendez, 204 S.W.3d 797, 800 (Tex. 2006).

14. *See* OWEN, *supra* note 11, § 6.2; Jerry J. Phillips, *A Synopsis of the Developing Law of Products Liability*, 28 DRAKE L. REV. 317, 342 (1978); William C. Powers, Jr., *The Persistence of Fault in Products Liability*, 61 TEX. L. REV. 777, 782 (1983).

15. RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2 (1998).

16. *Id.* § 2(a).

17. *Id.* § 2(b).

18. *Id.* § 2(c).

19. *See* Reed v. Tiffin Motor Homes, Inc., 697 F.2d 1192, 1196 (4th Cir. 1982); Singleton v. Int'l Harvester Co., 685 F.2d 112, 115 (4th Cir. 1981); Sturm, Ruger & Co. v. Day, 594 P.2d 38, 44-45 (Alaska 1979), *modified*, 615 P.2d 621 (Alaska 1980), *overruled by* Dura Corp. v. Harned, 703 P.2d 396 (Alaska 1985); Brady v. Melody Homes Mfr., 589 P.2d 896, 899 (Ariz. Ct. App. 1978). *But see* Indianapolis Athletic Club, Inc. v. Alco Standard Corp., 709 N.E.2d 1070, 1074-75 (Ind. Ct. App. 1999) (holding that statutory state-of-the-art defense applied to manufacturing

defect and failure-to-warn cases.[20]

1. Failure-to-Warn.—In many cases, particularly those involving toxic substances or prescription drugs, manufacturers have attempted to defend against failure-to-warn claims by contending that the risk in question was undiscoverable given the state of scientific knowledge at the time the product was manufactured.[21] Some courts have rejected these attempts to invoke state-of-the-art as a defense in failure-to-warn cases, while others have been more receptive. *Beshada v. Johns-Manville Products Corp.* is illustrative of the former position,[22] while *Feldman v. Lederle Laboratories* represents the latter approach.[23]

In *Beshada*, the plaintiffs were injured as the result of exposure to asbestos insulation products in the workplace between 1930 and 1980.[24] They claimed that the defendants failed to provide warnings on their products about the health hazards of asbestos exposure.[25] The defendants responded that the medical profession did not become aware of the potential danger of exposure to low concentrations of asbestos until the 1960s.[26] However, the New Jersey Supreme Court characterized state-of-the-art as a "negligence defense" and refused to allow the defendants to raise it.[27] Instead, the *Beshada* court declared that under strict liability in tort, liability was based on the condition of the product rather than what the defendant knew or could have known about the product's inherent risks.[28] The court supported its decision to reject the state-of-the-art defense by maintaining that it would advance various goals of products liability such as risk spreading, accident cost avoidance and facilitating the fact finding process in litigation.[29] In addition, the *Beshada* court emphasized that imposing liability on producers of dangerous products would encourage them to discover risks and improve product safety more rapidly.[30]

Although a few courts agreed with the *Beshada* court's hindsight approach and concluded that manufacturers had a duty to warn about scientifically unknowable risks,[31] the reaction of legal commentators to the New Jersey court's

---

defects).

20. *Reed*, 697 F.2d at 1196; *Brady*, 589 P.2d at 899.

21. *See, e.g.*, Feldman v. Lederle Labs., 479 A.2d 374, 386 (N.J. 1984); Beshada v. Johns-Manville Prods. Corp., 447 A.2d 539, 545-46 (N.J. 1982).

22. *Beshada*, 447 A.2d at 546-47.

23. *Feldman*, 479 A.2d at 386.

24. *Beshada*, 447 A.2d at 542-43.

25. *Id.* at 542.

26. *Id.*

27. *Id.* at 546.

28. *Id.*

29. *Id.* at 547-49.

30. *Id.* at 548.

31. *See* Kisor v. Johns-Manville Corp., 783 F.2d 1337, 1341-42 (9th Cir. 1986); Johnson v. Raybestos-Manhattan, Inc., 740 P.2d 548, 549 (Haw.), *certifying questions to sub nom. In re* Asbestos Cases, 829 F.2d 907 (9th Cir. 1987); Hayes v. Ariens Co., 462 N.E.2d 273, 277-78 (Mass. 1984), *abrogated by* Vassallo v. Baxter Healthcare Corp., 696 N.E.2d 909 (Mass. 1998); Elmore

decision was largely critical.[32] Many of them argued that the hindsight approach imposed an impossible burden on manufacturers, that it would cause financial hardship on producers, lead to unnecessary bankruptcies, and encourage undesirable corporate behavior.[33]

An opposing view is reflected by *Feldman v. Lederle Laboratories*,[34] ironically decided by the same New Jersey court that had decided *Beshada* two years earlier. In *Feldman*, a young girl's teeth became discolored as the result of taking Declomycin, a prescription tetracycline antibiotic manufactured by the defendant.[35] The plaintiff sued the drug manufacturer, claiming that it should have warned about the risk of discoloration when the drug was first marketed in 1959.[36] The defendant, on the other hand, insisted that the risk of tooth discoloration in humans did not become known until several years after the plaintiff was exposed to the drug.[37] After finding that strict liability applied to manufacturers of prescription drugs,[38] the court declared that the reasonableness of the defendant's conduct was a factor to consider in determining liability.[39] Moreover, the *Feldman* court concluded, the scientific knowledge available to the defendant was relevant to measuring the reasonableness of its conduct.[40] Although it refused to expressly overrule *Beshada*, the court restricted that case "to the circumstances giving rise to its holding," whatever those might have been.[41] However, in a minor concession to *Beshada*, the New Jersey Supreme Court declared that the defendant had the burden of proving that information about the particular product risk was not available and, therefore, it had neither actual nor constructive knowledge of the need for a warning.[42]

The *Feldman* court's foresight approach has been adopted by a majority of courts[43] and is also reflected in the *Products Liability Restatement*.[44] Although

---

v. Owens-Illinois, Inc., 673 S.W.2d 434, 438 (Mo. 1984).

32. *See, e.g.*, James A. Henderson, Jr. & Aaron D. Twerski, *Doctrinal Collapse in Products Liability: The Empty Shell of Failure to Warn*, 65 N.Y.U. L. REV. 265, 273-80 (1990); Victor Schwartz, *The Post-Sale Duty to Warn: Two Unfortunate Forks in the Road to a Reasonable Doctrine*, 58 N.Y.U. L. REV. 892, 901-05 (1983); John W. Wade, *On the Effect in Product Liability of Knowledge Unavailable Prior to Marketing*, 58 N.Y.U. L. REV. 734, 754-56 (1983).

33. *See* OWEN, *supra* note 11, § 6.2; Alan Schwartz, *Products Liability, Corporate Structure, and Bankruptcy: Toxic Substances and the Remote Risk Relationship*, 14 J. LEGAL STUD. 689, 736 (1985).

34. 479 A.2d 374 (N.J. 1984).

35. *Id.* at 376-77.

36. *Id.* at 377-78.

37. *Id.* at 377.

38. *Id.* at 380-84.

39. *Id.* at 385.

40. *Id.* at 386.

41. *Id.* at 388.

42. *Id.* at 388-89.

43. *See, e.g.*, Lohrmann v. Pittsburgh Corning Corp., 782 F.2d 1156, 1164-65 (4th Cir. 1986); Powers v. Taser Int'l, Inc., 174 P.3d 777, 783-84 (Ariz. Ct. App. 2008); Gourdine v. Crews,

this approach has much to recommend it from a fairness perspective, it is sometimes difficult to apply in practice because knowledge of a particular fact may be hard to pinpoint. For example, in *Feldman*, knowledge of the risk of tooth discoloration from ingestion of tetracycline evolved over a period of at least seven years.[45] The first study on the subject, published in 1956, revealed “that tetracycline accumulated in mineralized portions of growing bones and teeth of mice.”[46] Another study, published in 1957, reported that laboratory animals developed yellow fluorescents (not staining) in teeth and bones after receiving dosages of tetracycline.[47] Studies in 1959 and 1960 also revealed fluorescents, but not staining, in patients with cystic fibrosis following massive doses of tetracycline.[48] The link between tetracycline and permanent tooth discoloration in humans was not clearly established until 1963, when the manufacturer began to receive complaints from doctors that Declomycin (a tetracycline antibiotic) was causing tooth discoloration in patients.[49] The court let the jury determine at what point the manufacturer of Declomycin should have discovered the connection between tetracyclines and tooth discoloration.[50]

It appears that the *Feldman* court’s foresight test has carried the day. Since manufacturers do not have to warn about scientifically discoverable risks, state-of-the-art evidence will be critical to the issue of whether a particular risk was scientifically discoverable at the time the product was sold.

2. Design Defects.—The concept of state-of-the-art also applies to product design. In the early years of products liability, however, many courts refused to allow manufacturers to introduce evidence that a safer design was not within the state-of-the-art.[51] The reason for excluding state-of-the-art evidence was that it was only relevant to the reasonableness of the manufacturer’s conduct—not to the product’s condition. Thus, according to these courts, state-of-the-art evidence might be admissible in a negligence action but was irrelevant in a strict liability action where the focus was solely on the condition of the product.[52]

---

955 A.2d 769, 781 n.10 (Md. 2008); Owens-Illinois, Inc. v. Zenobia, 601 A.2d 633, 641 (Md. 1992); Vassallo v. Baxter Healthcare Corp., 696 N.E.2d 909, 922-24 (Mass. 1998); Young v. Key Pharm., Inc., 922 P.2d 59, 63-65 (Wash. 1996). *But see* Sternhagen v. Dow Co., 935 P.2d 1139, 1147 (Mont. 1997); Green v. Smith & Nephew AHP, Inc., 629 N.W.2d 727, 746-47 (Wis. 2001).

44. *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2(c) (1998).

45. *See Feldman*, 479 A.2d at 378-79 (noting evolution of discovery from 1956-1963).

46. *Id.* at 378.

47. *Id.*

48. *Id.*

49. *Id.* at 379.

50. *Id.* at 392-93.

51. *See, e.g.*, Dart v. Wiebe Mfg., Inc., 709 P.2d 876, 881-82 (Ariz. 1985); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Cryts v. Ford Motor Co., 571 S.W.2d 683, 686-89 (Mo. Ct. App. 1978); Carrecter v. Colson Equip. Co., 499 A.2d 326, 330-31 (Pa. Super. Ct. 1985).

52. *See, e.g.*, Rucker v. Norfolk & W. Ry. Co., 381 N.E.2d 715, 723-24 (Ill. App. Ct. 1978), *rev’d on other grounds*, 396 N.E.2d 534 (Ill. 1979); Matthews v. Stewart Warner Corp., 314 N.E.2d

Nowadays, many courts require plaintiffs to establish that the defendant's design is defective by showing that an alternative design was potentially available to the defendant that would have prevented the plaintiff's injuries or at least have reduced the severity of those injuries. When this occurs, it becomes important to know exactly what state-of-the-art means and what kind of evidence is necessary to determine whether an alternative design is, or is not, within the state-of-the-art.

B. The Various Meanings of "State-of-the-Art" in Products Liability Law

A number of courts[53] and commentators[54] have pointed out that state-of-the-art has various meanings and this had led to much confusion in the law. At various times, courts have defined state-of-the-art to include the following: (1) custom or common practices within an industry;[55] (2) standards promulgated by independent standards development organizations like the American National Standards Institute (ANSI);[56] (3) standards embodied in statutes and governmental regulations;[57] and (4) technical, mechanical or scientific knowledge reasonably feasible when a product is manufactured.[58] In addition, a number of states have enacted statutes that purport to define the meaning of state-of-the-

---

683, 692 (Ill. App. Ct. 1974); *Gelsumino*, 295 N.E.2d at 113; *Cryts*, 571 S.W.2d at 689; *Carrecter*, 499 A.2d at 329.

53. *See, e.g.*, Potter v. Chi. Pneumatic Tool Co., 694 A.2d 1319, 1345 (Conn. 1997); Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 298 (Iowa 1994) (Ternus, J., concurring).

54. *See* OWEN, *supra* note 11, § 10.4, at 706-07; Patrick R. Buckler, *State of the Art Evidence in Products Liability Suits in Maryland*, 28 U. BALT. L. REV. 117, 120 (1998); Gary C. Robb, *A Practical Approach to Use of State of the Art Evidence in Strict Products Liability Cases*, 77 NW. U. L. REV. 1, 2 (1982); Frank J. Vandall, *State-of-the-Art, Custom, and Reasonable Alternative Design*, 28 SUFFOLK U. L. REV. 1193, 1193 (1994).

55. *See* Smith v. Minster Mach. Co., 669 F.2d 628, 633 (10th Cir. 1982); Keogh v. W.R. Grasle, Inc., 816 P.2d 1343, 1349 (Alaska 1991); Sturm, Ruger & Co. v. Day, 594 P.2d 38, 44 (Alaska 1979), *modified*, 615 P.2d 621 (Alaska 1980), *overruled by* Dura Corp. v. Harned, 703 P.2d 396 (Alaska 1985); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Suter v. San Angelo Foundry & Mach. Co., 406 A.2d 140, 151 (N.J. 1979), *superseded by statutes as stated in* Dewey v. R.J. Reynolds Tobacco Co., 577 A.2d 1239 (N.J. 1990).

56. *See, e.g.*, AC&S, Inc. v. Asner, 686 A.2d 250, 254-55 (Md. Ct. App. 1996).

57. *See, e.g.*, Bruce v. Martin-Marietta Corp., 544 F.2d 442, 446-47 (10th Cir. 1976).

58. *See, e.g.*, Gosewisch v. Am. Honda Motor Co., 737 P.2d 365, 370 (Ariz. Ct. App. 1985), *opinion vacated in part by* 737 P.2d 376 (Ariz. 1987); Indianapolis Athletic Club, Inc. v. Alco Standard Corp., 709 N.E.2d 1070, 1074 (Ind. Ct. App. 1999); Falada v. Trinity Indus., Inc., 642 N.W.2d 247, 251 (Iowa 2002); *AC&S, Inc.*, 686 A.2d at 254; Wiska v. St. Stanislaus Soc. Club, Inc., 390 N.E.2d 1133, 1138 n.8 (Mass. App. Ct. 1979); Johnson v. Salem Corp., 477 A.2d 1246, 1251-52 (N.J. 1984); Crispin v. Volkswagenwerk AG, 591 A.2d 966, 973 (N.J. Super. Ct. App. Div. 1991); Lancaster Silo & Block Co. v. N. Propane Gas Co., 427 N.Y.S.2d 1009, 1016 n.2 (App. Div. 1980); Boatland of Hous., Inc. v. Bailey, 609 S.W.2d 743, 748 (Tex. 1980); Lenhardt v. Ford Motor Co., 683 P.2d 1097, 1099 (Wash. 1984).

art.[59]

1. Custom or Practice Within an Industry.—"Custom of the industry" is defined as the "usual practice of the manufacturer"[60] or others within a particular industry.[61] While a few courts have declared industry custom to be equivalent to state-of-the-art,[62] most have recognized that the two concepts are different.[63] According to these latter courts, custom of the industry evidence describes what manufacturers within an industry have actually achieved, while state-of-the-art evidence is concerned with what is feasible for manufacturers to achieve, whether they have done so or not.[64] At the same time, some of these courts have acknowledged that industry custom may be offered as evidence of what the state-of-the-art is.[65]

In a negligence case, a product is considered to be negligently designed when a manufacturer fails to exercise reasonable care when designing the product, thereby failing to make it safe for its intended uses.[66] Evidence of a custom or practice within an industry is usually admissible in a negligence action.[67] This is because one who acts like others in a trade or industry is arguably exercising reasonable care in that respect.[68] On the other hand, compliance with the custom

---

59. For an excellent discussion of these statutes, see DAVID G. OWEN ET AL., 1 MADDEN & OWEN ON PRODUCTS LIABILITY § 10:7, at 661-68 (3d ed. 2000).

60. Carter v. Massey-Ferguson, Inc., 716 F.2d 344, 347 (5th Cir. 1983).

61. *See Gosewisch*, 737 P.2d at 370; *Falada*, 642 N.W.2d at 250; Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 295 (Iowa 1994); Chown v. USM Corp., 297 N.W.2d 218, 221 (Iowa 1980); *Lenhardt*, 683 P.2d at 1099.

62. *See* Smith v. Minster Mach. Co., 669 F.2d 628, 633 (10th Cir. 1982); Keogh v. W.R. Grasle, Inc., 816 P.2d 1343, 1349 (Alaska 1991); Sturm, Ruger & Co. v. Day, 594 P.2d 38, 44 (Alaska 1979), *modified*, 615 P.2d 621 (Alaska 1980), *overruled by* Dura Corp. v. Harned, 703 P.2d 396 (Alaska 1985); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Suter v. San Angelo Foundry & Mach. Co., 406 A.2d 140, 151 (N.J. 1979); *see also* Walker v. Trico Mfg. Co., 487 F.2d 595, 600 (7th Cir. 1973) (characterizing expert's testimony relating to practice of the trade as "state-of-the-art" evidence).

63. *See Carter*, 716 F.2d at 347-48; *Gosewisch*, 737 P.2d at 370; *Falada*, 642 N.W.2d at 250; *Hughes*, 522 N.W.2d at 295-96; *Chown*, 297 N.W.2d at 221-22; Owens-Illinois, Inc. v. Zenobia, 601 A.2d 633, 640-41 (Md. 1992); Hancock v. Paccar, Inc., 283 N.W.2d 25, 35 (Neb. 1979); *Boatland of Hous., Inc.*, 609 S.W.2d at 748; *Lenhardt*, 683 P.2d at 1099; Cantu v. John Deere Co., 603 P.2d 839, 840 (Wash. Ct. App. 1979); *see also* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2, Reporters' Note, at 81-84 (1998).

64. *Carter*, 716 F.2d at 347 & n.6.

65. *See Hughes*, 522 N.W.2d at 296; *Hancock*, 283 N.W.2d at 35.

66. *See* Mather v. Caterpillar Tractor Corp., 533 P.2d 717, 719 (Ariz. Ct. App. 1975).

67. *See* Holloway v. J.B. Sys., Ltd., 609 F.2d 1069, 1073 (3d Cir. 1979); Collins v. Ridge Tool Co., 520 F.2d 591, 594 (7th Cir. 1975); Hillrichs v. Avco Corp., 514 N.W.2d 94, 98 (Iowa 1994); Spieker v. Westgo, Inc., 479 N.W.2d 837, 843-44 (N.D. 1992).

68. *See* Rexrode v. Am. Laundry Press Co., 674 F.2d 826, 831-32 (10th Cir. 1982); Garst v. Gen. Motors Corp., 484 P.2d 47, 61 (Kan. 1971).

of the industry is not conclusive evidence of reasonable care.[69] As Judge Learned Hand observed many years ago, industry custom is not necessarily determinative of due care because "a whole calling may have unduly lagged in the adoption of new and available devices."[70]

This principle is illustrated by *Hillrichs v. Avco Corp.*[71] The plaintiff in that case was injured when his hand was caught in the rollers of the husking bed of a twenty-year-old cornpicking machine manufactured by the defendant.[72] The plaintiff sued under negligence, breach of warranty, and strict liability, alleging that the cornpicker's design was defective because the machine was not equipped with an emergency stop device.[73] The first trial resulted in a jury verdict for the manufacturer.[74] The plaintiff appealed and the Iowa Supreme Court affirmed in part, but remanded the case back for a new trial on the plaintiff's negligence claim based on the enhanced injuries he suffered as a result of not being able to turn the machine off quickly enough.[75] At the second trial, the defendant contended that its design was consistent with the custom and practice of the farm implement industry because no cornpicker on the market was equipped with an emergency stop device.[76] However, the plaintiff responded by pointing out that at the time the cornpicker was manufactured, an emergency stop device could have been installed for less than fifty dollars and that other machines that used rollers, such as printing presses, were already equipped with such devices.[77] Declaring that compliance with industry custom was not the same as compliance with state-of-the-art, the court affirmed the lower court's judgment for the plaintiff on his enhanced injury claim.[78]

There is less agreement on the role of industry custom when the plaintiff's case is based on strict products liability instead of negligence. A number of courts allow defendants to introduce evidence of industry custom in strict liability cases.[79] As with negligence cases, such evidence is not conclusive on

---

69. *See* Sturm, Ruger & Co. v. Day, 594 P.2d 38, 45 (Alaska 1979), *modified*, 615 P.2d 621 (Alaska 1980), *overruled by* Dura Corp. v. Harned, 703 P.2d 396 (Alaska 1985); Caterpillar Tractor Co. v. Beck, 593 P.2d 871, 887 (Alaska 1979); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Montgomery Ward & Co. v. Gregg, 554 N.E.2d 1145, 1155-56 (Ind. Ct. App. 1990); Dudley Sports Co. v. Schmitt, 279 N.E.2d 266, 276 (Ind. App. 1972); *Garst*, 484 P.2d at 61; Cantu v. John Deere Co., 603 P.2d 839, 841 (Wash. Ct. App. 1979).

70. T.J. Hooper, 60 F.2d 737, 740 (2d Cir. 1932).

71. *Hillrichs*, 514 N.W.2d 94.

72. *Id.* at 96.

73. *Id.*

74. *Id.*

75. *Id.*

76. *Id.* at 98.

77. *Id.* at 97.

78. *Id.* at 100-01.

79. *See* Robinson v. Audi NSU Auto Union Aktiengesellschaft, 739 F.2d 1481, 1486 (10th Cir. 1984); Carter v. Massey-Ferguson, Inc., 716 F.2d 344, 348 (5th Cir. 1983); Sturm, Ruger & Co. v. Day, 594 P.2d 38, 45 (Alaska 1979), *modified*, 615 P.2d 621 (Alaska 1980), *overruled by*

the issue of defectiveness, but is only a factor for the jury to consider.[80]

In *Robinson v. Audi NSU Auto Union Aktiengesellschaft*, for example, several plaintiffs who were injured in a rear-end collision brought suit against Audi, the manufacturer of their automobile.[81] The crash caused the car's doors to be wedged shut and also caused the fuel tank to burst into flames.[82] The plaintiffs alleged that the automobile was defectively designed because the fuel tank was placed where it could be easily punctured by the trunk contents when struck from behind.[83] The jury found in favor of the defendant and the plaintiffs appealed.[84]

In their appeal, the plaintiffs argued, inter alia, that the trial court erroneously allowed the defendant to introduce evidence of customary fuel tank design to prove that their design complied with the "state-of-the-art."[85] The defendant, on the other hand, contended that evidence of the custom and practice within the industry was relevant in a strict liability case to determine the expectations of the ordinary consumer.[86] The appeals court agreed, pointing out that since the plaintiffs had introduced evidence of fuel tank designs in other motor vehicles to show that such alternative designs were feasible, the defendant should be allowed to rely on industry custom to establish the expectations of the ordinary consumer.[87]

In contrast, other courts maintain that evidence of industry custom and practices is irrelevant in a strict products liability case and distracts the jury's attention from the condition of the product to the reasonableness of the defendant's conduct.[88] For example, in *Holloway v. J.B. Systems, Ltd*., a federal appeals court, applying Pennsylvania law, concluded that evidence of industry custom was not admissible.[89] In that case, the plaintiff was struck in the head by a bolt which broke loose from a tanker truck as it was being pressurized to

---

Dura Corp. v. Harned, 703 P.2d 396 (Alaska 1985); Murphy v. Chestnut Mountain Lodge, Inc., 464 N.E.2d 818, 823-24 (Ill. App. Ct. 1984); Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 296 (Iowa 1994); Thibault v. Sears, Roebuck & Co., 395 A.2d 843, 850-51 (N.H. 1978).

80. *See* Smith v. Minster Mach. Co., 669 F.2d 628, 633 (10th Cir. 1982); Keogh v. W.R. Grasle, Inc., 816 P.2d 1343, 1349 (Alaska 1991); *Sturm, Ruger & Co.*, 594 P.2d at 45; Hancock v. Paccar, Inc., 283 N.W.2d 25, 35 (Neb. 1979).

81. *Robinson*, 739 F.2d at 1482-83.

82. *Id.*

83. *Id.* at 1483.

84. *Id.*

85. *Id.* at 1485.

86. *Id.*

87. *Id.* at 1485-86.

88. *See* Habecker v. Clark Equip. Co., 36 F.3d 278, 282-83 (3d Cir. 1994); Holloway v. J.B. Sys., Ltd., 609 F.2d 1069, 1073 (3d Cir. 1979); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Uxa v. Marconi, 128 S.W.3d 121, 131-32 (Mo. Ct. App. 2003); Lewis v. Coffing Hoist Div., Duff-Norton Co., Inc., 528 A.2d 590, 594 (Pa. 1987); Lenhardt v. Ford Motor Co., 683 P.2d 1097, 1099-1100 (Wash. 1984).

89. *Holloway*, 609 F.2d at 1073.

discharge its fuel.[90] In a strict liability case against the manufacturer of the tank, the plaintiff argued that the manufacturer should have provided a warning that the tank should not be subjected to internal pressurization.[91] At trial, the plaintiff objected to testimony that the defendant, in not providing a warning with the tank, had merely conformed to the custom of the industry when the tank was manufactured six years before the accident.[92]

On appeal, the court held that it was improper to admit testimony regarding trade custom to the effect that virtually no other tank manufacturer in 1969 provided a warning about pressurization.[93] Although the court ultimately concluded that the lower court had not committed reversible error, it affirmed that "negligence concepts such as 'trade custom' or 'reasonable care' have no place in suits brought under [*Restatement* §] 402A as that section has been interpreted by the Pennsylvania courts."[94]

2. Industry Standards.—Courts sometimes include industry standards within the meaning of state-of-the-art, while others distinguish the two concepts, typically concluding that industry standards are not as demanding as state-of-the-art requirements.[95] A Maryland court in *AC&S, Inc. v. Asner* defined "industry standards" as follows: "Industry standards are the practices common to a given industry. They are often set forth in some type of code, such as a building code or electrical code, or they may be adopted by the trade organization of a given industry."[96] Industry standards fall into two basic categories: those that are formulated by the industry itself, often through trade associations; and those that are promulgated by independent standard development organizations like the American National Standards Institute (ANSI), the American Society of Mechanical Engineers (ASME), Underwriters Laboratories (UL) or the American Society for Testing and Materials (ASTM).[97] Although standards created by industry trade associations are sometimes weak or self-serving,[98] standards formulated by independent standard development organizations tend to be more

90. *Id.* at 1070-71.

91. *Id.* at 1071.

92. *Id.* at 1072.

93. *Id.* at 1073.

94. *Id.*

95. *See* Mercer v. Pittway Corp., 616 N.W.2d 602, 622 (Iowa 2000); AC&S, Inc. v. Asner, 686 A.2d 250, 254-55 (Md. 1996); Cantu v. John Deere Co., 603 P.2d 839, 840 (Wash. Ct. App. 1979).

96. *AC&S, Inc.*, 686 A.2d at 254-55 (quoting Lohrmann v. Pittsburgh Corning Corp., 782 F.2d 1156, 1164 (4th Cir. 1986)).

97. *See, e.g.*, Milanowicz v. Raymond Corp., 148 F. Supp. 2d 525, 533 (D.N.J. 2001). Other standard setting organizations include the American Standards Association (ASA); the National Safety Council (NSC); the Society of Automotive Engineers (SAE); and the National Fire Protection Association (NFPA). *See* OWEN, *supra* note 11, § 2.3, at 83.

98. *See, e.g.*, Hall v. E.I. Du Pont de Nemours & Co., 345 F. Supp. 353 (E.D.N.Y. 1972) (holding blasting cap manufacturers and their trade association liable for developing inadequate safety standards with respect to warnings on blasting caps).

rigorous. These organizations require that their standards be technologically up to date, that they be developed in a transparent manner and that they reflect the consensus view of all interested parties.[99]

As is the case with industry custom and practice, evidence of compliance with industry standards is usually admissible in negligence actions as proof that a manufacturer exercised reasonable care in the design of the product.[100] However, there is a split of authority over whether such evidence should be admissible in a strict liability case. A large number of courts allow evidence of industry standards, particularly those that have been promulgated by independent standard development organizations, to be admitted in order to establish that a product is not defective.[101]

Some courts exclude evidence of industry standards in strict liability cases because they believe that such evidence unduly focuses attention on the conduct of the manufacturer instead of the condition of the product.[102] The Pennsylvania Supreme Court's decision in *Lewis v. Coffing Hoist Division, Duff-Norton Co.*[103] exemplifies this approach. The plaintiff in that case was injured while operating an overhead electric chain-hoist to lift into position a carriage assembly component of a machine being manufactured by his employer.[104] The hoist could be stopped and started by means of a "control pendant," which included a control box leading to a hoist motor overhead.[105] While attempting to fix a stuck chain, the plaintiff "stumbled and fell, causing his thumb to strike the 'down' button on the control box."[106] This, in turn, caused the carriage assembly to swing forward and hit the plaintiff in the legs.[107] The plaintiff brought a strict liability action against the manufacturer of the hoist, claiming that the control box was

---

99. *See* Robert H. Heidt, *Damned for Their Judgment: The Tort Liability of Standards Development Organizations*, 45 WAKE FOREST L. REV. 1227, 1263-64 (2010).

100. *See AC&S, Inc.*, 686 A.2d at 255; Back v. Wickes Corp., 378 N.E.2d 964, 970 (Mass. 1978).

101. *See* Sappington v. Skyjack, Inc., 512 F.3d 440, 454 (8th Cir. 2008); Murphy v. L & J Press Corp., 558 F.2d 407, 411 (8th Cir. 1977); Stanczyk v. Black & Decker, Inc., 836 F. Supp. 565, 567 (N.D. Ill. 1993); Clement v. Rousselle Corp., 372 So. 2d 1156, 1160 (Fla. Dist. Ct. App. 1979); Dugan v. Sears, Roebuck & Co., 447 N.E.2d 1055, 1057 (Ill. App. Ct. 1983); *AC&S, Inc.*, 686 A.2d at 256; *Back*, 378 N.E.2d at 970; Fabian v. Minster Mach. Co., 609 A.2d 487, 491 (N.J. Super. Ct. App. Div. 1992); Vermett v. Fred Christen & Sons Co., 741 N.E.2d 954, 971 (Ohio Ct. App. 2000); Hansen v. Abrasive Eng'g & Mfg., Inc., 831 P.2d 693, 697 n.5 (Or. Ct. App. 1992), *aff'd in part and rev'd in part*, 856 P.2d 625 (Or. 1993).

102. *See* Habecker v. Clark Equip. Co., 36 F.3d 278, 286-87 (3d Cir. 1994); Rexrode v. Am. Laundry Press Co., 674 F.2d 826, 831-32 (10th Cir. 1982); Matthews v. Stewart Warner Corp., 314 N.E.2d 683, 691-92 (Ill. App. Ct. 1974); Johnson v. Hannibal Mower Corp., 679 S.W.2d 884, 885 (Mo. Ct. App. 1984).

103. 528 A.2d 590 (Pa. 1987).

104. *Id.* at 590-91.

105. *Id.* at 591.

106. *Id.*

107. *Id.*

defectively designed because it did not have a guard or other protective feature to prevent the hoist from being accidently activated.[108]

The trial court granted the plaintiff's in limine motion to exclude an ASME publication which set forth standards for electric hoists and other industrial lifting equipment.[109] The court also ruled that the defendant could not present evidence that at least ninety percent of the control boxes made in the United States did not have guards over their activating buttons.[110] The jury rendered a verdict in the plaintiff's favor and the defendant appealed, claiming that it was error for the lower court to exclude evidence of industry standards.[111]

The Pennsylvania court declared that in order to determine the admissibility of evidence, it must first consider the relevance of the evidence to the issue in question.[112] Finding that industry standards were only relevant to the reasonableness of the defendant's conduct, and not to the condition of the product, the court stated that the introduction of this sort of evidence would have improperly brought concepts of negligence law into the case.[113] Furthermore, the court concluded that "such evidence would have created a strong likelihood of diverting the jury's attention from the appellant's control box to the reasonableness of the appellant's conduct in choosing its design."[114] Accordingly, the court affirmed the lower court's judgment for the plaintiff.[115]

3. Government Regulatory Standards.—Government standards may also serve as a measure of state-of-the-art. Government standards have traditionally been admissible in negligence cases as evidence of reasonable care although the approach is asymmetrical.[116] In the majority of states, failure to comply with applicable regulatory standards constitutes negligence per se, that is, the defendant's conduct is deemed to be negligent as a matter of law.[117] On the other hand, compliance with applicable regulatory standards is considered to be merely evidence of reasonable care and not conclusive.[118] The reason for this anomaly is that many courts believe that government regulations often do not establish optimal standards of care. For example, standards may be obsolete[119] or

108. *Id.*

109. *Id.*

110. *Id.*

111. *Id.* at 592.

112. *Id.*

113. *Id.* at 594.

114. *Id.*

115. *Id.*

116. *See* Richard C. Ausness et al., *Providing a Safe Harbor for Those Who Play by the Rules: The Case for a Strong Regulatory Compliance Defense*, 2008 UTAH L. REV. 115, 117.

117. *See* Elsworth v. Beech Aircraft Corp., 691 P.2d 630, 634 (Cal. 1984).

118. *See* Raymond v. Riegel Textile Corp., 484 F.2d 1025, 1028 (1st Cir. 1973); Burke v. Dow Chem. Co., 797 F. Supp. 1128, 1142 (E.D.N.Y. 1992); Blasing v. P.R.L. Hardenbergh Co., 226 N.W.2d 110, 115 (Minn. 1975); Feiner v. Calvin Klein, Ltd., 549 N.Y.S.2d 692, 693 (App. Div. 1990); Sherman v. M. Lowenstein & Sons, Inc., 282 N.Y.S.2d 142, 143-44 (App. Div. 1967).

119. *See* Mark DeSimone, Comment, *The State of the Art Defense in Products Liability:*

substantively inadequate[120] and, therefore, not reflective of an appropriate standard of due care.

Although there are exceptions,[121] most courts acknowledge that government standards involving products are relevant to the issue of defectiveness.[122] However, because they regard regulatory standards as no more than minimums, violations are treated differently than compliance. In some cases, a product that fails to comply with government standards is regarded as defective per se.[123] On the other hand, while manufacturers may introduce evidence that their products complied with government standards, this evidence is seldom conclusive.[124]

Bruce v. Martin-Marietta Corp.[125] provides a good illustration of these principles. The case arose out of a tragic airplane crash in 1970, in which many members of the Wichita State University football team and some of its supporters were killed.[126] The chartered plane crashed into a mountain in Colorado while traveling to a football game in Logan, Utah.[127] During the crash, seats in the aircraft broke loose from their floor attachments and blocked the exit.[128] When the aircraft caught fire, many of the passengers were unable to escape because of the blocked exit.[129] As a result, thirty-two of the forty passengers and crew were killed.[130]

Injured passengers and personal representatives of some of those killed in the accident brought suit against Martin-Marietta, the manufacturer of the airplane.[131] Their negligence and strict liability claims alleged that the airplane was not crashworthy because the seat attachments were inadequate and the airplane

---

"Unreasonably Dangerous" to the Injured Consumer, 18 DUQ. L. REV. 915, 923 (1980).

120. *See* Teresa Moran Schwartz, *The Role of Federal Safety Regulations in Products Liability Actions*, 41 VAND. L. REV. 1121, 1147, 1151 (1988).

121. *See* Sheehan v. Cincinnati Shaper Co., 555 A.2d 1352, 1354-55 (Pa. Super. Ct. 1989) (providing the admission of OSHA regulations).

122. *See* Bruce v. Martin-Marietta Corp., 544 F.2d 442, 447 (10th Cir. 1976); Clement v. Rousselle Corp., 372 So. 2d 1156, 1160 (Fla. Dist. Ct. App. 1979); Moehle v. Chrysler Motors Corp., 443 N.E.2d 575, 577-78 (Ill. 1982); Turney v. Ford Motor Co., 418 N.E.2d 1079, 1084-85 (Ill. App. Ct. 1981); Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 296 (Iowa 1994); Wilson v. Piper Aircraft Corp., 577 P.2d 1322, 1324-25 (Or. 1978).

123. *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 4(a) (1998); OWEN, *supra* note 11, § 6.4, at 396-98; *see also* Stanton v. Astra Pharm. Prods., Inc., 718 F.2d 553, 569-71 (3d Cir. 1983); Lukaszewicz v. Ortho Pharm. Corp., 510 F. Supp. 961, 965 (E.D. Wis.), *opinion amended by* 532 F. Supp. 211 (E.D. Wis. 1981).

124. *See Bruce*, 544 F.2d at 446; *Moehle*, 443 N.E.2d at 578; Brooks v. Beech Aircraft Corp., 902 P.2d 54, 63 (N.M. 1995).

125. *Bruce*, 544 F.2d 442.

126. *Id.* at 444.

127. *Id.*

128. *Id.*

129. *Id.*

130. *Id.*

131. *Id.*

lacked proper fire protection features.[132] To support their design defect claim with respect to the passenger seats, the plaintiffs offered evidence that at the time of the crash there were seats in use that would have withstood the crash.[133] In response, the defendant alleged that when the aircraft was manufactured in 1952, it met or exceeded "all applicable design requirements, safety requirements and other criteria prescribed by the Civil Aeronautics Administration and was manufactured and certificated in accordance with specified CAA regulations."[134]

The plaintiff argued that state-of-the-art evidence was not relevant in a strict liability case.[135] Accordingly, evidence that the aircraft complied with CAA requirements in 1952 had no bearing on whether it was defective in 1970.[136] In the plaintiffs' view, "a showing of a design defective in 1970 establishes that the plane was defective in 1952, the time of the original sale, absent a subsequent alteration of the plane."[137] However, the court rejected this argument, declaring that compliance with CAA regulations in 1952 was evidence that the aircraft design complied with the state-of-the-art at that time.[138] Furthermore, the court concluded that an ordinary consumer would not expect an airplane manufactured in 1952 to necessarily have the safety features of one that was made in 1970.[139] Therefore, the appellate court affirmed the lower court's summary judgment for the defendant.[140]

4. Technological Feasibility.—Many courts agree that technological feasibility is the principal focus of the state-of-the-art concept as it applies to product design.[141] Some state statutes also define state-of-the-art in terms of available technology at the time of manufacture.[142] Technological feasibility

132. *Id.*

133. *Id.* at 446.

134. *Id.* (internal quotation marks omitted).

135. *Id.* at 447.

136. *Id.*

137. *Id.*

138. *Id.* at 446-47.

139. *Id.* at 447.

140. *Id.* at 449.

141. *See, e.g.*, Gosewisch v. Am. Honda Motor Co., 737 P.2d 365, 370 (Ariz. Ct. App. 1985), *opinion vacated in part by* 737 P.2d 376 (Ariz. 1987); Indianapolis Athletic Club, Inc. v. Alco Standard Corp., 709 N.E.2d 1070, 1074 (Ind. Ct. App. 1999); Falada v. Trinity Indus., Inc., 642 N.W.2d 247, 251 (Iowa 2002); AC&S, Inc. v. Asner, 686 A.2d 250, 254 (Md. 1996); Wiska v. St. Stanislaus Soc. Club, Inc., 390 N.E.2d 1133, 1138 n.8 (Mass. App. Ct. 1979); Johnson v. Salem Corp., 477 A.2d 1246, 1251-52 (N.J. 1984); Crispin v. Volkswagenwerk AG, 591 A.2d 966, 973 (N.J. Super. Ct. App. Div. 1991); O'Brien v. Muskin Corp., 463 A.2d 298, 305 (N.J. 1983), *superseded by statute as stated in* Dewey v. R.J. Reynolds Tobacco Co., 577 A.2d 1239 (N.J. 1990); Lancaster Silo & Block Co. v. N. Propane Gas Co., 427 N.Y.S.2d 1009, 1016 n.2 (App. Div. 1980); Boatland of Hous., Inc. v. Bailey, 609 S.W.2d 743, 748 (Tex. 1980); Lenhardt v. Ford Motor Co., 683 P.2d 1097, 1099 (Wash. 1984).

142. *See, e.g.*, ARIZ. REV. STAT. ANN. § 12-681(10) (2012); NEB. REV. STAT. § 25-21,182 (2011).

would seem to involve more than a mere knowledge or understanding of basic scientific principles, although such knowledge or understanding is an essential first step. For example, as Heron of Alexandria’s “Sphere of Aeolus” illustrates, the ancient Greeks understood that steam could be used as a power source to operate machinery.[143] However, they were unable to put this knowledge to practical use because they did not have the technology to construct even the simplest steam engine.[144] Almost two thousand years elapsed before steam engines made an appearance.[145]

Technological feasibility also requires more than an ability to conceptualize in general terms how a particular device may be designed or constructed. For example, Leonardo da Vinci produced a number of clever drawings of helicopters and flying machines in the late fifteenth century.[146] However, without an internal combustion engine or some other lightweight power source, none of Leonardo’s flying machines would have gotten off the ground had he tried to construct one.[147] In other words, manned flight in heavier-than-air machines was not technologically feasible in Leonardo’s time and would not become so for another four hundred years.

Even the construction of working models may not be enough to constitute technological feasibility because scaling up can present serious challenges. Anyone who has watched the popular television series *Mythbusters*[148] has no doubt observed that full-scale devices do not always behave like smaller-scale prototypes. For this reason, at least one court has refused to recognize a small-scale model of a proposed safety device as evidence that a full-scale version is technologically feasible. In *Maxted v. Pacific Car & Foundry Co*.[149] the plaintiff lost control of a tractor trailer rig loaded with logs, causing it to jack-knife and overturn.[150] The plaintiff claimed that the vehicle was negligently designed because the manufacturer failed to equip it with a device that would have

143. *See* DERRY & WILLIAMS, *supra* note 1, at 313 (noting that Heron used a jet of steam to rotate a wheel).

144. *See generally id.* at 312-20 (describing the beginnings of the steam engine).

145. *See id.* at 321.

146. *See id.* at 396.

147. *See id.* (“Leonardo da Vinci’s inquiries and speculations about the problems of flight represent only a renaissance intensification of an interest . . . .”).

148. *See generally Mythbusters*, GOOGLE, http://www.google.com (query “Mythbusters”) (providing a variety of background information on the series). For example, in a recent episode, the Mythbusters, Jamie Hyneman and Adam Savage, scaled up a model of a Newton’s Cradle device in which five metal balls are suspended in a row from a frame. *See Mythbusters: Newton’s Crane Cradle* (Discovery Channel television broadcast Oct. 5, 2011), *available at* http://dsc.discovery.com/videos/mythbusters-newtons-crane-cradle/. The scale model transferred energy from the first ball to the last with ninety-eight percent efficiency. *Id.* However, this efficiency dropped to less than forty percent when the Mythbusters built a twenty foot version of the device with five one-ton balls. *Id.*

149. 527 P.2d 832 (Wyo. 1974).

150. *Id.* at 833.

jettisoned the trailer during an emergency.[151] At trial, the plaintiff's expert witnesses acknowledged that no manufacturer within the trucking industry had ever employed such a breakaway device.[152] Therefore, in order to bolster his testimony, one of the plaintiff's experts submitted a drawing of a proposed breakaway device, along with a small-scale model to show how the device would work.[153] However, the trial court rejected this proffer of evidence and granted summary judgment in favor of the manufacturer on the negligent design count.[154] This was affirmed on appeal, where the court declared that "[t]here was no safer design available at the time this unit was manufactured and there is absolutely no evidence of feasibility or any testing [of an alternative design]."[155]

Of course, a device may be technologically feasible for one use, but not for another. For example, the steam engines of the early eighteenth century were adequate to pump water out of coal mines, but they were too large, heavy and inefficient to be used for transportation purposes. It was not until almost a hundred years later, and after many improvements in steam engine technology that steamships appeared, and another twenty years passed before steam engines that were suitable to transport of passengers and freight on land were developed.[156]

Most courts allow both plaintiffs and defendants to raise the issue of technological feasibility in design defect cases, particularly when the risk-utility test is used.[157] The New Jersey Supreme Court addressed the relationship between technological feasibility and risk-utility analysis in *O'Brien v. Muskin Corp.*[158] Mr. O'Brien was injured when he dove into an above-ground swimming pool manufactured by the defendant.[159] The plaintiff alleged that the bottom of the pool was lined with slippery vinyl material which caused his outstretched hands to slide apart instead of breaking the force of the dive.[160] As a result, he

---

151. *Id.*

152. *Id.* at 834.

153. *Id.*

154. *Id.* at 835-36.

155. *Id.* at 836.

156. *See* DERRY & WILLIAMS, *supra* note 1, at 331-37.

157. *See, e.g.*, Caterpillar Tractor Co. v. Beck, 593 P.2d 871, 887 (Alaska 1979), *superseded by statute as stated in* Smith v. Ingersoll-Rand Co., 14 P.3d 990 (Alaska 2000); Gosewisch v. Am. Honda Motor Co., 737 P.2d 365, 370 (Ariz. Ct. App. 1985), *opinion vacated in part by* 737 P.2d 376 (Ariz. 1987); Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 296-98 (Iowa 1994); Wiska v. St. Stanislaus Soc. Club, Inc., 390 N.E.2d 1133, 1138 (Mass. Ct. App. 1979); Lancaster Silo & Block Co. v. N. Propane Gas Co., 427 N.Y.S.2d 1009, 1016 (App. Div. 1980); Bolm v. Triumph Corp., 422 N.Y.S.2d 969, 974 (App. Div. 1979); Boatland of Hous., Inc. v. Bailey, 609 S.W.2d 743, 749 (Tex. 1980); Cantu v. John Deere Co., 603 P.2d 839, 841 (Wash. Ct. App. 1979).

158. 463 A.2d 298 (N.J. 1983), *superseded by statute as stated in* Dewey v. R.J. Reynolds Tobacco Co., 577 A.2d 1239 (N.J. 1990).

159. *Id.* at 301.

160. *Id.* at 302.

struck his head of the bottom of the pool and suffered severe injuries.[161] The plaintiff brought suit against the manufacturer, contending that vinyl should not have been used to line the pool bottom.[162] The trial court removed the design defect issue from the jury’s consideration and the jury rendered a verdict for the defendant.[163] An intermediate appellate court reversed and the defendant appealed to the New Jersey Supreme Court.[164]

The *O’Brien* court affirmed that risk-utility analysis was appropriate “when the product may function satisfactorily under one set of circumstances, yet because of its design present undue risk of injury to the user in another situation.”[165] After enumerating some of the factors that were relevant to a risk-utility analysis, the court acknowledged that “[b]y implication, risk-utility analysis includes other factors such as the ‘state-of-the-art’ at the time of the manufacture of the product.”[166] The court then defined state-of-the-art as “the existing level of technological expertise and scientific knowledge relevant to a particular industry at the time a product is designed.”[167]

The court observed that state-of-the-art was relevant to both sides of the risk-utility equation.[168] According to the court, the risk side of the equation focuses on product-related risks that the manufacturer knew or should have known about, as well as the adequacy of any warnings that may have been provided.[169] On the other hand, the utility side is concerned with the necessity of the product and the feasibility of alternative designs.[170] At the same time, the court declared that a product could comply with the state-of-the-art but still be considered defective if its overall risks outweighed its utility.[171]

Applying this analysis to the facts of the case, the court held that even though there was no evidence that it was technologically feasible to use some other material to line the bottom of the pool, it was still possible for a jury to conclude that the swimming pool was defective if it determined that the risk of injury outweighed its utility.[172] Consequently, the *O’Brien* court affirmed the intermediate appellate court’s judgment and remanded the case back for a new trial on the design defect claim.[173]

One of the more controversial aspects of the *O’Brien* decision was that it

---

161. *Id.*

162. *Id.* at 302-03.

163. *Id.* at 301-02.

164. *Id.*

165. *Id.* at 304.

166. *Id.* at 304-05 (citing Cepeda v. Cumberland Eng’g Co., 386 A.2d 816 (N.J. 1978), *overruled by* Suter v. San Angelo Foundry & Mach. Co., 406 A.2d 140 (N.J. 1978)).

167. *Id.* at 305 (citing Robb, *supra* note 54, at 1, 4-5 & n.15).

168. *Id.*

169. *Id.*

170. *Id.*

171. *Id.*

172. *Id.* at 305-06.

173. *Id.* at 308.

authorized the jury to hold the manufacturer liable if it concluded that the risks of the pool as designed outweighed its overall utility, even if there was no safer material available to line the bottom of the pool.[174] This approach has largely been rejected.[175] Instead, the majority of courts expect the plaintiff to propose a safer alternative design with risks and utility which can then be compared with those of the original design.[176] In such cases, defendants may claim that a proposed alternative design is not technologically feasible. *Caterpillar Tractor Co. v. Beck*[177] provides a good illustration of this. The decedent in *Beck* was killed in 1973 when his Caterpillar 944 front-end loader rolled over an embankment and crushed him.[178] The decedent's widow contended that her husband would not have been killed if the loader had been equipped with a roll-over protective shield (ROPS).[179] The parties disagreed about whether it would have been feasible for Caterpillar to have installed a ROPS when the front loader was manufactured in 1964.[180] The plaintiff's expert testified that auxiliary manufacturers began selling ROPS as early as 1961, thereby implying that these protective devices were technologically feasible at the time the front-end loader was manufactured.[181]

The defendant, however, argued that these after-market ROPS did not have sufficient structural integrity to protect operators against rollovers.[182] According to the defendant, it began testing ROPS for its own vehicles in 1966 and first installed them on front loaders in 1969.[183] On appeal from a judgment for the plaintiff, the Alaska Supreme Court declared that the jury should consider a number of factors in its risk-utility analysis, including "the mechanical feasibility of a safer alternative design, the financial cost of an improved design, and the adverse consequences to the product and to the consumer that would result from an alternative design."[184]

The feasibility of the plaintiff's alternative design was also an issue in

---

174. *See id.* at 305-06; *see also* James A. Henderson, Jr. & Aaron D. Twerski, *Closing the American Products Liability Frontier: The Rejection of Liability Without Defect*, 66 N.Y.U. L. REV. 1263, 1316 (1991) [hereinafter Henderson & Twerski, *Closing the Frontier*].

175. *See, e.g.*, Jones v. Amazing Prods., Inc., 231 F. Supp. 2d 1228, 1248-51 (N.D. Ga. 2002); Parish v. Jumpking, Inc., 719 N.W.2d 540, 543-45 (Iowa 2006). Indeed, the *O'Brien* case itself was promptly superseded by statute. *See* N.J. REV. STAT. § 2A:58C-3(3) (2012); Henderson & Twerski, *Closing the Frontier*, *supra* note 174, at 1315 & n.195.

176. *See* Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers' Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061, 1079-93 (2009).

177. 593 P.2d 871 (Alaska 1979), *superseded by statute as stated in* Smith v. Ingersoll-Rand Co., 14 P.3d 990 (Alaska 2000).

178. *Id.* at 874-75.

179. *Id.* at 875.

180. *Id.*

181. *Id.*

182. *Id.* at 876.

183. *Id.*

184. *Id.* at 886 (quoting Barker v. Lull Eng'g Co., 573 P.2d 443, 455 (Cal. 1978)).

Murphy v. Chestnut Mountain Lodge, Inc.[185] The plaintiff, who was injured in a skiing accident, argued that the skis provided by the lodge were defective because they did not have an anti-friction device allowing the bindings to release when the skier's legs were twisted during a fall.[186] The plaintiff's expert testified that anti-friction devices were feasible at the time of the accident.[187] However, the defendant's expert claimed that he had tested several of the anti-friction devices in development at the time and found that they were either not feasible or not effective.[188] Upholding a lower court judgment for the defendant, the Illinois Appellate Court held that testimony regarding the feasibility or infeasibility of the plaintiff's proposed safer design was admissible.[189]

5. Commercial Feasibility.—Most definitions of state-of-the-art focus solely on the technological aspects of feasibility and ignore the commercial aspects of technological development.[190] However, new technologies should not be judged solely on whether or not they work, but also whether they find acceptance in the marketplace. *Boatland of Houston, Inc. v. Bailey*[191] is one of the few cases to take commercial feasibility into account when discussing state-of-the-art.[192] In *Boatland*, the decedent was killed while operating a sixteen-foot bass boat manufactured the defendant.[193] The plaintiff was thrown from the boat when it "struck a partially submerged tree stump."[194] The boat circled back and hit the decedent, killing him.[195] The decedent's wife and children sued Boatland, claiming that the boat was defectively designed because it was not equipped with a "kill switch," which would have automatically shut off the engine when Bailey was thrown into the water.[196] A jury verdict for the defendant was reversed by an intermediate appellate court and the plaintiffs appealed to the Texas Supreme Court.[197]

The Texas Supreme Court declared that when the plaintiff claimed that a product was defective because it did not have a particular safety feature, the focus should be on whether the manufacturer had the ability "to provide the feature without greatly increasing the product's cost or impairing usefulness."[198]

---

185. 464 N.E.2d 818 (Ill. App. Ct. 1984).

186. *Id.* at 820-21.

187. *Id.*

188. *Id.* at 821.

189. *Id.* at 823-24.

190. *See* discussion *supra* Part I.B.4.

191. 609 S.W.2d 743 (Tex. 1980).

192. *See* Flock v. Scripto-Tokai Corp., 319 F.3d 231, 239 (5th Cir. 2003); Artis v. Corona Corp. of Japan, 703 A.2d 1214, 1217 (D.C. Cir. 1997); McDaniel v. Trail King Indus., Inc., 248 F. Supp. 2d 749, 757 (N.D. Ill. 2002).

193. *Boatland*, 609 S.W.2d at 745.

194. *Id.*

195. *Id.*

196. *Id.* at 746.

197. *Id.* at 745.

198. *Id.* at 746.

According to the court, the feasibility of a safer alternative could be shown by evidence that it was used by or available to, the industry at the time of manufacture.[199] The plaintiff could also show that a safer alternative design was feasible by providing evidence that the industry had the economic and technological capacity to develop this alternative.[200]

In *Boatland*, the plaintiff's experts stated that the concepts behind kill switches were not new and that homemade kill switches had been used on racing boats for more than thirty years.[201] However, the defendant responded that kill switches for boats were not *commercially* available at the time at the time of the accident.[202] Based on this evidence, the court concluded that while it was technologically possible to fabricate a kill switch at the time of the accident, it was not feasible for the defendant to have installed one on Bailey's boat at the time of sale because they were not available for purchase to the trade at that time.[203] Accordingly, the Texas Supreme Court reversed the intermediate appellate court and affirmed the trial court's judgment for the defendant.[204]

Although the *Boatland* court's holding about commercial feasibility seems correct, it should be noted that the defendant in that case was a retail seller, not a manufacturer, and apparently assembled boats to meet the needs of individual customers.[205] Therefore, unlike large-scale boat manufacturers, Boatland probably could not have developed a kill switch on its own, but instead had no choice but to wait for them to become available commercially.

C. Procedural Effects

One of the most basic procedural issues is whether a court will admit state-of-the-art evidence at all. As discussed earlier, some courts consider state-of-the-art to inject negative principles and as a result, have refused to allow the parties to use such evidence in strict liability cases, even when defined in terms of technological feasibility.[206] Presently, however, the great majority of courts permit either party to submit state-of-the-art evidence to prove that a product either was, or was not, defectively designed.[207] A number of state statutes also

---

199. *Id.*

200. *Id.*

201. *Id.* at 747.

202. *Id.*

203. *Id.* at 749.

204. *Id.* at 750.

205. *Id.* at 752 (Campbell, J., dissenting).

206. *See, e.g.*, Dart v. Wiebe Mfg., Inc., 709 P.2d 876, 881-82 (Ariz. 1985); Matthews v. Stewart Warner Corp., 314 N.E.2d 683, 692 (Ill. App. Ct. 1974); Gelsumino v. E.W. Bliss Co., 295 N.E.2d 110, 113 (Ill. App. Ct. 1973); Cryts v. Ford Motor Co., 571 S.W.2d 683, 689 (Mo. Ct. App. 1978); Carrecter v. Colson Equip. Co., 499 A.2d 326, 329 (Pa. Super. Ct. 1985).

207. *See, e.g.*, Reed v. Tiffin Motor Homes, Inc., 697 F.2d 1192, 1198 (4th Cir. 1982); Caterpillar Tractor Co. v. Beck, 593 P.2d 871, 887 (Alaska 1979), *superseded by statute as stated in* Smith v. Ingersoll-Rand Co., 14 P.3d 990 (Alaska 2000); Gosewisch v. Am. Honda Motor Co.,

provide for the admission of state-of-the-art evidence.[208]

Nevertheless, courts differ on how state-of-the-art evidence may be used. For example, some courts only allow defendants to introduce state-of-the-art evidence to rebut testimony by the plaintiff.[209] Thus, in *Murphy v. Chestnut Mountain Lodge, Inc*., the court ruled that once the plaintiff alleged certain anti-friction devices for skis were available and within state-of-the-art, the defendant would be allowed to rebut this claim by showing that the devices proposed by the plaintiff were ineffective.[210] Other courts seem to put the burden on the defendant to prove that its product complied with state-of-the-art at the time it was manufactured.[211] This is also true of some state statutes.[212]

Placement of the burden of proof, as opposed to the burden of production, seems to depend on whether compliance with state-of-the-art is regarded as an affirmative defense or not. Under the traditional approach, the plaintiff has the burden of proving that the product is defective.[213] If the plaintiff makes a prima facie case, usually by submitting evidence of a safer alternative design, the defendant will usually respond by arguing that the proposed alternative design is "not within the state of the art."[214] Failure on the part of the defendant to effectively respond to the plaintiff's evidence of feasibility will probably result

---

737 P.2d 365, 370 (Ariz. Ct. App. 1985), *opinion vacated in part by* 737 P.2d 376 (Ariz. 1987); Hughes v. Massey-Ferguson, Inc., 522 N.W.2d 294, 296-98 (Iowa 1994); Wiska v. St. Stanislaus Soc. Club, Inc., 390 N.E.2d 1133, 1138 (Mass. App. Ct. 1979); Lancaster Silo & Block Co. v. N. Propane Gas Co., 427 N.Y.S.2d 1009, 1016 (App. Div. 1980); Bolm v. Triumph Corp., 422 N.Y.S.2d 969, 974-75 (App. Div. 1979); *Boatland*, 609 S.W.2d at 749; Cantu v. John Deere Co., 603 P.2d 839, 841 (Wash. Ct. App. 1979).

208. *See* FLA. STAT. ANN. § 768.1257 (West 2012); MICH. COMP. LAWS ANN. § 600.2946 (West 2012), *recognized as unconstitutional by* White v. Smith Kline Beecham Corp., 538 F. Supp. 2d 1023 (W.D. Mich. 2008); S.D. CODIFIED LAWS § 20-9-10.1 (2011); TENN. CODE ANN. § 29-28-105 (West 2012); WASH. REV. CODE ANN. § 7.72.050 (West 2012).

209. *See* Murphy v. Chestnut Mountain Lodge, Inc., 464 N.E.2d 818, 823-24 (Ill. App. Ct. 1984); *Boatland*, 609 S.W.2d at 749; *Cantu*, 603 P.2d at 841.

210. *Murphy*, 464 N.E.2d at 823-24.

211. *See* O'Brien v. Muskin Corp., 463 A.2d 298, 305 (N.J. 1983), *superseded by statute as stated in* Dewey v. R.J. Reynolds Tobacco Co., 577 A.2d 1239 (N.J. 1990). *But see* Fabian v. Minster Mach. Co., 609 A.2d 487, 494 (N.J. Super. Ct. App. Div. 1992) (holding that plaintiff has the burden of proving nonconformity with state-of-the-art).

212. ARIZ. REV. STAT. ANN. § 12-683(1) (2012); IOWA CODE ANN. § 668.12 (West 2012); LA. REV. STAT. ANN. § 9:2800.59(3) (West 2012); MO. ANN. STAT. § 537.764(2) (2012); NEB. REV. STAT. § 25-21,182 (2011); N.H. REV. STAT. ANN. § 507:8-g (2011).

213. *See, e.g.*, Prentis v. Yale Mfg. Co., 365 N.W.2d 176, 181-82 (Mich. 1984) (declaring that "the plaintiff must, in every case, in every jurisdiction, show that the product was defective").

214. *See Boatland*, 609 S.W.2d at 749; *Cantu*, 603 P.2d at 841. In addition, a number of states have enacted statutes that create a state-of-the art affirmative defense, thereby requiring the defendant to prove that a proposed alternative design was not within the state-of-the-art. *See supra* note 212.

in a verdict for the plaintiff.[215]

Furthermore, most courts hold that compliance with state-of-the-art is not a complete defense, but merely a relevant factor for the jury to consider.[216] As far as proof is concerned, an Arizona court in *Gosewisch v. American Honda Motor Co*.[217] identified several forms of proof that a defendant could offer to determine state-of-the-art. *Gosewisch* involved a three-wheeled All Terrain Cycle (ATC), which flipped over and injured the plaintiff.[218] The plaintiff argued the ATC was defectively designed because it was equipped with very low-pressured, collapsible tires, it lacked a mechanical suspension, was inherently unstable because of its high center of gravity, had weak front forks, and its front brake was prone to accidental engagement.[219] The court declared that the manufacturer could show state-of-the-art by offering evidence about the thoroughness of its testing and research prior to manufacture.[220] The manufacturer presented evidence that the ATC model in question was thoroughly tested in the laboratory and with test riders, including novices, who rode the vehicle over "some of the most diverse terrain in the world."[221] The court also pointed out that state-of-the-art can be established by demonstrating the infeasibility of each of the testing and design alternatives proposed by the plaintiff.[222]

In this case, the defendant offered evidence to rebut the plaintiff's claims that it should have used computer simulations to test the safety of its ATC.[223] The defendant responded by showing "that computer simulations could not account for the variables involved in a rider-active vehicle."[224] The plaintiff also maintained that mechanical suspension should have been included in the ATC.[225] However, the defendant's experts tested ATCs with the type of mechanical suspensions suggested by the plaintiffs and concluded that their handling was inferior to the defendant's vehicles.[226] In addition, the front forks in the original design proved to be stronger than the replacement forks proposed by the

---

215. *See* Rucker v. Norfolk & W. Ry. Co., 381 N.E.2d 715, 725 (Ill. App. Ct. 1978), *rev'd on other grounds*, 396 N.E.2d 534 (Ill. 1979).

216. *See, e.g.*, Caterpillar Tractor Co. v. Beck, 593 P.2d 871, 887 (Alaska 1979), *superseded by statute as stated in* Smith v. Ingersoll-Rand Co., 14 P.3d 990 (Alaska 2000); *O'Brien*, 463 A.2d at 305; Crispin v. Volkswagenwerk AG, 591 A.2d 966, 973 (N.J. Super. Ct. App. Div. 1991); Patterson v. Ravens-Metal Prods., Inc., 594 N.E.2d 153, 161-62 (Ohio Ct. App. 1991); Owens-Corning Fiberglass Corp. v. District Court (*Caldwell*), 818 S.W.2d 749, 752 (Tex. 1991).

217. 737 P.2d 365 (Ariz. Ct. App. 1985), *opinion vacated in part by* 737 P.2d 376 (Ariz. 1987).

218. *Id.* at 367.

219. *Id.*

220. *Id.* at 370.

221. *Id.*

222. *Id.*

223. *Id.*

224. *Id.*

225. *Id.*

226. *Id.*

plaintiff.[227] Finally, the defendant provided comparisons of its tires with those suggested by the plaintiff.[228] All of this state-of-the-art evidence persuaded the jury to conclude that the ATC was not defective.[229]

The various applications of the state-of-the-art concept discussed above are confusing to say the least. The best approach to use would seem to be the following: The Plaintiff must show that the product is defective in design by proposing an alternative design that is technologically feasible. The defendant can then try to rebut the plaintiff's claim by showing that the alternative design is not technologically or commercially feasible.

D. Products Liability Restatement

Section 2(b) of the *Products Liability Restatement*, which defines defective design, does not specifically mention state-of-the-art. Commentary in section 2 does discuss the state-of-the-art concept, but its analysis is somewhat unclear. After acknowledging that the term state-of-the-art has a number of different meanings, the comment declares that if the defendant can show the existing design of "its product was the safest in use at the time of sale, it may be difficult for the plaintiff to prove that an alternative design could have been practically adopted."[230] In other words, the quality of the *defendant's* design is offered as proof that the plaintiff's alternative design was not feasible. However, the comment then seems to equate state-of-the-art with industry practice by declaring that "[t]he defendant is thus allowed to introduce evidence with regard to industry practice that bears on whether an alternative design was practicable."[231] Commentary also suggests that industry practice may be relevant to the issue of whether the defendant's failure to adopt a safer alternative design caused the product to be "not reasonably safe."[232] It is also suggested that evidence of industry practice is admissible, but not necessarily dispositive.[233] According to the commentary, if the plaintiff introduces expert testimony that a reasonable alternative design was practical and could have been adopted by the defendant, then the trier of fact may seemingly conclude that the product was defective even though no one in the industry had adopted, or even considered adopting, the alternative design at the time of sale.

Further commentary lists the various factors that the fact-finder may take into account in determining whether a proposed alternative design is reasonable.[234] For the most part, these factors are not weighted in any way and appear to be

---

227. *Id.*
228. *Id.*
229. *Id.* at 367.
230. RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2 cmt. d, at 20 (1998).
231. *Id.*
232. *Id.*
233. *Id.*
234. *Id.* cmt. f.

nothing more than a laundry list, originally formulated by Dean John Wade[235] that courts have taken into account to determine whether a product's utility, as designed, outweighs its risks. In its discussion of these factors, the *Restatement* notes that the relative advantages and disadvantages of the product as designed, and as it could have been designed, may be considered, including the effects of an alternative design on longevity production, maintenance, and repair costs.[236] The listing of these factors seems to suggest that any alternative design proposed by the plaintiff must feasible in order to meet the *Restatement*'s reasonableness requirement.

II. TECHNOLOGICAL DEVELOPMENT

The foregoing discussion concluded that courts generally associate state-of-the-art concept with technological feasibility. The issue of technological feasibility often arises in connection with evidence of a safer alternative design offered by the plaintiff. Although courts generally agree that an alternative design must be technologically feasible in order to be within state-of-the-art, they also point out that an alternative design does not necessarily have to be in existence when the product in question is originally marketed, so long as it was technologically feasible to produce it at that time. However, this approach may be inappropriate when technological development in a particular area is dependent upon developments in other industries. In other words, it is not always correct to assume that industry necessarily controls the pace of technological development. The following discussion will describe the various ways in which technology develops.

A. Linear Development

With the exception of products that are discovered entirely by accident, such as penicillin[237] or vulcanized rubber,[238] most technologies develop over time in either a linear fashion or an interdependent fashion. The linear model usually involves an established technology that is further developed over time without significant interaction with any other technology.[239] Some examples of this model include violins from Brescia and Cremona and mid-nineteenth century clipper ships developed for the China tea trade.

1. Violins.—One example of linear development comes from the golden age of violin making in northern Italy from 1550 to 1750.[240] Violin makers (or luthiers) of that period changed the basic design of the violin only slightly,

---

235. *See* John W. Wade, *On the Nature of Strict Tort Liability for Products*, 44 MISS. L.J. 825, 837-38 (1973).

236. RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2, cmt. f, at 22-25.

237. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 118-19.

238. *Id.* at 46.

239. *See supra* pages 670-71 and accompanying footnotes.

240. *See* William Bartruff, *The History of the Violin*, BARTRUFF, http://www.bartruff.com/history/ (last visited Feb. 7, 2012) (providing a history of early violin makers) [hereinafter *Bartruff*].

although they did experiment with various woods, glues, and varnishes to improve the quality and sound of their instruments.[241] The oldest existing instrument is one of a set of twelve constructed by Andrea Amati for the French King, Charles IX, in 1560.[242] To the casual observer, this violin looks very much like a contemporary instrument. The basic sound of the violin did not change dramatically over the years either.[243] Nevertheless, the luthiers of Brescia and Cremona did make changes in the construction of their instruments.[244] For example, Andrea Amati's sons, Antonio and Girolamo, perfected the shape of the violin's *f*-holes in the early seventeenth century.[245] In the late seventeenth century, Antonio Stradivari began to cut a more distinct bevel and also began to outline the heads of his instruments in black.[246] Later, Stradivari began to make his violins in a larger pattern than before and also began to use a darker, richer varnish to improve the appearance and tone of his instruments.[247] As these examples suggest, violin technology was perfected by Italian craftsmen over the course of two centuries by small incremental changes without any significant technological innovation.

2. Clipper Ships.—Another example of linear technological development is the evolution of the clipper ship. In 1833, the East India Company lost its monopoly over the China tea trade, thereby creating an opportunity for the ships of other countries to transport tea from China to Great Britain.[248] Since the tea-drinking public of that time believed that tea was better when it was fresh, speed became a greater consideration than it had been previously.[249] The design for tea

---

241. *See id.*

242. *Id.*

243. For example, it is interesting to note that violin virtuoso, Ole Bull, used an instrument built by Gasparo da Salo in 1574 in thousands of concerts between 1840 and 1880. *See History of the Violin*, CLASSIC VIOLINSHOP.COM, http://classicviolinshop.com/ArticleView.asp?Article ID=1 (last visited Mar. 9, 2012). The sound of Bull's sixteenth century instrument apparently did not sound out-of-place to the ears of his nineteenth century audience. *See id.* (noting the popularity of da Salo's violin).

244. *See Bartruff*, *supra* note 240.

245. *Amati Family Violins*, THE OLD VIOLIN, http://www.collectibleviolins.com/amati.html (last visited Mar. 9, 2012).

246. Marshall C. St. John, *Chapter Two, Stradivari's Violins*, ANTONIO STRADIVARI, http://www.cello.org/heaven/hill/two.htm (last visited Mar. 9, 2012).

247. *See id.*

248. *See* DERRY & WILLIAMS, *supra* note 1, at 367.

249. *See id.* Prior to the mid-nineteenth century, tea and other commodities from the Far East were transported by East Indiamen. These well-armed sailing ships averaged only five or six knots an hour. *See* STEPHEN TAYLOR, STORM AND CONQUEST: THE CLASH OF EMPIRES IN THE EASTERN SEAS, 1809, at 58 (2007). Furthermore, they often "snugged down" by reducing sail during the night. DERRY & WILLIAMS, *supra* note 1, at 367; *see also* BRIAN GARDNER, THE EAST INDIA COMPANY: A HISTORY 98 (1971). As a result, it took almost two years to complete a voyage from the Far East to Great Britain. *See* DERRY & WILLIAMS, *supra* note 1, at 365 (noting ships "would complete three voyages in six years").

clippers may have originated in two-masted schooners developed in the United States for privateering during the War of 1812.[250] In any event, by the mid-nineteenth century, clipper ships were much larger—exceeding 2000 tons and were built to get the most out of the light winds of the China seas.[251] The Ariel, which sailed from China to London in ninety-nine days, provides a good example of mid-nineteenth century clipper ship design.[252] The ship had three masts with four sails on the foremast, five sails on the mainmast, and four sails on the mizzenmast.[253] It was 197 feet long with a beam of about thirty-four feet.[254] It was "built of teak planking laid over iron frames" and relied on iron masts to support the pressure exerted on its sails.[255] The Ariel averaged fourteen knots per hour at full sail.[256] During the heyday of tea clippers, American and English shipbuilders competed with each other to build the fastest ships.[257] However, they did not change the clipper's basic design and most of their improvements were subtle and barely noticeable.[258]

B. The Ping-Pong Model

Sometimes two technologies develop in tandem, with one reacting to the other. For example, this sort of interaction occurred in the sixteenth century between warships and naval artillery.[259] At the beginning of the sixteenth century, the carrack represented the ultimate in shipbuilding technology.[260] The carrack was about 100 feet long with a beam of forty feet and a displacement of 600 to 800 tons.[261] It was a three-masted, fully rigged vessel with a foremast and

---

250. *See* DERRY & WILLIAMS, *supra* note 1, at 367.

251. *Id.* at 368.

252. *See id.*

253. *Id.* fig.172.

254. *Id.*

255. *Id.* at 368-69.

256. *Id.* at 368.

257. *See id.* at 364-70.

258. *See id.*

259. *See* ANGUS KONSTAM, SOVEREIGNS OF THE SEA: THE QUEST TO BUILD THE PERFECT RENAISSANCE BATTLESHIP 128 (2008).

260. Another vessel, popular at this time, was the caravel. Of Islamic origin, the caravel was used by Portuguese explorers in the fifteenth century to map the west coast of Africa. *See* ARCHIBALD R. LEWIS & TIMOTHY J. RUNYAN, EUROPEAN NAVAL AND MARITIME HISTORY, 300-1500, at 159 (1985). Carvel-built, the caravel usually had two masts that were fitted with lateen sails. *Id.* The ship was small, but maneuverable and easy to operate. *Id.* Two of the ships that accompanied Columbus on his first voyage to America, the Niña and the Pinta, were caravels of fifty to sixty tons displacement. *Id.* The larger Santa Maria was probably a nao, a larger vessel, also of Portuguese origin, that was somewhat similar to a carrack. *See* KONSTAM, *supra* note 259, at 176. However, it should be noted that some historians believe that the Santa Maria was actually a carrack. *See, e.g.*, LEWIS & RUNYAN, *supra*, at 159.

261. *See* KONSTAM, *supra* note 259, at 38. Carracks could be either clinker-built or carvel-

mainmast, fitted with square mainsails and topsails, and a mizzenmast, which carried a triangular lateen sail.[262] Carracks typically also had a bowsprit sail on front[263] and they were known for their very high forecastle and sterncastle structures.[264]

Their roomy hulls and wide beams made carracks excellent cargo vessels, but they were also used as warships.[265] In the fifteenth and early sixteenth century, carrack warships were usually equipped with a few breech-loading iron cannons, along with a much larger number of smaller weapons mounted on the forecastle and sterncastle decks.[266] The heavy guns of this period were not very powerful because their principal function was to assist in boarding enemy ships by disabling their rigging rather than trying to sink them.[267] Heavy cannons of the day were made of wrought iron bars welded together into a tube and reinforced with iron hoops.[268] They were breech loading because it was not possible at that time to fabricate muzzle loading cannons without casting them.[269] These iron cannons generally fired stone cannonballs and had an effective range of only seventy-five yards.[270]

Several developments of significance occurred in the early sixteenth century that eventually led to the replacement of the carrack by its more effective successor, the galleon. The first innovation was the gunport.[271] This involved the cutting of ports or openings (with waterproof covers) into the hull of the ship, as opposed to the superstructure, so that heavy cannons could be placed on the main deck.[272] This not only allowed ship designers to add more guns, it also enabled them to use heavier artillery pieces without adversely affecting the vessel's stability.[273] During the same time period, two other developments greatly increased the firepower of large guns. First, ship designers began to phase out breech-loading wrought iron cannons and replaced them with muzzle-loading cast bronze guns.[274] These weapons had a range of 400 yards and could

---

built. *See id.* at 40.

262. *Id.* at 69. Some carracks had a fourth mast, known as a bonaventure mast, which also carried a lateen sail. *Id.*

263. *See* LEWIS & RUNYAN, *supra* note 260, at 158.

264. *See* KONSTAM, *supra* note 259, at 38.

265. *See* JAMES BURKE, CONNECTIONS 189 (1978) [hereinafter BURKE, CONNECTIONS].

266. For example, when Henry VII's warship, Sovereign, was launched in 1488, she was equipped with thirty-two heavy wrought iron cannons and 110 swivel guns designed for close range anti-personnel use. *See* KONSTAM, *supra* note 259, at 41.

267. *Id.* at 43-44.

268. *See* CARLO M. CIPOLLA, GUNS, SAILS AND EMPIRES: TECHNOLOGICAL INNOVATION AND THE EARLY PHASES OF EUROPEAN EXPANSION 1400-1700, at 23 (1965).

269. *Id.* at 24 n.1.

270. *See* KONSTAM, *supra* note 259, at 41.

271. *See* JEREMY BLACK, TOOLS OF WAR 72 (2007).

272. *Id.*

273. *See* CIPOLLA, *supra* note 268, at 82.

274. *See* BLACK, *supra* note 271, at 90.

fire up to a sixty pound shot.[275] Firepower was also increased by the replacement of stone shot with iron cannonballs, which had much greater penetrating power.[276]

However, even with these improvements in ship design and armament, the carrack had serious deficiencies as a warship. First of all, the improvements in gunnery made carracks easy targets because of their high profile.[277] Moreover, their height also made them top-heavy and vulnerable to rolling, which reduced the accuracy of their gunfire.[278] Finally, carracks, with their high superstructures, were difficult to handle "as the wind tended to push them to leeward as they sailed," making it difficult to keep a straight course.[279]

In the 1530s, Spanish and Portuguese shipbuilders searched for a design that would combine "the speed and maneuverability of the [caravel] with the cargo capacity of the" carrack.[280] The result was that icon of pirate movies, the Spanish galleon. Unlike the carrack, the galleon was purpose-built as a warship, primarily to protect Spanish treasure ships returning to Europe from the Americas.[281] Galleons were narrower than carracks, with a typical length-to-beam ratio of 4-to-1, as compared with the 3.5-to-1 ratio for carracks.[282] This made for greater speed and better handling.[283] The earliest galleons were rather small, about 350 tons, but gradually increased in size during the course of the sixteenth century.[284]

The next development in ship design occurred in England (which was not yet Great Britain) in 1570 when a master shipwright, Richard Chapman, teamed up with an ex-pirate and slave trader, Sir John Hawkins, to produce the first race-built galleon, a 300-ton vessel called the Foresight.[285] The ship had a sleek hull,

---

275. *See* KONSTAM, *supra* note 259, at 43. These bronze cannons came in various sizes. For example, a curtow could fire a sixty pound ball; a demi-curtow could fire a thirty pound ball; and a culverin could fire a twenty pound shot. *Id.*

276. *See* BLACK, *supra* note 271, at 90.

277. KONSTAM, *supra* note 259, at 175.

278. *Id.*

279. *Id.*

280. *Id.* at 176.

281. *Id.* at 180. The need for protection was demonstrated by an encounter off the southwest coast of Portugal in 1523. Three Spanish caravels, sailing without an escort and loaded with treasure from South America, were overhauled and captured by French privateers, resulting in a severe financial loss to the Spanish Crown. *Id.* at 177-78.

282. *Id.* at 181.

283. *See id.*

284. *Id.* at 181-82. Portuguese galleons tended to be larger, sometimes reaching 1000 tons displacement. *Id.* at 182.

285. *Id.* at 188. In 1568, while commanding an ancient carrack, the Jesus of Lubeck, Hawkins, accompanied by his young cousin, Francis Drake, had an opportunity to compare first hand the fighting qualities of these two kinds of ships and barely escaped with his life. *See* ERNLE BRADFORD, DRAKE 45-59 (1965). Many crewmembers were forced to surrender to the Spanish after the ship was destroyed. *Id.* at 57. One of the scenes from Errol Flynn's 1940 film, *The Sea*

with a relatively deep draft, a low superstructure, a long beak in front, and a narrow stern. The foremast was placed farther forward than normal and slanted forward slightly, which improved the vessel's handling.[286] Heavily armed for its size, the Foresight carried twenty-eight heavy bronze muzzle-loading cannons arrayed on one continuous gundeck.[287] Faster, more maneuverable, and more heavily armed than Spanish warships, race-built galleons proved their worth when war finally broke out between England and Spain in 1588.[288]

C. The Pinball Model

Technological development often follows a path where changes are interdependently linked. In other words, technological change does not occur in a vacuum, but relies on developments in other fields.[289] Three historical examples help to illustrate this process: the printing press, the railroad, and the automobile.

1. Printing.—The development of printing required innovations in at least four areas, the press, paper, oil-based ink, and movable metal type.[290] The printing press itself was easily adapted from the linen and paper presses of the time.[291] However, another innovation, paper, was also necessary to the development of printing. Prior to the introduction of paper in Europe, the only material available was parchment, which was made from the skins of sheep. A 200 quarto-page codex would require the skins of twelve sheep, which was very expensive and also hard on the sheep.[292] Of Chinese origin, paper was introduced into Spain by the Moors in the twelfth century, and had spread to the rest of Europe by the fourteenth century.[293] High-quality paper could be made from discarded linen. Fortunately, most of the population was able to afford linen undergarments by the fourteenth century because the Black Death had caused wages to rise sharply.[294] This meant that worn out linen clothing could be collected by itinerant "rag-and-bone" men and sold to paper mills.[295] Paper,

---

Hawk, was loosely based on this incident. For general information regarding *The Sea Hawk*, see *The Sea Hawk*, IMDB, http://www.imdb.com/title/tt0033028/ (last visited May 1, 2012).

286. KONSTAM, *supra* note 259, at 188-89.

287. *Id.* at 189.

288. *See id.* at 198-203.

289. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 36.

290. DIANA CHILDRESS, JOHANNES GUTENBERG AND THE PRINTING PRESS 60 (2008) (describing Gutenberg's contributions to the printing press).

291. *See* DERRY & WILLIAMS, *supra* note 1, at 236.

292. *See id.* at 232. In ancient times, sheets of parchment were sewn together to produce lengthy documents, which were then rolled up for storage. *Id.* Around the second century A.D. scribes began to fold rectangular sheets of parchment into pages and then bound them together to form a codex or book. *Id.*

293. *Id.* at 232-33.

294. *See* BURKE, CONNECTIONS, *supra* note 265, at 100.

295. *See id.*

produced by water-powered mills, was much cheaper than parchment.[296]

Another piece of the puzzle was the invention of movable metal type by Johann Gutenberg in the mid-fifteenth century. Prior to that time, wood blocks had been used in Europe to print playing cards, calendars and prayers.[297] However, wood blocks were expensive to carve and wore out quickly.[298] Gutenberg's solution was to manufacture a movable type that was durable, that would be uniform in size and that would lie side-by-side in a holder so as to produce an even line of print.[299] Furthermore, movable type could be used repeatedly, and when a letter wore out, a new one could be reproduced in a mold.[300] Interchangeability of type required standardization of the size and shape of the letters, greatly enhancing readability.[301] Not surprisingly, printing spread rapidly throughout Europe during the late fifteenth century.[302]

The final requirement for the commercialization of printing was the development of a suitable ink. At first printers used a water-based ink similar to that used by scribes to copy manuscripts.[303] However, because of its low viscosity, this type of ink made a poor impression on the absorbent paper that was required for printing.[304] As a result, the impression was often smudged and showed through the opposite side of the page.[305] Fortunately, these problems were eventually solved with the introduction of an oil-based ink which was made by combining linseed oil with lampblack or powdered charcoal.[306]

2. Railroads.—The development of the railroad was made possible, first and foremost, by the invention of the steam engine. Originally developed to pump water from coal mines, this versatile device was adapted for many uses in the nineteenth century.[307] However, the steam engine took a long time to develop and many other technologies were essential to its use as a power source for railroads. The first requirement was an understanding of the principles of atmospheric pressure and vacuums.[308] The second was to apply these principles to a piston and cylinder mechanism in order to produce power.[309] The third

---

296. For example, as early as 1300, paper sold for one-sixth of the price of parchment. *Id.* at 101.

297. *See id.*

298. *Id.*

299. *Id.* at 102. To create a letter of type, a steel punch was used to hammer the letter's impression into a copper matrix. *Id.* A lead alloy was poured into a mold in order to produce a piece of typeface that was raised on a shoulder and stalk of uniform height. *Id.*

300. *See* DERRY & WILLIAMS, *supra* note 1, at 237.

301. *See id.*

302. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 274.

303. *See* DERRY & WILLIAMS, *supra* note 1, at 235.

304. *Id.* at 235-36.

305. *Id.*

306. *Id.* at 236.

307. *See id.* at 312.

308. *Id.* at 323-25.

309. *Id.* at 324-25.

breakthrough was the lathe-type boring machine and other machine tools (borrowed from clockmakers) that were necessary to fabricate steam engine parts to the precise specifications necessary to generate power efficiently.[310] The last requirement was the sun-and-planet gear, developed by James Watt, that converted the up-and-down motion of the piston to a rotary motion that could be used to drive wheels.[311]

The story begins in the mid-seventeenth century when scientists discovered that a vacuum could be produced when air pressure was allowed to drop in an enclosed space.[312] This discovery was prompted by the fact that European silver and iron mines were being flooded because the suction pumps of the day could not lift water more than thirty-two feet.[313] Subsequent experiments by Otto von Guericke suggested that if a device could be made that would repeatedly create a vacuum, atmospheric pressure that could be used to operate a pump.[314] Some years later, Denis Papin put this principle to work by constructing a machine which used steam power to move a piston up and down in a cylinder.[315] In 1712, Thomas Newcomen applied this piston and cylinder concept to construct a steam powered pump that was installed at a coal mine in Straffordshire, England.[316] Finally, in 1765, James Watt greatly improved the efficiency of the Newcomen steam engine by adding a condenser so that water did not have to be sprayed directly into the cylinder to condense the steam.[317]

The steam engine might have been limited to pumping water out of coal mines if it had not been for two other developments in the eighteenth century. The first was the invention of the sun-and-planet gear by James Watt in 1784.[318] This device allowed vertical motion of the piston and cylinder to be converted into a rotary motion, enabling steam engines to turn wheels carrying belts and

---

310. *Id.* at 322.

311. *Id.* at 323.

312. *See* BURKE, CONNECTIONS, *supra* note 265, at 75.

313. *See id.* at 72-73. Experiments by Evangelista Torricelli and Gasparo Berti in the 1640s established the existence of atmospheric pressure and showed that the weight of the air pressing down on a pool of water at the foot of a mineshaft prevented the water from rising more than thirty-two feet. BURKE, PINBALL EFFECT, *supra* note 10, at 133. Then in 1648, a group of French scientists climbed a mountain with a tube of mercury suspended in a dish of mercury. *See id.* at 134. As they climbed higher, the mercury level in the tube fell, indicating that atmospheric pressure was lessening. *Id.* This was expected; however, the scientists were surprised to find a space between the mercury and the top of the tube. *Id.* Since air could not enter the tube because it was blocked by the mercury, they concluded that the mysterious space in the tube could only be a vacuum. *Id.*

314. *See* DERRY & WILLIAMS, *supra* note 1, at 314.

315. *Id.* at 315.

316. CHRISTOPHER MCGOWAN, RAIL, STEAM AND SPEED: THE “ROCKET” AND THE BIRTH OF STEAM LOCOMOTION 39-40 (2004).

317. DERRY & WILLIAMS, *supra* note 1, at 320-21.

318. *Id.* at 323-24.

thereby operate machines in mills, breweries, and ironworks.[319] The second major innovation was John Wilkinson's lathe-type boring machine, originally developed in 1774 to bore bronze cannons.[320] The cylinders in Watt's 1765 versions of steam engines needed to be constructed with great precision. However, this was not possible until 1775, when Wilkinson adapted his machine to bore cylinders for Watt's steam engines.[321] This set the stage for the development of railroads, steamships and steam-powered farm machinery in the nineteenth century.

The origins of the railroad can be traced to Richard Trevithick, who built a "steam carriage" known as Captain Dick's Puffer in 1801.[322] Unfortunately, the vehicle exploded four days after its debut because Trevithick forgot to turn off the boiler.[323] In 1803, he built another steam carriage for use in London, but it was unable to cope with the poor roads of the time.[324] This caused Trevithick to turn his attention to rail travel.[325] The next year, he constructed a locomotive which transported a ten-ton load of cast iron from the Pen-y-darren Iron Works to the Glamorganshire Canal nine miles away at a speed of five miles per hour.[326] Trevithick's last project was the 1808 steam locomotive Catch-me-who-can which briefly carried passengers along a circular track in London's Euston Square.[327]

Meanwhile, responding to the high cost of horse fodder caused by the Napoleonic wars, coal mine owners and textile manufacturers began to envision the railroad as a cheaper alternative to horse-drawn transportation.[328] The first public railroad was the Stockton & Darlington Railway Company, founded in 1824 to transport coal from the Durham coal mines to the wharves of Stockton on the River Tees, about thirty miles away.[329] Several years later, the Liverpool & Manchester Railway was formed to ship cotton goods from Manchester to the

---

319. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 24.

320. *Id.* at 250.

321. DERRY & WILLIAMS, *supra* note 1, at 350. Unlike other steam engines, Watt's did not have a water seal on top of the piston because it would have cooled the cylinder down and reduced the steam engine's power; however, this meant that the cylinder and piston had to fit perfectly in order to prevent steam from escaping. *See* BURKE, CONNECTIONS, *supra* note 265, at 175.

322. *See* Richard Cavendish, *Richard Trevithick's First Steam Carriage*, 51 HISTORY TODAY (2001), *available at* www.historytoday.com/richard-cavendish/richard-trevithicks-first-steam-carriage; *see also* MCGOWAN, *supra* note 316, at 49-50. In 1769, Nicholas Cugnot, a French military officer, constructed a steam-powered wagon to haul field artillery. *Id.* at 50. It worked, but was so slow and inefficient that it was soon abandoned. *See id.*

323. *Id.*

324. *Id.* at 52.

325. *Id.*

326. *Id.* at 57-58.

327. *Id.*

328. *See* MCGOWAN, *supra* note 316, at 86.

329. *See id.* at 2-3.

seaport of Liverpool.[330] Liverpool and Manchester held a series of trials at the village of Rainhill, near Liverpool, in October 1829 to determine who would supply them with locomotives.[331] The winner of this competition was a locomotive named Rocket, built by George Stephenson.[332] The secret of Rocket's success was a system of copper tubes that Stephenson designed to carry hot water from the boiler through the hot gases escaping up the chimney from the firebox.[333] The tubes provided a much greater heating surface and consequently generated more high pressure steam to drive the engine's cylinders than the competition.[334] The Liverpool & Manchester Railway began operation on September 15, 1830 with the Duke of Wellington as its reluctant guest of honor.[335]

Although the steam engine was critical to the development of the railroad, there were other elements as well. For example, developments in track construction greatly aided the development of the railroad. At first, railroads used cast iron rails, which had been developed for use in coal mines during the eighteenth century.[336] However, cast iron rails were brittle and did not hold up well, leading to their replacement by wrought iron rails in 1830.[337] Eventually, iron rails were replaced by steel rails.[338] In addition to rail construction, the invention of the telegraph made it possible to safely run trains in opposite directions along the same track.[339] Contemporaneous advances in bridge-building, such as the fabrication of iron and steel bridges, and developments in tunnel construction techniques also contributed to the commercial success of railroads in the nineteenth century.[340]

3. Motor Vehicles.—Like the printing press and the railroad, a number of technological developments had to occur before motor vehicles could be successfully introduced into the stream of commerce. Among these developments were the internal combustion engine, the carburetor, the spark plug, pneumatic rubber tires and gasoline fuel.

The motor vehicle's most important feature is the internal combustion engine. Unlike a steam engine, which burns fuel externally in a firebox, an internal combustion engine burns fuel inside its cylinders.[341] First, a mixture of

---

330. *See id.* at 4.

331. *See id.* at 6-7, 221.

332. *Id.* at 221-22.

333. *Id.* at 181-82.

334. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 70.

335. *See* MCGOWAN, *supra* note 316, at 73, 244-45. The Duke was not fond of rail travel. *See id.* at 73 ("[H]e abhorred the very idea of locomotives.").

336. *See* DERRY & WILLIAMS, *supra* note 1, at 377-78.

337. *Id.* at 378.

338. *Id.* at 381.

339. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 29.

340. *See* DERRY & WILLIAMS, *supra* note 1, at 448-62.

341. *Id.* at 602.

gasoline and air is sprayed into the cylinder and compressed by a piston.[342] At the point of maximum compression, an electrical spark ignites the fuel.[343] The explosion produces hot gases which push the piston down.[344] These combustion gases are then vented and more of the fuel-air mixture is injected to run a second stroke.[345] Valves control the intake of the fuel-air mixture and vent exhaust gases at the appropriate time in the cycle.[346]

The basic concept of the internal combustion engine was not new at the turn of the nineteenth century. As early as the seventeenth century, Denis Papin (mentioned earlier in connection with the development of the steam engine) experimented with a device to produce power from exploding gunpowder.[347] It was not much of a success.[348] In 1859, Étienne Lenoir invented a stationary internal combustion engine that used an electric spark to ignite a mixture of coal gas and air.[349] Finally, in 1876, Nikolaus Otto patented a four-stroke engine which he modestly called the Otto Cycle Engine.[350] This engine, with subsequent improvements developed by Gottlieb Daimler and Wilhelm Maybach, became the standard power source for gasoline-powered motor vehicles in the twentieth century.[351]

However, a number of other inventions also contributed to the successful operation of motor vehicles. One such device was the carburetor, invented by Maybach in 1893, used to spray the right amount of air and fuel into the cylinder.[352] This gadget was based on mid-nineteenth century perfume atomizers, which used a drop in air pressure to convert a liquid into a fine mist.[353] This phenomenon is known as the Venturi Effect.[354] Another invention was the spark plug, which can be traced back to Alessandro Volta's eudiometric pistol.[355] This device, developed in 1776, consisted of a glass pistol-like container filled with gas and corked at one end.[356] A spark resulted inside the pistol when Volta touched one of two electrically charged wires, inserted into the pistol, while making contact with the other wire and the pistol's electrophore lid.[357] Volta unknowingly made another contribution to the development of the automobile by

---

342. *Id.*
343. *Id.*
344. *Id.*
345. *Id.*
346. *Id.*
347. *Id.* at 600.
348. *See id.*
349. *Id.* at 601-02.
350. *Id.* at 602-03.
351. *Id.* at 605.
352. BURKE, PINBALL EFFECT, *supra* note 10, at 109-10.
353. *See* BURKE, CONNECTIONS, *supra* note 265, at 181-82.
354. *Id.*
355. *See id.* at 178-79.
356. *Id.*
357. *Id.* at 178.

inventing a chemical fuel cell, the prototype of the rechargeable car battery.[358]

Another vital component to the development of the motor vehicle was the pneumatic rubber tire. In 1839, Charles Goodyear discovered the process of vulcanization by which natural rubber was toughened by being heated in a mixture of sulfur and white lead.[359] In 1846, Thomas Hancock began manufacturing solid rubber road tires for horse-drawn carriages.[360] Pneumatic rubber tires also appeared at this time, but were not a commercial success.[361] In 1888, J.B. Dunlop, an Irish surgeon, developed pneumatic rubber tires for bicycles.[362] Dunlop's invention came along at just the right time to serve the needs of the newly-developed automobile.[363]

The final requirement was to find a suitable fuel to power the automobile's internal combustion engine. The solution was a petroleum product called gasoline. Oil was first discovered by Edwin Drake in Pennsylvania in 1859.[364] It could be distilled into various components for use in heating, lighting or lubrication.[365] At first, the lighter, more volatile products of the distillation process were thrown away as waste.[366] However, toward the end of the nineteenth century, this waste product, now known as gasoline, was found to be the perfect fuel to power the newly-developed automobile.[367] More than a century later, gasoline-powered motor vehicles remain the principal means of transportation in the world.

4. Modern Technologies.—The interdependent model of technological development is not limited to the nineteenth century and earlier; it is also applicable to more modern inventions such personal computers, cellular telephones, and global positioning systems.

Various technologies have contributed to the development of the personal computer. For example, computers require one or more Central Processing Units (CPU) or microprocessors in order to execute software instructions.[368] Software includes application software which is used to carry out tasks such as word processing, sending and receiving e-mail, internet browsing, faxing, and playing computer games. In addition, computers also require system software, which interfaces with hardware to support application software.[369] All of this software

---

358. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 146-47.

359. *Id.* at 46.

360. *See* DERRY & WILLIAMS, *supra* note 1, at 528-29.

361. *Id.* at 529.

362. *Id.* at 392.

363. *Id.*

364. BURKE, CONNECTIONS, *supra* note 265, at 179.

365. *Id.* at 179-80.

366. *Id.* at 180.

367. *See id.*

368. *The Central Processing Unit*, COMPUTER SPECIALIST, http://www.ispcp.org/the-central-processing-unit.html (last visited May 2, 2012).

369. *See How Computers Work—Software*, STAN. U., http://www.stanford.edu/class/cs101/how-computers-work-software.html (last visited May 3, 2012) (discussing programs and operating

relies on a language system of ones and zeroes called binary code. Binary code was derived from office tabulators and calculators, which in turn adapted it from the Jacquard loom.[370] The computer keyboard is descended from the typewriter keyboard.[371] The monitor was originally based on cathode ray tube technology, which was also used in television and medical monitoring machines.[372]

Mobile telephones, especially modern smart phones, also incorporate many distinct technologies. The most important of these technologies is voice radio transmission, which originated with ship-to-shore transmissions by Reginald Fessenden in the early twentieth century.[373] All mobile telephones are powered by a rechargeable battery, which can be traced back to Alessandro Volta's experiments with chemical batteries in the late eighteenth century.[374] Mobile telephones also have a CPU similar to that of a computer to run all of the telephone's software.[375] Additionally, modern smart phones rely on many other technological innovations to play music, take photographs, download video and audio data, send and receive text messages and email, and access the Internet.[376] Obviously, if these technologies had not already been in place when mobile telephones came on the scene, the mobile telephone that emerged would look very different from today's version.

The global positioning system (GPS), maintained by the United States Government, provides highly accurate location and time information for airplanes, cars, boats, submarines and even pedestrians.[377] The system was developed in 1973 and became fully operational in 1994.[378] The technologies

---

systems).

370. *See* BURKE, CONNECTIONS, *supra* note 265, at 108-13.

371. *See* TYPING THROUGH TIME: KEYBOARD HISTORY, DASKEYBOARD, http://www.daskeyboard.com/blog/?page_id=1329 (last visited May 3, 2012) (discussing the transition from typewriters to modern computer keyboards).

372. *See Cathode Ray Tube (CRT) Monitor*, HIST. OF COMPUTERS, http://history-computer.com/ModernComputer/Basis/cathode.html (last visited May 3, 2012). Modern computer screens are now generally based on liquid crystal display technology. *See Liquid Crystal Display (LCD)*, HIST. OF COMPUTERS, http://history-computer.com/ModernComputer/Basis/lcd.html (last visited May 3, 2012).

373. Courtney Livingston Quale, Note, *Hear an [Expletive], There an [Expletive], But[t] . . . the Federal Communications Commission Will Not Let You Say an [Expletive]*, 45 WILLAMETTE L. REV. 207, 214 (2008).

374. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 146-47.

375. *See* Bonnie Cha, *Smartphones Unlocked: Understanding Processors*, CNET (Aug. 8, 2011, 12:00 AM), http://www.cnet.com/8301-17918_1-20088704-85/smartphones-unlocked-understanding-processors/ (describing the various processors used in cell phone technology).

376. *See Smartphone Features*, PC MAG., http://www.pcmag.com/encyclopedia_term/0,1237,t=smartphone+features&i=6433,00.asp (last visited June 21, 2012).

377. *See* NAT'L RESEARCH COUNCIL, THE GLOBAL POSITIONING SYSTEM: A SHARED NATIONAL ASSET: RECOMMENDATIONS FOR TECHNICAL IMPROVEMENTS AND ENHANCEMENTS 13-14 (3d ed. 1997) (describing the government's role in the creation of GPS systems).

378. *Id.* at 13-14, 16.

that were needed to create the GPS system included those associated with satellites, rockets, radios, radar, semiconductors, solar batteries, computers and atomic clocks.[379]

III. Commercial Feasibility

Although the issue has not arisen very often, it would seem that the state-of-the-art concept should include commercial as well as technological feasibility. Some technological innovations achieve immediate commercial success. For example, printing was quickly accepted by literate consumers in fifteenth century Europe. By 1480, printing presses were located in at least 110 towns.[380] By 1500, there were more than 35,000 editions and 20 million individual books in print.[381] Railroads were also an immediate success, both in Great Britain and in the United States.[382] For example, in its first full year of operation, the Liverpool & Manchester Railway carried about 445,000 passengers and 43,000 tons of freight.[383] By 1860, railroad companies had laid down nearly 15,000 miles of track in Great Britain and 30,000 miles of track in the United States.[384] In more recent times, personal computers and cell phones also achieved commercial success relatively quickly. For example, personal computer sales went from 48,000 in 1977[385] to more than 300 million in 2008.[386] Mobile telephones were even more successful as cell phone subscribers increased worldwide from 12.4 million in 1990 to more than 4.6 billion in 2010.[387]

---

379. *See generally Real-World Relativity: The GPS Navigation System*, OHIO ST. U. (Apr. 27, 2009), http://www.astronomy.ohio-state-edu/~pogge/Ast162/Unit5/gps.html (providing background on GPS technology).

380. *See* BURKE, CONNECTIONS, *supra* note 265, at 105.

381. *See* BURKE, PINBALL EFFECT, *supra* note 10, at 274. Book sales were boosted by the practice, pioneered by the Venetian printer, Aldus Manutius, of publishing inexpensive octavo-sized editions that could be easily carried in a customer’s saddlebag. *See id.*

382. Construction began on the first railroad in America, the Baltimore & Ohio Railroad in 1826. *See* Susan P. Fino, *A Cure Worse Than the Disease? Taxation and Finance Provisions in State Constitutions*, 34 RUTGERS L.J. 959, 965-67 (2003); Jerry L. Mashaw, *Administration and “The Democracy”: Administrative Law from Jackson to Lincoln, 1829-1861*, 117 YALE L.J. 1568, 1570 n.4 (2008).

383. *See* MCGOWAN, *supra* note 316, at 265.

384. *See* DERRY & WILLIAMS, *supra* note 1, at 302 (“The length of the railways [in the United States] grew in that time from 10,000 to 30,000 miles, which was more than twice the entire network of the United Kingdom.”).

385. *Disruptive Technology Stocks for Smart Small-Cap Investors*, MONEY MORNING (Apr. 6, 2012), http://www.moneymorning.com.au/20120406/disruptive-technology-stocks-for-smart-small-cap-investors.html.

386. *See Worldwide PC Shipments in 2008*, ZDNET (Mar. 16, 2009, 4:56 AM), http://www.zdnet.com/blog/itfacts/worldwide-pc-shipments-in-2008/15672 (listing total 2008 shipments at 302,207,500).

387. *The Birth of the Mobile Phone*, CISCO COMMUNITIES (Jan. 12, 2012, 10:52 AM),

A. Catalysts to the Commercialization of Technology

There are a number of events or conditions that can create a receptive environment for new products and technologies, including sudden changes in the physical environment, depletion of natural resources, military competition among nations, popular dissatisfaction with the state of existing technology, as well as changing demographic and social conditions. Many of these changes are beyond a manufacturer's control.

1. Environmental Changes.—Over the centuries, changes in the physical environment have often provided an economic incentive for the introduction of new products and technologies. For example, in the fourteenth century a drop in temperature caused by the Little Ice Age led to the invention of fireplaces and chimneys to provide better home heating.[388] Also, the introduction of knitting and buttons during that period enabled tailors to make warmer clothing.[389] In our own century, concern about global warming has generated a potential market for the development of "green" technologies like recycling, solar power, and alternative fuels for cars and trucks.

2. Depletion of Natural Resources.—Depletion of natural resources also creates a demand for substitute products. The destruction of English forests in the sixteenth and seventeenth centuries due to glassblowing and shipbuilding activities triggered improvements in coal mining technology when coal replaced wood as a fuel source in many industries.[390] In the nineteenth century, the excessive killing of African elephants led to a shortage of natural ivory and thereby created a market for celluloid billiard balls and other ivory substitutes.[391] In the twenty-first century, the depletion of fossil fuels has encouraged the development of alternative fuel technologies for both transportation and power generation.

3. Military Competition.—An arms race can also provide a ready market for new technologies.[392] For example, in the mid-nineteenth century, when exploding shells fired by rifled guns took the place of smooth-bore cannons and solid shot, navy officials realized that they needed to replace slow-moving wooden sailing ships with something better.[393] As a result, they were receptive to the introduction of ironclads and iron-hulled steam powered warships once they became available.[394] During this same period, European and American armies were also quick to abandon their traditional smoothbore muskets for more

---

https://communities.cisco.com/community/solutions/sp/mobility/blog/2012/01/12/the-birth-of-the-mobile-phone.

388. *See* BURKE, CONNECTIONS, *supra* note 265, at 157-59.

389. *Id.* at 161.

390. DERRY & WILLIAMS, *supra* note 1, at 145-47.

391. *See* BURKE, CONNECTIONS, *supra* note 265, at 279.

392. *See* DERRY & WILLIAMS, *supra* note 1, at 278.

393. *See* JOHN BEELER, BIRTH OF THE BATTLESHIP 70-71 (2001).

394. *See id.*

rapid firing breech-loading rifles.[395] Later, military strategists spotted the potential of airplanes. Within eight years of the Wright brothers' first flight in 1903, the Italian Army employed airplanes to bomb Turkish positions during the Italian-Turkish war.[396] In more recent times, radar, rocket, helicopter and jet engine technology developed at an increased pace during World War II.[397]

4. Dissatisfaction with Existing Technology.—Dissatisfaction with existing technology creates opportunities for those who can "build a better mousetrap."[398] As mentioned earlier, printed books were commercially successful because they were much cheaper and much easier to read than hand written codices. Likewise, railroads quickly replaced stage coaches because they were faster, cheaper to operate, safer and more comfortable than horse powered transportation. Railroads also offered a better means of transporting bulk goods than canals and eventually put most of the canal companies out of business. Ocean-going steamships replaced wooden sailing ships for the same reason. Unlike sail powered packet ships, steamships could maintain fixed schedules because they were not dependent on the wind. After the introduction of compound-expansion engines, high-pressure boilers, iron and steel hulls and the screw propeller, steamships became larger, faster and more reliable than sailing ships and almost entirely replaced them by the latter decades of the nineteenth century.

Another example of the triumph of one technology over another involves illumination.[399] Originally, homes and factories were illuminated by candles, oil lamps, and later by coal gas. However, once electric lighting appeared on the scene, it displaced these other forms of illumination because it was safer, more convenient to use and provided more illumination. Finally, the gasoline powered automobile not only supplanted the horse and carriage, it also displaced electric and steam powered vehicles, which were more expensive and less convenient than their gasoline powered competitors.[400]

5. Demographic Changes.—Demographic changes may also encourage commercial acceptance of new products. For example, in the late fourteenth

---

395. *See* BLACK, *supra* note 271, at 111-12.

396. *Id.* at 154.

397. *See* BLACK, *supra* note 271, at 174-76 (radar); *id.* at 188 (jet fighters); ERNST KLEE & OTTO MERK, THE BIRTH OF THE MISSILE: THE SECRETS OF PEENEMÜNDE 10-24 (T. Schoeters trans., 1965) (German rockets); J. GORDON LEISHMAN, PRINCIPLES OF HELICOPTER AERODYNAMICS 1-49 (2d ed. 2006) (helicopters); *20th Century and Beyond*, NASA, http://www. grc. nasa.gov/WWW/k-12/rocket/BottleRocket/20thBeyond.htm (rockets) (last visited June 21, 2012).

398. Ralph Waldo Emerson is generally attributed with saying, "If you build a better mousetrap, the world will beat a path to your door." Thomas Cooper & Tom Kelleher, *Better Mousetrap? Of Emerson, Ethics, and Postmillennium Persuasion*, 16 J. MASS MEDIA ETHICS 176, 176 (2001).

399. *See generally* Walter Hough, *The Development of Illumination*, 3 AM. ANTHROPOLOGIST 342 (1901) (providing history of illumination).

400. *See* William W. Bottorff, *What Was the First Car? A Quick History of the Automobile for Young People*, AUSTIN BUS. COMPUTERS, INC., http://www.ausbcomp.com/~bbott/cars/carhist.htm (last visited May 3, 2012).

century, population losses from the Black Death caused wages to increase, thereby creating a demand for linen undergarments (made cheaper by the invention of the horizontal loom and the spinning wheel).[401] Later, in Western Europe, an increase in population during the nineteenth century created a need to produce more food and, thus, stimulated an interest in mechanized farm equipment like steam tractors and reapers.[402] Increased population also led to greater urbanization, which made it possible to popularize gas lighting and movie theaters.[403]

6. Social Changes.—Finally, changes in social conditions may create a demand for new products and technologies. For example, increases in the literacy rate during the Renaissance fueled a strong demand for inexpensive books just at a point in time when printing by means of movable type was invented.[404] The expanding number of female drivers contributed to the commercial success of the self-starter for automobiles when it was introduced in 1912.[405] In more recent times, the phenomenon of the two-earner household has no doubt helped to sustain a ready market for household gadgets.[406]

B. Impediments to Commercial Acceptance of Technology

However, sometimes an otherwise promising technology does not catch on right away. There can be many reasons for this, including high cost, competition from other products or technologies, lack of demand for the product or technology, resistance from special interest groups, cultural resistance or opposition from religious groups, and lack of existing infrastructure support. These conditions are also largely beyond a manufacturer's control.

1. High Costs.—In some cases, an innovative product may be too expensive in either absolute or relative terms to achieve commercial success when it is first introduced. A product that is too expensive in absolute terms is one which the public would like to purchase, but cannot afford. For example, when automobiles were built to order, only comparatively wealthy persons could afford them. Eventually, however, mass production techniques were introduced which made automobiles more affordable.[407] Likewise, prior to World War II, air travel

---

401. *See* BURKE, CONNECTIONS, *supra* note 265, at 98-100.

402. *See* DERRY & WILLIAMS, *supra* note 1, at 70-74.

403. *See id.* at 278.

404. *See* BURKE, CONNECTIONS, *supra* note 265, at 105-06.

405. *See* THE AUTOMOBILE AND AMERICAN CULTURE 139 (David L. Lewis & Laurence Goldstein eds., 1983).

406. *See* Steven Lugauer et al., *The Effect of Household Appliances on Female Labor Force Participation: Evidence from Micro Data*, 17 LABOUR ECONS. 503 (2010); Susan Guibert, *Increase in Number of Working Woman Due to Appliance Ownership, Notre Dame Study Shows*, NOTRE DAME NEWS, Jan. 19, 2011, http://newsinfo.nd.edu/news/18152-increase-in-number-of-working-women-due-to-appliance-ownership-notre-dame-study-shows.

407. *See* JULIAN PETTIFER & NIGEL TURNER, AUTOMANIA: MAN AND THE MOTOR CAR 173 (1984).

was very expensive and only the wealthy could afford to fly.[408] After the War, as passenger planes became larger and faster, flying became cheaper and more popular with the general public.[409] The same may be occurring today with regard to space travel. Recently, Virgin Galactic has begun to take reservations on its spaceship WhiteKnightTwo for space flights.[410] Unfortunately, the cost is $200,000 per passenger.[411] Although this cost may eventually come down, space travel will probably not become an established means of mass transportation anytime soon.

2. *Competition*.—A product may also encounter resistance in the marketplace because it is more expensive to purchase or operate than its competition. For example, ocean-going steamships did not supersede wooden sailing ships until the middle of the nineteenth century in part because they consumed too much coal in relation to the amount of power that they generated.[412] It was not until compound-expansion engines were installed, along with more efficient boilers, iron and steel hulls, and screw propellers, that oceangoing steamships were able to compete successfully with sailing ships.[413] A similar problem seems to have stymied the commercialization of electric cars in the twenty-first century. Due to the high cost of rechargeable batteries, it is estimated that an electric automobile costs at least $5000 more than a comparable gasoline powered vehicle.[414] For this reason, the public has been slow to embrace electric cars although this may change if the price of gasoline increases dramatically.[415]

3. *Lack of Demand*.—Lack of demand may also be based on the perception that the product or technology is not needed. For example, carriage owners shunned pneumatic tires in the 1840s, though they readily purchased solid rubber tires, because they felt that the suspension system of existing carriages was adequate to handle bumpy road conditions.[416] It was not until bicycles (which had no suspension systems) came into vogue in the 1880s that the market for

---

408. *See* CARL SOLBERG, CONQUEST OF THE SKIES: A HISTORY OF COMMERCIAL AVIATION IN AMERICA 221 (1979) (declaring that "[t]he well-to-do flew").

409. *Id.* at 345 (stating that "[s]tarting around 1948, commercial flying began to turn into the kind of travel that the general public could afford").

410. *See* Mark Flores, *Blast Off?—Strict Liability's Potential Role in the Development of the Commercial Space Market*, 17 RICH. J.L. & TECH. 1, 1-2 (2010).

411. *See id.* at 4.

412. *See* BEELER, *supra* note 393, at 32.

413. *See id.* at 54-60.

414. Press Release, J.D. Power & Assocs., Future Global Market Demand for Hybrid and Battery Electric Vehicles May Be Over-Hyped; Wild Card Is China (Oct. 27, 2010), *available at* http://businesscenter.jdpower.com/news/pressrelease.aspx?ID=2010213.

415. *See id.*

416. *See* DERRY & WILLIAMS, *supra* note 1, at 529 ("But in the same decade of the 1840's a patent was taken out for pneumatic rubber tyres [sic] with an outer casing of leather: these 'aerial wheels,' as they were called, were tested in Hyde Park and ran successfully for as much as 1,200 miles; they even spread to New York, yet the invention was soon forgotten.").

pneumatic tires improved.[417]

4. *Resistance from Special Interest Groups*.—Commercial competitors and workers have also occasionally blocked the introduction of new technology. For example, in the early nineteenth century, canal companies lobbied hard to prevent the construction of railroad lines.[418] Concern about the effect of new technology on existing labor patterns also may delay the introduction of technology. For instance, the Emperor Vespasian opposed the use of waterwheels in the Roman Empire because he believed that it might cause unemployment.[419] In the eighteenth century, the silk weavers of Lyons rioted for the same reason when the Jacquard loom was introduced.[420]

5. *Cultural or Religious Opposition*.—Religious or social opposition may delay a new technology or prevent its introduction altogether. For example, during the Middle Ages, the Catholic Church banned (without much success) the use of crossbows and gunpowder.[421] More recently, religious groups have opposed the use of birth control pills and other contraceptive devices.[422] New technology may also clash with social or cultural norms. For example, when automobile manufacturers first introduced automatic transmission, the response from the driving public was not very enthusiastic.[423] Years later, there was considerable reluctance by some demographic groups to use seat belts until laws were passed to make seat belt use mandatory.[424] On a smaller scale, for many years, professional baseball players resisted using batting helmets because it conflicted with their macho image.[425]

6. *Lack of Infrastructure*.—Finally, commercial success may elude a new technology because no infrastructure exists to support it. The electric light bulb

---

417. *Id.*

418. *See* David E. Lilienthal & Irwin S. Rosenbaum, *Motor Carrier Regulation by Certificates of Necessity and Convenience*, 36 YALE L.J. 163, 187 n.95 (1926).

419. DERRY & WILLIAMS, *supra* note 1, at 252.

420. *See* BURKE, CONNECTIONS, *supra* note 265, at 111.

421. *See* JOHN NORRIS, ARTILLERY: A HISTORY 4 (2000).

422. *See* Susan J. Stabile, *State Attempts to Define Religion: The Ramifications of Applying Mandatory Prescription Contraceptive Coverage Statutes to Religious Employers*, 28 HARV. J.L. & PUB. POL'Y 741, 749-55 (2005); Kate Spota, Comment, *In Good Conscience: The Legal Trend to Include Prescription Contraceptives in Employer Insurance Plans and Catholic Charities' "Conscience Clause" Objection*, 52 CATH. U. L. REV. 1081, 1084-87 (2003).

423. *See* Preston Taylor, *History of the Automotive Automatic Transmission*, DIABLO TRANSMISSIONS, http://www.diablotransmissions.com/history_automatic_transmission.html (last visited June 21, 2012).

424. *See* Tina Wescott Cafaro, *You Drink, You Drive, You Lose: Or Do You?*, 42 GONZ. L. REV. 1, 5-6 (2006); Barry L. Huntington, Comment, *Welcome to the Mount Rushmore State! Keep Your Arms and Legs Inside the Vehicle at All Times and Buckle Up . . . Not for Your Safety, but to Protect Your Constitutional Rights*, 47 S.D. L. REV. 99, 101 n.14 (2002) (stating that fewer than fifteen percent of drivers used seat belts when they were first introduced in the 1950s).

425. *See* ROBERT M. GORMAN & DAVID WEEKS, DEATH AT THE BALLPARK: A COMPREHENSIVE STUDY OF GAME-RELATED FATALITIES, 1862-2007, at 10-12 (2009).

was a commercial success because Thomas Edison and others constructed power stations and power lines to supply electricity to their customers.[426] Infrastructure also contributed to the success of the gasoline-powered automobile. In contrast to the early nineteenth century, when the pothole covered roads of London thwarted the efforts of Richard Trevithick to introduce steam carriages to Regency England, by the end of that century a macadamized road system existed, at least in urban areas that ensured the success of gasoline-powered automobiles.[427] In addition, oil refineries were already in operation making kerosene and lubricating oil when automobiles were first introduced.[428] The distillation process being used in these refineries could be readily adapted to manufacture gasoline.[429] In contrast, electric cars failed to catch on at the beginning of the twentieth century, at least in part, because of the lack of charging or battery swapping facilities outside of urban areas.[430]

IV. RETHINKING THE STATE-OF-THE-ART CONCEPT

State-of-the-art is a confusing concept in the law of products liability. Its rationale is indeterminate; its meaning is muddled; and, its procedural effects are highly variable. Accordingly, this portion of the Article will offer some suggestions for clarifying the state-of-the-art concept and defining its proper role in products liability litigation.

A. Rationale

The principal function of the state-of-the-art doctrine is to provide a practical limit to the concept of defectiveness and, thereby, limit the scope of strict products liability. Almost fifty years ago, the drafters of section 402A declared that liability would only be imposed on sellers for injuries that were caused by *defective* products.[431] This defectiveness requirement has also been retained by the *Products Liability Restatement*.[432] For this reason, a manufacturer should not be held liable simply because the plaintiff has proposed a safer alternative design. Instead the state-of-the-art concept requires the plaintiff to prove that it was actually possible for the manufacturer to have adopted the proposed design when the product was sold.[433] This same principle supports a requirement that any

---

426. *See* DERRY & WILLIAMS, *supra* note 1, at 615-21.

427. This new process for paving roads was developed by James McAdam in the 1820s and 1830s. *Id.* at 432-36.

428. *See* BURKE, CONNECTIONS, *supra* note 265, at 180.

429. *See id.*

430. *See* Robert Bryce, *Five Myths About Green Energy*, WASH. POST, Apr. 25, 2010, http://www.washingtonpost.com/wp-dyn/content/article/2010/04/23/AR2010042302220.html (discussing unreliability of electric car batteries).

431. *See* OWEN, *supra* note 11, § 6.1, at 343.

432. *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2 (1998).

433. *See id.*

alternative design submitted by a plaintiff be commercially feasible.[434] A requirement that plaintiffs show that their proposed alternative designs are commercially feasible helps to ensure that manufacturers are not punished for making efficient design decisions.

B. Terminology

At the present time, the term state-of-the-art has no fixed meaning in products liability law. To some courts, it refers to customs or practices of the industry.[435] Other courts take a broader view and include safety standards promulgated by trade associations, independent standards development organizations, or government regulatory agencies within the state-of-the-art concept.[436] A third alternative defines state-of-the-art as that which is technologically feasible.[437] Finally, a few courts have included commercial feasibility within the definition of state-of-the-art.[438]

Considering that there is no consensus about the meaning of state-of-the-art, it might be better to avoid the phrase entirely and rely instead on terms like "technological feasibility" and "commercial feasibility," which have more generally accepted meanings. Under this approach, technological feasibility would include: (1) knowledge of the underlying scientific principles; (2) the ability to convert these scientific principles into working models and prototypes of the product or safety device in question; and (3) the capacity to manufacture the product or safety device on a commercial scale.[439] In some cases, technological feasibility would also include the ability to purchase necessary raw materials or component parts from other vendors. Commercial feasibility, on the other hand, means that: (1) there is an existing or potential consumer interest in the product or safety device; (2) the manufacturer has the resources to market and distribute the product or safety device to the public; and (3) the technology has developed to the point where the product or safety device can be offered for sale at an affordable price.[440]

C. The Procedural Effect of State-of-the-Art Evidence in Product Liability Litigation

Until recently, defendants often relied on state-of-the-art evidence to establish that their product designs were not defective.[441] However, requiring a

---

434. *See id.*

435. *See id.*

436. *See* discussion *supra* Part I.B.1.

437. *See* OWEN, *supra* note 11, § 10.4, at 710-11.

438. *See, e.g.*, Boatland of Hous., Inc. v. Bailey, 609 S.W.2d 743, 746 (Tex. 1980).

439. OWEN, *supra* note 11, § 10.4, at 711-12.

440. *Id.*

441. *See* O'Brien v. Muskin Corp., 463 A.2d 298, 305 (N.J. 1983), *superseded by statute as stated in* Dewey v. R.J. Reynolds Tobacco Co., 577 A.2d 1239 (N.J. 1990); Crispin v. Volkswagenwerk AG, 591 A.2d 966, 973 (N.J. Super. Ct. App. Div. 1991); Patterson v. Ravens-

manufacturer to prove that its product conforms to the state-of-the-art is the wrong approach. On one hand, if conformity with state-of-the-art is equivalent to feasibility, then any product that was produced and sold commercially would be "feasible" and, therefore, would conform to the state-of-the-art. Such a test would be meaningless. On the other hand, if conformity with the state-of-the-art requirement meant that the defendant's design had to reflect the highest level of technology available, many products would fail to meet this standard. However, this standard seems wrong because a product's design does not have to be the best or the safest; it simply has to be non-defective.[442]

The *Product Liability Restatement*'s approach is more logical. Section 2(b) of the *Products Liability Restatement* requires the plaintiff to prove that a product is defectively designed by showing that a reasonable alternative design would have prevented or reduced his or her injuries.[443] Arguably, for a design to be "reasonable," it must be technologically and commercially feasible.[444] Of course, once the plaintiff offers proof by expert testimony that his or her proposed alternative design was technologically and commercially feasible at the time of sale, the defendant will then need to rebut this claim with evidence that the plaintiff's proposed design was not in fact technologically or commercially feasible at that time.[445]

How does a defendant go about rebutting the plaintiff's claim that a proposed alternative design is technologically and commercially feasible? This will be difficult to do if some or all of the defendant's competitors have already adopted the plaintiff's alternative design at the time of sale. However, if the alternative design was not in use at the time of sale by the defendant, but some or all of the defendant's competitors have adopted it since then, the defendant would have little choice but to argue that the design was not technologically feasible at the time of sale.[446] If the alternative design had not been adopted by any of the

---

Metal Prods., Inc., 594 N.E.2d 153, 161-62 (Ohio Ct. App. 1991); Owens-Corning Fiberglass Corp. v. District Court (*Caldwell*), 818 S.W.2d 749, 752 (Tex. 1991).

442. See Davis v. Caterpillar Tractor Co., 719 P.2d 324, 326-27 (Colo. App. 1985) (holding that tractor was not defective because it was not equipped with rollover protective device even though such devices were available).

443. *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2(b) (1998).

444. *See* OWEN, *supra* note 11, § 10.4, at 737.

445. *See* Rucker v. Norfolk & W. Ry. Co., 381 N.E.2d 715, 725 (Ill. App. Ct. 1978), *rev'd on other grounds*, 396 N.E.2d 534 (Ill. 1979).

446. *See* Bruce v. Martin-Marietta Corp., 544 F.2d 442, 446-47 (10th Cir. 1976) (rebutting evidence that passenger planes were equipped with better seats when accident occurred in 1970 by arguing that these seats were not in use when the airplane was manufactured in 1952); Brady v. Melody Homes Mfr., 589 P.2d 896, 898 (Ariz. Ct. App. 1978) (responding to evidence that mobile homes were currently equipped with smoke detectors and pop-out windows by pointing out that no manufacturer provided these safety features ten years earlier when the mobile home in question was manufactured); Wiska v. St. Stanislaus Social Club, Inc., 390 N.E.2d 1133, 1137 (Mass. App. Ct. 1979) (excluding evidence of safer windshield glass that was adopted by the automobile industry two years after the accident).

defendant's competitors, but had been adopted by other industries at the time of sale; the defendant would likely contend that it was not possible to transfer the technology from one industry to another.[447] Finally, if the plaintiff's alternative design is purely hypothetical and has not been adopted by anyone, either at the time of sale or the time of trial, the defendant would likely argue that the proposed design either was not commercially feasible, but like Leonardo's helicopters, would have to await additional developments before it would become technologically feasible.[448]

D. A Statutory Approach

Assuming that these observations have merit, what is the best mechanism for changing the existing state-of-the-art doctrine? The best approach would probably be to codify this version of the state-of-the-art doctrine by statute. A number of states have already adopted statutes that purport to protect products sellers from liability when their products have been designed in accordance with the prevailing state technological development in the industry at the time of sale.[449] Some of these statutes require that the defendant be able to conform to the state-of-the-art or they make conformance to the state-of-the-art an affirmative defense.[450] Others provide that products that conform to the state-of-the-art are presumed to be non-defective.[451] Finally, a few statutes have adopted the *Product Liability Restatement*'s approach and require the plaintiff to offer proof of a feasible alternative design.[452]

I believe that any statutory codification of the state-of-the-art doctrine should

447. *See* Boatland of Hous., Inc. v. Bailey, 609 S.W.2d 743, 747 (Tex. 1980) (concluding that evidence of use of kill switches in racing boats did not necessarily prove that such devices were within the state-of-the-art for bass boats). *But see* Hillrichs v. Avco Corp., 514 N.W.2d 94, 97 (Iowa 1994) (allowing plaintiff whose hand was caught in rollers of corn husking machine to present evidence that manufacturers of printing presses equipped their machines with emergency stop devices).

448. *See, e.g.*, Stanczyk v. Black & Decker, Inc., 836 F. Supp. 565, 566-68 (N.D. Ill. 1993) (excuding proposed safer design that had not been tested for feasibility); Maxted v. Pac. Car & Foundry Co., 527 P.2d 832, 834 (Wyo. 1974) (refusing to allow as evidence drawings and model of a safety device for tractor-trailers that had never been tested or constructed at the time of manufacture).

449. *See* OWEN, *supra* note 11, § 10.4, at 734-36.

450. ARIZ. REV. STAT. § 12-683 (2012); IOWA CODE ANN. § 668.12 (West 2012); LA. REV. STAT. ANN. § 9:2800.59 (2012); MICH. COMP. LAWS ANN. § 600.2948(3) (West 2012); MISS. CODE ANN. §§ 11-1-63(b) to (c) (West 2011); MO. ANN. STAT. § 537.764 (2012); NEB. REV. STAT. § 25-21,182 (2011); N.H. REV. STAT. ANN. § 507:8-g (2011).

451. COLO. REV. STAT. ANN. § 13-21-403(1) (West 2012); IND. CODE § 34-20-5-1 (2011); KY. REV. STAT. ANN. § 411.310(2) (West 2011).

452. LA. REV. STAT. ANN. § 9:2800.56; MISS. CODE ANN. § 11-1-63; N.J. STAT. ANN. § 2A:58c-3 (West 2012); N.C. GEN. STAT. 99B-6 (West 2011); TEX. CIV. PRAC. & REM. CODE ANN. § 82.005 (West 2011); WASH. REV. CODE ANN. § 7.72.030 (West 2012).

contain the following provisions. First, the statute should make it clear that a state-of-the-art analysis should only be applied to evaluate a safer alternative design proposed by the plaintiff. Second, although the plaintiff's proposed alternative design does not have to be actually adopted by others in the industry, the plaintiff must prove that the proposed alternative design was technologically and commercially feasible at the time of sale. Third, the defendant is allowed to rebut the plaintiff's evidence by offering evidence that the proposed alternative design was not technologically or commercially feasible at the time the product was sold. Fourth, the defendant should be allowed to prove that the plaintiff's proposed alternative design was not technologically feasible by alleging that it could not control its development. If the plaintiff fails to prove that his or her proposed alternative design is technologically and commercially feasible, the case against the defendant must be dismissed. On the other hand, even if the plaintiff proves that his or her proposed design is technologically and commercially feasible, the defendant should still be able to avoid liability by showing that its existing design is not defective because it is more aesthetically pleasing or offers more convenience or consumer choice.

CONCLUSION

It seems inherently unfair to hold a manufacturer liable for failing to do something that is beyond the scope of existing technology. At the same time, it is desirable to have a liability rule that encourages producers to make optimal investments in product safety. The state-of-the-art concept is one device to induce manufacturers to achieve what is technologically and commercially possible in the area of product safety. However, some courts and statutes incorrectly focus on the technological and commercial feasibility of the manufacturer's design choices instead of looking at the feasibility of alternative designs. Ideally, the plaintiff should be required to prove that the defendant's product is defectively designed by offering evidence of a safer alternative design. The defendant should then be allowed to show that the proposed design was not capable of being produced and marketed using the technology that was available at the time.[453]

Furthermore, the defendant should be permitted to show that the technology behind the plaintiff's proposed design is interdependent in nature and, therefore, not solely within the defendant's control. In the event a safer design was technologically feasible at the time the product was sold, the manufacturer should be allowed to argue that the design was not commercially feasible at that time. Finally, even if the plaintiff proves that his or her proposed design is technologically and commercially feasible, the defendant should still be able to show that its design is cheaper, more convenient, or offers a greater range of consumer choice.

---

453. *See* Flock v. Scripto-Tokai Corp., 319 F.3d 231, 239 (5th Cir. 2003); Potter v. Chi. Pneumatic Tool Co., 694 A.2d 1319, 1346-47 (Conn. 1997).

CONGRESSIONAL WILL AND THE ROLE OF THE EXECUTIVE IN *BIVENS* ACTIONS: WHAT IS SPECIAL ABOUT SPECIAL FACTORS

ANYA BERNSTEIN*

I. INTRODUCTION

Say a government actor violates a person's constitutional right. In an odd federalist twist, what recourse the person has depends on which government level the violator acted for. If the government actor was cloaked in the mantle of state authority, the victim can sue under the Civil Rights Act of 1871.[1] If, in contrast, the violator acted on behalf of the federal government, the victim has no explicit statutory cause of action. She may, however, be able to sue the government actor directly under the U.S. Constitution on the theory articulated in *Bivens v. Six Unknown Named Agents of Federal Bureau of Narcotics*.[2] Both § 1983 and *Bivens* impose monetary liability on errant government agents who violate individuals' constitutional rights.[3] But, as a judge-made remedy, *Bivens* has been

---

* Bigelow Teaching Fellow and Lecturer in Law, The University of Chicago Law School. J.D., Yale Law School; Ph.D., The University of Chicago; B.A., Columbia University. Many thanks to Michael Coenen, Nicholas Day, William Eskridge, Aziz Huq, Jonathan Masur, James Pfander, Alexander Reinert, and Lior Strahilevitz for helpful commentary.

1. 42 U.S.C. § 1983 (2006) ("Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law . . . ."); *see* Monroe v. Pape, 365 U.S. 167, 172 (1961) (concluding that § 1983's use of the phrase "under color of" does not limit the statute to acts undertaken in the authorized, legally correct pursuit of state law, but encompasses acts cloaked with state power undertaken by "those who carry a badge of authority of a State and represent it in some capacity, whether they act in accordance with their authority or misuse it"), *overruled by* Monell v. Dep't of Soc. Servs. of N.Y.C., 436 U.S. 658 (1978).

2. 403 U.S. 388 (1971). James Pfander reveals the provenance of the case's unusual name, noting that the government attorneys' "suggest[ion] that the agents were 'unknown' at the time of filing but were later identified and served with process, and thus 'named'" does not explain why the complaint itself, written *pro se*, "referred to 'Unknown Named' agents." James E. Pfander, *The Story of* Bivens v. Six Unknown-Named Agents of the Federal Bureau of Narcotics, *in* FEDERAL COURTS STORIES 275, 281-82 (Vicki C. Jackson & Judith Resnik eds., 2010) [hereinafter Pfander, *The Story of* Bivens]. Pfander presents a "more plausible" explanation, confirmed by Webster Bivens himself: "from the plaintiff's perspective, the agents simply had 'unknown names.'" *Id.* at 282.

3. Section 1983 provides remedies for violations of any federal right, not just those guaranteed by the federal Constitution. Somewhat analogously, the Federal Tort Claims Act (FTCA) of 1946, 28 U.S.C. § 1346(b)(1) (2006), provides a remedy against the United States (not the individual actor) "for injury . . . caused by the negligent or wrongful act or omission of any employee of the Government while acting within the scope of his office or employment, under

subject both to skeptical scrutiny by commentators[4] and to somewhat ad hoc development by the Supreme Court, which has allowed a case-by-case determination of whether "special factors" preclude constitutional damages.[5]

These "special factors" have become a central concept in *Bivens* jurisprudence, yet no court has ever defined the term. This does not mean, however, that "special factors" has no definition; by developing the special factors inquiry over time and through case law, courts have delimited the concept's contours quite clearly; they just have not explicitly articulated what the concept is. This Article provides that articulation. By illuminating the rationale of Supreme Court special factors jurisprudence and canvassing circuit court cases to see how this rationale has been applied, I derive the definition of special factors from existing case law.

This is possible because, as I show, special factors jurisprudence has not been as *ad hoc* as it seems at first glance. Building on Supreme Court precedent, federal courts of appeals have developed a coherent approach widely shared among the circuits. Identifying that approach and explaining its underlying rationale, as I do here, gives courts an explicit, workable model for applying the special factors analysis in *Bivens* cases. We need such a model because without it, outlier cases can fail to recognize that special factors is a term of art rather than a simple descriptive phrase—a bounded legal analysis, not an appeal to judges' sensibilities. Such misrecognition has led to occasional incongruous opinions whose special factors analysis comports neither with the accepted way of conducting the inquiry, nor with its underlying rationale.

A. The Contours and Rationale of the Special Factors Analysis

This Article shows that the special factors analysis has evolved into an inquiry as to whether Congress has indicated that it wishes to reserve decisionmaking about remediation in some area for itself.[6] If Congress has

---

circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred."

4. *See, e.g.*, James E. Pfander & David Baltmanis, *Rethinking* Bivens*: Legitimacy and Constitutional Adjudication*, 98 GEO. L.J. 117, 121 (2009) (arguing that courts should change their approach to evaluating a *Bivens* claim).

5. *See* Wilkie v. Robbins, 551 U.S. 537, 550-51 (2007).

6. I am, of course, not the first to connect *Bivens* remedies to congressional will. But because I focus on explaining the development and rationale of special factors, I can provide a more thorough analysis of the case law that has shaped the doctrine and expressed its underlying reasoning. Most scholarship that makes this connection instead presupposes the relationship and makes reform proposals on that basis. *See, e.g.*, Betsy J. Grey, *Preemption of* Bivens *Claims: How Clearly Must Congress Speak?*, 70 WASH. U. L. Q. 1087, 1122-23 (1992) (proposing a clear statement requirement for congressional preclusion of constitutional damages remedies); Kristina A. Kiik, Comment, *Quantum of Competence: Balancing* Bivens *During the War on Terror*, 62 SMU L. REV. 1945, 1963-64 (2009) (proposing that courts treat legislative action as more of a special factor than executive action).

indicated that it has already provided all the remedies it thinks are due, or that it prefers that no remedies at all be provided, courts find that special factors preclude recognizing a constitutional damages remedy.[7] If Congress has not so indicated, courts feel confident recognizing a remedy directly under the Constitution.[8] As this description suggests, the special factors analysis maintains the balance of powers between two specific branches of the federal government: the judicial and the legislative.

It may be helpful to think of this process as analogous to federal preemption analysis: while preemption analysis primarily maintains the balance of powers between Congress and the states, special factors analysis maintains the balance of powers between Congress and the federal judiciary.[9] The analogy is not perfect because the vertical concerns of federalism do not map perfectly onto the horizontal concerns of the separation of powers. However, it can be a helpful shorthand because special factors analysis employs techniques drawn from preemption analysis.

In its early version, the special factors analysis looked for explicit congressional preclusion of *Bivens* remedies, requiring something like explicit preemption.[10] Subsequently, it incorporated a version of field preemption,[11] where thorough occupation of a legal area by Congress indicated that there was no room for judge-made remedies. Finally, the special factors analysis has broadened in scope to include something resembling implicit preemption[12]: if the statutory language indicated that Congress wished to reserve the creation of remedies for itself, courts refrained from recognizing constitutional damages remedies. As this description suggests, throughout its development, the special factors analysis has focused on whether congressional legislation precluded constitutional damages remedies against federal actors.

---

7. *See* Schweiker v. Chilicky, 487 U.S. 412, 420-21 (1988).

8. *See* Carlson v. Green, 446 U.S. 14, 18-19 (1980).

9. Preemption analysis often addresses the relative authority of executive agencies and the states, but executive agency rules preempt state law only to the extent that Congress delegates that preemptive power to the Executive. *See, e.g.*, Crosby v. Nat'l Foreign Trade Council, 530 U.S. 363, 374-75 (2000) ("'When the President acts pursuant to an express or implied authorization of Congress, his authority is at its maximum, for it includes all that he possesses in his own right plus all that Congress can delegate.'" (quoting Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 635-36 n.2 (1952))). The power of preemption, in other words, usually lies with Congress even when others exercise that power for Congress.

10. *See* Pennsylvania v. Nelson, 350 U.S. 497, 505 (1956) (explicit preemption occurs when federal law and state law directly conflict).

11. *Id.* at 504 (field preemption occurs when "federal statutes 'touch a field in which the federal interest is so dominant that the federal system (must) be assumed to preclude enforcement of state laws on the same subject'" (quoting Rice v. Santa Fe Elevator Corp., 331 U.S. 218, 230 (1947))).

12. *Id.* at 502 (implicit preemption is when the "'federal regulation [is] so pervasive as to make reasonable the inference that Congress left no room for the States to supplement it'" (quoting *Rice*, 331 U.S. at 230)).

Through the special factors analysis, *Bivens* has come to allow relief in the limited set of situations for which Congress has neither legislated remedies nor expressed a preference regarding their existence.[13] *Bivens* fills the gap when Congress has not indicated how a particular situation ought to be handled. By inquiring into congressional will, the special factors analysis keeps courts sensitive to whether there is a gap to be filled. In this way, *Bivens* remedies allow courts to vindicate constitutional guarantees of individual rights while coordinating their authority with the legislative branch. The meaning of the term "special factors" has changed substantially since its first, somewhat offhand, use in *Bivens*. But throughout those changes, the term has consistently concerned the distribution of authority between the judiciary and the legislature.

This approach is not only grounded in precedent, it fits the logic of judge-made remedies. By allowing a damages action for a constitutional violation, *Bivens* attempts to act out *Marbury*'s conviction that a right ought to come with a remedy.[14] But, as the *Bivens* opinion intuited, it also threatens the separation of powers between the judiciary and the legislature, as courts assume the power to create causes of action and provide remedies—a power normally exercised by Congress.[15] Through the special factors analysis, courts ensure that their vindication of constitutional rights does not infringe on Congress's authority.

Of course, the judiciary and the legislature are not the only branches that have a hand in crafting remedies. The modern executive branch, with its administrative remedial schemes and its prominent role in the process of legislation, also plays a part. However, as the *Bivens* case discussed throughout this Article indicates, the Executive's role in remedy-creation is still subordinate to that of Congress. Administrative remedial schemes must be authorized through a delegation of congressional power to the Executive and are subject to legislative strictures and specifications. Although the President often plays a significant role in the crafting of legislation and must sign a bill into law, it is still Congress that debates and passes it.[16] Responding to these realities, case law regarding constitutional damages consistently looks to congressional will to ensure that judge-made remedies do not disturb the balance of authority between the judiciary and the legislature.

B. Why This Analysis Now

Special factors analysis developed into its present form gradually, over many cases and courts. This is why, in contrast with much of the scholarship on *Bivens*, I look not just to the Supreme Court but also to the circuit courts to

13. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 395-96 (1971).

14. Marbury v. Madison, 5 U.S. (1 Cranch) 137, 163 (1803) ("'[I]t is a general and indisputable rule, that where there is a legal right, there is also a legal remedy by suit or action at law, whenever that right is invaded.'" (quoting 3 WILLIAM BLACKSTONE, COMMENTARIES *23)).

15. *Bivens*, 403 U.S. at 395-97.

16. *See* U.S. CONST. art. I, § 7, cl. 1-2.

understand the meaning of "special factors." The Supreme Court has sporadically indicated what direction it has wished special factors analysis to take, but much of the doctrine has been developed, refined, and cohered by the federal circuit courts.

This is not to say that the courts of appeal have failed to follow Supreme Court precedent, or even that appellate case law is in tension with that of the Supreme Court. On the contrary, *Bivens*'s special factors jurisprudence instantiates a common phenomenon in doctrinal development: The Supreme Court sketches broad parameters for approaching an issue, and the appellate courts follow that lead to fill in the details. This Article illuminates this doctrinal history at both court levels and explains what the special factors analysis has come to mean—something courts need to understand to determine whether to recognize a constitutional damages remedy in each particular case.

As I have noted, in most cases, the appellate courts have implicitly or unconsciously recognized the thrust of the special factors analysis and applied it appropriately. Why, then, is the explicit discussion I present here necessary? Without an explicit understanding of the doctrine's development and rationale, courts may fail to understand that special factors analysis focuses specifically on the separation of judicial from *legislative* authority, not from that of the executive. This distinction becomes especially important as the courts see increasing numbers of lawsuits challenging harms incurred through national security projects. The legal context of such lawsuits often fits squarely within the traditional *Bivens* domain, but there are important differences. First, these cases implicate national security policy; second, they often inspire publicity. These non-legal aspects of some contemporary *Bivens* cases can make the legal issues seem new, as well.

Because of the political prominence of the Executive in these cases, courts that do not understand the rationale of the special factors analysis sometimes invert the inquiry, focusing on the balance of judicial and *executive*, rather than legislative, powers. By tradition, the executive branch is considered to have special authority in the area of national security.[17] But this competence does not give the executive branch the authority to legislate or preclude remedies for those whose rights are violated in the course of such projects. That authority belongs primarily to Congress, which remains the proper focus of special factors analysis even in cases involving national security projects. By explaining the doctrine's rationale and detailing how it has developed, I give courts facing *Bivens* claims a principled and precedent-based model for performing the special factors analysis.

This Article differs from much of the scholarship on constitutional damages precisely because my aim is to give courts a useful tool for dealing with *Bivens* claims now, rather than to reform *Bivens* jurisprudence. My analysis derives

---

17. *See, e.g.*, Ctr. for Nat'l Sec. Studies v. U.S. Dep't of Justice, 331 F.3d 918, 926-27 (D.C. Cir. 2003) ("It is equally well-established that the judiciary owes some measure of deference to the executive in cases implicating national security, a uniquely executive purview." (internal citations omitted)).

from my survey of case law. I do not claim that this is the way *Bivens* ought to work. On the contrary, constitutional damages are ripe for reform, and a number of scholars have proposed vehicles for improvement. James Pfander and David Baltmanis, for instance, have suggested eliminating the special factors analysis altogether and, instead, advocate treating *Bivens* cases like those arising under § 1983.[18] Pfander and Baltmanis point out that, while Congress has not legislated a *Bivens* remedy, it has entrenched constitutional damages in American law by passing legislation that presupposed the availability of a *Bivens* remedy.[19] They suggest matching the statutory preclusion of *Bivens* remedies to the Supreme Court's "relatively narrow view of the implied displacement of § 1983 claims."[20] Others have proposed letting *Bivens* suits run against the federal government itself, rather than only against federal actors in their individual capacities.[21]

---

18. Pfander & Baltmanis, *supra* note 4, at 121 (explaining that under their proposal, "the 'special factors' that the Court has taken into account in deciding whether to allow an action would no longer operate as a threshold barrier to litigation"). Not everyone finds the parallel convincing: George D. Brown has argued that § 1983, concerned as it is with the relations between states and the federal government, is not an appropriate analogue to *Bivens*'s exclusive focus on federal government issues; additionally, *Bivens*'s basis in the federal Constitution weakens the contention that Congress always has superior authority in crafting remedies. *See* George D. Brown, *Letting Statutory Tails Wag Constitutional Dogs—Have the* Bivens *Dissenters Prevailed?*, 64 IND. L.J. 263, 266 (1989). As I discuss, though, courts have not followed this lead.

19. Pfander & Baltmanis, *supra* note 4, at 132-33 ("[N]ew remedies under the FTCA were designed to supplement, not displace, the *Bivens* action.").

20. *Id.* at 143. Pfander and Baltmanis specifically propose the § 1983 preclusion standards enunciated in *Fitzgerald v. Barnstable School Committee*, 555 U.S. 246, 250-60 (2009), which held that Title IX of the Education Amendments of 1972, 20 U.S.C. § 1681(a) (2006), did not displace a § 1983 action challenging a school's inadequate response to allegations that a child was being sexually harassed. Pfander & Baltmanis, *supra* note 4, at 143. *Fitzgerald* concluded that the "remedial schemes [were] parallel rather than inconsistent" because Congress had not expressed a "legislative intent to displace § 1983 claims[;] . . . had not put in place detailed or more restrictive remedies that would suggest the inapplicability of constitutional tort litigation[;]" had tailored Title IX to a different "substantive scope of coverage" than that of the Equal Protection clause; and had limited Title IX remedies to the "responsible governmental entities" rather than to "individual officers." *Id.*

21. *See* Cornelia T.L. Pillard, *Taking Fiction Seriously: The Strange Results of Public Officials' Individual Liability under* Bivens, 88 GEO. L.J. 65, 65-68 (1999). Pillard claims that, insofar as the government indemnifies officials sued in their individual capacities under *Bivens*, the individual liability requirement of *Bivens* "undercut[s] *Bivens*'s remedial efficacy" by allowing a range of defenses that often defeat otherwise meritorious cases and creating an "illusion of an opportunity to obtain damages" that allows Congress and the courts to avoid dealing with the very real problem of bad conduct by federal actors. *Id.* at 67-68. James Pfander has suggested that federal government indemnification for officials sued under *Bivens* is not automatic: "[l]egal representation" for personal capacity suits "requires a written request from the agency head and a finding that representation will further the interests of the federal government." James E. Pfander, *Resolving the Qualified Immunity Dilemma: Constitutional Tort Claims for Nominal Damages*,

Undoubtedly, other ways still remain to make constitutional damages actions more predictable, sensible, and congruent with other areas of law. However, lower courts do not have the luxury of making such changes themselves: Reforms like these can be instituted only by the Supreme Court or by Congress. For now, lower courts must employ the special factors analysis, and this Article explicitly delineates what this analysis entails.

C. *Why* Bivens *Is Important*

The idea of permitting remedies against government actors who violate individuals' rights did not start with *Bivens* but had been around for centuries. Either the government itself or the officers acting in its name were amenable to suit in England beginning in the thirteenth century; by the eighteenth and nineteenth centuries, English "reports . . . abound[ed] with cases against inferior officers."[22] In the United States, by the time *Bivens* was decided, such remedies usually took the form of injunctions granted under the theory of *Ex Parte Young*.[23] Based on the long-term availability of injunctions in such cases, Justice Harlan's concurrence in *Bivens* noted that its result was hardly radical.[24] Since federal courts already granted injunctions in such situations, judge-made

---

111 COLUM. L. REV. 1601, 1609 n.36 (2011) [hereinafter Pfander, *Resolving the Qualified Immunity Dilemma*]. Still, the actual frequency of such designations, and of indemnification, remains unclear; most commentators assume that federal employees are routinely indemnified. A number of other works have proposed government, rather than individual, liability under *Bivens*. *See, e.g.*, PETER H. SCHUCK, SUING GOVERNMENT: CITIZEN REMEDIES FOR OFFICIAL WRONGS 100 (1983) ("[I]f enterprise liability for official misconduct is substantially expanded, society can enjoy significant and unambiguous gains."); William P. Kratzke, *Some Recommendations Concerning Tort Liability of Government and its Employees for Torts and Constitutional Torts*, 9 ADMIN. L. REV. AM. U. 1105, 1110 (1996) ("Congress should amend the FTCA [to extend] respondeat superior and exclusive liability of the United States to *Bivens* claims.").

22. Louis L. Jaffe, *Suits Against Governments and Officers: Sovereign Immunity*, 77 HARV. L. REV. 1, 15 (1963) (collecting cases). Jaffe argues that suing the government was easier in England than in the early Republic because the doctrine of sovereign immunity did not prevent suits against the King.

> [M]any claims affecting the Crown could be pursued in the regular courts if they did not take the form of a suit against the Crown. And when it was necessary to sue the Crown *eo nomine* consent apparently was given as of course. . . . By a magnificent irony, this body of doctrine and practice, at least in form so favorable to the subject, lost one-half of its efficacy when translated into our state and federal systems. Because the King had been abolished, the courts concluded that where in the past the procedure had been by petition of right there was now no one authorize to consent to suit!

Id. at 1-2.

23. 209 U.S. 123 (1908).

24. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 404 (1971) (Harlan, J., concurring).

remedies for federal employees' constitutional violations were nothing new.[25] In recognizing a damages remedy, he reasoned, the Court did not create a new judicial power so much as apply an existing power to a class of people whose injury could not be addressed through an injunction.[26]

In fact, Jerry Mashaw has demonstrated that the *Bivens* remedy was even less novel than this reasoning implies.[27] In late eighteenth and early nineteenth century America, people whose rights were violated by federal employees had "broad opportunities for common law damage actions" against them.[28] Such opportunities came not through special judicial dispensation but simply as a matter of common law tort: "private citizens objecting to official action by an agent of the United States had ample legal weapons at their disposal."[29] Legislators and ordinary American residents shared the presumption that damages actions were available in such situations—and not just available but necessary to keep the government in check. "Americans valued common law actions and criminal prosecutions, subject to trial by jury, as protections against the depredations of federal officials."[30] Indeed, because federal agents were subject to suit in their individual capacities, and because their only affirmative defense was that "they had acted entirely properly pursuant to the laws of the United States,"[31] working as a "federal administrative agent may have been legally quite treacherous."[32] Damages remedies against federal employees were once the norm.

Such private remedies remained available throughout most of the nineteenth century: While "immunity had been somewhat precariously established for cabinet level officials"[33] in the postbellum period, "lower level officers were protected only by a reasonableness defense in a few instances provided by statute. And other suits involving private rights, but indirectly challenging prior official determinations, followed the forms of private actions that hardly acknowledged their public law implications."[34] At the time the U.S. Constitution was written, a common law cause of action was simply presumed to exist, and for at least a century after the Constitution was framed, individuals could sue public officials who had violated their constitutional rights for damages.

The growth of the administrative state forced courts to balance the importance of smooth and continued government functioning against traditional

---

25. *Id.* One could point also to suppression of evidence and invalidation of a statute.

26. *Id.*

27. *See generally* Jerry L. Mashaw, *Recovering American Administrative Law: Federalist Foundations, 1787-1801*, 115 YALE L.J. 1256 (2006).

28. *Id* at 1341.

29. *Id.* at 1328.

30. *Id.* at 1322.

31. *Id.* at 1331.

32. *Id.* at 1325.

33. Jerry L. Mashaw, *Federal Administration and Administrative Law in the Gilded Age*, 119 YALE L.J. 1362, 1379 (2010) [hereinafter Mashaw, *Federal Administration*].

34. *Id.* (internal quotation marks omitted).

remedies in tort and spurred the development of official immunities.[35] This act of reframing around the necessity of effective governmental operations, rather than individual right, obstructed the existing common law remedies for constitutional violations.[36] *Bivens*, then, did not invent a new remedy so much as it reinstated an old one in a new version. As people's practical ability to sue errant federal actors diminished, the Court reintroduced a remedy that had been universally available when the Constitution was written.[37]

Bivens thus reinstated a previously available remedy, and it has come to "play[] a central role in our system of constitutional remedies."[38] As recent work by Alexander Reinert suggests, while commentators seem convinced that *Bivens* actions are futile and the Supreme Court appears hostile toward them, *Bivens* actions continue to be viable at lower court levels and secure approximately the same success rate as other kinds of civil rights cases.[39]

There are also other reasons to think that our system of government requires some version of constitutional damages. Justice Harlan's concurrence in *Bivens* famously noted that injunctions have no remedial effect for violations that are not ongoing or likely to be repeated, he stated, "For people in Bivens' shoes, it is damages or nothing."[40] The question becomes whether we prefer nothing over damages. That question leads squarely to what we think constitutes a right and what it entails—an issue the full exploration of which falls far beyond the scope of this Article. I will note, however, that if we think of a right as something that is real only insofar as it exists in the world—that is, to the extent as it has effects felt by individuals[41]—then a right fairly requires a remedy.[42] As Daryl Levinson

---

35. *See* Sina Kian, *The Path of the Constitution: The Original System of Remedies, How It Changed, and How the Court Responded*, 87 N.Y.U. L. REV. 132, 138-39 (2012).

36. *Id.* at 149-62.

37. *Id.* at 192.

38. Pfander & Baltmanis, *supra* note 4, at 117.

39. Alexander A. Reinert, *Measuring the Success of* Bivens *Litigation and Its Consequences for the Individual Liability Model*, 62 STAN. L. REV. 809, 842 (2010) ("The data suggest that if *Bivens* claims survive [the sua sponte dismissal of frivolous claims], their rate of success is somewhere in between the previously reported success rates for prisoner civil rights litigation and nonprisoner civil rights litigation."); *see also id.* at 833-36, 843-46 (describing the study's methodology for determining what constituted a *Bivens* case and what qualified as success). Reinert examined filings in New York (Eastern and Southern Districts), Texas (Southern District), Pennsylvania (Eastern District), and Illinois (Northern District) from 2001 to 2003. *Id.* at 832.

40. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 410 (1971) (Harlan, J., concurring). There may of course be other ways of ensuring that the guarantees of the Bill of Rights are enforced. One can imagine a comprehensive administrative system devoted to preserving individual constitutional rights and remediating breaches by government employees. Such an approach might even lead to more consistent remedies for individuals and lower violation rates among federal actors. So requiring some remedial options does not necessarily mean requiring damages actions. Still, in our adversarial, individualist system of rights enforcement, damages actions are the natural choice.

41. American pragmatism was perhaps the philosophical tradition that most clearly

has argued, if a right is defined by its existence in the world, then it not only requires, but is, in fact, *defined by* its remedy: "Rights are dependent on remedies not just for their application to the real world, but for their scope, shape, and very existence."[43] The basic idea is not new in American law—it appeared in *Marbury v. Madison*,[44] which itself quoted Blackstone for the proposition that "'it is a general and indisputable rule, that where there is a legal right, there is also a legal remedy . . . whenever that right is invaded.'"[45]

Alternatively, one can choose to see rights as pure ideas with no necessary worldly existence: noumenal, extrasensory forms. But insofar as rights in a democracy are seen to inhere in individuals and to be guaranteed by the government, this view creates a dilemma. If rights are only abstractions to which governments profess allegiance and not facts on the ground subject to enforcement and remediation, it becomes difficult to distinguish a right from a public relations strategy. And it becomes difficult for a government to credibly claim to guarantee a right whose breach gives rise to no remedies.[46] As Justice Harlan noted in *Bivens*, constitutional rights represent "social policies . . . aimed predominantly at restraining the Government as an instrument of the popular

---

enunciated a definition of reality as that which has effects in the world, but realists of many stripes would subscribe to this general approach. *See, e.g.*, Charles Sanders Peirce, *The Essentials of Pragmatism*, *in* PHILOSOPHICAL WRITINGS OF PEIRCE 251, 252 (Justus Buchler ed., 1955) ("[I]f one can define accurately all the conceivable experimental phenomena which the affirmation or denial of a concept could imply, one will have therein a complete definition of the concept, and *there is absolutely nothing more in it*."); LUDWIG WITTGENSTEIN, TRACTATUS LOGICO-PHILOSOPHICUS 7 (A.J. Ayer ed., D.F. Pears & B.F. McGuinness trans., 1922) ("The world is all that is the case."). Perhaps the best known American theorist of rights and remedies, Wesley Newcomb Hohfeld, also saw rights as defined by their effects in the world: He defined a right as something that creates a duty in another. WESLEY NEWCOMB HOHFELD, FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING AND OTHER LEGAL ESSAYS 36-38 (Walter Wheeler Cook ed., 1923).

42. One can imagine a system in which a right's effect in the world implicated an internal administrative law more than an individual adversarial process. *See, e.g.*, Mashaw, *Federal Administration*, *supra* note 33, at 1470. Our own system, however, is more focused on external enforcement by individuals.

43. Daryl J. Levinson, *Rights Essentialism and Remedial Equilibration*, 99 COLUM. L. REV. 857, 858 (1999); *see also id.* at 874 ("Rights are often shaped by the nature of the remedy that will follow if the right is violated. The definition of most or all rights incorporates 'remedial' prophylactic rules. And, perhaps most obviously, the cash value of any right is a function of the remedial consequences attached to its violation.").

44. 5 U.S. (1 Cranch.) 137 (1803).

45. *Id.* at 163 (quoting 3 WILLIAM BLACKSTONE, COMMENTARIES *23).

46. *See, e.g.*, David S. Law & Mila Versteeg, Sham Constitutions 6-7 (Nov. 1, 2001) (unpublished manuscript) (on file with author) ("[E]valuating the extent to which countries actually uphold the rights found in their constitutions" by investigating "the gap between what rights are guaranteed on paper and what rights are respected in practice," and concluding that those with the largest gap between promise and worldly effect have "sham constitutions.").

will,"[47] and it would be "at least anomalous" if the very entity restrained were also the only one allowed to create remedies for its own lack of restraint.[48] For Justice Harlan, the *Bivens* remedy rested on the premise that foxes are not the only acceptable guards for henhouses.[49]

While *Bivens* actions are not the only way constitutional guarantees can be vindicated, they fit both the history of constitutional rights vindication in America and the individualist approach to rights enforcement that our adversarial system fosters. And they fulfill a primary criterion for recognizing a right: providing a remedy for its violation.[50]

D. Article Outline

In what follows, I first trace the development of special factors analysis in the Supreme Court, from its origin as vague, off-hand dictum, to its culmination into a standard inquiry about congressional will. Part II presents, in some detail, how *Bivens* itself introduced the special factors, and then progresses through the Supreme Court's several stages of subsequent special factors analyses. Moving on to the federal courts of appeal, Part III shows that the common understanding of *Bivens*'s central problem as one of the separation of judicial from legislative powers has led to a fairly standard, though never fully articulated, approach. Almost uniformly, circuit courts treat the special factors analysis as an inquiry into whether Congress has indicated that it wishes to preclude *Bivens* remedies.

Part IV turns to outliers. One kind of outlier applies special factors analysis like everybody else, by looking for indications of legislative preclusion, but mistakenly finds congressional intent where none could have been implied. That kind of outlier stays within the parameters of the accepted doctrine and follows the special factor rationale, but it comes to unreasonable conclusions. The second kind of outlier deviates more seriously from the mainstream by failing to recognize that special factors analysis inquires into *legislative* will. The second kind of outlier turns *Bivens*'s separation of powers concern on its head by focusing instead on the Executive. *Arar v. Ashcroft*,[51] a Second Circuit case, exemplifies this mistake, and I examine it in detail to demonstrate how it deviates from the Supreme Court's established form of special factors analysis and

---

47. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 404 (1971) (Harlan, J., concurring). One could point also to evidence suppression and statute invalidation.

48. *Id.* at 403. In this passage, Justice Harlan lumped together the executive and the legislative branches as the popular government restrained by constitutional guarantees. Later *Bivens* jurisprudence made this distinction more clearly.

49. *Bivens* remedies are usually applied against executive officers, while it is Congress that creates the remedies. Justice Harlan's concurrence read the Constitution as restraining the will of both political branches. *Id.*

50. *See* Thomas R. Phillips, *The Constitutional Rights to a Remedy*, 78 N.Y.U. L. REV. 1309, 1310-14 (2003).

51. 585 F.3d 559 (2d Cir. 2009) (rehearing en banc).

misconstrues its underlying separation of powers concern. I also show that what the *Arar* court thought were *Bivens* special factors are, in fact, separate concerns that come with legal doctrines of their own, and I discuss how those concerns would have looked if the court had used the appropriate legal frameworks. Finally, the conclusion synthesizes the elements of special factors analysis generally shared among the circuit courts and reviews some of the stakes of getting it right.

As the Supreme Court has moved away from recognizing implied rights of action generally, it has also expressed ambivalence about the wisdom and the extent of *Bivens*. At least two sitting justices have strongly implied that *Bivens* was a mistake.[52] But the Supreme Court—those two Justices included—has also reserved for itself "the prerogative of overruling its own decisions," cautioning others against treating its own moods and apparent preferences as law.[53] In this vein, it has instructed lower courts to "follow [a] case which directly controls" even when that case "appears to rest on reasons rejected in some other line of decisions."[54] *Bivens* actions remain viable; this Article gives lower courts a principled way to assess them.

II. SPECIAL FACTORS ANALYSIS AT THE SUPREME COURT

A. Special Factors in Bivens

Bivens famously held that a "violation of [the Fourth Amendment guarantee against unreasonable searches and seizures] by a federal agent acting under color of his authority gives rise to a cause of action for damages" against the individual agent.[55] Webster Bivens was allowed to sue for damages, in their personal capacities,[56] the federal agents who had entered his home without a warrant and

---

52. *See, e.g.*, Wilkie v. Robbins, 551 U.S. 537, 568 (2007) (Thomas, J., concurring) ("*Bivens* is a relic of the heady days in which this Court assumed common-law powers to create causes of action." (quoting Correctional Servs. Corp. v. Malesko, 534 U.S. 61, 75 (2001) (Scalia, J., concurring))). As my brief historical discussion indicates, of course, the *Bivens* decision came in the wake of the Court's exercise of its common-law powers to expand official immunities.

53. Rodriguez de Quijas v. Shearson/Am. Express, Inc., 490 U.S. 477, 484 (1989).

54. *Id.*; *see also* Agostini v. Felton, 521 U.S. 203, 237 (1997).

55. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 389 (1971).

56. As Reinert explains:

> Personal capacity claims are brought against government officials individually, almost always for damages. In theory, defendants who are found liable in their personal capacity are responsible for paying damages out of their own pockets, although the federal government, like most states and municipalities, usually indemnifies employees for the damages awarded in constitutional tort actions. . . . Official capacity claims, by contrast, are brought nominally against government officials, but typically seek injunctive relief against a government entity that would otherwise be immune from suit in federal court.

arrested him.[57] Rejecting the defendants' contention that Bivens should sue for tort in state court, the Court held that a federal right should be vindicated in federal court.[58] In a seeming afterthought, the *Bivens* opinion added that "[t]he present case involves no special factors counselling hesitation in the absence of affirmative action by Congress,"[59] that is, in the absence of a congressionally created remedy.

Bivens did not elaborate on what it meant by special factors; it just gave a couple of examples.[60] Later cases would develop the idea more fully. Looking closely at this original presentation of special factors helps us see both how much the analysis has changed since *Bivens* was decided, and how its animating concerns have remained similar.

In the first example, *United States v. Standard Oil Co. of California*,[61] the Court had refused to create a new cause of action to allow the government to sue a corporation whose employee had injured a soldier.[62] Rejecting a tort law analogy, the Court concluded that the case primarily raised issues of the "federal fiscal policy" of expenditures on injured soldiers.[63] Tort-inflicted costs were hardly new, the Court reasoned, yet Congress, "the custodian of the national purse,"[64] which knew how to "take steps to prevent interference with federal

---

Reinert, *supra* note 39, at 811 n.2 (citations omitted).

57. Pfander notes that although the Court assumed Bivens's innocence, it is not clear that police lacked probable cause for the arrest, and it is likely that the warrantless entry and arrest were lawful under New York law at the time. Pfander, *The Story of* Bivens, *supra* note 2, at 290, 292-93. The larger point, of course, remains undisturbed: The Court was deciding only whether Bivens had a cause of action against allegedly unconstitutional treatment, not whether the officers had in fact acted unconstitutionally, nor whether New York law countenanced unconstitutional action.

58. Specifically, the Court explained that, were Bivens to sue in state court, the Fourth Amendment would not preserve an affirmative right; but, it would only "limit the extent to which the agents could defend . . . their actions [as] a valid exercise of federal power," leaving them to "stand before the state law merely as private individuals" if they could be "shown to have violated" it. *Bivens*, 403 U.S. at 390-91. But officers are not, the Court emphasized, merely private individuals: The power of the federal officer to enter a private home and enact an arrest differs from the authority of a private individual, and a person's ability to resist trespass by a private individuals differs from his ability to resist trespass by a federal agent—resistance that may itself be criminal. *Id.* at 391-95. Moreover, the Court stated, since the rights guaranteed by the Fourth Amendment are federal and do not depend on the provisions of state law, it made no sense to rely on state law to remedy their violation. *Id.* at 392-93. Finally, regarding its recognition of a remedy for Bivens, the Court noted that "[h]istorically, damages have been regarded as the ordinary remedy for an invasion of personal interests in liberty." *Id.* at 395.

59. *Id.* at 396. The standard spelling of "counseling" has apparently changed since 1971, and it appears spelled both with one "l" and with two in the case law.

60. *Id.*

61. 332 U.S. 301 (1947).

62. *Id.* at 314.

63. *Id.*

64. *Id.*

funds," had not devised a cause of action for that situation.[65] Creating a cause of action would improperly supplant congressional authority with an executive "fiscal and regulatory polic[y]."[66] *Standard Oil* rebuffed the Executive's request that the Court step outside its normal role as "arbiter[] between citizen[]" parties, each equally powerless to create a cause of action.[67] The Court reasoned that since the federal government could, itself, create the cause of action it requested, it was inappropriate for the Court to do so.[68] *Bivens* approvingly referred to *Standard Oil*'s refusal to intervene in "federal fiscal policy"[69] and its conclusion that "the United States [was] the party plaintiff to the suit. And the United States has the power at any time to create the liability."[70] *Bivens* also cited, but did not discuss, *United States v. Gilman*,[71] where the Court declined to extend common-law respondeat superior indemnification to the Federal Tort Claims Act (FTCA) when the federal government sued its own employee for indemnification.[72] The *Gilman* Court focused on Congress's expertise in the area, stating "[t]he relations between the United States and its employees have presented a myriad of problems with which the Congress over the years has dealt."[73]

Bivens did not explain what constituted special factors in these cases. However, the cases suggest that the Court declined to do for the other branches what they ought to do for themselves—deferring to congressional authority over the federal purse and deferring to congressional expertise in the area of federal employment.

Bivens's second example of a situation involving special factors was *Wheeldin v. Wheeler*,[74] in which the recipient of a House Un-American Activities Committee subpoena sued a committee investigator for improperly using his subpoena power.[75] The Court held that the allegations did not constitute a

---

65. *Id.* at 315-16.

66. *Id.* at 314-15.

67. *Id.* at 316.

68. *Id.* at 316-17. Although *Standard Oil* started out distinguishing the legislative from the executive branch, this later passage lumped them together. At the same time, *Standard Oil* also assumed that courts *could* themselves create remedies, especially in "the constitutional area." *Id.* at 313.

69. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 396 (1971) (quoting *Standard Oil*, 332 U.S. at 311).

70. *Id.* (alteration in original) (quoting *Standard Oil*, 332 U.S. at 316).

71. 347 U.S. 507 (1954).

72. *Id.* at 509-13.

73. *Id.* at 509.

74. 373 U.S. 647 (1963).

75. *Id.* at 648. Alleging a violation of his Fourth Amendment rights, the recipient claimed that the investigator was not authorized by the committee to subpoena him specifically, but had instead filled in, without receiving properly delegated authority to do so, a blank subpoena signed by the committee chairman, and that the statute authorizing the House Un-American Activities Committee to subpoena witnesses was itself unconstitutional. *Id.* at 648-49. He further alleged that the receipt of the subpoena, which was delivered to his workplace, caused him to lose his job and

constitutional claim because "there was neither a search nor a seizure of him. He was neither arrested nor detained . . . . In short, the facts alleged do not establish a violation of the Fourth Amendment."[76] As with any inquiry into whether a plaintiff has stated "a claim upon which" the relief demanded "can be granted,"[77] this example merely required that the remedy fit the right: Constitutional damages require a constitutional violation.[78] Later cases in the *Bivens* line adopt this requirement, not as a special factor, but as a simple element of the claim.[79]

Finally, *Bivens* noted that congressional action might obviate a constitutional damages remedy.[80] Rejecting the argument that a judge-made remedy was appropriate only if "*necessary* to enforce" a constitutional right,[81] the Court reasoned that the question was "merely whether petitioner . . . is entitled to redress his injury through a particular remedial mechanism normally available in the federal courts."[82] Presuming this normal availability, the Court furthermore found "no explicit congressional declaration" that people in Mr. Bivens's situation "may not recover money damages . . . but must instead be remitted to another remedy, equally effective in the view of Congress."[83] In *Bivens*, overcoming the presumption that constitutional damages were available required an "explicit congressional declaration" both making such a remedy unavailable and requiring the petitioner to pursue "another remedy [that was] equally effective in the view of Congress."[84] If Congress barred a damages remedy and directed harmed individuals to some other relief that it viewed as an equivalent to damages, the Court should pause and consider that fact before deciding whether to grant a damages remedy.[85]

While *Bivens* did not label this third situation as a special factor, its concerns are congruent with those of the first. Together, the two provide the enduring rationale behind special factors analysis: a deference to congressional authority and expertise over the federal budget, federal employees, and the creation of remedies.

---

to be stigmatized, leading to economic and emotional damages. *Id.* at 648.

76. *Id.* at 649-50.

77. FED. R. CIV. P. 12(b)(6).

78. *See* Carey v. Piphus, 435 U.S. 247, 254 (1978).

79. *See, e.g.*, Davis v. Passman, 442 U.S. 228, 248 (1979) ("Moreover, a plaintiff seeking a damages remedy under the Constitution must first demonstrate that his constitutional rights have been violated. We do not hold that every tort by a federal official may be redressed in damages.").

80. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 396 (1971)

81. *Id.* at 397 (emphasis added).

82. *Id.*

83. *Id.*

84. *Id.*

85. As discussed below, the term "counselling hesitation" has come to mean something more like "prohibiting" a damages remedy. *Id.* at 396. This later meaning, however, which is the product of decades of case law, is hard to read back into *Bivens* itself—"counsel," after all, does not usually mean command, and "hesitation" does not usually mean cessation.

B. Explicit Preclusion (1979-1983)

In the decade following *Bivens*, the Court explained that there were "two situations" in which a *Bivens* claim "may be defeated": when "defendants demonstrate 'special factors counselling hesitation,'"[86] and when they "show that Congress has provided an alternative remedy which it explicitly declared to be a *substitute* for recovery directly under the Constitution and viewed as equally effective."[87] At this stage, the special factor label primarily applied to the

86. Carlson v. Green, 446 U.S. 14, 18 (1980) (quoting *Bivens*, 403 U.S. at 396).

87. *Id.* at 18-19 (citing *Bivens*, 403 U.S. at 396). This Article focuses specifically on the development of the special factors analysis, rather than on the development of *Bivens* jurisprudence as a whole. Scholarship dealing with the remedy more broadly tends to present its history as one of a brief initial decade of expansion, during which, in addition to *Bivens*'s initial recognition of constitutional damages for Fourth Amendment violations, the Court recognized remedies for violations of Fifth Amendment due process rights to equal treatment for women and Eighth Amendment rights regarding medical attention while incarcerated. *See id.*; Davis v. Passman, 442 U.S. 228, 245-48 (1979). After 1980, however, the Court's tide turned against *Bivens*, and from then on, *Bivens* cases primarily declined to recognize, rather than expand, individuals' remedies against federal employees. In 1983, the Court held unanimously that military personnel had no *Bivens* claim against their superior officers for racial discrimination. Chappell v. Wallace, 462 U.S. 296, 304 (1983). Additionally, and again unanimously, the Court held that federal employees had no *Bivens* claim against superiors who violated their First Amendment free speech rights. Bush v. Lucas, 462 U.S. 367, 388-89 (1983). In 1988, the Court declined to recognize that people wrongly denied social security disability benefits had a *Bivens* due process claim under the Fifth Amendment. Schweiker v. Chilicky, 487 U.S. 412, 425, 428-29 (1988). In 1994, the Court held that only individuals, not agencies, could be sued under *Bivens*. FDIC v. Meyer, 510 U.S. 471, 485-86 (1994). And in 2001 it held, analogously, that a private corporation was not liable under *Bivens* suit either. Correctional Servs. Corp. v. Malesko, 534 U.S. 61, 66 (2001). In 2007, the Court declined to recognize *Bivens* actions for violations of the Fifth Amendment's Takings Clause, although it did not unequivocally deny *Bivens* remedies to all takings claims. Wilkie v. Robbins, 551 U.S. 537 (2007). And most recently, in *Minneci v. Pollard*, the Court held that employees of a privately ran federal prison were not subject to *Bivens* suits because state law provided alternative remedies. Minneci v. Pollard, 132 S. Ct. 617 (2012); *see also* Alexander A. Reinert & Lumen N. Mulligan, *Asking the First Question: Reframing* Bivens *After* Minneci, 90 WASH. U. L. REV. (forthcoming 2013), *available at* ssrn.com/abstract=2042175.

Bivens jurisprudence is often described as initially expanding to include certain constitutional provisions and then ceasing to expand to others. It is not clear, however, why, if a damages remedy is available for violations of some constitutionally guaranteed individual rights, it should not be presumptively available for all. After all, the Constitution does not guarantee some individual rights less vigorously than others. (Instead, it would make more sense to draw the line at the Constitution's structuring and constraint of government action, which does not guarantee individual rights.) This history is perhaps better interpreted as successive determinations about the kinds of *situations*—the kinds of relations between plaintiffs and defendants—that *Bivens* may encompass, rather than about the particular constitutional *provisions* it can vindicate.

question of whether defendants "enjoy[ed] such independent status in our constitutional scheme . . . that judicially created remedies against them might be inappropriate."[88]

Thus in *Carlson v. Green*, the Court determined that prison guards receive no constitutional protection from suit beyond qualified immunity, which ensures that a lawsuit does not "inhibit their efforts to perform their official duties."[89] Although in *Davis v. Passman*[90] the defendant's role as a former Congressman was a "special concern[] counseling hesitation," the Court determined that his immunity from suit for "actions taken in the course of his official conduct" was sufficiently addressed by qualified immunity and "the protections . . . [of] the Speech or Debate Clause" of the Constitution.[91] Although the defendant's unusual position in *Davis* counseled hesitation, it did not preclude a remedy.

The Court also continued to assess whether an "*explicit* congressional declaration" precluded constitutional damages.[92] *Davis*, for instance, reasoned that a plaintiff excluded from explicit statutory remedies must have a constitutional damages remedy.[93] The *Davis* plaintiff was a congressional employee not subject to Title VII protections.[94] The Court determined that there was "no evidence . . . that Congress meant . . . to foreclose alternative remedies . . . . On the contrary, [the statute] leaves undisturbed whatever remedies petitioner might otherwise possess."[95] Similarly, *Carlson* held that the plaintiff inmate's ability to sue the government under the FTCA did not preclude his constitutional damages suit against the individuals who injured him.[96] On the contrary, the Court concluded that the Senate Report accompanying the 1974 amendments to the FTCA "made it crystal clear that Congress views FTCA and *Bivens* as parallel, complementary causes of action."[97]

During this stage of *Bivens* jurisprudence, in sum, the Court presumed that a constitutional damages remedy was available unless (1) the defendant was so "independent" within the "constitutional scheme" that judicial remedies against

88. Carlson v. Green, 446 U.S. 14, 19 (1980) (holding that an inmate's ability to bring an action under the FTCA did not preclude a *Bivens* suit for violations of his Eighth Amendment rights and addressing what law governed such an action by an inmate's estate).

89. *Id.*

90. 442 U.S. 228 (1979).

91. *Id.* at 246.

92. *Id.* at 246-47 (internal citations omitted).

93. *Id.* at 245.

94. *See* 42 U.S.C. § 2000e-16(a) (2006) (protecting only employees in the competitive civil service).

95. *Davis*, 442 U.S. at 247. As a non-competitive employee (that is, not a regular civil service employee but someone hired and fired by the Congressman), the plaintiff had no statutory or administrative remedy. *Id.* at 247 n.26. And because the defendant Congressman had since left office, the plaintiff could seek no equitable relief either. *Id.* at 231.

96. Carlson v. Green, 446 U.S. 14, 20 (1980); *see also* Pfander & Baltmanis, *supra* note 4, at 132-33.

97. *Carlson*, 446 U.S. at 20; *see also* Pfander & Baltmanis, *supra* note 4.

him would upset the federal government's balance of powers;[98] or (2) Congress had explicitly provided some other, equally effective, remedy.[99] While either of these factors could preclude a *Bivens* remedy, only the first was explicitly called a special factor. The second still constituted its own inquiry into the extent of "affirmative action by Congress."[100] Later cases, in contrast, began to treat the existence of congressional remedies as a special factor as well.[101]

C. Implicit Preclusion (1983-2007)

Special factors analysis changed dramatically over the following decades. The Court abandoned its interest in a defendant's place in the constitutional scheme and in the power differences between the parties. It focused instead on the separation of powers and relative institutional competence between the legislative and the judicial branches. Instead of requiring an explicit legislative substitute for constitutional damages, the Court began to look for implicit indications of legislative preference. The Court also abandoned the presumption of constitutional damages, turning instead to case-by-case evaluations. Instead of a consideration that counseled hesitation, special factors became an obstacle that commanded cessation.[102]

Two unanimous 1983 decisions started these new trends. One, *Bush v. Lucas*,[103] declined to recognize a *Bivens* remedy for a federal employee, holding that the "elaborate, comprehensive scheme" protecting such employees, within which the plaintiff's claims were "fully cognizable," provided sufficient recourse.[104] In deferring to Congress's special authority over federal employees,

---

98. *Carlson*, 446 U.S. at 19.

99. *Id.* at 18-19.

100. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 396 (1971).

101. *See, e.g.*, Moore v. Glickman, 113 F.3d 988, 991 (9th Cir. 1997) ("Implied preclusion of a *Bivens* action can be found when defendants can demonstrate the existence of 'special factors counselling hesitation in the absence of affirmative action by Congress.'" (quoting Schweiker v. Chillicky, 487 U.S. 412, 421 (1988))).

102. There is another, more legal realist, way to tell this same history. Simply put, after the expansive rights-creation of the Warren years, the Supreme Court limited and constrained constitutional damages just as it limited and constrained other rights created or enhanced in the 1960s and 1970s. Though I focus on a more doctrinally intricate explanation, I also agree with the legal realist version of the history: The two are not mutually contradictory. What I explain in shorthand is that the legal realist approach looks to the result—the contraction, or at least the stopped expansion, of rights. I look to how the Court gets there—drawing on preemption analysis to create a somewhat analogous approach at the federal level, it retained its focus on the separation of judicial and *legislative* powers.

103. 462 U.S. 367 (1983).

104. *Id.* at 385-86. The plaintiff in *Bush v. Lucas* was a NASA employee who had been demoted for publicly criticizing the agency. *Id.* at 369. He first sought administrative remedies available to him under the civil service laws, including a final layer of review, completed several

Bush was in line with *Gilman*, one of the cases cited by *Bivens* when it introduced special factors, which refused to intrude on the federal employee relationship.[105] But *Bush* added something new as well: implicit preclusion. Congress, the Court stated, "may . . . indicate its intent, by statutory language, by clear legislative history, or perhaps even by the statutory remedy itself."[106] After *Bush*, courts were responsible for discerning an implicit congressional desire to preclude constitutional damages. Such implicit preclusion would decisively constitute a special factor.

In this same process, the Court abandoned *Bivens*'s concern with power differentials between the parties and focused instead on the relative institutional competence of different branches of government. The Court refused to "augment[]" Congress's "elaborate remedial system . . . constructed step by step, with careful attention to conflicting policy considerations" with a judge-made remedy.[107] Deferring to congressional competence in the federal employee realm, *Bush* also compared the institutional competence of Congress and the Executive.[108] Discussing *Standard Oil*, which was the original example of special factors relied on in *Bivens*, *Bush* explained the Court has "refused to create a damages remedy" because such a remedy would have been "the instrument for . . . establishing the federal fiscal and regulatory policies which the . . . executive . . . thinks should prevail" but which was, in fact, for Congress to determine.[109] *Bush* thus emphasized that the executive branch may have policy preferences, but it is Congress, not the judiciary, that implements them in laws.

The other 1983 decision marked an even more radical change in the special factors analysis. *Chappell v. Wallace*[110] announced for the first time that "a [*Bivens*] remedy will not be available when 'special factors counselling hesitation' are present."[111] This announcement was styled as a description, not a holding.[112] Whether the very existence of special factors precludes a *Bivens* remedy or merely gives rise to a more careful inquiry on the part of courts—as it had in pre-*Chappell* cases—does not yet have a definitive answer. But,

---

years after the demotion, in which an agency review board recommended his reinstatement with back pay; NASA followed the recommendation. *Id.* at 371. While this final administrative appeal was pending, Bush sued the superiors who had demoted him for violating his First Amendment right to free speech. *Id.*

105. *See supra* note 72 and accompanying text.

106. *Bush*, 462 U.S. at 378.

107. *Id.* at 388.

108. *Id.* at 379.

109. *Id.* (quoting United States v. Standard Oil Co. of Cal., 332 U.S. 301, 314 (1947)).

110. 462 U.S. 296 (1983).

111. *Id.* at 298 (quoting Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 396 (1971)).

112. The *Chappell* opinion asserted that "[t]he Court, in Bivens and its progeny, has expressly cautioned" that special factors would have this preclusive effect, *id.*, but as my discussion analysis demonstrates, this was not accurate. In fact, before *Chappell*, *Bivens* and its progeny treated special factors as something that a court should consider, not something that should bring a court up short.

building on *Chappell*'s confident proclamation, courts lean toward preclusion.[113]

Chappell also introduced what might be termed a field preclusion approach to special factors. In *Chappell*, military personnel sued their superiors for discrimination on the basis of race.[114] *Chappell* declined to recognize a remedy for military servicepeople against military superiors by analogy to the *Feres* doctrine, which exempts the federal government from suit under the FTCA for harms sustained "incident to [military] service."[115] The *Feres* doctrine has been justified through the requirements of military discipline and its necessity to national security.[116] *Chappell* further emphasized Congress's plenary control over, and wide-ranging regulation of, military affairs, which had effectively occupied the field of remedy creation in the military context.[117] "[T]he unique disciplinary structure of the [m]ilitary [e]stablishment and Congress' activity in the field constitute 'special factors'" making a *Bivens* remedy "inappropriate."[118] The Court thus read congressional intent off of congressional silence on the issue, and interpreted its own authority with reference to the arena occupied by Congress.[119]

The Court did not explicitly acknowledge the changes these cases wrought in the special factors analysis, but treated them as mere explications of an existing approach. Of course, unacknowledged changes to existing doctrine are not unusual. They are, rather, how our precedential system of law-making works. In this case, however, the changes wrought by the 1983 cases may have been so

---

113. *Compare, e.g.*, Arar v. Ashcroft, 585 F.3d 559, 574 (2d Cir. 2009) (asserting that all that is required to preclude a *Bivens* suit is a single factor counseling hesitation, and concluding that the "relevant threshold" for preclusion is thus "remarkably low"), *with id.* at 583 (Sack, J., concurring in part and dissenting in part) ("[T]he applicable test is not whether 'special factors' exist, but whether after 'paying particular heed to' them, a *Bivens* remedy should be recognized with respect to at least some allegations in the complaint."). It is especially difficult to reconcile *Chappell*'s characterization of special factors with *Davis*, where a *Bivens* action was approved despite the acknowledged existence of a special factor. *Chappell* did not purport to overrule *Davis*.

114. *Chappell*, 462 U.S. at 297.

115. *Id.* at 299 (alteration in original) (citing Feres v. United States, 340 U.S. 135, 138 (1950)).

116. The reasoning of *Chappell* was extended in *United States v. Stanley*, 483 U.S. 669 (1987), which held that *Bivens* was unavailable not only to subordinates suing their military superiors, but in any case that implicated military discipline where the injury occurred incident to military service. *Id.* at 683-84. *Stanley* effectively equated the *Bivens* special factor analysis in cases implicating the military with *Feres*'s 'incident to service' analysis. *Id.* at 678-84.

117. *Chappell*, 462 U.S. at 299.

118. *Id.* at 304.

119. This approach, of course, contrasts sharply with that of *Davis*, in which the Court held that the omission of the injured party from a remedial scheme left her with access to any otherwise existing remedies. Davis v. Passman, 442 U.S. 228 (1979); *see supra* note 79 and accompanying text. It also pairs oddly with *Bush*, handed down the same day, which denied a *Bivens* remedy precisely because the injured party *was* included in the remedial scheme. Bush v. Lucas, 462 U.S. 367 (1983); *see supra* note 104 and accompanying text.

subtle that even some members of the Court did not recognize them at the time.

These changes would be cemented five years later in *Schweiker v. Chilicky*.[120] Unlike *Bush* and *Chappell*, the plaintiffs in *Chilicky* were not federal employees. They were disabled people who had been wrongfully denied federal disability benefits under a program that put pressure on administrators to terminate ongoing disability claims.[121] Through administrative redress procedures, their benefits had been reinstated, with back pay.[122] The plaintiffs claimed, however, that because they depended on these benefits to sustain themselves, restitution and back pay alone could not remedy the damage caused by going without basic life necessities for months on end.[123] They also argued that administrative process could not address the constitutional violations of administrators "adopt[ing] illegal policies that led to the wrongful termination of benefits by state agencies."[124] Although the plaintiffs were not federal employees, the Court concluded that the case "cannot reasonably be distinguished from *Bush v. Lucas*"[125] because the Social Security Act provided a comprehensive scheme governing social security benefits.[126] Because this "suggest[ed] that Congress ha[d] provided what it consider[ed] adequate remedial mechanisms for constitutional violations that may occur in the course of [the program's] administration," *Bivens* was not available even if the scheme did not provide remedies for the particular kind of wrong at issue.[127]

Chilicky cemented the new special factors implicit preclusion approach, stating that "the concept of 'special factors counselling hesitation in the absence of affirmative action by Congress' has proved to include an appropriate judicial

---

120. 487 U.S. 412 (1988).

121. *Id.* at 417.

122. *Id.*

123. *Id.* at 418.

124. *Id*. The opinion details a congressional investigation of the Social Security Agency's Continuing Disability Review (CDR), under which claimants who appealed the denial of benefits were denied benefits pending administrative review. Administrators reviewing disability claims had been pressured to deny them even though "benefits were too often being improperly terminated by state agencies, only to be reinstated by a federal administrative law judge." *Id.* at 415. The inquiry found that the program had had disastrous consequences, improperly denying support to "hundreds of thousands of truly disabled Americans." *Id.* at 416 (quoting 130 CONG. REC. 26000, 13234 (1984) (statement of Sen. Moynihan)). "There is little doubt," *Chilicky* concluded, "that CDR led to many hardships and injuries that could never be adequately compensated." *Id.* at 417. *Chilicky*'s named plaintiff, for instance, "was in the hospital recovering from open-heart surgery when he was informed that his heart condition was no longer disabling." *Id*. at 418. The *Chilicky* plaintiffs alleged that administrators had knowingly failed to use public and equal criteria to judge their claims, ignored evidence of their disabilities, selected biased doctors to perform review, and imposed quotas requiring that a certain number of people be denied the continuation of benefits regardless of the actual state of their disability. *Id.* at 418-19.

125. *Id.* at 425.

126. *Id.* at 424-25.

127. *Id.* at 423.

deference to indications that congressional inaction has not been inadvertent."[128] After *Chilicky*, courts focused less on the sufficiency and explicitness of existing remedies and more on the simple question of whether Congress had "enacted a comprehensive . . . scheme governing the area involved," and on "Congress'[s] expertise and authority in the field in question."[129] If there was a comprehensive scheme, then a "failure to provide a damages remedy for constitutional violations" would be interpreted as "deliberate rather than inadvertent."[130] This effectively reversed the earlier presumption that constitutional damages would be available in situations where Congress had "heavily regulated a certain subject . . . but ha[d] said nothing about a right of action for constitutional violations."[131]

Dissenting from *Chilicky*, Justice Brennan, who had written the opinions in *Bivens*, *Davis*, and *Carlson* and had joined the opinions in *Bush* and *Chappell*, acknowledged for the first time the changes wrought by the 1983 cases. "I find it inconceivable," he wrote, "that Congress meant by mere silence to bar all redress for such injuries."[132] There were two important distinctions, he explained, between the *Bush* and *Chappell* situations and the one in *Chilicky*.[133] First, the remedial schemes at issue in *Bush* and *Chappell* provided relief for constitutional violations.[134] In *Chilicky*, "by stark contrast," the disability "recipients cannot even raise constitutional challenges to agency action . . . and . . . can recover no consequential damages whatsoever."[135] Second, military personnel and federal employees stood in a unique relation to the legislature: Congress has plenary power over the first, and "enjoys a special expertise" regarding the second.[136] It was the unique relationship of Congress to the military and the civil service, not simple "congressional authority over a given subject," that were the special factors in *Chappell* and *Bush*.[137] But Congress has no special relationship to, and no special expertise with, disabled people and their benefits. It was "simply competent to legislate in [that] area" as in any other.[138]

As Justice Brennan's dissent demonstrates, *Chilicky* entrenched the new style of special factors analysis introduced in *Chappell* and *Bush*. No longer would the Court require an explicit legislative substitute to preclude constitutional damages, and no longer would it even require Congress to have any special expertise in, or relationship to, the subject or people covered by a remedial scheme. This new approach greatly expanded the range of situations in which courts would find *Bivens* remedies precluded. At the same time, the new approach rested on the

---

128. *Id.*
129. Jones v. Tenn. Valley Auth., 948 F.2d 258, 264 (6th Cir. 1991).
130. *Id.*
131. McIntosh v. Turner, 861 F.2d 524, 525 (8th Cir. 1988).
132. *Chilicky*, 487 U.S. at 432 (Brennan, J., dissenting).
133. *See id.* at 436-42.
134. *Id.* at 436-37.
135. *Id.* at 437.
136. *Id.* at 441.
137. *Id.* at 442.
138. *Id.*

same foundation as the previous one: The Court still looked to Congress to provide an alternative remedial scheme or to delegate the provision of such a scheme to an administrative agency. The separation of powers issue at the heart of *Bivens* remained the separation of judicial from legislative authority.

D. Manageability (2007)

One final case added a new concern to *Bivens* decisions while retaining the components of special factors analysis. *Wilkie v. Robbins*[139] has caused considerable confusion, but did not disturb the existing framework for special factors preemption analysis. *Wilkie* introduced for the first time what it called a "familiar sequence" for determining whether a *Bivens* remedy should be available.[140] First, a court should ask "whether any alternative, existing process for protecting the interest amounts to a convincing reason for the [j]udicial [b]ranch to refrain from providing a new and freestanding remedy in damages."[141] This first step harks back to older cases' requirement that Congress provide some adequate alternative remedy if *Bivens* is to be precluded. If no such process exists, *Wilkie*'s second step requires a court to "make the kind of remedial determination that is appropriate for a common-law tribunal, paying particular heed, however, to any special factors counselling hesitation before authorizing a new kind of federal litigation."[142] Appending special factors analysis as a subsection of the common law determination in this way can lead to the confusion of courts that do not realize that each inquiry has its own jurisprudential tradition. A careful reading of *Wilkie*, however, reveals that the special factors analysis remains the same as it was before the Court announced its decision.

In *Wilkie*, a landowner sued agents of the Bureau of Land Management (BLM) for trying to force him to grant an easement over his land in abrogation of his Fifth Amendment right to just compensation for a taking of property.[143] The previous owner had granted the BLM an easement, but BLM officials neglected to record the grant; under local law, the easement was therefore not passed on to the new landowner.[144] Robbins, the plaintiff, alleged that when he refused to simply grant the easement with no compensation, BLM officials

---

139. 551 U.S. 537 (2007).

140. *Id.* at 550.

141. *Id.* (citing Bush v. Lucas, 462 U.S. 367, 378 (1983)).

142. *Id.* (internal quotation marks omitted) (citing *Bush*, 462 U.S. at 378).

143. *Id.* at 541. *Wilkie* declined to recognize a cause of action for a Fifth Amendment Takings Clause violation, but it did not categorically exclude the Takings Clause from *Bivens*'s ambit. Rather, it found *Bivens* inappropriate in that particular case. *Id.* at 562. *Wilkie* did not, for instance, question, or even mention, a Fourth Circuit case that recognized a Takings Clause *Bivens* remedy. *See* Dunbar Corp. v. Lindsey, 905 F.2d 754 (4th Cir. 1990) (recognizing a constitutional damages remedy under the Fifth Amendment's Takings Clause for a private company with a possessory interest in property taken over by the government).

144. *Wilkie*, 551 U.S. at 542-43.

launched a multi-year campaign to force him to do so or to force him out of business.[145] The campaign included both acts that were themselves illegal (such as trespass) and acts that were otherwise legal but that were undertaken for an illegal purpose (such as selectively refusing to grant Robbins standard permits necessary to run his business).[146]

After determining that Robbins had no statutory or administrative remedy,[147] *Wilkie* went on to the new second step of the *Bivens* analysis, weighing the interests at stake while remaining cognizant of the possibility of special factors.[148] This part of *Wilkie*[149] has sometimes been read to be *about* special factors, but the opinion itself did not purport to find any special factors. In fact, it never mentioned the term again. This makes sense, as the Court expressly found that no remedial scheme encompassed the wrongs of which Robbins complained but did not find that Congress, in failing to legislate remedies for these kinds of harms, had purposely excluded plaintiffs like Robbins from access to remedies.[150] Thus, the Court's analysis showed that the existing requirements for finding special factors had not been met.

Rather than rely on special factors, *Wilkie* upheld the dismissal of Robbins's claims on two grounds. First, the Court concluded that allowing the suit would enable a flood of lawsuits from people disgruntled with their interactions with bureaucrats.[151] Justice Ginsburg's dissent called this a new kind of "special factor."[152] But, as a prudential concern, the floodgate argument is better seen as part of the weighing of interests "that is appropriate for a common-law tribunal."[153] There is, as Justice Ginsburg herself pointed out, no precedent for the floodgate argument in a special factors analysis.[154] And the opinion itself did

145. *Id.* at 543.

146. *Id.* at 543-47.

147. *Id.* at 554 (summarizing the argument for granting a *Bivens* remedy as "the inadequacy of discrete, incident-by-incident remedies").

148. *Id.*

149. *Id.* at 577 (Ginsburg, J., dissenting).

150. *Id.* at 555 (majority opinion).

151. While "[i]t is true that the Government is no ordinary landowner," the Court concluded, allowing constitutional damages against agents acting on the government-as-landowner's behalf "would invite claims in every sphere of legitimate governmental action affecting property interests, from negotiating tax claim settlements to enforcing Occupational Safety and Health Administration regulations." *Id.* at 558, 561.

152. *Id.* at 577 (Ginsburg, J., dissenting).

153. *Id.* at 550 (majority opinion) (quoting Bush v. Lucas, 462 U.S. 367, 378 (1983)).

154. *Id.* at 577 (Ginsburg, J., dissenting) (calling the floodgates argument "a special factor counseling hesitation quite unlike any we have recognized before"). Indeed, *Bivens* itself had rejected a floodgates argument without suggesting that it might be a special factor. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 391 n.4 (1971) ("In estimating the magnitude of any such 'avalanche [of new cases],' it is worth noting that a survey of comparable actions against state officers under 42 U.S.C. § 1983 found only 53 reported cases in 17 years (1951-1967) that survived a motion to dismiss. Increasing this figure by 900% to allow for

not refer to it as such. Rather, the floodgate argument appeared in the part of the opinion devoted to common-law weighing of relevant interests.

Second, the Court concluded that a *Bivens* remedy would force courts to draw hard-to-see lines, having to distinguish between bureaucratic conduct that zealously pursues the public interest and that which oversteps zeal and violates constitutional rights.[155] Such line-drawing would tax the Court's institutional competence because Robbins challenged "actions that, on their own, fall within the [g]overnment's enforcement power," and because he challenged a multi-year accrual of pressure rather than discrete acts.[156] The Court lacked the competence to make a remedy in this case because of "the elusiveness of a limiting principle for Robbins's claim."[157] In that context, the Court concluded, "'Congress is in a far better position than a court to evaluate the impact of a new species of litigation' against those who act on the public's behalf."[158]

This second ground concludes that courts are ill-equipped to deal with claims against long-term campaigns of largely lawful action taken with unlawful intent: It is simply too difficult to define what the violation is. *Wilkie*'s second ground thus says less about whether a particular *remedy* is appropriate and more about a court's ability to recognize a particular kind of *claim*. The problem with Robbins's suit was not that it would be inappropriate to provide a remedy for a Fifth Amendment wrong but that it was unclear how to define, or even identify, the wrong that had been committed against him.[159] The *Wilkie* opinion does not call this a special factor. If anything, this definitional issue harks back to *Bivens*'s requirement that constitutional damages remedies be given only where plaintiffs successfully allege violations of a constitutional magnitude—more a threshold pleading requirement than a special factor.

The two grounds on which the Court resolved *Wilkie*, then, were both aimed at making claims practically and conceptually manageable. Practically, recognizing a remedy should not give rise to a large and unnecessary amount of new litigation. Conceptually, a remedy should be recognized only for injuries that fit within the normal parameters of a legal claim. *Wilkie* did not discuss the relative authority of the courts and the legislature, and it never doubted that the Court had the authority to recognize a remedy. Rather, it doubted the Court's *ability* to handle the remedy it could create.[160] This concern fits more comfortably into the common-law weighing of interests that *Wilkie* introduced than the separation of powers concerns of special factors analysis. *Wilkie*, in sum, found that intra-judicial concerns raise sufficient grounds to bar a *Bivens*

---

increases in rate and unreported cases, every federal district judge could expect to try one such case every 13 years." (internal citation omitted)).

155. *Wilkie*, 551 U.S. at 562.

156. *Id.* at 560.

157. *Id.* at 561 n.11.

158. *Id.* at 562 (quoting Bush v. Lucas, 462 U.S. 367, 389 (1983)).

159. *Id.* at 561 n.11.

160. *Id.* at 562.

remedy.[161] It did not turn to the inter-branch concerns that special factors implicate and, in the end, simply did not perform a special factors analysis. It may be that the Court will eventually subsume the special factors analysis into *Wilkie*'s common-law weighing, but if that is the case, the Court has not made any such movement clear. Thus, the special factors analysis after *Wilkie* remains unchanged.

III. Consolidating Special Factors Analysis in the Courts of Appeal

Part II showed how special factors emerged in *Bivens* as an inchoate reference, loosely oriented around ensuring that judicial remedies did not encroach on the prerogatives of Congress. The concept gained definition over the ensuing decades, coming into sharper focus in the 1980s, when the Court shaped its inquiry into one of congressional will, with techniques drawn from preemption analysis. When Congress explicitly provided alternative remedies, legislatively occupied the field, or implicitly indicated that no additional remedies should be provided, the Court refused to recognize a remedy. Because the Court developed this approach over only a few cases, the special factors analysis at the Supreme Court level retains somewhat vague outlines. The latest case to touch upon it, *Wilkie*, mentioned special factors without actually utilizing the concept.[162]

The Federal Courts of Appeal have lent coherence and specificity to these somewhat vague outlines by keeping the focus of special factors analysis consistently on congressional will. In the past two decades, the circuit courts have consistently based the availability of *Bivens* remedies on whether a comprehensive statutory remedial scheme addressed the situation at issue. While the courts have been consistent in how they actually approach special factors, courts of appeal have *described* the analysis in a variety of ways. Some explicitly look to remedies available by statute. Others say that they consider the categories of parties in the case. Still others rely on the simple fact that some law touches on the case without examining whether it provides a relevant comprehensive scheme. Sometimes courts combine these phrasings. As I demonstrate in this Part, however, what appear as different approaches are, in fact, merely different ways of describing the same underlying question: whether Congress has indicated that it wishes the judiciary not to create remedies in a particular situation. Sometimes explicitly, sometimes unwittingly, courts of appeal have converged on and developed the implicit preclusion style of special factors analysis that the Supreme Court introduced in the 1980s.

A. *Self-Conscious Implicit Preclusion Analysis*

Most appellate courts recognize that the special factors analysis serves to prevent courts from treading on ground already claimed by Congress. Some courts find it important to determine whether a statutory remedial scheme

161. *Id.* at 577 (Ginsburg, J., dissenting).
162. *See supra* Part II.D.

specifically addresses constitutional violations while others do "not require a foray into the meaningfulness of [a plaintiff's] remedies under the [statutory scheme]," and simply "focus . . . on the comprehensive nature of the administrative system protecting the rights of the plaintiff, as well as Congress'[s] expertise and authority in the field in question."[163] Generally, courts do not ask whether the statutory scheme affords relief comparable to what a plaintiff would have under *Bivens* or even whether it affords any relief at all. The question instead is whether Congress appears to have considered the kind of claim at issue and has indicated a preference about how to handle it.[164] Courts interpret both inclusion in, and exclusion from, a statutory remedial scheme as a statement of congressional policy. Thus, for instance, because the statutory remedial scheme relevant to a particular case *addressed* violations of constitutional rights but did not provide damages for them, the Seventh Circuit concluded that Congress had intended to preclude a constitutional damages remedy.[165]

A small number of statutory schemes preclude *Bivens* most frequently. Chief among these is the Civil Service Reform Act (CSRA), which provides a complex remedial framework for public employees.[166] Courts agree that Congress has occupied the field of federal employment and find that the CSRA precludes *Bivens* remedies both where it provides federal employees with redress and where it does not. For instance, the CSRA has precluded a *Bivens* suit by a federal educator, a kind of employee not covered by the full range of CSRA protections[167] and by an employee of the federal judiciary, to whom the CSRA

---

163. Jones v. Tenn. Valley Auth., 948 F.2d 258, 264 (6th Cir. 1991) (citing Feit v. Ward, 886 F.2d 848, 854-55 (7th Cir. 1989)) (providing examples from the U.S. Supreme Court, as well as the Eighth, Ninth, and D.C. Circuits).

164. Dotson v. Griesa, 398 F.3d 156, 166 (2d Cir. 2005).

165. Feit v. Ward, 886 F.2d 848, 853-54 (7th Cir. 1989).

166. Pub. L. No. 95-454, 92 Stat. 1111 (1978) (codified as amended in scattered sections of 5 U.S.C.). Scholarship suggests that CSRA remedies are in fact inadequate to address constitutional claims, especially under the First Amendment. *See, e.g.*, John F. Preis, *Constitutional Enforcement by Proxy*, 95 VA. L. REV. 1663, 1714 (2009) ("A recent study of First Amendment claims by federal employees suggests that the 'administrative scheme [utilized in such cases] is not vindicating the First Amendment . . . rights of federal employees.'" (alterations in original) (quoting Paul M. Secunda, *Whither the* Pickering *Rights of Federal Employees?*, 79 U. COLO. L. REV. 1101, 1103 (2008))). As has become clear through subsequent *Bivens* jurisprudence, courts are loath to investigate the *adequacy* of remedies that a statutory scheme provides as long as the statutory scheme indicates congressional attention to the matter. One could, however, argue that inadequate remedies indicate that Congress did not, in fact, exercise its expertise or imply its will in a particular area.

167. *See* Volk v. Hobson, 866 F.2d 1398, 1403-04 (Fed. Cir. 1989); *see also* Sarullo v. U.S. Postal Serv., 352 F.3d 789, 795-99 (3d Cir. 2003) (holding that postal service workplace investigation of plaintiff who was suspected of selling drugs at work constituted a work-related process for which the Civil Service Reform Act provided a comprehensive remedial scheme); Bennett v. Barnett, 210 F.3d 272, 275 (5th Cir. 2000) (holding that *Bush* had "left it up to Congress

does not apply at all.[168] As the Fifth Circuit stated, while "[t]he Supreme Court has not held expressly that the CSRA in all cases precludes federal employees from bringing *Bivens* actions arising out of their employment relationship," the circuit courts "have held that the CSRA precludes at least some *Bivens* actions brought in such circumstances."[169]

Other kinds of federal employment have been treated similarly. A Veterans Administration (VA) employee, for instance, was precluded from suing superiors under *Bivens* because his employment was governed by the Department of Medicine and Surgery (DMS) regulations.[170] Since Congress had delegated to the department the authority to use internal procedures to address rights violations, the court concluded that Congress intended that "DMS personnel should be subject to a less protective grievance system than is available to ordinary civil service employees" through the CSRA.[171]

At the same time, courts can allow federal employee cases to go forward when they determine that Congress did not intend a particular application of federal employment laws. Thus, when an employee in the department of Agriculture's Agricultural Stabilization and Conservation Service sued his superiors, the Eighth Circuit noted that the service employed a nonstandard hiring practice that left employees outside the CSRA's protective ambit.[172] Since Congress had not envisioned nonstandard hiring practices, the court concluded, it could not have intended such employees to be deprived of both CSRA protections and constitutional damages.[173] The plaintiff did have recourse in the form of an internal appeals process, but unlike the appeals process in the CSRA, this nonstandard version allowed him to challenge his supervisor's unconstitutional conduct only through an appeal to that supervisor himself.[174] Although courts normally view administrative appeals systems as the delegated part of a statutory remedial system, the Eighth Circuit recognized that "[o]nly Congress has the power to decide that a statutory or administrative scheme will foreclose a *Bivens* action."[175] But since Congress could not have foreseen the service's employment scheme, the court could not infer that Congress had meant

---

to provide the appropriate remedy" for federal employee constitutional violations and that the Postal Reorganization Act provided the appropriate remedy for a postal worker claiming constitutional violations).

168. *Dotson*, 398 F.3d at 163-65. The court determined that Congress was aware of the judiciary's internal review procedures for employees and purposely excluded such employees from the CSRA's ambit. *Id.* at 161.

169. Rollins v. Marsh, 937 F.2d 134, 139 (5th Cir. 1991); *see also* Karamanos v. Egger, 882 F.2d 447, 453 (9th Cir. 1989) (refusing to recognize a *Bivens* action by an Internal Revenue Service employee on the basis of the CSRA's comprehensive remedial scheme).

170. Heaney v. U.S. Veterans Admin., 756 F.2d 1215, 1218-19 (5th Cir 1985).

171. *Id.* at 1219.

172. Krueger v. Lyng, 927 F.2d 1050, 1056 (8th Cir. 1991).

173. *Id.* at 1056-57.

174. *Id.* at 1052-53.

175. *Id.* at 1055.

to exclude people in the plaintiff's position from *Bivens* remedies.[176] Cutting employees off from CSRA protections, the court concluded, was an executive policy with no legislative support.[177] It did not indicate that Congress had intended that result.[178] Thus, the mere presence of a comprehensive statutory scheme does not suffice to preclude a constitutional damages remedy. As with implicit preemption, the court must determine that Congress had considered the kind of situation at issue in the case.

The other consistent impediment to recovery under *Bivens* is the Internal Revenue Code. A number of circuits have found that Congress's thorough regulation of taxes and the Code's redress mechanisms demonstrate congressional occupation of the field.[179] Thus, plaintiffs who were tricked by an IRS agent into claiming improper deductions could not sue under *Bivens*.[180] The court noted that the circuits consistently find "*Bivens* actions inapplicable for claims arising from federal tax assessment or collection."[181] Circuit courts also consistently decline to recognize *Bivens* remedies against the IRS even where the Internal Revenue Code does not provide complete relief for the violation at issue. Such cases hold that the existence of that comprehensive remedial scheme, even

---

176. *Id.* at 1056; *see also* Carpenter's Produce v. Arnold, 189 F.3d 686, 689 (8th Cir. 1999) ("The remedial scheme in *Krueger*, however, unlike that in *Chilicky* . . . was created entirely by regulation, and that was the point of our holding.").

177. *Krueger*, 927 F.2d at 1056 ("Congress did not direct, or in any way imply, that these employees should be excluded from the CSRA; their exclusion is solely the result of the [s]ecretary's *sua sponte* decision to use a 'non-traditional' hiring method. It is not a 'non-inadvertent' act by Congress that excludes Krueger from civil service protection; instead, this exclusion results solely from the [s]ecretary's having used standard enabling language as the basis for adopting an uncommon hiring practice.").

178. *Id.* Six years later, however, the Ninth Circuit concluded that Congress had indicated its awareness of these kinds of employees when it permitted their participation in the federal retirement plan, in addition to other civil service benefits, and that their exclusion from the CSRA's remedial ambit had thus been (or become) an intentional omission. Moore v. Glickman, 113 F.3d 988, 992-93 (9th Cir. 1997) (explicitly recognizing the "non 'employee'" status of this type of USDA worker). The next year, the Eleventh Circuit followed *Moore*'s reasoning but rested its conclusion on the Administrative Procedure Act rather than the CSRA. *See* Miller v. U.S. Dep't of Agric. Farm Servs. Agency, 143 F.3d 1413, 1416 (11th Cir. 1998); *see also infra* note 208.

179. *See* Shreiber v. Mastrogiovanni, 214 F.3d 148, 152 (3d Cir. 2000) ("[A] *Bivens* action should not be inferred to permit suits against IRS agents accused of violating a taxpayer's constitutional rights in the course of making a tax assessment."); Fishburn v. Brown, 125 F.3d 979, 982-83 (6th Cir. 1997) (summarizing extra-circuit case law and concluding that *Bivens* actions may not be brought against IRS agents for due process violations arising out of tax collection activities), *abrogated by* Hoogerheide v. IRS, 637 F.3d 634 (6th Cir. 2011); McMillen v. U.S. Dep't of Treasury, 960 F.2d 187, 190 (1st Cir. 1991) (per curiam) (stating that even if alleged IRS actions violated plaintiff's constitutional rights, "we doubt that the creation of a *Bivens* remedy would be an appropriate response").

180. Adams v. Johnson, 355 F.3d 1179, 1184 (9th Cir. 2004).

181. *Id.*

if incomplete or not comparable, precludes *Bivens*.[182]

At the same time, courts can be careful to ascertain not just that the defendant is an IRS agent, but that the Internal Revenue Code could realistically be seen to apply to the situation at hand. When taxpayers sued an IRS agent for allegedly outrageous harassment, the Fifth Circuit concluded the Internal Revenue Code and its remedial scheme addressed *property*, not *liberty*, interests.[183] If the plaintiff taxpayers could show that the auditor's conduct implicated liberty interests, the court determined, they would have a *Bivens* claim despite the case implicating the Internal Revenue Code.[184]

Other, less frequently cited, statutes have precluded *Bivens* on these same grounds. Veteran's claims against the Veteran's Association have been rejected because Congress set up an "elaborate remedial structure" for veteran's benefits, including an "administrative process . . . [that] provid[ed] for a comprehensive review of veterans' benefits disputes," but explicitly prohibited "judicial review of [those] disputes."[185] National banking regulation has precluded a *Bivens* claim

182. *See, e.g.*, Judicial Watch, Inc. v. Rossotti, 317 F.3d 401, 410-11 (4th Cir. 2003) (denying a *Bivens* claim for retaliatory taxation because tax laws provide a comprehensive regulatory scheme, even if they do not provide complete relief); *see also* Dahn v. United States, 127 F.3d 1249, 1254 (10th Cir. 1997) ("[I]n light of the comprehensive administrative scheme created by Congress to resolve tax-related disputes, individual agents of the IRS are also not subject to *Bivens* actions."); Vennes v. An Unknown Number of Unidentified Agents of the United States, 26 F.3d 1448, 1454 (8th Cir. 1994) ("Congress has provided specific and meaningful remedies for taxpayers who challenge overzealous tax assessment and collection activities. . . . These carefully crafted legislative remedies confirm that . . . Congress's refusal to permit unrestricted damage actions by taxpayers has not been inadvertent."); Baddour, Inc. v. United States, 802 F.2d 801, 807-09 (5th Cir. 1986) ("[C]reation of a damages remedy under circumstances where Congress has provided for corrections of tax collection errors could wreck [sic] havoc with the federal tax system."); Cameron v. IRS, 773 F.2d 126, 129 (7th Cir. 1985) ("[I]t would make the collection of taxes chaotic if a taxpayer could bypass the remedies provided by Congress simply by bringing a damage action against Treasury employees.").

183. Rutherford v. United States, 702 F.2d 580, 583 (5th Cir. 1983), *superseded by statute* I.R.C. §§ 7432-33 (2006), *as recognized in* Barron v. United States, 998 F. Supp. 117 (D.N.H. 1998). The plaintiffs alleged that the IRS auditor "willfully and maliciously assess[ed] them for taxes they did not owe, harass[ed] them into paying those taxes, and forc[ed] them to sue for a refund." *Id.* at 581. The auditor allegedly "invented" over $100,000 of gross income; assessed the plaintiffs twice for the same income; demanded "useless documentation"; accused the plaintiffs of hiding money; "once insisted that [one plaintiff] empty his pockets of money and let him count it"; told one plaintiff, "You don't think I am going to spend this much time on this audit and not come up with a considerable sum of money due and owing"; and "arranged for his audit report to be delivered to the [plaintiffs'] home at 4:30 p.m. on Christmas Eve." *Id.*

184. *Id.* at 584-85 (remanding for a district court determination of whether allegedly outrageous conduct by an IRS auditor constituted a deprivation of a Fifth Amendment liberty interest, in which case a *Bivens* claim could proceed).

185. Zuspann v. Brown, 60 F.3d 1156, 1161 (5th Cir. 1995); *see also* Hicks v. Small, 69 F.3d 967, 969 (9th Cir. 1995) (holding that *Bivens* was unavailable against the plaintiff veteran's doctor

by a bank's owner against the Office of the Comptroller of the Currency employees,[186] and congressional regulation of federal disaster relief has precluded applicants' *Bivens* claims against the agents reviewing eligibility for relief for race discrimination.[187] The Parole Commission and Reorganization Act of 1978 has precluded a parolee's *Bivens* action against his parole officer.[188] And when a group of people who had spent time in lower Manhattan sued the Environmental Protection Agency for injuries allegedly arising from the Agency's false statements about air quality and safety in the wake of the World Trade Center attacks of September 11, 2001, the Second Circuit held that the Air Transportation Safety and System Stabilization Act,[189] which "provided a statutory cause of action for claims 'arising out of' the airplane crashes that destroyed the WTC towers," precluded *Bivens* remedies.[190]

One can always question whether Congress's foray into some particular field really indicates its desire to preclude all remedies it did not happen to construct. Did Congress really consider all the possible harms that could come to banks when it created banking regulations? Did it really mean to leave people with no way to enforce their rights to equal protection under the federal disaster relief statute? Regardless of whether the courts' interpretations seem realistic, these cases show the central thread running through special factors analysis: In all these examples, Congress had considered the area of law and the kind of plaintiff, at issue. Focusing on implicit indications of legislative will, courts have consistently found this sufficient to infer that Congress's silence was purposeful.

B. Categorical Bars as Shorthand for Implicit Preclusion Analysis

While most courts explicitly treat special factors analysis as an inquiry into legislative preclusion, some talk about certain kinds of plaintiffs as categorically barred from *Bivens* remedies.[191] In such cases, courts do not analyze how comprehensive a statutory scheme is. They simply determine whether the plaintiff belongs to a category that they consider barred from constitutional damages. These sound like different inquiries—statutory scheme versus plaintiff category—but, in fact, in both kinds of cases, the courts' conclusion rests on an inquiry into implicit preclusion of remedies by the legislature. The difference is

---

because Congress had comprehensively regulated veterans' relations with the Veterans Administration through the Veterans Judicial Review Act of 1988).

186. Sinclair v. Hawke, 314 F.3d 934, 940-42 (8th Cir. 2003).

187. Carpenter's Produce v. Arnold, 189 F.3d 686, 687, 689 (8th Cir. 1999).

188. Rauschenberg v. Williamson, 785 F.2d 985, 988 (11th Cir. 1986).

189. Pub. L. No. 107-42, 115 Stat. 230 (2001) (codified at in scattered sections of 49 U.S.C.).

190. Benzman v. Whitman, 523 F.3d 119, 126 (2d Cir. 2008). *Benzman* also described the area of federal disaster relief as thoroughly occupied by congressional regulation in a way analogous to Congress's control over the military as described in *Chappell* and held that the plaintiffs had failed to plead a constitutional violation. *Id.* at 126-29.

191. *See, e.g.*, Corr. Servs. Corp. v. Malesko, 534 U.S. 61, 70-71, 74 (2001) (holding that *Bivens* remedies would not be extended to private corporations).

simply that courts using categorical bars do not conduct the legislative preclusion inquiry themselves. Instead, they take as their starting point the *conclusion* of an inquiry already conducted by the Supreme Court. The categorical approach thus depends entirely on the question of legislative preclusion. This is because the kinds of plaintiffs that have been barred from constitutional damages have been barred precisely because the Supreme Court has found that the comprehensive remedial schemes available to them indicate Congress's occupation of their field. As discussed above, for instance, civil servants have been barred from *Bivens* because of the CSRA and members of the military because of Congress's occupation of that field.[192] In reality, then, this categorical style of special factors inquiry still focuses on congressional preemption of judicial remedies: asking about plaintiff categories is just a shorthand for asking about remedial schemes. The inconsistency among courts of appeal probably stems from the Supreme Court's own *Bivens* cases, which have described statutory schemes in both ways at different times. In practice, the special factors inquiry, while phrased in varying ways, remains an inquiry into legislative preclusion.

In this vein, a *Bivens* suit by civilian employees at a naval base was precluded solely because the claims arose from a "federal employment relationship,"[193] a decision that treated the Supreme Court's preemption analysis in *Bush v. Lucas*[194] as a categorical bar of federal employee *Bivens* suits. Other courts have combined the categorical bar approach with reference to a comprehensive statutory scheme, as when the Third Circuit determined that "in light of the existence of [statutory] remedies the employer-employee relationship . . . is a special factor."[195]

Similarly, for cases involving members of the military, brief references to the *Feres* doctrine sometimes stand in for analysis. Thus, a plaintiff's status as a member of the military precluded his *Bivens* suit challenging his superiors' racially discriminatory failure to promote him.[196] Additionally, a discharged serviceman was prevented from challenging his military jailers' Eighth

192. *See supra* notes 167, 185 and accompanying text.

193. Zimbelman v. Savage, 228 F.3d 367, 370 (4th Cir. 2000); *see also* Palermo v. Rorex, 806 F.2d 1266, 1270-73 (5th Cir. 1987) (holding that the congressionally regulated nature of federal employment weighed against judicially created remedies). *Palermo* preceded *Chilicky*; the Supreme Court analysis at that point still focused on the special relationship of the federal employer and federal employees, rather than on the existence of a statutory remedial scheme per se. *Id.* at 1271.

194. 462 U.S. 367 (1983).

195. Purtill v. Harris, 658 F.2d 134, 138 (3d Cir. 1981) (holding that the Age Discrimination in Employment Act of 1967, 29 U.S.C. § 633a, provided a comprehensive remedial scheme for a federal employee's complaint); *see also* Hall v. Clinton, 235 F.3d 202, 206 (4th Cir. 2000) (denying a *Bivens* claim because of federal employment and comprehensive regulation by the CSRA, even though the CSRA did not provide relief for the plaintiff's particular claims); Yokum v. Frank, No. 90-2196, 1991 WL 118008, at *12-13 (4th Cir. July 3, 1991) (holding that the combination of a plaintiff's federal employment and the fact that a comprehensive remedial scheme was available for his complaints created a special factor barring *Bivens* remedies).

196. Randall v. United States, 95 F.3d 339, 340 (4th Cir. 1996).

Amendment violations because the conduct and the conviction that landed him in prison occurred while he was still in the military.[197] While some courts treat the *Feres* doctrine as part of the special factors analysis,[198] others eschew the analysis altogether. Instead, they treat *Chappell* as importing the *Feres* bar on military suits into the *Bivens* arena. For example, holding that a military serviceman could not challenge a hazing incident, the First Circuit mentioned special factors but performed its analysis exclusively under *Feres*.[199] Courts of appeal have also built on the military personnel analysis to bar *Bivens* suits by employees of the Public Health Service (PHS). The PHS is not part of the military, but it is similarly structured and in many ways treated similarly by Congress.[200] Two circuits have relied on the *Feres* doctrine to bar a PHS employee's discrimination suit against PHS superiors.[201]

Thus, even when courts cite a plaintiff's membership in a particular category as the reason for refusing to recognize a *Bivens* claim, the analysis underlying that decision still rests on whether the *Bivens* remedy has been precluded by legislation. And that makes perfect sense, since the central problem with *Bivens* remedies revolves around the judiciary's attempt not to infringe on the remedy-creating prerogative of the legislature.

IV. OUTLIERS

The preceding Part illustrated that the majority of courts agree that the special factors analysis inquires into legislative will regarding constitutional damages remedies, even though they may refer to it by different terms in different cases. In this Part, I turn to courts that have deviated from the mainstream. The first kind of deviation involves mistaken results: Some courts unreasonably interpret statutes that are only peripherally relevant to the situation in a particular case as indicating a legislative desire to preclude *Bivens* remedies. The second kind of deviation is more serious: Courts may mistakenly look to executive will as determining the fate of constitutional damages. In this situation, the court not only deviates from the accepted doctrine of the special

197. Ricks v. Nickels, 295 F.3d 1124, 1130-31 (10th Cir. 2002).

198. *Id.* at 1128-30.

199. Day v. Mass. Air Nat'l Guard, 167 F.3d 678, 682, 684 (1st Cir. 1999); *see also* Wright v. Park, 5 F.3d 586, 589-91 (1st Cir. 1993) (holding that the claims of a part-military, part-civilian, technician were barred because they arose from events incident to military service, and noting *Chappell*'s reasoning without performing a special factors analysis, resting instead on the *Feres* doctrine alone).

200. The PHS provides medical care in prisons and immigration detention centers. The Supreme Court recently barred an inmate's *Bivens* action against the PHS for Eighth Amendment violations, but did so on statutory interpretation grounds, without reference to *Bivens* special factors. Hui v. Castaneda, 130 S. Ct. 1845, 1852-55 (2010). The special factors analysis mostly focused on the ability of PHS employees, not inmates, to sue superior PHS personnel. *Id.* at 1851.

201. *See* Middlebrooks v. Leavitt, 525 F.3d 341, 350 (4th Cir. 2008); Diaz-Romero v. Mukasey, 514 F.3d 115, 118-19 (1st Cir. 2008).

factors analysis, it also fails to recognize that the underlying separation of powers concern in constitutional damages is the relationship between the judiciary and the legislature. Such cases should be read narrowly to prevent these mistakes from setting broad precedent that undermines the logic of the special factors inquiry.

A. Mistaking Peripheral Statutes for Implicit Preclusion

A few courts have mistakenly found *Bivens* precluded by statutes that touch only peripherally on the matters at issue in a case. Instead of looking for indications that Congress considered the situation presented in a case, such courts have settled for statutory schemes that merely implicate some of the interests or issues involved. Taking the existence of *some* remedial scheme to preclude a *Bivens* remedy without ensuring that the scheme indicates that Congress considered the situation at stake, such courts mistakenly find a congressional policy preference where none could have been implied. Even this mistaken approach, however, shows that courts look to congressional will to determine whether to recognize a *Bivens* action.

For instance, in *Downie v. Middleburg Heights*,[202] the Sixth Circuit held that the Privacy Act precluded a constitutional damages suit by a Customs Service informant against his handler.[203] The handler had allegedly asked the Customs Service to "blacklist[]" the plaintiff from future work; arranged to have his firearms license revoked and his firearms seized; arranged for a false arrest record to be entered about him; and threatened to stop working with the plaintiff's boss if the plaintiff were not fired.[204] The Sixth Circuit assumed that because the case involved records created by government agents, the Privacy Act addressed the issues it raised.[205] However, the plaintiff's claim was not primarily that the defendant collected incorrect information or even created false records—issues the Privacy Act addresses. Rather, the plaintiff claimed that a federal actor requested retaliatory action by the agency (rather than merely

---

202. 301 F.3d 688 (6th Cir. 2002).

203. *Id.* at 696. The Privacy Act of 1974 regulates how government agencies collect, keep, use, and disseminate information about individuals. 5 U.S.C. § 552a (2006 & Supp. 2010). With certain exceptions, it requires agencies to announce the kinds of information they collect and the "routine uses" they put it to, maintain accurate records; and allow individuals to access and contest inaccurate records. *See id.* § 552a(a)(3) (defining "maintain" to include "maintain, collect, use, or disseminate"); *id.* § 552a(a)(4) (defining "record" as "item, collection, or grouping of information about an individual that is maintained by an agency"); *see also id.* §§ 552a(b), (d), (e)(4). The Act allows individuals to sue agencies for damages for failure to adhere to its requirements. *Id.* § 552a(g). The Privacy Act makes no provisions for constitutional violations, but as discussed above, after *Chilicky* the alternative statutory remedial scheme no longer needed to provide comparable relief to *Bivens*; in a preemption-like approach, the fact that Congress has considered the issue suffices, even if Congress then decided not to provide remedies for constitutional violation.

204. *Downie*, 301 F.3d at 690-91.

205. *Id.* at 692.

provided an incorrect record); provided false information to law enforcement (rather than failed to properly maintain a record); and threatened third parties with sanctions to achieve his retaliatory goal.[206]

There is no indication that Congress ever considered these kinds of wrongs relevant to the area that the Privacy Act regulates: government agency record maintenance.[207] The Sixth Circuit's reasoning perversely suggests that a federal employee can immunize himself from *Bivens* suits simply by creating records relating to his conduct. The Privacy Act, in short, may be peripherally related to the issues in *Downie*, but it cannot reasonably be read to indicate a congressional preference regarding them. *Downie* amply demonstrates how a peripherally relevant statute—a statute that has something to do with the facts of a case but does not primarily regulate the parties or situations at issue—can be recruited to preclude a *Bivens* remedy by a court that fails to perform the special factors analysis correctly.[208] At the same time, such decisions demonstrate that, even

---

206. *Id.* at 696.

207. *See* 5 U.S.C. § 552a(g)(1) (providing grounds for suit against agency); *see also* Quinn v. Stone, 978 F.2d 126, 131 (3d Cir. 1992) (discussing the elements of a claim brought under the Privacy Act). In addition, the Privacy Act provides only for suit against the agency, not against the individual agent. *Id.* at 135.

208. *Downie*, 301 F.3d at 696. A similar mistake motivated the decision in *Wilson v. Libby*, which declined to recognize a *Bivens* remedy for Valerie Plame Wilson, the undercover CIA operative whose identity was revealed in a newspaper column by Robert Novak. Wilson v. Libby, 535 F.3d 697, 702 (D.C. Cir. 2008). Novak obtained the information from then-Deputy Secretary of State Richard Armitage, and employees in the Offices of the President and Vice President also revealed Ms. Wilson's identity to the press. *Id.* The D.C. Circuit declined to recognize Wilson's claim against those employees, holding that because the Privacy Act, which "regulates the collection, maintenance, use, and dissemination of information about individuals by federal agencies," excluded the Offices of the President or the Vice President from its definition of "agency," the Act itself must preclude Wilson's *Bivens* claim. *Id.* at 704-09 (citation omitted) (internal quotation marks omitted). By its own description, *Wilson* thus concluded that a comprehensive statutory scheme regulating agencies precluded claims not brought under that statute against individuals who were not part of an agency. A more logical approach might conclude that the Privacy Act does not control claims against persons it does not regulate. The *Wilson* court suggested that there was a connection as the plaintiffs' *Bivens* claims all "alleg[ed] damages from the improper disclosure of information covered by the Privacy Act." *Id.* at 707. But the Privacy Act does not purport to regulate *information*; it regulates agencies. The *Wilson* court thus unreasonably inferred a congressional intent from the mere fact that the Privacy Act, like the lawsuit, dealt with information.

In a final example of peripheral-statute preclusion, a federal employee excluded from CSRA protections could not bring a *Bivens* suit because he had recourse to remedies under the Administrative Procedure Act (APA), which the court held was a comprehensive statutory remedial scheme. Miller v. U.S. Dep't of Agric. Farm Servs. Agency, 143 F.3d 1413, 1416 (11th Cir. 1998). Rather than holding, like other circuits, that Congress had occupied the field of federal employment, the court asserted that "the existence of a right to judicial review under the APA is, alone, sufficient to preclude a federal employee from bringing a *Bivens* action." *Id.* at 1416. It thereby treated the when they mistakenly read irrelevant statutes to preclude *Bivens* remedies, courts confronting constitutional damages cases consistently attempt to discern how Congress wants such cases to be addressed.

B. Mistaking the Executive for the Legislature

Another kind of outlier may become more prominent as *Bivens* cases emerging from national security projects become more common.[209] The publicized, politicized nature of such cases may blind courts to the doctrinal coherence and rationale of the special factors analysis. Instead of asking whether Congress has indicated a preference for the availability of remedies, such a decision asks about the preferences of the Executive. As this Article has detailed, however, executive preferences do not fit the logic, the rationale, or the doctrinal history of the special factors analysis. As a matter of doctrinal history, I have demonstrated a broad, cross-circuit agreement that special factors analysis looks for implicit legislative preclusion of remedies. The rationale of the special factors analysis focuses on maintaining a balance between the Judiciary and the Legislature, not the Executive. Logically, because the Executive is usually named as a defendant in *Bivens* cases, it makes little sense to hinge a case's justiciability on the defendant's desire to allow the plaintiff a remedy.[210]

Arar v. Ashcroft,[211] a prominent example of this mistaken approach, declined to recognize a constitutional damages remedy for a plaintiff secretly detained and then rendered to Syria, where he was tortured for a year.[212] United States

---

APA's provisions for review of agency action as though they regulated the federal employment relationship. However, other courts have disagreed that the APA is a sufficient ground. *See* Munsell v. Dep't of Agric., 509 F.3d 572, 590 (D.C. Cir. 2007) ("[W]e are unaware of any Supreme Court decision holding that APA review *alone* is sufficient to eliminate the need for a *Bivens* remedy.").

209. *See, e.g.*, Rumsfeld v. Padilla, 542 U.S. 426 (2004); Vance v. Rumsfeld, 653 F.3d 591 (7th Cir. 2011), *reh'g en banc granted, vacated* (Oct. 28, 2011).

210. This is not to say that the Executive is irrelevant to remedy creation. As I point out above, the modern Executive plays an important role in crafting legislation, and the comprehensive remedial schemes that preclude *Bivens* suits often come partly from administrative agencies. Even in these cases, the creation of remedies is, in the end, authorized or delegated by Congress. It may be that Congress can, and maybe even should, delegate the power to create and to preclude statutorily implied causes of action to administrative agencies. *See* Matthew C. Stephenson, *Public Regulation of Private Enforcement: The Case for Expanding the Role of Administrative Agencies*, 91 VA. L. REV. 93, 126-27 (2005). Even if that were the case, the delegation, and thus the decision, would come from Congress. One can also question whether any such congressional delegation would be valid insofar as it allowed the executive to preclude constitutional, rather than only statutory, causes of action. The point for my purposes is that, for special factors analysis, the power to decide lies with Congress.

211. 585 F.3d 559 (2d Cir. 2009).

212. *Id.* at 563-64, 566. *Arar* is a complex case with a complex opinion and several dissents. I outline it here but limit my discussion to the use of special factors to deny the plaintiff's *Bivens*

officials seized Arar, a Canadian-Syrian dual national, while he was changing planes at John F. Kennedy Airport in New York on his way to Canada. Held in New York without access to an attorney for nearly two weeks of interrogation—Arar claimed that federal agents lied to his lawyer to prevent her from finding him—he was subsequently deported to Syria without a hearing, despite his Canadian passport and his assertion that he feared torture in Syria.[213] In an en banc decision, the Second Circuit concluded that special factors precluded Arar's *Bivens* suit.[214]

The court first decided that the case presented a new "context" for *Bivens*: "extraordinary rendition," a term referring to "the *extrajudicial* transfer of a person from one [country] to another."[215] Despite determining that the Immigration and Nationality Act (INA), a comprehensive statutory scheme regulating aliens in the United States, did not address the kind of situation of which Arar complained,[216] the court concluded that special factors barred relief.

---

claims. The facts of Arar's experience have been widely reported. *See, e.g.*, Benjamin Weiser, *Appeals Court Rejects Suit by Canadian Man Over Detention and Torture Claim*, N.Y. TIMES, Nov. 2, 2009, http://www.nytimes.com/2009/11/03/nyregion/03arar-web.html; Editorial, *The Unfinished Case of Maher Arar*, N.Y. TIMES, Feb. 18, 2009, at A26, *available at* http://www.nytimes.com/2009/02/18/opinion/18wed2.html.

213. *Arar*, 585 F.3d at 565-66. Under the Convention Against Torture (CAT), to which the United States is a party, "[n]o State Party shall expel, return . . . or extradite a person to another State where there are substantial grounds for believing that he would be in danger of being subjected to torture," regardless of the ground on which such torture would occur. Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, G.A. Res. 39/46, art. 3, U.N. Doc. A/RES/39/46 (Dec. 10, 1984); *see also* Foreign Affairs Reform and Restructuring Act of 1998, Pub. L. No. 105-277, § 2242, 112 Stat. 2681-822 (codified at 8 U.S.C. § 1231 (2006 & Supp. 2010)); *see also* 8 U.S.C. § 1231(b)(3)(A) ("[T]he Attorney General may not remove an alien to a country if the Attorney General decides that the alien's life or freedom would be threatened in that country because of the alien's race, religion, nationality, membership in a particular social group, or political opinion."). Arar's complaint further alleged that his fear of torture in Syria was justified. He was incommunicado for a year in an underground cell measuring "six feet by three, and seven feet high." *Arar*, 585 F.3d at 566. Furthermore, he was beaten with electrical wire and threatened with permanent injury. Complaint and Demand for Jury Trial at 16, *Arar*, 585 F.3d 559 (No. CV-04-0249). Arar also made a claim under the Torture Victim Protection Act, which I do not discuss here. *Arar*, 585 F.3d at 563; *see* 28 U.S.C. § 1350(a)(1) (2006).

214. *Arar*, 585 F.3d at 563.

215. *Id.* at 564 & n.1. The dissenters in *Arar* disputed that the case presents a "new context." *Id.* at 583 (Sack J., concurring in part and dissenting in part).

216. *Id.* at 571. The court noted that Arar was not attempting to enter the United States, but was merely passing through, and concluded that "it is not clear that the INA's judicial review provisions govern circumstances of involuntary rendition." *Id*. It also noted that, given the alleged conduct of the United States employees who detained Arar, such as serving him "with the removal order while he was already en route to Amman, the INA could have afforded him no relief" at the time of the events in question. *Id*. Although *Arar* is clearly not an immigration case—insofar as

The opinion did not clearly distinguish all the special factors it pointed to, naming some and implying several others. Examining the concerns the court raised, however, shows that none of the concerns constituted a special factor under the accepted doctrine. Had the court analyzed its concerns within the parameters of the applicable case law, moreover, it would have found Arar's suit was not barred.[217] *Arar*, in short, mistakenly treated special factors not as a legal term of art, but as an appeal to the political intuitions of judges.[218]

Arar is important for two main reasons. First, it may set precedent that would confuse the special factors doctrine and render its rationale incoherent, encouraging other courts to believe that executive desires and judges' own inchoate misgivings suffice to "counsel hesitation"[219] in *Bivens* cases. For this reason, it should be read narrowly. Second, by analyzing concerns that belong to diverse legal doctrines, it may encourage other courts to forego a legally grounded analysis of the issues raised by a *Bivens* suit, stymieing the development of other doctrines by lumping unrelated issues under the special factors label.

The Second Circuit cited the lack of clarity in the law regarding extraordinary rendition as one reason to uphold dismissal.[220] Lack of clarity in

---

Arar was not attempting to enter the United States but merely changing planes in it—this description does suggests another interpretation: that the INA *did* provide a remedial scheme for at least the parts of Arar's experience that occurred while in United States custody, but United States employees actively obstructed Arar's recourse to that remedial scheme. It is unclear why the court did not discuss this possibility, which suggests that federal employees who breached statutorily imposed constraints violated Arar's right to due process.

217. For a similar reading of the *Arar* court's mistake, see Carlos M. Vázquez & Stephen I. Vladeck, *State Law, the Westfall Act, and the Nature of the* Bivens *Question After* Minneci v. Pollard, 161 U. PA. L. REV. (forthcoming 2012) (manuscript at 10), *available at* http://ssrn.com/abstract=2038641 (explaining that the *Arar* court's concerns were "potentially relevant . . . to questions of immunity or privilege or preemption" but were "not relevant to the decision whether to recognize a *Bivens* claim").

218. The *Arar* court affirmed the dismissal of the case on a number of grounds. Specifically, it listed the following reasons as special factors: the law surrounding extraordinary rendition is not clear; classified information would be relevant to adjudicating the claims, and the potential revelation of classified or sensitive information would allow the plaintiff to pressure the government to settle the case irrespective of the merits ("graymail"); because classified and sensitive information would be relevant to adjudicating the claims, the case would likely require closed hearings and sealed evidence; Arar's lawsuit is actually aimed not at individual actors but at the executive policy of extraordinary rendition; the court would have to determine whether Syria offered reliable assurances that Arar would not be tortured there, and whether federal employees acted in good faith on the basis of any such assurances; and the case would force the court to adjudicate issues of national security and foreign policy. *Arar*, 585 F.3d at 573-77. In what follows, I examine each factor to show why it is not a *Bivens* special factor and what doctrine it properly belongs to.

219. *Id.* at 573.

220. *Id.* at 581. The court did concede that this lack of clarity "may or may not amount to a special factor." *Id.* at 580.

the law governing defendants' conduct is, however, amply addressed by qualified immunity, an affirmative defense which provides that a federal agent is only liable for conduct that violates rights that were clearly established at the time of the events at issue.[221] The special factors analysis, in contrast, attends to the availability of remedies for plaintiffs, not the legal regime governing the conduct of defendants. The distinction matters. The qualified immunity inquiry asks whether a defendant's conduct violated clearly established laws, and whether a law is clearly established turns on social and legal expectations that may change over time. Courts assessing that question "promote[] the development of constitutional precedent."[222] The Supreme Court recently reaffirmed the desirability of this developmental process, noting that if courts' qualified immunity inquiry does not serve to specify existing and emerging rights, "[q]ualified immunity . . . may frustrate . . . the promotion of law-abiding behavior" as "[c]ourts fail to clarify uncertain questions, fail to address novel claims, fail to give guidance to officials about how to comply with legal requirements."[223] Special factors analysis, which focuses on Congress's views about *remedies* rather than on the contours of *rights*, has no such developmental process built in. By mistaking qualified immunity for a special factor, *Arar* thus impeded the development of constitutional precedent that the Supreme Court has said is central to "the promotion of law-abiding behavior."[224]

The *Arar* court also upheld dismissal because the case would inevitably involve classified information.[225] This concern does not relate to Congress's views on the availability of remedies; it is a prudential concern, not one of special factors. It has been addressed by Congress in the Classified Information Procedures Act, which lets courts balance their need for information with the Executive's need to keep certain information secret through techniques like protective orders and discovery privileges.[226] Congress's regulation of classified information illustrates that Congress assumed that judicial proceedings would sometimes use such information, and did not intend to allow the mere presence of classified information to preclude remedies.[227]

---

221. *See, e.g.*, Rasul v. Myers, 563 F.3d 527, 530 (D.C. Cir. 2009) (dismissing a *Bivens* suit on qualified immunity grounds because "[n]o reasonable government official would have been on notice that plaintiffs had any Fifth Amendment or Eighth Amendment rights" at the time of the events at issue).

222. Pearson v. Callahan, 555 U.S. 223, 236 (2009).

223. Camreta v. Greene, 131 S. Ct. 2020, 2031 (2011).

224. *Id.*

225. *Arar*, 585 F.3d at 576.

226. Classified Information Procedures Act, Pub. L. No. 96-456, 94 Stat. 2025 (1980) (codified at 18 U.S.C. app. §§ 1-16 (2006 & Supp. 2010)).

227. Strangely, the *Arar* court explicitly recognized that governments, like courts, routinely announce the results of their investigations while keeping classified information secret. As the opinion notes, the Canadian government completed a thorough investigation of Arar's claims and "paid [him] compensation for [his] role in the events surrounding this lawsuit, but has *also* asserted the need for Canada itself to maintain the confidentiality of certain classified materials related to

Arar further cited the possibility of "graymail,"[228] or a "'lawsuit[] brought to induce the [government] to settle a case . . . out of fear that any effort to litigate the action would reveal classified information that may undermine ongoing covert operations,' or otherwise compromise foreign policy efforts."[229] This factor merely reiterated the court's concern with classified information, and the government's susceptibility to graymail is simply one potential reaction to the possibility that information will be revealed. While the term "graymail" raises the specter of impropriety due to its color-based proximity to *blackmail*, pressure to settle is a normal byproduct of our legal system.[230] The court's discussion revealed no indication of a congressional preference as to the availability of remedies.

The *Arar* court also grounded its finding of preclusion on the possibility that "[t]he court's reliance on information that cannot be introduced into the public record" would necessitate closing judicial proceedings to the public.[231] The Supreme Court has held the public's right to witness judicial proceedings is "implicit in the guarantees of the First Amendment"[232] but has maintained that this right is not absolute. Instead, the Court has laid out, with some specificity, the requirements for closing a court to the public.[233] It is clear, then, that the Supreme Court does not consider the possibility that a case may require closing the court a reason for precluding the case itself,[234] and there is no indication that Congress disagrees.

The ostensible policy-oriented nature of the lawsuit also posed a concern in *Arar*. "Although this action is cast in terms of a claim for money damages against the defendants in their individual capacities," the court stated, "it operates as a constitutional challenge to policies promulgated by the executive."[235] But

---

[his] claims." *Arar*, 585 F.3d at 576.

228. *Id.* at 578.

229. *Id.* at 578-79 (second alteration in original) (quoting Tenet v. Doe, 544 U.S. 1, 11 (2005)).

230. *See, e.g.*, Dodds v. Am. Broad. Co., 145 F.3d 1053, 1066 (9th Cir. 1998) ("Exerting pressure on parties is part of the settlement process . . . .").

231. *Arar*, 585 F.3d at 577.

232. Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 580 (1980) (plurality opinion).

233. Waller v. Georgia, 467 U.S. 39, 48 (1984) (citing Press-Enter. Co. v. Super. Ct. of Cal., 464 U.S. 501, 509-12 (1984)). The Second Circuit, which decided *Arar*, has specifically adopted the *Waller* standard. *See* Williams v. Artuz, 237 F.3d 147, 152 (2d Cir. 2001).

234. *See Arar*, 585 F.3d at 609-10 (Sack, J., concurring in part and dissenting in part) ("The presumption of openness . . . can be, and routinely is, overcome. We regularly hear, on the basis of partially or totally sealed records, not only cases implicating national security or diplomatic concerns, but those involving criminal defendants' cooperation with prosecutors, other criminal matters, . . . probation department reports, upon which federal criminal sentences are to a significant extent typically based, . . . trade secrets, and any manner of other criminal and civil matters. Hardly a week goes by, in our collective experience, in which some document or fact is not considered by a panel of this Court out of the public eye." (internal citations omitted)).

235. *Id.* at 574 (majority opinion).

neither Congress nor the courts have ever conditioned the availability of remedies for constitutional wrongs on the plaintiff's motivation.[236] Indeed, according to the dissenting justices in *Arar*, a broader policy orientation is typical of civil rights actions,[237] and Congress has never indicated its intent to preclude civil rights remedies on this basis.[238]

---

236. A litigant's motivation may sometimes influence how a court deals with a lawsuit. For example, a government prosecutor who pursues criminal charges for improper purposes may be civilly liable to that defendant on a *Bivens* theory of retaliatory prosecution. *See, e.g.*, Hartman v. Moore, 547 U.S. 250, 263-66 (2006) (holding that a *Bivens* plaintiff suing for malicious prosecution, based on retaliation against the plaintiff's exercise of First Amendment rights, must show that the prosecutor had no probable cause on which to proceed). Similarly, the common-law torts of malicious prosecution and abuse of process "provided causes of action against private defendants for unjustified harm arising out of the misuse of governmental processes." Wyatt v. Cole, 504 U.S. 158, 164 (1992). In contrast, the *Arar* court took the opposite approach and treated a *plaintiff's* purportedly improper motivation as a bar to the underlying process itself. It is unclear what legal theory supported this approach.

Arar also did not explain why it would be wrong for a plaintiff to seek vindication of his rights through a judicial determination that they had been violated, notwithstanding any desire for monetary damages. *See, e.g.*, Pfander, *Resolving the Qualified Immunity Dilemma*, *supra* note 21, at 1620-26 (proposing that plaintiffs seeking to vindicate rights through judicial determinations that rights have been violated should sue for nominal damages, thus eliminating the need for qualified immunity, which is meant only to alleviate the threat of individual liability by public officials). Pfander notes that the violation of individual rights is, itself, a cognizable injury quite apart from any monetary damages. *Id.* at 1620 ("Today, no one questions the power of the federal courts to declare the rights of the parties in a case of actual controversy.").

237. *Arar*, 585 F.3d at 605 (Sack, J., concurring in part and dissenting in part); *id.* at 634-35 (Calabresi, J., dissenting).

238. Aside from being ungrounded in legislation or case law, precluding a lawsuit because it targets a policy implies that a plaintiff's desire to affect executive policy immunizes the federal employees who carry out that policy from suit. The *Arar* court appeared to view this kind of plaintiff-induced immunity as a form of judicial deference to executive policy. *Id.* at 574-75 (majority opinion). But a court's ability to determine whether a federal employee violated a constitutional right does not depend on an evaluation of the policy the employee was working under. *Id.* at 574. Moreover, such immunity presents a practical problem: It burdens courts with evaluating the motivation of a lawsuit to determine whether it seeks individual remedy or policy change. *See, e.g.*, *id.* at 576. This burden is unmanageable not only because the court has little access to a litigant's inner states, but also because litigants often act on a variety of motivations, some of which they may not even recognize. Empirical studies of people involved in the civil justice system consistently reveal that litigants have multiple, often conflicting, attitudes toward the cases in which they are involved. *See, e.g.*, CAROL J. GREENHOUSE ET AL., LAW AND COMMUNITY IN THREE AMERICAN TOWNS 1-2 (1994) ("[W]e address the irony that, although courts and law are central to the ways some residents construct their sense of community, these same individuals disparage the courts as having been captured by the 'wrong' people for the 'wrong' kinds of cases."); SALLY ENGLE MERRY, GETTING JUSTICE AND GETTING EVEN: LEGAL CONSCIOUSNESS AMONG WORKING-CLASS AMERICANS 5-6 (1990).

Executive control over national security and foreign policy also led the *Arar* court to find that *Bivens* actions were precluded.[239] Executive policy control, however, does not indicate how Congress would treat claims by those whose legal rights are violated in the course of carrying out a policy. Using executive policy as a rationalization for precluding *Bivens* actions implies that Congress's delegation of policy powers to the Executive would also delegate an authority to violate constitutional rights. Of course, Congress has no such authority to delegate.[240] Thus, this concern does not belong in a special factors analysis.

Rather, courts' relations to executive policy properly fit the framework of the political question doctrine.[241] Although the political question doctrine does not traditionally include conduct related to national security, courts often bundle national security and foreign affairs into related areas that fall particularly under the Executive's control.[242] The political question doctrine has caused courts to refrain from matters which include: determining the start and end points of military hostilities;[243] recognizing foreign powers and their representatives;[244] ratifying and interpreting treaties;[245] and, adjudicating presidential decisions to engage in hostilities abroad.[246] Arar, in contrast, alleged that he was denied due

---

239. *See Arar*, 585 F.3d at 575 ("A suit seeking a damages remedy against senior officials who implement an extraordinary rendition policy would enmesh the courts ineluctably in an assessment of the validity and rationale of that policy and its implementation in this particular case, matters that directly affect significant diplomatic and national security concerns.").

240. The U.S. Constitution itself grants the President some foreign policy powers, but just as Congress has no authority to violate constitutional rights and therefore cannot delegate any such authority, the Constitution provides no authority to violate itself. 16 C.J.S. *Constitutional Law* § 265 (2012) ("[T]he legislature may not delegate to the people the power to enact a law which the legislature itself is forbidden by the [C]onstitution to enact.").

241. *See* ERWIN CHEMERINSKY, FEDERAL JURISDICTION § 2.6 (5th ed. 2007).

242. *See* Haig v. Agee, 453 U.S. 280, 291-92 (1981).

243. *See, e.g.*, Commercial Trust Co. of N.J. v. Miller, 262 U.S. 51, 57 (1923) (holding that Congress alone determines when military hostilities have ended); Martin v. Mott, 25 U.S. (12 Wheat.) 19, 29-30 (1827) (holding that Congress delegated authority to the President to determine when hostilities have begun and to call up the militia in response).

244. *See, e.g.*, United States v. Belmont, 301 U.S. 324, 330 (1937) (holding that the President had the authority to recognize and pursue diplomatic relations with the Soviet Union).

245. *See, e.g.*, Terlinden v. Ames, 184 U.S. 270, 288-90 (1902) (holding that courts cannot decide whether a treaty survives when a signatory country becomes part of another country).

246. *See generally* Crockett v. Reagan, 720 F.2d 1355, 1356 (D.C. Cir. 1983) (per curiam) (In a challenge to the United States' "presence in, and military assistance to, El Salvador," the "war powers issue presented a nonjusticiable political question."); Holtzman v. Schlesinger, 484 F.2d 1307, 1309 (2d Cir. 1973); DaCosta v. Laird, 471 F.2d 1146, 1147, 1152 (2d Cir. 1973) ("[J]udicial inquiry with respect to the Vietnam War into the domain of tactical and strategic military decisions [of the President presents a] nonjusticiable political question."); Ange v. Bush, 752 F. Supp. 509, 517-18 (D.D.C. 1990) (addressing the Persian Gulf); *see also* ANTHONY A. D'AMATO & ROBERT M. O'NEIL, THE JUDICIARY AND VIETNAM 51-58 (1972) (collecting cases and discussing the dilemma of political questions in determining whether justiciable).

process while in the United States and that U.S. employees conspired to have him tortured. Actions governed under U.S. law are far removed from the kinds of foreign relations considerations that arise from military action in foreign states.[247] As the Second Circuit, where *Arar* was heard, noted shortly before the *Arar* decision came down, the mere fact that a case implicates foreign policy or national security does not bring the political question doctrine into play.[248] The claims in *Arar* would be justiciable under the political question doctrine. As the Supreme Court recently held, "an appeal of the President presenting free-ranging assertions of foreign policy consequences," but "unaccompanied by a persuasive legal claim," cannot in itself suffice to sustain a decision in the Executive's favor.[249]

---

247. *See, e.g.*, Solomon B. Shinerock, Samantar v. Yousuf*: Recent Development in the Laws Governing Civil Torture Claims in U.S. Courts*, 17 BUFF. HUM. RTS. L. REV. 155, 159 (2011) ("The central issue before the Court in Samantar was whether an individual sued for conduct undertaken in his official capacity is immune as a 'foreign state' within the meaning of the [Foreign Sovereign Immunities] Act [of 1976].").

248. Connecticut v. Am. Elec. Power Co., 582 F.3d 309, 322-23 (2d Cir. 2009), *rev'd on other grounds*, 131 S. Ct. 2527, 2529-30, 2535 (2011). *American Electric* held that the political question doctrine did not bar a nuisance suit based on harms arising from climate change. *Id.* The Supreme Court did not disturb that holding, but reversed the Second Circuit on the substantive issue of preemption by the Environmental Protection Act. *Am. Elec. Power Co.*, 131 S. Ct. at 2540. In *American Electric*, the Second Circuit emphasized that the political question doctrine should not be applied lightly, even in cases that involved important political issues; rather, courts should examine whether settled legal rules could address the actual issues presented in the case. *Am. Elec. Power Co.*, 582 F.3d at 323. The Second Circuit cited approvingly another decision, involving a wrongful death case in which an American was killed by the Palestinian Liberation Organization (PLO). *Id.* at 328-29. The PLO argued that the political question doctrine barred the case "because it raised 'foreign policy questions and political questions in a volatile context[, i.e., international terrorism,] lacking satisfactory criteria for judicial determination.'" *Id.* at 329 (alteration in original) (quoting Klinghoffer v. S.N.C. Achille Lauro Ed Altri-Gestione Motonave Achille Lauro in Amministrazione Straordinaria, 937 F.2d 44, 49 (2d Cir. 1991)). The Second Circuit, however, "looked beyond '[t]he fact that the issues before us arise in a politically charged context,'" to determine that the claims themselves were those of "an ordinary tort." *Id.* (alteration in original) (quoting *Klinghoffer*, 937 F.2d at 49).

249. *See* Garcia v. Texas, 131 S. Ct. 2866, 2868 (2011) (internal quotation marks and citation omitted) (noting the petitioner in *Garcia* was a Mexican national who had not been given access to his consulate, as required under the Vienna Convention on Consular Relations, before being convicted of murder and sentenced to death); *see also* Sanchez-Llamas v. Oregon, 548 U.S. 331, 337 (2006) (addressing Article 36 of the Vienna Convention). The United States, as amicus curiae, asked the Supreme Court to stay the execution pending the passage of congressional legislation implementing the Convention. *Garcia*, 131 S. Ct. at 2867. Acknowledging the United States' warning "of the grave international consequences that will follow from" allowing the execution to go forward, the Court concluded, "Congress evidently did not find these consequences sufficiently grave to prompt its enactment of implementing legislation" for the Vienna Convention. *Id.* at 2868 (citations omitted). The Court held there was "no authority to stay an execution in light of an

Finally, the *Arar* opinion concluded that *Bivens* remedies were not available because hearing evidence in the case might disturb the "general allocation of authority over foreign relations to the political branches."[250] Arar claimed that he was wrongly removed to a country where he would likely be tortured; federal regulations "authorize the removal of an alien to a foreign country following receipt from that country of sufficiently reliable assurances that the alien will not be tortured."[251] A district court hearing the case would thus have to "determine whether any such assurances were received from" Syria and "whether the relevant defendants relied upon them in good faith."[252] According to the *Arar* opinion, such a determination would intrude upon the foreign relations authority of the political branches and therefore posed a special factor.[253] American tradition grants the Executive extra authority over foreign relations and national security, but how federal employees may act in furtherance of executive policies is still subject to the strictures of law created by Congress.[254] The fact that receiving and acting upon foreign assurances can implicate sensitive political areas does not indicate a congressional preference as to how claims *relating to* such assurances should be treated, which is the focus of the special factors inquiry.[255] In fact, this concern properly belongs to the state secrets privilege,

---

appeal of the President presenting free-ranging assertions of foreign policy consequences, when those assertions come unaccompanied by a persuasive legal claim." *Id.* (citation omitted) (internal quotation marks omitted).

250. Arar v. Ashcroft, 585 F.3d 559, 578 (2d Cir. 2009).

251. *Id.*; *see also* Implementation of the Convention Against Torture, 8 C.F.R. § 208.18(c) (2011) ("(1) The Secretary of State may forward to the Attorney General assurances that the Secretary has obtained from the government of a specific country that an alien would not be tortured there if the alien were removed to that country. (2) If the Secretary of State forwards assurances described in paragraph (c)(1) of this section to the Attorney General for consideration by the Attorney General or her delegates under this paragraph, the Attorney General shall determine, in consultation with the Secretary of State, whether the assurances are sufficiently reliable to allow the alien's removal to that country consistent with Article 3 of the Convention Against Torture."). Aliens may also be removed absent any assurances, but "[a] government report state[d] that [*Arar*] involve[d] assurances received from other governments in connection with the determination that Arar's removal to Syria would be consistent with Article 3 of the CAT." *Arar*, 585 F.3d at 577-78.

252. *Arar*, 585 F.3d at 578.

253. The *Arar* opinion does not specify the particular elements of this situation that it considered to be a special factor. Insofar as the court was concerned with the use of sensitive information in itself, that issue resembled the classified information issue discussed above. *See id.* Insofar as the issue was one of foreign relations, it implicated the political question doctrine, also discussed above. *See id.*

254. *Id.* at 575. Moreover, some of Arar's claims, such as the allegation that federal employees denied him access to counsel by lying to his attorney and that federal employees denied him access to a neutral adjudicator to determine his country of destination by serving his Notice to Appear while he was already in transit, did not implicate foreign relations. *Id.* at 563.

255. *Id.* at 577-78.

which implicates the distribution of authority between the Judiciary and the Executive by limiting discovery or, in the extreme case, barring a suit altogether if producing the necessary evidence would pose a grave danger to national security.[256] In *Arar*, the Government did not invoke that privilege.

One can imagine, of course, national security dangers attendant upon producing any assurances the United States received from Syria, and it may be impossible, as well as inappropriate, for a district court to determine if any such assurances were "sufficiently reliable."[257] It is not clear that either the production or the determination would have been necessary in prosecuting Arar's case. A court could certainly determine *whether* assurances were received without inquiring further into their content, as well as determining *whether* the relevant officials determined that they were "sufficiently reliable," as required by the regulations,[258] without entering either assurances or official determinations into the record. Similarly, whether federal employees acted in good faith is a fairly standard judicial question about the internal states of individuals subject to U.S. law.[259] Inquiring into good faith does not necessarily require evaluating the objective merit of that faith. Absent the government's proper invocation of state secrets privilege, then, there is little reason for a court to decide sua sponte that it applies, as the *Arar* court effectively did.[260]

None of the concerns cited in *Arar* to preclude constitutional damages fit the existing jurisprudence of special factors. *Arar* thus misunderstood, and muddied, the doctrinal coherence that other courts and cases had achieved. More importantly, *Arar*'s concerns focused on the relationship between the courts and the Executive.[261] As I have detailed, the point of a special factors analysis is to maintain the proper balance of power between the courts and Congress. Misunderstanding that rationale leads to more than just a muddier doctrine: It eliminates the whole reason for special factors analyses to exist at all. Plenty of

---

256. *See, e.g.*, *id.* at 605 (Sack, J., concurring in part and dissenting in part); *id.* at 634 (Calabresi, J., concurring in part and dissenting in part). In the Second Circuit, where *Arar* was heard, the state secrets privilege requires the "head of the department with control over the matter in question" to personally assert that "disclosure [of the information in question] would be inimical to national security." Zuckerbraun v. Gen. Dynamics Corp., 935 F.2d 544, 546 (2d Cir. 1991).

257. *Arar*, 585 F.3d at 578.

258. *See* Implementation of the Convention Against Torture, 8 C.F.R. § 208.18(c) (2011) (instructing that "[i]f the Secretary of State forwards assurances" regarding torture received in foreign countries "to the Attorney General for consideration by the Attorney General . . . , the Attorney General shall determine, in consultation with the Secretary of State, whether the assurances are sufficiently reliable to allow the alien's removal to that country," and providing that "[t]he Attorney General's authority under this paragraph may be exercised by the Deputy Attorney General or by the Commissioner, Immigration and Naturalization Service, but may not be further delegated").

259. *See, e.g.*, Mark J. Loewenstein, *The Diverging Meaning of Good Faith*, 34 DEL. J. CORP. L. 433, 442-43 (2009).

260. *Arar*, 585 F.3d at 567.

261. *See id.* at 578.

legal doctrines balance the relationship between the Judiciary and the Executive, and those doctrines could have amply addressed *Arar*'s concerns. The special factors analysis, in contrast, is a peculiarity of *Bivens*, designed to overcome the particular separation of powers issue *Bivens* raises.[262] Substituting the Executive's prerogatives for those of Congress will eliminate the justification for doing special factors analysis at all. Rather than allow the *Arar* holding to confuse the doctrine and eliminate the rationale of special factors, other courts should treat the *Arar* decision narrowly, as a statement about the context of extraordinary rendition, rather than about the proper way to perform the special factors analysis.

CONCLUSION

The premise of a *Bivens* action is that a constitutional "violation . . . by a federal agent acting under color of his authority gives rise to a cause of action for damages."[263] The point is that the Constitution guarantees certain rights even in the absence of a statutory remedy. To allow this judge-made cause of action, courts must determine that there is, in fact, no statutory remedy that addresses the harms of the plaintiff before them.[264] Courts make this determination through the special factors analysis, which asks whether Congress has afforded any remedy, or has decided to provide no remedy, for the violation of a constitutional right.[265] *Bivens* actions cover those cases for which Congress has neither provided a remedy itself, nor indicated how courts should address a particular kind of harm.[266] *Arar* is the rare outlier, having applied this reasoning backwards and concluding that the *absence* of any indications of congressional preference bars *Bivens* remedies.[267] In fact, it is precisely the absence of any indications of congressional preference that *allows for* remedies for constitutional damages.

As this Article has shown, special factors analyses in *Bivens* cases have developed into an inquiry of legislative will regarding remedies. The inquiry draws on the tools of federal preemption analysis. There are good reasons to eliminate the special factors analysis entirely,[268] but lower courts cannot do this on their own. To avoid the ad hoc (and post hoc) invention of special factors, which is so tempting in the absence of clear judicial definition, courts should

---

262. *See, e.g.*, Doe v. Rumsfeld, 800 F. Supp. 2d 94, 109 (D.D.C. 2011).

263. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 389 n.4 (1971).

264. *Id.* at 396.

265. *See* Schweiker v. Chilicky, 487 U.S. 412, 421 (1988).

266. *See* Carlson v. Green, 446 U.S. 14, 19-20 (1980).

267. Arar v. Ashcroft, 585 F.3d 559, 581 (2d Cir. 2009) ("[I]f Congress wishes to create a remedy for individuals like Arar, it can enact legislation that includes enumerated eligibility parameters, delineated safe harbors, defined review processes, and specific relief to be afforded. Once Congress has performed this task, *then* the courts in a proper case will be able to review the statute and provide judicial oversight . . . .").

268. *See* Pfander & Baltmanis, *supra* note 4, at 141-48.

adhere to the pattern of the special factors analysis set forth by the Supreme Court, which is usually implemented by most circuit courts and addresses the rationale of having courts determine whether to recognize a *Bivens* remedy in a particular case. This rationale is the proper distribution of authority between the judiciary and the legislature.

The public debate that sometimes goes along with high-profile *Bivens* cases, such as *Arar*, should not blind courts to the legislative focus of the special factors analysis. Public debate may, of course, itself spur Congress to legislate about such situations. Under the special factors analysis, such legislation would render the *Bivens* remedy superfluous. In this sense, figuring out how far *Bivens* goes—and figuring out how to understand how far it goes—may be particularly important now, as the ramifications of executive actions taken in the name of national security increasingly make their way into public consciousness and into the courts.

In a principled, precedent-based special factors analysis, a court uses the tools available for a federal preemption analysis to determine whether Congress has indicated its intent to prohibit a plaintiff's remedy for constitutional damages against individual federal employees. To do so, the court first asks whether a statutory scheme addresses the area at issue, including the kind of harm and the kind of plaintiff involved. If it identifies such a statutory scheme, the court ascertains that the statute does not merely touch on some of the issues or interests in the case: The statute must actually address, not just peripherally relate to, the area at issue in the case. The court also considers whether the statute indicates a congressional presumption that *Bivens* actions *would* be available for constitutional violations not covered in the statutory scheme.

Finally, the court should not mistake the Executive for Congress. As with statutory preemption, whereby administrative agencies can only issue regulations having a preemptive effect if Congress has delegated them the power to do so,[269] in *Bivens* cases, it is the *statutory* scheme that must indicate *Congress's* wish regarding remedies.[270] Without proper delegation, the Executive cannot immunize itself from *Bivens* remedies. Understanding this fundamental aspect of *Bivens* jurisprudence can help courts undertake the special factors analysis in a way that is consistent, in line with mainstream precedent, and faithful to the underlying concern animating the special factors inquiry.

---

269. *See, e.g.*, Thomas W. Merrill, *Preemption and Institutional Choice*, 102 NW. U. L. REV. 727, 740-41, 760, 768-69 (2008).

270. *See Chilicky*, 487 U.S. at 421-22.

THE LAW AND ECONOMICS OF NETWORK NEUTRALITY

THOMAS W. HAZLETT*
JOSHUA D. WRIGHT**

TABLE OF CONTENTS

INTRODUCTION

The Federal Communications Commission (FCC) released a Network

---

* Professor of Law & Economics and Director, Information Economy Project, George Mason University School of Law.

** Professor of Law, George Mason University School of Law and Department of Economics. We thank Judd Stone for extensive comments on an earlier draft, and Angela Diveley and Stephanie Greco for superb research assistance.

Neutrality Order on December 23, 2010 ("NN Order") that regulates broadband Internet Service Providers (ISPs).[1] The NN Order regulates via provisions concerning *transparency*, governing how broadband networks explain their services to customers; a *no blocking* provision, mandating that subscribers be permitted to deploy whatever computers, mobile devices, or applications they like for use with the network access service they purchase; and a *no unreasonable discrimination* rule for network management actions, such that ISP efforts to maintain service quality (e.g., mitigating congestion), or to price and package their services do not burden rival applications.[2] The policy is deemed *network neutrality* (NN), and the NN Order passed on a 3-2 vote (both Republican commissioners dissenting).[3]

The first item in this regulatory trio generates only modest controversy; indeed, opponents of NN often suggest that full and frank disclosure of ISP practices are all that are needed for a well-functioning market.[4] Moreover, the NN Order mitigated potential opposition to such rules by declining to mandate any specific disclosure format, noting that "the best approach is to allow flexibility in implementation of the transparency rule, while providing guidance regarding effective disclosure models."[5] The second and third rule provisions, however, are intensely controversial both with respect to the agency's legal jurisdiction and to their ultimate economic effect.[6] Verizon and others have already issued challenges to the FCC's jurisdictional authority to execute the NN Order.[7] If the regulations are found to fall under the FCC's statutory charter, the rules will ostensibly reduce the discretion of broadband ISPs in how they price and bundle their services. Operators will generally not be allowed to impose the following "vertical restrictions" on their customers:

- the outright blocking of certain legally available Internet content;
- subscriptions that include services or applications delivered at lower prices and/or better quality than competing applications;
- provision of different levels of transport speed or reliability to differing applications; and

---

1. *In re* Preserving the Open Internet Broadband Indus. Practices, 25 FCC Rcd. 17,905 (2010) [hereinafter FCC NN Order].

2. *Id.* para. 1, at 17,906.

3. *See* Press Release, FCC, FCC Acts to Preserve Internet Freedom and Openness (Dec. 21, 2010), *available at* http://online.wsj.com/public/resources/documents/fcc122110.pdf.

4. *See, e.g.*, *In re* Economic Issues in Broadband Competition a National Broadband Plan for Our Future, 2010 WL 45550, at *11-12 (Jan. 4, 2010) [hereinafter DOJ 2010] (*Ex parte* Submission of Dep't of Justice).

5. FCC NN Order, *supra* note 1, para. 56, at 17,938.

6. *See* Gerald R. Faulhaber, *The Economics of Net Neutrality: A Review*, 3 COMMC'NS & CONVERGENCE REV. 7, 7 (2011); Aaron K. Brauer-Rieke, Note, *The FCC Tackles Net Neutrality: Agency Jurisdiction and the Comcast Order*, 24 BERKELEY TECH. L.J. 593, 599-605 (2009).

7. *See* Josh Smith, *Verizon Files Lawsuit Against FCC's Net Neutrality Rules*, NAT'L J., Sept. 30, 2011, http://www.nationaljournal.com/tech/verizon-files-lawsuit-against-fcc-s-net-neutrality-rules-20110930.

- charging of fees to content providers accessing end users on their broadband network.[8]

NN restrictions are imposed on both fixed and mobile (wireless) broadband networks, although rules for the latter are stated in narrower terms.[9] How tightly the regulations will be enforced is unclear, as the FCC has given itself wide latitude in enforcement.[10] For instance, "network management" is barred only if it is "unreasonable."[11]

Enforcement complications are nicely illustrated in the first complaint filed under the rules, received by the FCC on January 10, 2011.[12] The petition alleges that MetroPCS, the country's fifth largest mobile telephone network, violated NN by discriminatorily favoring one video site over another.[13] This allegation stems from the pricing schedule set by MetroPCS:

- a $60 per month "all you can eat" plan for unlimited voice, texting, and data over its advanced Fourth Generation (4G) network;
- a $40 plan for its older 2G network, unlimited voice calls, texting, and email/web browsing—but excluding video streaming, except YouTube videos, which are available to subscribers without limit.[14]

According to MetroPCS, YouTube videos are included in the cheaper package because they are very popular with MetroPCS customers, and Google, the owner of YouTube, constructed a special compression technique permitting the 2G network to transfer video files without the congestion spillovers normally incurred by video streaming.[15] Rival video sites are disadvantaged by the arrangement, but MetroPCS gains no benefit from that outcome (it has no ownership interest in Google and receives no compensation from the content

---

8. Gregory L. Rosston & Michael D. Topper, *An Antitrust Analysis of the Case for Wireless Network Neutrality* 1, 3-14 (Stanford Inst. for Econ. Policy Research, Discussion Paper No. 08-40, 2009), *available at* http://businessinnovation.berkeley.edu/Mobile_Impact/Rosston_Topper_Wireless_Net_Netruality_11_02_09.pdf.

9. The *no unreasonable discrimination* rule is applied only to fixed networks. FCC NN Order, *supra* note 1, para. 1, at 17,906. While the *no blocking* rule applies to mobile as well as fixed operators, it only prohibits blocking of services competing with mobile network voice or video products. *Id.*

10. *See* Dawn C. Nunziato, *By Any Means Necessary? The FCC's Implementation of Net Neutrality*, 8 FIRST AMENDMENT L. REV. 138, 154-55 (2009).

11. FCC NN Order, *supra* note 1, para. 6, at 17,908 ("Network management practices are reasonable if they are appropriate and tailored to achieving a legitimate network management purpose. Transparency and end-user control are touchstones of reasonableness.").

12. M. Chris Riley, *Ex Parte* Presentation from Free Press to Julius Genachowski, Chairman, FCC (Jan. 10, 2011), *available at* http://www.mediaaccess.org/wp-content/uploads/MetroPCSLetter011011.pdf.

13. *Id.* at *3.

14. *Id.* at *1, *2 n.5.

15. Carl W. Northrop, *Ex Parte* Presentation from MetroPCS Communications, Inc. to Julius Genachowski, Chairman, FCC, at *6 (Feb. 14, 2011), *available at* http://apps.fcc.gov/ecfs//document/view.action?id=7021029361.

provider); MetroPCS only benefits from the enhanced satisfaction of its customers.[16] While the case against anticompetitive foreclosure is overwhelming,[17] it is unclear whether the FCC will dismiss the complaint. Indeed, the NN Order *invites* the allegation, pointedly using the MetroPCS price schedule to illustrate the type of business arrangement it considers to be problematic discrimination.[18] It has quickly turned into an example of the complexity of the NN trade-offs—imposing restrictions inhibiting an innovative, low-cost competitor in broadband access markets so as to protect an "open" flow of traffic—inherent in NN policy enforcement.

While Internet growth and innovation are significant, the FCC finds that the marketplace "faces real threats."[19] Left unregulated, the FCC believes broadband providers will inevitably be tempted to bias the access service provided to end users by favoring applications that they own or are paid to support.[20] This would force upstart service suppliers to bargain with a "gatekeeper," and this undermined the ability of users "at the edge" of the "open internet" to freely communicate with all others.[21] The result would be a disruption of the virtuous circle—infrastructure builders creating demand for content and applications, and then content and applications driving demand for more infrastructure investment—fueling Internet growth. "Restricting edge providers' ability to reach end users, and limiting end users' ability to choose which edge providers to patronize, would reduce the rate of innovation at the edge and, in turn, the likely rate of improvements to network infrastructure."[22]

This Paper critiques the NN policy—specifically, the no blocking and no unreasonable discrimination rules. After a short legal analysis evaluating the likelihood that the FCC's rules are likely to be declared beyond the scope of the agency's charter in Part I, the Paper focuses on the economic impact of net neutrality regulations. In Part II, the Paper explains the regulatory status of the Internet. It is beyond paradoxical that the FCC argues that it is imposing new regulations to preserve the Internet's current economic structure; a structure that has developed, thus far, in an unregulated environment where firms are free to experiment with business models—and vertical integration—at will.

---

16. *Id.* at *12.

17. *See* Thomas W. Hazlett, *FCC Net Neutrality Rules and Efficiency*, FIN. TIMES, Mar. 29, 2011, http://www.ft.com/cms/s/0/f75fd638-5990-11e0-baa8-00144feab49a.html.

18. The NN Order states, "These dangers to Internet openness are not speculative or merely theoretical." FCC NN Order, *supra* note 1, para. 35, at 17,925. It then lists several examples. In the next paragraph, it offers, "a major mobile broadband provider prohibits use of its wireless service for 'downloading movies using peer-to-peer file sharing services' and VoIP applications." *Id.* para. 36, at 17,926. The footnote cites MetroPCS, which in addition to limiting video streaming on its 2G network, also (and for similar reasons) limited peer-to-peer voice calls. *Id.* at 17,926 n.114.

19. *Id.* para. 4, at 17,907.

20. *Id.* para. 32, at 17,923.

21. *Id.* at 18,043-46.

22. *Id.* para. 14, at 17,911.

Part III explores the widespread use of "non-neutral" business forms by ISPs, Internet backbone providers, and application developers. Far from the Internet being an architectural construction, the network of networks is an evolving ecosystem in which key linkages between the "transport layer" and the "content/application layer" are efficiently deployed, advancing innovation, serving consumers, and driving Internet growth. "Walled gardens" are an essential part of the Internet and exist (in varying forms) throughout the market.[23] Indeed, they have since the first government-run interconnected data networks in the U.S. defense establishment precluded unauthorized users, uses, and—categorically—all commercial enterprises.[24]

Walled gardens have enabled the emergence of mass market e-commerce via innovative business models deployed by AOL, NTT DoCoMo, and Apple iPhone, among others. They are adapted by content providers, like ESPN3, to create new models for delivering high-quality programs by selling only to ISPs and not to end users. They have ushered upstarts into the market, as when—in 2002—a fledgling Google wagered its future by paying dominant ISP AOL to feature its search utility as a default application on its subscribers' start-up page. They have also proven pivotal in attacking and overcoming the mother of all monopolies in plain old telephone service (POTS).[25]

In this episode, cable operators now offer "digital voice" fixed line phone service to over ninety-five percent of U.S. households—don't cry for Ma Bell—with guaranteed quality-of-service (QoS) for calls, using dedicated, congestion-free bandwidth on their own local systems—an advantage unavailable to independent voice-over-Internet (VoIP) providers and, hence, both discriminatory and procompetitive.[26] Such "gardens" have not hampered the evolution of networks or killer applications but bolstered incentives for investors and produced the very "innovation commons" upon which the FCC marvels today.[27]

Part IV lays out the economic problem that the NN rules aim to counter: anticompetitive foreclosure. Actions by firms resulting in this outcome are already illegal under the antitrust laws, where the "rule of reason" is employed to separate socially beneficial practices from those that are harmful.[28] NN goes far further than existing law, categorically prohibiting various forms of economic integration in a manner equivalent to antitrust's per se rule, properly reserved for conduct that is so likely to cause competitive harm that the marginal benefit of a fact-intensive analysis cannot be justified.[29] In this case, the NN Order bans

---

23. *Id.* para. 94, at 17,956.

24. *See id.* at 17,956-57.

25. *See* Salil K. Mehra, *Paradise Is a Walled Garden? Trust, Antitrust, and User Dynamism*, 18 GEO. MASON L. REV. 889, 903-13 (2011).

26. *See* Christopher S. Yoo, *Network Neutrality, Consumers, and Innovation*, 2008 U. CHI. LEGAL F. 179, 202.

27. FCC NN Order, *supra* note 1, at 18,041.

28. *See* Cont'l T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36, 50-51 (1977).

29. *See* Yoo, *supra* note 26, at 245-47.

conduct that is typically highly efficient, promoting investment and innovation, as has been demonstrated in the Internet space repeatedly.[30] While the FCC purports to examine instances to the contrary, neither the economic literature concerning vertical contracting practices, such as those banned by the NN Order, nor the FCC's collection of anecdotal allegations of anticompetitive foreclosure can withstand scrutiny. Part V deconstructs these speculative claims of anticompetitive ISP conduct.

Part VI then deals with the economic arguments marshaled by the FCC to support its claim that anticompetitive foreclosure threatens to disrupt broadband market gains. On the one side, the FCC ignores compelling evidence that "open access" regulations have distorted broadband build-out in the United States by reducing subscriber growth when imposed and increasing subscriber growth when repealed.[31] On the other hand, the FCC manages to cite just one study—not of the broadband market—to support its claims of widespread foreclosure threats.[32] This empirical study, upon closer scrutiny than the FCC appears to have given it, actually shows *no evidence* of anticompetitive foreclosure. This fatal analytical flaw constitutes a smoking gun in the FCC's economic analysis of net neutrality. Part VII reviews evidence from U.S. broadband markets that both demonstrates the competitiveness of those markets and undermines the economic basis of the NN Order.

I. FCC JURISDICTION

Some critics refer to the FCC's net neutrality policy as "unprecedented."[33] This proves an excessively charitable summary, as salient precedent rebukes the FCC's overtures towards far-reaching ancillary jurisdiction over the Internet itself. The FCC unsuccessfully attempted to claim such jurisdiction merely eight months prior to the NN Order in *Comcast Corp. v. FCC*.[34] In *Comcast*, several internet end users noticed their internet service provider—Comcast—reduced the traffic speed to certain peer-to-peer file sharing applications.[35] Two net neutrality advocacy groups petitioned the FCC to enjoin Comcast from managing network traffic by differentiating data speeds.[36] Comcast capitulated to these

30. *See* Lee L. Selwyn & Helen E. Golding, *Revisiting the Regulatory Status of Broadband Internet Access: A Policy Framework for Net Neutrality and an Open Competitive Internet*, 63 FED. COMM. L.J. 91, 102 (2010).

31. Reply Comments from Thomas W. Hazlett to the FCC, *In re* Preserving the Open Internet Broadband Industry Practices, at *13-14 (Apr. 23, 2010), *available at* http://www.arlington economics.com/studies/NN-FCC-TWH-4-23-10.pdf.

32. *See* John B. Horrigan, *Broadband Adoption and Use in America*, 1-51 (FCC Omnibus Broadband Initiative, Working Paper No. 1, 2009), *available at* http://online.wsj.com/public/ resources/documents/FCCSurvey.pdf.

33. *See* Kay Bailey Hutchison & Fred Upton, *U.S. Innovation Is Hostage to Regulatory Overreach*, POLITICO, Feb. 17, 2011, http://www.politico.com/news/stories/0211/49652.html.

34. 600 F.3d 642 (D.C. Cir. 2010).

35. *See id.* at 644.

36. *Id.*

demands in light of an impending adverse order; the FCC ordered Comcast make required disclosures indicating Comcast's development of nondiscriminatory network management policies but indicated an injunction mandating neutrality would follow if Comcast failed to comply with the FCC's requirements.[37]

Comcast appealed to the D.C. Circuit, challenging the FCC's order on jurisdictional, amongst other, grounds—specifically, Comcast claimed that the FCC lacked jurisdiction to regulate Internet network practices.[38] The FCC conceded Congress had not granted it express authority to regulate Internet network management but instead claimed that regulation of Comcast's network management fell within its ancillary jurisdiction[39]—its power to "perform any and all acts, make such rules and regulations, and issue such orders, not inconsistent with [the Communications Act], as may be necessary in the execution of its functions."[40] The FCC cited two Congressional policy statements emphasizing the "continued development of the Internet" and the growth of a "rapid, efficient, Nation-wide . . . communication service" at "reasonable charges" as providing grants of authority, thereby enabling the FCC's regulation.[41] The FCC also offered a handful of patchwork statutory sections as express grants of authority in the alternative, most notably § 706 of the Telecommunications Act, providing the FCC "shall encourage the deployment on a reasonable and timely basis of advanced telecommunications capability to all Americans."[42]

The D.C. Circuit unequivocally rejected the FCC's interpretation, calling it "flatly inconsistent" with precedent and noting that "if accepted it would virtually free the Commission from its congressional tether."[43] The court first laid out the relevant test from Supreme Court precedents: To support ancillary jurisdiction, the FCC must demonstrate its action is "reasonably ancillary to . . . effective performance of its statutorily mandated responsibilities."[44] The court then highlighted one of the FCC's most extreme concessions at oral argument—that under the FCC's rationale, it could subject Comcast's internet service to "pervasive rate regulation" in order to ensure a "reasonable" price—in demonstrating the breadth of the FCC's interpretation of its ancillary jurisdiction.[45] The D.C. Circuit unequivocally held that simple Congressional policy statements, while useful in delineating the bounds of expressly delegated authority, did not constitute "statutorily mandated responsibilities" which could

37. *Id.* at 645.

38. *Id.*

39. *See id.*

40. 47 U.S.C. § 154(i) (2006).

41. *Comcast*, 600 F.3d at 651-62; *see also* 47 U.S.C. § 230(b) (2006); *id.* § 151 (2006 & Supp. 2010).

42. 47 U.S.C. § 1302(a) (Supp. 2010); *see also Comcast*, 600 F.3d at 658.

43. *Comcast*, 600 F.3d at 655.

44. *Id.* at 646.

45. *Id.* at 655.

ground the FCC's ancillary jurisdiction alone.[46] The court dispensed summarily with the FCC's § 706 argument, pointing to a prior FCC § 706 interpretation acknowledging the provision granted no regulatory authority.[47] The court thereby vacated the FCC's order against Comcast,[48] inspiring some legal speculation as to whether the FCC could implement net neutrality absent additional congressional permission whatsoever.[49]

The FCC's recent net neutrality rulemakings demonstrate it shares no academic equivocation as to the breadth of its authority.[50] "In an act of superior confidence or of sheer foolishness," the FCC employed substantially similar ancillary jurisdiction theories, which have already been rebuked by *Comcast*.[51] The NN Order invokes a pastiche of statutory provisions in order to justify its three net neutrality rules: parts of Titles II, III, and VI of the Communications Act and, most candidly, an open reinterpretation of § 706 to enable a jurisdictional interpretation consistent with the FCC's understanding "before the *Comcast* decision."[52] As one dissenting FCC Commissioner remarked, the FCC "discover[ed]" § 706, a regulatory "superpower, unlocked only after an adverse court opinion and political pressure to find some legal foundation" to justify the NN Order[53]—the language of which closely tracked a failed bill before Congress.[54] *Comcast* appears to have taught the FCC little in the way of interpretative humility in the FCC's search for an unbounded grant of regulatory authority over the Internet.

Yet the historical parallels between the FCC's jurisdictional assertions

---

46. *Id.*

47. *Id.* at 659.

48. *Id.* at 661.

49. *See, e.g.*, FCC NN Order, *supra* note 1, at 18,052 (McDowell, Comm'r, dissenting); Patric M. Verrone, *The* Comcast *Case and the Fight for Net Neutrality*, L.A. LAW., May 2011, at 9, 9 ("The FCC . . . has been at the center of the [net neutrality] debate, most recently issuing [the NN Order] . . . . How, and even if, it can enforce that order in light of the *Comcast* case goes to the core of the FCC's rule-making authority over the Internet." (footnote omitted)).

50. *See* FCC NN Order, *supra* note 1, para. 115, at 17,967 ("Broadband Internet access services are clearly within the Commission's subject matter jurisdiction and historically have been supervised by the Commission. . . . [O]ur adoption of the basic rules of the road for broadband providers implements specific statutory mandates in the Communications Act and the Telecommunications Act of 1996." (footnote omitted)).

51. *See* Babette E.L. Boliek, *FCC Regulation Versus Antitrust: How Net Neutrality Is Defining the Boundaries*, 52 B.C. L. REV. 1627, 1631-32 (2011). Many academics expressed confidence that the FCC had jurisdictional authority for its *Comcast* decision, later rejected by the D.C. Circuit. *See, e.g.*, Letter from Lawrence Lessig, Professor, Stanford Law School to the FCC (Aug. 20, 2008), *available at* http://lessig.org/blog/2FCC.pdf ("Whether or not the Commission has the authority it claims in this particular case (and I am confident that it does), no company has the right to mislead the Commission in its proceedings.").

52. *See* FCC NN Order, *supra* note 1, at 17,968-78, 18,095.

53. *Id.* at 18,096 (Baker, Comm'r, dissenting).

54. *Id.*

underlying the NN Order and prior FCC errors neither begin nor end with *Comcast*. Cable television's rise in the late 1960s inspired substantial fear in broadcast television companies; in turn, broadcast companies sought FCC regulation of cable companies.[55] This presented a jurisdictional conundrum. While the Communications Act expressly granted the FCC power to regulate broadcasting companies—and the FCC's jurisdiction over cable systems supported by microwave antennas was widely accepted—the Communications Act failed to contemplate, much less regulate, non-broadcast cable signal transmission.[56] The FCC asserted jurisdiction over comprehensive non-broadcast cable regulation as "ancillary" to its express power to regulate broadcast transmissions.[57] The Supreme Court upheld the FCC's regulation as necessary to effect its textually enumerated responsibilities,[58] leading to a wave of cable-company regulations including common ownership requirements, sponsorship disclosures, and the now-infamous Fairness Doctrine.[59] In short order, however, the expansive use of FCC jurisdiction led to absurd results, with one Seventh Circuit case having to go so far as to expressly delineate that the FCC lacked authority to regulate building construction simply on account of some relationship to broadcast interference.[60] As will likely occur with the NN Order, the protectionist pedigree of the 1960s and 1970s cable regulatory regime grew apparent in retrospect.[61]

The FCC's NN Order also presents serious constitutional problems. As several federal district courts have noted, Broadband ISPs likely enjoy First Amendment speech protections,[62] which the NN Order casually dismisses.[63] Indeed, the NN Order asserts with little explanation and even less precedent that broadband ISPs' network regulation serves no editorial function within the First

---

55. *See* Glen O. Robinson, *The New Video Competition: Dances with Regulators*, 97 COLUM. L. REV. 1016, 1019 (1997).

56. *See* Joseph R. Fogarty & Marcia Spielholz, *FCC Cable Jurisdiction: From Zero to Plenary in Twenty-Five Years*, 37 FED. COMM. L.J. 113, 115 (1985).

57. *Id.* at 115-18.

58. United States v. Sw. Cable Co., 392 U.S. 157 (1968).

59. *See Evolution of General Cable Television*, FCC, http://www.fcc.gov/encyclopedia/evolution-cable-television (last updated Mar. 14, 2012) (discussing regulations after *United States v. Southwestern Cable Co.*).

60. Ill. Citizens Comm. for Broad. v. FCC, 467 F.2d 1397, 1401 (7th Cir. 1972).

61. *See* Robinson, *supra* note 55, at 1019.

62. *See, e.g.*, Ill. Bell Tel. Co. v. Vill. of Itasca, 503 F. Supp. 2d 928, 947-49 (N.D. Ill. 2007) (recognizing cable and satellite providers' right to free speech); Pac. Bell Tel. Co. v. City of Walnut Creek, 428 F. Supp. 2d 1037, 1051-52 (N.D. Cal. 2006) (recognizing cable company's right to free speech); Comcast of Cal. I, Inc. v. City of Walnut Creek, 371 F. Supp. 2d 1147, 1157 (N.D. Cal. 2005) (recognizing cable operator's free speech rights); Comcast Cablevision of Broward Cnty., Inc. v. Broward Cnty, 124 F. Supp. 2d 685, 690-91 (S.D. Fla. 2000) ("It is now well established that regulation of cable operators implicates [free speech abuses] . . . of the First Amendment.").

63. FCC NN Order, *supra* note 1, paras. 143-44, at 17,983 & n.458.

Amendment's purview.[64] This approach overlooks substantial federal First Amendment jurisprudence imposing little to no editorial requirement to entitle a publisher or carrier some First Amendment protection when filtering the content of others.[65] Even this relaxed approach ignores the robust First Amendment protection ISPs enjoy when providing more content-related services, such as video programming, which likely fall under the NN Order's ambit.[66] Net neutrality advocates often cite the potential harms of an ISP squelching a rival product's traffic—or "favoring" its own traffic—through lower or higher data speeds, respectively.[67] It is relatively simple to envision a potential First Amendment conflict when the NN Order prevents an ISP from carrying its own traffic—its own speech—in its preferred method.[68]

Further defects in the NN Order illustrate the ad hoc and conceptually incoherent qualities inherent in the FCC's approach. The cause of some dissenting ridicule,[69] and with historical echoes to construction regulation, the FCC disclaimed any intent to regulate retail distributors of broadband ISP access, such as coffee shops, bookstores, and airlines.[70] The FCC instead parses these retailers out of the proposed regulations, deeming them "premise operators."[71] Yet the conceptual justification for net neutrality—grounded in fears that Internet providers may favor their own content or disfavor content of rivals—broaches no obvious exception.[72] A hypothetical illustrates the absurdity of the distinction:

---

64. *Id.* at 17,983.

65. *See* Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 653 (1994).

66. *See* Randolph J. May, *Net Neutrality Mandates: Neutering the First Amendment in the Digital Age*, 3 INFO. SOC'Y J.L. & POL'Y 197, 202 (2007).

67. *See, e.g.*, Comcast Corp. v. FCC, 600 F.3d 642, 644-45 (D.C. Cir. 2010) (discussing FCC complaints filed by Free Press and Public Knowledge); *Net Neutrality 101*, SAVE INTERNET, http://www.savetheinternet.com/net-neutrality-101 (last visited July 1, 2012) ("Net Neutrality means that Internet service providers may not discriminate between different kinds of content and applications online. . . . The biggest cable and telephone companies . . . . believe they should be able to charge Web site operators, application providers and device manufacturers for the right to use the network. Those who don't make a deal and pay up will experience discrimination: Their sites won't load as quickly . . . .").

68. *See* May, *supra* note 66, at 204 ("Even though [neutrality laws] do not literally 'restrict' an ISP from publishing content of its own choosing, they would compel the ISP to convey or make available content that, in its editorial judgment, it would otherwise choose not to convey or make available.").

69. *See* David Eldridge, *FCC Chief Defends New "Rules of Road" on Net Neutrality*, WASH. TIMES, Feb. 16, 2011, http://www.washingtontimes.com/news/2011/feb/16/fcc-chairman-defends-net-neutrality-rules/ (quoting Congressman Greg Walden's sarcastic remark: "I am relieved, however, that the FCC declined under its newfound authority to regulate coffee shops, bookstores, airlines and other entities" (internal quotation marks omitted)).

70. FCC NN Order, *supra* note 1, para. 52, at 17,935.

71. *Id.*

72. *See generally Net Neutrality 101*, *supra* note 67 (discussing the drawbacks of altered Internet).

What competitive problems inhere in AT&T degrading network traffic to Comcast that do not similarly infect Starbucks blocking Internet traffic to Caribou Coffee?

The above issues illustrate the limits of categorical mandates more than the limits of the FCC. Absurdities in both jurisdictional assertions and substance arise in applying categorical prohibitions to certain classes of conduct without reference to their actual causes and effects, both beneficial and malign. The FCC attempts to ameliorate these distinctions by creating ad hoc exceptions that prevent certain absurd outcomes without acknowledging that the NN Order inherently promotes other equally absurd outcomes. The counter-productivity of categorical prohibitions may prove a novel problem to the FCC—and an inherent one to the administrative rule-making process—but as it turns out, it is an exceptionally old dilemma to antitrust, which evolved a sophisticated balancing process for investigating and adjudicating these claims on a case-by-case basis: the Rule of Reason.[73]

II. THE NON SEQUITUR: SAVING THE (UNREGULATED) INTERNET BY REGULATING IT

> "[T]here is little dispute in this proceeding that the Internet should continue as an open platform."[74]

> "Broadband Providers Have the Incentive and Ability to Limit Internet Openness."[75]

The FCC's net neutrality policy perches on irony: If the new rules are needed to preserve the salubrious structure of the Internet, why has the asserted threat failed—by the FCC's own analysis—to yet undermine the "open platform"?[76] Why have broadband ISPs resisted the easy profits available from foreclosing competition among applications, squeezing their subscribers, and profiting from the very actions feared? The FCC does not ask this question, but it is worth answering.

The NN Order posits that "[t]oday, broadband providers have incentives to interfere with the operation of third-party Internet-based services that compete with the providers' revenue-generating telephony and/or pay-television services."[77] In a fundamental sense, that is undeniably true—firms are always tempted to extract additional consumer surplus, given the opportunity, and indeed have a fiduciary obligation to shareholders to pursue such returns. This may even be a defensive imperative in the marketplace, as the FCC notes: "[O]nline edge services appear likely to continue gaining subscribers and market

73. *See* Nat'l Soc. of Prof'l Eng'rs v. United States, 435 U.S. 679, 688-92 (1978) (discussing the rule of reason).

74. FCC NN Order, *supra* note 1, para. 19, at 17,915.

75. *Id.*

76. *See id.* para. 1, at 17,906 (describing the Internet as "open").

77. *Id.* para. 22, at 17,916.

significance, which will put additional competitive pressure on broadband providers' own services."[78]

Exactly—spotlighting the rivalrous tension promoting customer interests. These "competitive pressures" spring from market forces unconstrained by network neutrality regulation.[79] "[B]roadband providers have the ability to act as gatekeepers,"[80] and if they thereby "have the incentive and ability" to pre-empt these proconsumer options, they have taken a different path, creating an outcome so robust as to create consensus that "the Internet should continue as open platform."[81] The marketplace that the FCC recommends preserving stands as an ongoing experiment as to whether the dangerous consequences the FCC warns of will obtain without new rules. Instead of reporting on those harms materializing, the regulator instead reports that unregulated gatekeepers do just what the FCC would like them to.[82]

Actually, the market process is far more nuanced, and market structure far more interesting, than the FCC describes. The FCC sees the Internet as a constructed edifice, engineered to produce a particular flow of communications.[83] This vision is incorrect in technical terms. Moreover, it is immediately misapplied by extending the perceived structural template as a descriptor of *economic relations*.[84] Whatever the engineering designs of networks or the interfaces between them, the terms of trade on which demanders and suppliers transact are *economic*. Those terms are not, and have never been, "open end-to-end."[85] They are the standard building blocks of markets: property and contracts, layered upon a general legal regime enabling ownership, production, and trade.

Whatever is argued about the manner in which networks operate, the creation of networks is a financial investment. So, too, the wide array of inputs and complements supporting the ecosystem—from website applications, to online services, to content, to private networks, to virtual private networks, to servers, routers, and the hardware and software employed by subscribers to access their broadband ISP. Backbone networks, transporting high-volume data flows over long-distance links, send and receive traffic at prices negotiated with other

---

78. *Id.* para. 22, at 17,917-18.

79. *See id.*

80. *Id.* para. 24, at 17,919.

81. *Id.* para. 19, at 17,915.

82. *See id.* para. 1, at 17,906.

83. *See id.* para. 13, at 17,909 ("The Internet's founders intentionally built a network that is open, in the sense that it has no gatekeepers limiting innovation and communication through the network.").

84. *See id.* para. 23, at 17,918 (discussing how broadband providers may exclude rivals of paying edge providers).

85. *See generally* James Kempf & Rob Austein, *The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture*, INTERNET SOC'Y (Mar. 2004), http://tools.ietf.org/html/draft-iab-e2e-futures-04 (discussing the "end-to-end" principle).

networks.[86] Large backbones often use "bill-and-keep" contracts, where traffic is exchanged without payments either way, a practice known in the Internet as "peering."[87] Such arrangements reduce transaction costs when traffic flows are roughly commensurate.[88] When smaller networks connect to larger ones, however, it pays to keep track, and the smaller typically pays the larger.[89] The unregulated system creates an accounting system with valuable properties, matching supply with demand, supporting Internet build-out, and yielding incentives for the creation of larger, faster networks.

The market is "open" or "neutral" in that entry is free, and costs and consumer demand interact to set prices. But this is distinctly not the "neutrality" advanced by the FCC, where suppliers—transport networks on one side, applications providers on the other—rigidly adhere to "layers" with strict boundaries.[90] In the FCC's view, data networks are "dumb pipes" that stick to their assigned task, treating all traffic, all applications, and all other carriers alike.[91] In reality, this system cannot be saved by network neutrality because it never existed.

Networks discriminate against other networks, refusing to accept traffic from those which do not offer satisfactory payments.[92] Were smaller networks not paying discriminatorily higher prices to send their traffic to larger networks, the incentive to invest more to grow larger would be greatly diminished, reducing Internet performance via free riding—"tragedy of the commons."[93] Content providers, as well, pay to play. Google, Microsoft and Yahoo! monetize ad revenues from intention-based advertising surrounding "free" key word searches offered to end users.[94] Each of these applications is created by large investments not only in web crawlers that catalogue millions of websites and in proprietary

---

86. Michael Kende, *The Digital Handshake: Connecting Internet Backbones*, 1, 5 (FCC, Office of Plans & Policy, Working Paper No. 32, 2000), *available at* http://transition.fcc.gov/Bureaus/OPP/working_papers/oppwp32.pdf.

87. *Id.* at 8.

88. *See id.* at 8 n.26 ("If traffic flow [is] . . . balanced, the net settlement that each pays is zero . . . .").

89. *See id.* at 7 (noting that a transit system is used between larger and smaller backbones).

90. *See* FCC NN Order, *supra* note 1, para. 7, at 17,908 (supporting a separation between broadband providers and application providers).

91. *See id.* at 18,091 (Baker, Comm'r, dissenting) (accusing majority of viewing data networks as "dumb pipes").

92. *See* Kende, *supra* note 86, at 16-17 (listing examples of networks discriminating against one another).

93. *See generally* Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243 (1968); *see also* MICHAEL HELLER, THE GRIDLOCK ECONOMY 17 (2008) (describing tragedy of the commons as a situation when everyone is too motivated by personal interest to care about the sustainability of a shared resource).

94. *See, e.g.*, *How Does Google Make Money? What Is Driving Googles Growth?*, GOOGLE: INVESTOR RELATIONS, http://www.investor.google.com/corporate/faq.html#toc-money (last visited July 1, 2012).

search algorithms that attempt to deliver matches most valuable to the user, but via bundling (integration) with high-quality transport service.[95] The Internet's layers naturally mix. When customers get their search results faster, they are happier, and more likely to return for more. Speed is therefore a key competitive advantage, not just in search but in virtually every application on the Internet.

Hence, firms integrate. Content companies like Google construct their own global delivery networks;[96] others purchase such speed-enhancements through content delivery networks (CDNs) like Akamai, BitGravity, or Limelight Networks.[97] Some ISPs have attempted to compete with these CDNs by providing application vendors local caching services (storing commonly requested data on services nearer to end users, speeding delivery) for an extra fee.[98] This constitutes a pay-for-fast-delivery option that improves service for customers, allowing entrants to better compete with incumbents. When an upstart search engine takes on Google, the opportunity to pay an ISP for faster service—closing the gap with Google's own global CDN—is closed.

The FCC is more or less right when it says that "[t]he Internet is a level playing field. Consumers can make their own choices about what applications and services to use . . . ."[99] But it is wrong when it attributes that outcome to a structure that quarantines ISPs, keeping them from actively managing networks, creating content alliances, or exercising "gatekeeper" functions. As shown in the next Part, ISPs commonly engage in such "non-neutral" behavior, always have, and in so doing advance the Internet's "innovation commons."[100] Nonetheless, a market structure has emerged that exhibits a striking degree of transport specialization by ISPs, which elect to leave development of most services and applications to third party suppliers. The mass-market "walled gardens" of years past, including those of AOL and Excite@Home, have faded.[101] This result has not been produced by engineering design but by profit calculus under competitive market (including capital market) constraints.[102] The choices made by ISPs

95. *See Benefits of Bundling: Integrating Cable and Online Advertising*, COX MEDIA, http://www.coxmedia.com/basic.jsp?page=82&navigation=27 (last visited July 1, 2012).

96. *See* Bradley Zarich, *Google-Led Consortium Vows to Multiply Worldwide Internet Speed*, WEB TRICKS (Sept. 12, 2011), http://thewebtricks.com/cdn-usage-and-technology/.

97. *See Content Delivery and Distribution Services*, WEB CACHING, http://www.web-caching.com/cdns.html (last visited July 1, 2012).

98. *See, e.g.*, Matt Evans, *New Time Warner Pricing Could Boost ISP Competition*, BUS. J., Apr. 20, 2009, http://www.bizjournals.com/triad/stories/2009/04/20/story5.html.

99. FCC NN Order, *supra* note 1, para. 3, at 17,907.

100. *See* discussion *infra* Part III.

101. *See* Dan Frommer, *Amazingly, AOL Still has 3.5 Million Dialup Subscribers*, SPLATF (Nov. 3, 2011, 10:09 AM), http://www.splatf.com/2011/11/aol-dialup-charts/ (showing a decline in AOL users); Ben Heskett & Rachel Konrad, *Excite@Home Files for Bankruptcy*, CNET, Oct. 1, 2011, http://www.news.cnet.com/2100-1033-273689.html.

102. *See* Jennifer Bosavage, *The Rise and Fall (and Rise?) of AOL*, INFO. WEEK, Oct. 4, 2006, http://www.informationweek.com/news/193104723 (attributing AOL's decline to DSL companies that offered free services).

reflect trade-offs in garnering additional revenues (say, by "blocking or degrading content" to favor affiliated services)[103] against revenues lost due to subscriber defections.

Curiously, the FCC sees this, understanding that firm self-interest provides an efficiency check.[104] It argues for new NN rules based, in part, upon the fact that "the market has already spoken in favor of nondiscriminatory access by turning away from 'walled gardens' such as AOL, Genie, Delphi, Prodigy, and Compuserve."[105] This is, firstly, a highly incomplete rendition of history. The market *turned to* "walled gardens" during an important time, and the model succeeded because consumers were well served (and therefore had a higher demand for the ISP subscriptions) by the proprietary content that the "gardens" grew. This enabled a critical extension of e-commerce into the mass-market, both by encouraging AOL's "carpet bombing" of America with millions of easy-to-use dial-up sign-up disks,[106] a marketing investment of considerable scale, and then by driving an enormously positive response to the campaign by consumers. As Ken Auletta describes it: "Webheads would sneer that using AOL was 'the Internet on training wheels.' Yet it was AOL's user-friendliness that helped popularize the Web—and which attracted thirty-four million paid subscribers in 2002."[107] That the integrated model became markedly less useful, as content markets flourished throughout the (non-AOL) Internet, is reflected in the reality that AOL and other ISPs migrated to new modes.[108] The correct takeaway is that markets reflect efficiencies, not that a given structure, at a given point in time, is the "correct" model to freeze into place by law. The marketplace reveals efficiencies by continually testing new options and discovering what innovations might improve upon extant operations. Indeed, the recent purchase of the *Huffington Post* (a news and opinion website) by AOL is emblematic of the ongoing search for optimality.[109] It is a $315 million wager that the ISP ownership of content may indeed have some new, or remaining, efficiencies.[110]

---

103. FCC NN Order, *supra* note 1, para. 4, at 17,907.

104. *Id.* paras. 24-25, at 17,919 (describing changing fees to edge providers as inefficient if too high).

105. Comments of Verizon and Verizon Wireless, *In re* Preserving the Open Internet Broadband Industry Practices, at *37-38 (Jan. 14, 2010), *available at* https://prodnet.www.neca.org/publicationsdocs/wwpdf/0114verizonnn.pdf.

106. In 1996 alone, AOL distributed some 250 million subscription sign-up disks in the United States. KARA SWISHER, AOL.COM: HOW STEVE CASE BEAT BILL GATES, NAILED THE NETHEADS, AND MADE MILLIONS IN THE WAR FOR THE WEB 99 (1998).

107. KEN AULETTA, GOOGLED: THE END OF THE WORLD AS WE KNOW IT 94-95 (2009).

108. Bosavage, *supra* note 102 (noting that AOL has abandoned its "walled garden" approach in past years).

109. *See* Rebecca Kaplan, *AOL Buys* Huffington Post *in $315M Deal*, NAT'L J., Feb. 7, 2011, http://www.nationaljournal.com/whitehouse/aol-buys-huffington-post-in-315-million-deal-20110207.

110. *See* Jeremy W. Peters & Verne G. Kopytoff, *Betting on News, AOL Is Buying the* Huffington Post, N.Y. TIMES, Feb. 7, 2011, http://www.nytimes/2011/02/07/business/media/

Second, the evidence that the market migrated away from "walled gardens" is one which suggests regulation is unnecessary. The FCC here argues that the outcome of an unregulated market process was efficient, but then argues to disrupt that process to mandate administratively designed outcomes[111]—the recurring non sequitur.

Third, the argument reflects the FCC's underlying assumption that firms respond rationally to economic challenges. The methodology is uncontroversial; the "market test" provides subtle, essential, and far-reaching information. But the FCC then seeks to advance a new regime under which such valuable data cannot be revealed. Business models will be regulated, and—should certain types of economic integration become more efficient—rigid structures preclude experimentation.

In fact, there are often strong economies to integration, but also many productive gains from specialization. When the latter outweigh the former, these services are generally uneconomic for the ISP to supply. The alternative possibility is that ISPs inefficiently provide certain services, and/or impose various vertical restrictions, because they are able to quash competitive forces in the process. As witnessed by the FCC's lack of evidence to this effect—the FCC, even while claiming jurisdiction, has prosecuted virtually no actual instances of anticonsumer conduct, and argues for NN rules on the basis of a *looming* threat—there is no sign that anticompetitive foreclosure is driving ISP structural choices.[112] Conversely, when ISPs do integrate into complementary

---

070101.html ("AOL is betting that it can sell, alongside [Huffington Post's] content, local advertising and display advertising, areas that Google does not dominate.").

111. *See* FCC NN Order, *supra* note 1, para. 1, at 17,906 ("To provide . . . continued freedom and openness of the Internet, we adopt three basic rules . . . .").

112. It has been widely noted in the NN proceeding that the FCC repeatedly cites just two instances of egregious "gatekeeper" conduct by broadband ISPs. *See id.* para. 37, at 17,926-27. In 2005, a small telephone company in North Carolina, Madison River, blocked the use of VoIP services for its DSL customers. A $15,000 fine was imposed on the operator, which agreed to discontinue the practice. *See id.* para. 35, at 17,925. This is the only such example noted by the Commission among the more than 1,000 fixed line telephone carriers in the United States. "There are approximately 1,300 companies that have historically provided local telephone service in the United States." FCC, STATISTICS OF COMMUNICATIONS COMMON CARRIERS (2006-2007), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/ DOC-301505A1.pdf. In 2008, Comcast, the country's largest cable TV operator, blocked certain users' peer-to-peer downloads without adequately disclosing the practice to customers. *In re* Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation for Secretly Degrading Peer-to-Peer Applications, 23 FCC Rcd. 13,028, 13,030-31 (2008), *order vacated by* Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010). Comcast reached an agreement with BitTorrent, the peer-to-peer software vendor whose users were being blocked, which was then approved by the Commission. *Id.* at 13,067 (Martin, Chairman, opinion). The FCC, nonetheless, issued an Order, which was then overturned by the D.C. Circuit Court of Appeals as being beyond the scope of its legislative charter. Comcast Corp. v. FCC, 600 F.3d 642, 644 (D.C. Cir. 2010). Neither instance was likely to constitute anticompetitive foreclosure, but even if they had, they would constitute little justification for a

services, and the foray survives, it signals that the ISP profits from the integration. The FCC, which supposes that any such integration is prima facie evidence of foreclosure, short-circuits the analysis: It must show, not only that ISPs integrate, but that they do so to achieve anticompetitive, inefficient outcomes. Indeed, it must show far more: that such anticompetitive outcomes are ubiquitous rather than rare. Otherwise, a categorical prohibition would likely incur far more net costs than a case-by-case adjudication under antitrust law. The FCC must also show that the collateral damage inflicted by its rules—a first-order decrease in investment incentives, as networks lose property rights to manage their systems—does not outweigh the social gains.

In fact, the FCC's articulated goal of "preserving" what it observes to be an "open Internet" is all that is needed to reject the policy advanced.[113] Were the anticompetitive opportunities ubiquitous under a regime permitting integration, then the Internet would not constitute either an "open Internet" nor one worth "preserving." The cognitive dissonance of regulating an unregulated market to protect what has emerged is, however, of longstanding and high pedigree. For over a decade, we have been living under the dark cloud of Internet death, even as amazing innovations from the network of networks rock our world and then rock it again. Harvard Law Professor Lawrence Lessig wrote in 2001:

> The Internet revolution has ended just as surprisingly as it began. None expected the explosion of creativity that the network produced; few expected that explosion to collapse as quickly and profoundly as it has. The phenomenon has the feel of a shooting star, flaring unannounced across the night sky, then disappearing just as unexpectedly. Under the guise of protecting private property, a series of new laws and regulations are dismantling the very architecture that made the Internet a framework for global innovation.[114]

---

sweeping new regulatory policy. As Georgetown University economist Marius Schwartz writes, "This is a remarkably thin record on which to even contemplate the far-reaching regulation. Furthermore, both incidents were swiftly addressed in the absence of the proposed rules, further spotlighting the glaring gap between the proffered rationale for intervention and the proposed rules." Exhibit 3: Declaration of Marius Schwartz, at 4, *In re* Preserving the Open Internet Broadband Indus. Practices, GN Docket No. 09-191 (FCC Jan. 14, 2010) [hereinafter Schwartz 2010]. U.C. Berkeley economist Michael Katz cites additional FCC examples of potential neutrality violations and quotes the FCC's words in stating: "*Even if all of these assertions were correct*, which is far from evident, they would not establish that the *NPRM*'s proposed rule against discrimination would promote consumer welfare. In fact, these assertions contribute nothing toward 'distinguishing socially beneficial discrimination from socially harmful discrimination in a workable manner.'" Attachment B: Declaration of Michael L. Katz, at 38, *In re* Preserving the Open Internet Broadband Indus. Practices, GN Docket No. 09-191 (FCC Jan. 9, 2010) [hereinafter Katz 2010] (quoting Preserving the Open Internet, 74 Fed. Reg. 62,638, 62,646 (Oct. 22, 2009) (to be codified at 47 C.F.R. pt. 8)).

113. *See* FCC NN Order, *supra* note 1, para. 1, at 17,906.

114. Lawrence Lessig, *The Internet Under Siege*, FOREIGN POL'Y, Nov.-Dec. 2001, at 56, 56.

Lessig articulated just the model of the Internet that the FCC relies on today. He added, "Policymakers need to understand the importance of this architectural design to the innovation and creativity of the original network."[115] This structure, which allegedly baked in "end-to-end" data flows to avoid frictions imposed by transport networks,[116] was being violated by the "walled gardens" of the Internet. AOL had emerged as the dominant U.S. ISP, in large measure because it offered its customers' proprietary content.[117] In 2000 several major cable operators had entered into exclusive agreements with ISPs such as Excite@Home and RoadRunner,[118] companies that—like AOL—sought to optimize the user's experience by providing some of the content and applications that they typically would access.[119] This is just the form of discrimination feared by the FCC today.

So the development of the broadband market over the past decade affords an opportunity for an 'out of sample' forecast of the Lessig hypothesis. What has emerged? Broadband was nascent when Lessig wrote; not one percent of U.S. households subscribed.[120] Today, over seventy percent do—one index of "the explosive growth in the use of broadband."[121] As the Commission itself describes it:

> [B]roadband is transforming American life . . . . Parents on business trips use their smartphones to check e-mail or watch short videos of their children playing soccer, hundreds, if not thousands, of miles away. Americans work together in real time on complex documents from different desks in the same office, and workers in different offices around the world collaborate via videoconferencing technology. . . . Students draw on the richness of the Internet to research historical events or watch simulations of challenging math problems. People are using broadband in ways they could not imagine even a few years ago.[122]

"Death," or "anticompetitive foreclosure," is difficult to find in the FCC's description of the Internet ecosystem, one decade on from Lessig's diagnosis.[123]

---

115. *Id.*

116. *See* Mark A. Lemley & Lawrence Lessig, *The End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era*, 48 UCLA L. REV. 925 (2001).

117. *See id.* at 944.

118. *See* Jerry A. Hausman et al., *Cable Modems and DSL: Broadband Internet Access for Residential Customers*, 91 AM. ECON. REV. 302, 302 (2001).

119. Lemley & Lessig, *supra* note 116, at 927-28.

120. *See* U.S. DEP'T OF COMMERCE, DIGITAL NATION: EXPANDING INTERNET USAGE 7 (2011).

121. FCC, CONNECTING AMERICA: THE NATIONAL BROADBAND PLAN 16 (2010) [hereinafter CONNECTING AMERICA], *available at* http://download.broadband.gov/plan/national-broadband-plan.pdf.

122. *Id.* at 15.

123. Professors Mark Lemley and Lawrence Lessig made further predictions as to the development of broadband markets. *See generally* Lemley & Lessig, *supra* note 116. Most striking was the fear that cable TV operators would suppress online video streaming to protect their cable TV service revenues. *Id.* at 944. The explosion of Internet video sites, like YouTube, Hulu, and

Indeed, the patient has become so robust that the FCC wants to institute rules to protect its current strapping, posthumous structure.[124] The "architectural design" model failed to predict the market's evolution—as evidenced by the FCC's own view of the matter.

III. THE NETWORK OF NETWORKS IS NOT NEUTRAL

> "The network is not neutral and never has been," Clark said, dismissing as "happy little bunny rabbit dreams" the assumptions of net neutrality supporters that there was once a "Garden of Eden" for the Internet. NSFnet, an early part of the Internet backbone, gave priority to interactive traffic, he said: "You've got to discriminate between good blocking and bad blocking."[125]

The FCC misunderstands the economic nature of the network of networks, presenting a stylized history of the Internet that is highly misleading. For instance, the NN Order asserts that there is no historic practice of "pay to play," wherein delivery networks collect fees from content suppliers for access (or superior access) to the ISP's customers by stating:

> First, pay for priority would represent a significant departure from historical and current practice. Since the beginning of the Internet, Internet access providers have typically not charged particular content or application providers fees to reach the providers' retail service end users or struck pay-for-priority deals, and the record does not contain evidence that U.S. broadband providers currently engage in such arrangements. Second[,] this departure from longstanding norms could cause great harm to innovation and investment in and on the Internet. . . . [P]ay-for-priority arrangements could raise barriers to entry on the Internet by requiring fees from edge providers . . . .[126]

This statement, taken as a predicate for categorical rules outlawing certain pricing practices or deals between ISPs and content providers, is both dubious and dangerous. It implicitly concedes that "Internet access providers have . . . struck pay-for-priority deals," but spins the market description by inserting the modifier, "typically."[127] This protects the FCC from a straightforward falsehood, but opens the path to rules barring what have been, and are today, important business models advancing innovation. Not only have "walled gardens" generated network growth and, therein, network externalities—from AOL's dial-

---

Netflix, strongly supports rejection of the Lessig hypothesis. *See* Tim Lee, *A Look Back at Lessig and Lemley,* TECH. LIBERATION FRONT, Dec. 12, 2007, http://techliberation.com/2007/12/12/a-look-back-at-lessig-and-lemley/.

124. *See* FCC NN Order, *supra* note 1, para. 10, at 17,908.

125. Greg Piper, *Internet Architect Suggests 'Futures Market' to Avoid Policy Disputes*, 29 COMMC'NS DAILY, Feb. 5, 2009, at 9.

126. FCC NN Order, *supra* note 1, para. 76, at 17,947.

127. *Id.*

up ISP to the proprietary, vertically integrated cable ISPs (like @Home, RoadRunner) that forged the path in residential broadband) but "discriminatory" economic arrangements help a multitude of backbone networks, edge apps, and mass market ISPs today.

It is not a departure from "longstanding norms" for app vendors to strike deals for preferential treatment. ISPs have long sold prime real estate on their start-up pages—charging app providers for preferential treatment—in deals putting together Google/AOL,[128] Rogers Cable/Yahoo!,[129] and Disney/Comcast.[130]

"Non-neutrality" also widely exists in the very lively CDN space, where popular app sellers buy faster access to the customer's screen by paying for local caching (supplied globally) by a company like Akamai or Limelight Networks.[131] And it intensifies competitive forces when a broadband ISP is allowed to compete for this business, caching content for those applications that pay extra.[132] These payments are not, as characterized by the FCC, simply extractions that have one-way impacts, increasing barriers to entry.[133] This single-entry bookkeeping overlooks the fact that, in offering to deliver content better and faster, ISPs take money to *deliver content better and faster*. The "dirt road fallacy" the FCC advances is bogus.[134] Marius Schwartz criticizes this thinking

---

128. *See* DAVID A. VISE & MARK MALSEED, THE GOOGLE STORY 112 (2005).

129. Rogers is a major Canadian cable operator. *See* Craig McTaggart, *Was the Internet Ever Neutral?* (paper prepared for 34th Research Conference on Commc'n, Info & Internet Policy, Sept. 30, 2006).

130. *See* Claire Atkinson, *Comcast Snares Mouse House Deal*, N.Y. POST, Jan. 5, 2012, http://www.nypost.com/p/news/business/comcast_snares_mouse_house_deal_NIrN3PEQOXmJ5QNBfAN3JL.

131. Ex Parte Filing by U.S. Dep't of Justice, at 8, *In re* Broadband Indus. Practices, WC Docket No. 07-52 (FCC Sept. 6, 2007) [hereinafter DOJ 2007] (footnotes omitted).

> [C]ommercial content distribution networks, such as Akamai, Limelight Networks, and Internap Network Services, operate thousands of servers throughout the world that cache content and services to provide faster and more reliable access to specific Internet websites. . . . [T]hese arrangements allow participating content and access providers to pay for a higher quality of service . . . . In addition to these caching services, the Department [of Justice Antitrust Division] believes that there can be significant benefits in allowing broadband providers to manage their networks and differentiate among some traffic on the Internet.

Id.

132. *See id.* at 10.

133. *See* FCC NN Order, *supra* note 1, para. 76, at 17,947.

134. Professors J. Gregory Sidak and David J. Teece deconstruct the claim:

> It is not credible that a network operator would intentionally degrade its best-effort delivery of packets in hopes of inducing suppliers of content and applications to buy prioritized delivery of packets. The empirical evidence confirms that broadband ISPs have, in fact, been investing billions of dollars annually to *increase* the speed and *improve* the quality of best-effort Internet service, even while many broadband ISPs also

as the "simplistic notion, associated with crude versions of the so-called 'end-to-end principle,' that the Internet should be a dumb network with rigidly uniform service quality and pricing."[135]

ISPs routinely prioritize traffic so as to improve customer experience. This happens both when CDNs allow app vendors to "pay to play," and in standard network management functions. Service providers with no conceivable anticompetitive motive, including non-profit organizations and firms lacking market power, routinely restrict customers' devices and use of the network in their "acceptable use policies" (AUPs). For example, Virginia Broadband (VABB), a Wireless Internet Service Provider (WISP) serving about 3200 subscribers[136] requires that subscribers refrain from "excessive" use of the network.[137] VABB, in competing for subscribers, has evidently determined that

---

> provide prioritized delivery of video and voice packets over the same physical infrastructure. That outcome is exactly what economics would predict under real-world conditions of platform competition and complementarity between content availability and performance and demand for broadband Internet access services. Even if ISPs were to consider relegating traffic from content and applications providers who did not choose to pay for enhanced QoS to a full-time dirt road—as distinct from the beneficial prioritization of delay-sensitive traffic at times and places of congestion through packet-scheduling algorithms, which is the real issue here—the risk of loss of subscribers would mean, among other things, that ISPs would have no net incentive to do so. Charging different prices for different levels of service promotes inclusion, not exclusion.

J. Gregory Sidak & David J. Teece, *Innovation Spillovers and the "Dirt Road" Fallacy: The Intellectual Bankruptcy of Banning Optional Transactions for Enhanced Delivery over the Internet*, 6 J. COMPETITION L. & ECON. 521, 532 (2010).

135. Schwartz 2010, *supra* note 112, at 2.

136. This is the subscriber level reported in an undated documented submitted to the National Telecommunications & Information Administration, U.S. Department of Commerce, requesting funds from the American Recovery and Reinvestment Act of 2009 ("stimulus" monies) to extend its rural network (which stretches across seventeen counties in Virginia). Virginia Broadband, LLC, *Executive Summary of BIP and BTOP*, NAT'L TELECOMM. & INFO. ADMIN. (2009), http://www.ntia.doc.gov/broadbandgrants/applications/summaries/1220.pdf.

137. The AUP states:

> VABB shall have the right to monitor Customer's "bandwidth consumption" (i.e. aggregate volume of data that may be sent or received) at any time and on an on-going basis, and to limit excessive bandwidth consumption by Customer (as determined by VABB) by any means available to VABB, including suspension or termination of Services. VABB reserves the right to implement specific limits on the maximum amount of bandwidth consumption available to Customer per month—defined as 30 consecutive days, beginning on the first day of service for the level of ISP Service subscribed for by Customer. If Customer exceeds the bandwidth consumption limits assigned to the level of ISP Service for which Customer has subscribed in any month, VABB has the right to limit bandwidth consumption by Customer in excess of such level by any means available to VABB, including to impose an additional fee of

the losses associated with the proscribed options are exceeded by the value of improved opportunities for network users overall. These limits help VABB create a competitive network; indeed, their rules are *productive inputs* into the supply of new broadband options. Lariat Wireless, a small ISP in Laramie, Wyoming, forbids its customers from operating servers, another effort to reduce network congestion and preserve the utility of the system for other users.[138] Entrepreneur Brett Glass, who launched the WISP in the early 1990s as a co-op, took it over as a for-profit venture in 2003—making returns of "less than" $5 per customer per month.[139] Glass testified:

> Our most popular residential service plan comes with a minor restriction; it does not allow the operation of servers.
>
> Now, Mr. Chairman, most Internet users would not know what a server was if it bit them, and they have no problem uploading content to a Web site such as YouTube for distribution. This means customers that do need to operate a server could obtain that capability by paying a bit more to cover the additional cost [of expensive rural bandwidth]. But if [the rules take effect and] the FCC decides against MetroPCS, we will almost certainly be forced to shift everyone to the more expensive plan. We will therefore be less competitive, offer less value to consumers and especially less value to economically disadvantaged ones.[140]

An even starker example of usage restrictions is observed with respect to the (fixed) local area network owned by Ohio University.[141] In a policy that went into effect in April 2007, students and faculty were prohibited from using any peer-to-peer application.[142] The intent is clearly not to suppress competition. According to Chief Information Officer Brice Bible, "[t]he network is a shared resource, and we must ensure that it is available to all users . . . . Peer-to-peer file-sharing consumes a disproportionate amount of resources, both in bandwidth and human technical support."[143] Other universities have banned high-bandwidth

---

> $.05/Megabyte and/or suspension of Services.

Terms and Conditions, VA. BROADBAND, http://vabb.com/terms.php#uacceptableuse (last visited July 1, 2012).

138. *See Ensuring Competition on the Internet: Net Neutrality and Antitrust: Hearing Before the Subcomm. on Intellectual Prop., Competition, and the Internet of the H. Comm. on the Judiciary*, 112th Cong. 55 (2011) [hereinafter *NN Hearing*] (prepared testimony of Laurence Brett ("Brett") Glass, Owner & Founder, Lariat).

139. *Id.*

140. *Id.* at 56. MetroPCS is the fifth largest wireless carrier and the firm inspiring the first NN complaint to the FCC under the new rules. In brief, MetroPCS allows subscribers on an inexpensive plan to access some video content, but not to use their phones for unlimited video streaming. *See generally* Riley, *supra* note 12.

141. *Ohio University Announces Changes in File-sharing Policies*, OHIO UNIV. (Apr. 25, 2007), http://www.ohio.edu/students/filesharing.cfm.

142. *Id.*

143. *Id.*

communications programs like Skype, at least for some period of time. These include U.C. Santa Barbara,[144] San Jose State,[145] and the University of Minnesota.[146] More recently, Oxford University banned the music service Spotify, citing network bandwidth concerns.[147] The point of these examples is not to suggest that the IT administrators are right (or wrong), or that better methods for managing networks (than outright bans on certain devices or applications) will become available. It is to read the very strong evidence that reasonable experts charged with keeping networks running—and in no position to extract monopoly profits from vertical foreclosure strategies—might engage in actions that the FCC identifies as departing from "longstanding norms."[148]

Even if such norms did exist, it would not be clear why regulators should lock the market into them. Indeed, the FCC attempts to make its argument by appealing to the acceptability of such business models to unregulated firms, but then arguing for regulatory enforcement, slipping back into the non sequitur.[149] Not only is the FCC's structural argument wrong, it is clear to some that the structure of the Internet is, and ought to be, in flux. Many network engineers, including Internet pioneer David Clark, co-author of the oft-cited "end-to-end" paper,[150] argue that with broadband networks displacing dial-up Internet, it is appropriate and efficient that "large content networks" (where applications are supplied to the web via high-density access providers) send monetary payments to "large eyeball networks" (where residential customers are served in less dense configurations and, therefore, at generally higher average costs).[151] This is because broadband networks involve substantial infrastructure projects, while in contrast, the dial-up ISPs largely piggybacked on existing systems.[152] The policy conclusion is a normative appeal—"pay to play" is welfare-enhancing—but it is based on a positive observation. The emergence of CDNs is already affecting such transactions, and the integration of "large eyeball networks" into the CDN

---

144. *See* Ryan Paul, *More Universities Banning Skype*, ARS TECHNICA, Sept. 24, 2006, http://arstechnica.com/old/content/2006/09/7814/.

145. *Id.*

146. *See* Elizabeth Redden, *Skype Skirmishes on Campus*, INSIDE HIGHER ED (Sept. 29, 2006), http://www.insidehighered.com/news/2006/09/29/skype (noting that although not completely banned, Skype use is discouraged by university policy).

147. Mike Butcher, *Oxford University Takes a Dislike to Spotify, Bans It*, TECHCRUNCH (Jan. 18, 2010), http://techcrunch.com/2010/01/18/oxford-university-takes-a-dislike-to-spotify-bans-it/.

148. FCC NN Order, *supra* note 1, para. 76, at 17,947.

149. *See id.*

150. J.H. Saltzer et al., *End-to-End Arguments in System Design*, 2 ACM TRANSACTIONS ON COMPUTER SYS. 277 (1984).

151. P. Faratin et al., *Complexity of Internet Interconnections: Technology, Incentives, and Implications for Policy*, at 1, 12 (2007), *available at* http://mitas.csail.mit.edu/papers/clark_Lehr_Faratin_Complexity_interconnection_TPRC_2007.pdf (paper prepared for 35th Annual Telecomm. Policy Research Conference).

152. *See id.* at 3 n.2.

space is a natural development well under way.[153] A new norm has arrived. To pre-empt this evolving market niche on the view that the Internet is an "open end-to-end" network lacking "gatekeepers" would be to subvert that development, undermining Internet growth.[154]

It has been postulated that "innovation at the edge" is far more robust than "innovation at the core" of the Internet.[155] Exciting new applications that ride over the network are therefore seen as generating more economic value than the pipes that carry bits to their destination. "[U]nco-ordinated innovation at the edge of the network . . . has taught us that, at least sometimes, decentralised innovation trumps innovation at the core."[156] Yet comparing one set of innovations to the other[157] is not only problematic because we lack a metric to scale the rival contributions, but it is conceptually flawed. Applications at the edge rely on investments in the core, and vice versa. The sets of services are complements, precisely the argument for net neutrality rules—which seek to reduce barriers to edge innovation by attempting to impose rules that purportedly best maintain this complementarity. The implication of that position is that a flourishing edge is indicative of a flourishing core. To separate developments based on appearances is to arbitrarily unpack a team effort.

Similarly, it is an error to categorically favor one set of investment activities over the other as a matter of law. Restrictions placed on advanced data transport networks will predictably harm edge innovators where the result of such regulation is to materially forestall investments in complementary capital (i.e., broadband build-out). Rules constraining network business models are liable to do just this, as they impose rigidities on a changing and unpredictable market environment. Economists and business strategy experts have focused on the general problem for innovators as one where those creating productive platforms may be left without economic gain, even as other firms extract returns. "It is quite common for innovators . . . to lament the fact that competitors/imitators have profited more from the innovation than the firm" that took the original risks.[158]

This dilemma may undermine the deployment of advanced communications networks. As one recent study laments, "The broadband value chain is headed

153. *Id.* at 21.

154. *See generally* David D. Clark & Marjory S. Blumenthal, *The End-to-End Argument and Application Design: The Role of Trust*, 63 FED. COMM. L.J. 357 (2011).

155. Lawrence Lessig, *Do You Floss?*, LONDON REV. BOOKS, Aug. 18, 2005, at 24, 25.

156. *Id.* at 25.

157. "The Internet has inspired a wide range of innovation. Because of its particular architectural design, that innovation has come primarily from the 'edge' or 'end' of the network through application competition." *Hearing on "Network Neutrality" Before the S. Comm. on Commerce, Science and Transportation*, 109th Cong. 55 (2006) (testimony of Lawrence Lessig, C. Wendell and Edith M. Carlsmith Professor of Law, Stanford Law School).

158. David J. Teece, *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RES. POL'Y 285, 285 (1986).

for a train wreck."[159] The source of this dire forecast is that network builders will not recoup sufficient returns from the value yielded network applications, disrupting feedback loops and leading to market failure. "The 'all you can eat' pricing models that are common today create incentives for providers to limit usage growth rather than invest to support it."[160] The study concludes that "[g]ood solutions to this problem need to align the incentives of network operators and upstream stakeholders, for example by enabling monetization of usage that imposes costs on providers."[161] The policy conclusion may be right or wrong. The more fundamental point is that the analysis properly sees the success of edge and core as inextricably linked, and it properly sees that dynamic adjustments to business models may well improve the mechanisms by which the market supports newer, faster, and better services over time. Rather than protecting one class of economic activity by imposing restrictions on competition from other parts of the value chain, it sees a balancing of interests as key to progress for the system as a whole. The following examples of efficient non-neutrality highlight this economic view.

A. DoCoMo's "Walled Garden"[162]

NTT's DoCoMo, the leading cellular carrier in Japan, first brought web access to customers in February 1999, before cellular systems were engineered for broadband (3G) applications.[163] The carrier, NTT Mobile Communications Network,[164] launched i-mode as "the first packet-based, always-on, mobile Internet service available anywhere in the world."[165] "Official" i-mode vendors are featured on the phone's menu, enabling customers to easily access their content. Billing is handled exclusively through DoCoMo, which lists transactions on subscribers' monthly statements, and charges content providers nine percent of revenues for the service. DoCoMo also allows "unofficial sites" to be accessed by i-mode users, although such vendors suffer a severe competitive disadvantage.[166]

DoCoMo constructed a "walled garden" which, critics charged, limited

---

159. BROADBAND WORKING GRP., MIT COMMC'NS FUTURES PROGRAM & CAMBRIDGE UNIV. COMMC'NS RES. NETWORK, THE BROADBAND INCENTIVE PROBLEM 2 (2005) [hereinafter BROADBAND INCENTIVE PROBLEM], *available at* http://cfp.mit.edu/docs/incentive-wp-sept2005.pdf.

160. *Id.* at 11.

161. *Id.*

162. This sub-section is based on Thomas W. Hazlett, *Modular Confines of Mobile Networks: Are iPhones iPhony?*, 19 SUP. CT. ECON. REV. 67 (2011) [Hazlett, *Modular Confines*].

163. *Id.* at 96.

164. "Originally NTT Mobile Communications Network. Renamed NTT DoCoMo in April 2000." *Id.* at 96 n.106.

165. John Ratliff, Dep't of Sociology Santa Clara Univ., DoCoMo as National Champion: I-Mode, W-CDMA and NTT's Role as Japan's Pilot Organization in Global Telecommunications 12 (Sept. 13, 2000), *available at* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1. 22.9078.

166. *Id.*

customer choice.[167] Yet, i-mode created an innovative hot-house for content. By enabling a platform that limited application prices via vertical restraints, included payments to the ISP, and excluded non-compliant services (specifications set by the carrier), content providers have been given access to a more valuable platform and endowed with more productive opportunities.[168]

> At the heart of all this is a paradox: i-mode depends on outside providers for everything from handsets to content, yet it's managed so carefully that nothing is left to chance. Critics see a walled garden, more mobile mall than wireless Web. But in fact, i-mode's success comes less from being walled than from being obsessively tended.[169]

I-mode has proven exceptionally popular with third party applications developers. Kazutomo Robert Hori, CEO of Cybird, has twenty-three sites connected to i-mode.[170] "For a company like us," Hori said, "the i-mode environment has proven very profitable."[171] The result has been a steady stream of content innovation.[172] DoCoMo's vertical control has favored certain technologies, formats, or business models. The carrier decided, for example, to support Linux and Symbian software for i-mode applications but to exclude Microsoft.[173] Customer acceptance was so pronounced that DoCoMo became Japan's leading ISP, fixed or mobile.[174] By March 2007, it served 52.6 million cell phone subscribers, of which 47.6 million bought i-mode services.[175] This success prompted Japan's other wireless networks, KDDI and Softbank,[176] to each offer competing platforms. DoCoMo responded by extending its proprietary platform into e-commerce.[177] The upshot is that Japan is noted as the leading wireless data services market globally.[178]

---

167. *See* Frank Rose, *Pocket Monster: How DoCoMo's Wireless Internet Service Went from Fad to Phenom—and Turned Japan into the First post-PC Nation*, WIRED, Sept. 2001, http://www.wired.com/wired/archive/9.09/docomo.html.

168. *See id.*

169. Hazlett, *Modular Confines*, *supra* note 162, at 97 (citation omitted).

170. Rose, *supra* note 167.

171. *Id.*

172. *See* Jack Qiu, *NTT DoCoMo: Review of a Case*, JAPAN MEDIA REV. (Oct. 2004), http://www.ojr.org/japan/research/1097446811.php.

173. Wireless Watch, *Microsoft Excluded from DoCoMo's Ecosystem*, REGISTER, Nov. 26, 2004, http://www.theregister.co.uk/2004/11/26/Microsoft_excluded_from_docomo/.

174. *See, e.g.*, John Boyd, *Here Comes the Wallet Phone*, IEEE SPECTRUM, Nov. 2005, at 12; Bruce Einhorn, *DoCoMo's "New Business Model*,*"* BLOOMBERG BUSINESSWEEK, Apr. 18, 2004, http://www.businessweek.com/stories/2004-04-18/docomos-new-business-model.

175. *Annual Operating Data*, NTT DOCOMO, http://www.nttdocomo.co.jp/english/corporate/mir/finance/index.html (last visited July 1, 2012).

176. Softbank acquired the assets of Vodafone Japan in 2006. Vodafone had purchased J-Phone in 2001. Hazlett, *Modular Confines*, *supra* note 162, at 97 n.114.

177. *See* sources cited *supra* note 174.

178. *See* sources cited *supra* note 174.

B. Dedicated Cable Bandwidth for Cable Telephony

For years, U.S. regulators grappled with the challenges presented in the local telephone market. Thought to be a natural monopoly at the time of the AT&T divestiture in 1984, the objective of gaining rivalry between competing services formed the basic motivation for the Telecommunications Act of 1996.[179] While great efforts were expended in network sharing mandates, ultimately overturned by federal courts as inimical to the stated objectives of the Act, success was finally had: local phone competition emerged when cable TV operators provided high-quality wireline voice services nationwide, head to head with local phone companies.[180]

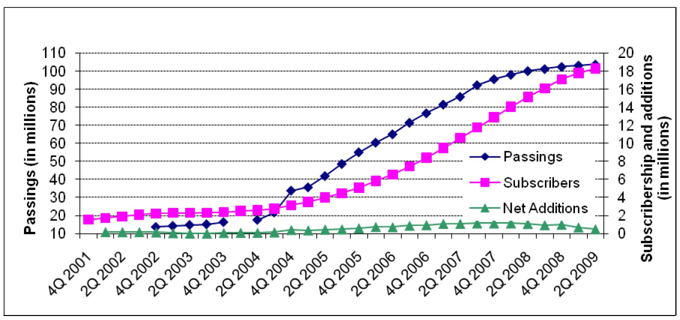

Figure 1. U.S. Cable Telephone Deployment, 2001-09[181]

The technology deployed by cable operators was voice-over-Internet (VoIP), a product generally called "digital voice." Build-out and subscribership began exploding in about 2004.[182] Two points are directly relevant to the discussion here. First, cable operators use dedicated bandwidth within their local area networks (LANs) to provide a premium service for cable VoIP subscribers.[183] Independent VoIP service providers such as Vonage or Skype are free to market their services to cable modem subscribers, but cannot gain access to the LAN

179. Telecommunications Act of 1996, 47 U.S.C. § 609 (2006 & Supp. 2010); *see also* White House: Office of Commc'ns, President Statement on the Telecom Bill Signing, 1996 WL 54454 (Feb. 9, 1996).

180. *See infra* text accompanying notes 181-85.

181. LOCAL TELEPHONE COMPETITION: STATUS AS OF JUNE 30, 2009, FCC Rcd. (2009); LEICHTMAN RESEARCH GRP., INC., RESEARCH NOTES (3Q 2002); LEICHTMAN RESEARCH GRP., INC., RESEARCH NOTES (1Q 2002 through 3Q 2009), *available at* http://www.leichtmanresearch.com/research.html#notes.

182. *See* fig.1.

183. *See* Ryan Leatherbury, *Dedicated Bandwidth over Cable: Simplifying the Migration to VoIP Service*, CONNECTED PLANET (May 21, 2003), http://connectedplanetonline.com/access/infocus/telecom_dedicated_bandwidth_cable/.

'fast lane' reserved for the cable operator's voice customers.[184] This evinces vertical control of a complementary application, with discrimination in favor of network-provided services. Second, the discrimination is clearly procompetitive. The leveraging of the network is what, in part, entices cable operators to create and vigorously market voice services. It yields higher returns, and more aggressive competitive risk taking, in extending competitive telephony. With over 100 million U.S. homes now having a choice between rival fixed line phone operators,[185] a key, longstanding competitive policy objective has been met. It owes much to the vertical integration of cable data and cable voice.

C. Clearwire's Network Discrimination

An emergent wireless broadband network is being built by Clearwire, a public company whose investors have included Sprint, Intel, Motorola, Google, Comcast, and Time Warner Cable.[186] The system, now offering 4G services at the cutting edge of wireless technology, is investing billions of dollars in an effort to challenge the leading fixed and wireless broadband ISPs. One of the interesting structural features of its operations is that it seeks to leverage the competencies of its partners, favoring their vertical services over unaffiliated rivals.[187] For instance, network access devices embed default applications provided by partner Google.[188] Marketing deals extend to partners Sprint, Comcast, and Time Warner Cable.[189] In some respects, such discriminatory operations are unexceptional—even *de rigueur*. Investors often seek, and obtain, preferential terms in exchange for their financial support.[190] With Clearwire an upstart (with just 688,000 subscribers at year-end 2009,[191] as compared to more than 100 million fixed and mobile ISP subscribers) presumably possessing no market power, there is nothing to suggest that these preferential business terms are anything but procompetitive. But this says much more than that whatever NN

---

184. "The telephone service that Comcast and the telephone companies sell uses dedicated bandwidth, while the over-the-top VoIP service that Vonage and Skype offer uses shared bandwidth. I certainly hope that native phone service outperforms ad hoc VoIP; I pay good money to ensure that it does." Richard Bennett, *Damned if You Do, Screwed if You Don't*, RICHARD BENNETT: A PERSONAL BLOG (Jan. 20, 2009), http://bennett.com/blog/2009/01/damned-if-you-do-screwed-if-you-dont/.

185. CONNECTING AMERICA, *supra* note 121, at 24 n.30.

186. *See* Ted Hearn, *Clearwire Quiet on Google Stake*, MULTICHANNEL NEWS, June 19, 2008, http://www.multichannel.com/article/106191-Clearwire_Quiet_On_Google_Stake.php.

187. *See id.*

188. "According to published reports, Google has invested $500 million to secure its place as Clearwire's default Internet search engine—which probably means Clearwire users will automatically rely on Google unless they know how to manipulate their handheld device's software to select another search application." *Id.*

189. *Id.*

190. *See id.*

191. Press Release, Clearwire, Clearwire Reports Fourth Quarter and Full Year 2009 Results (Feb. 24, 2010), *available at* http://corporate.clearwire.com/releasedetail.cfm?ReleaseID=551255.

rules are adopted should make allowance for firms without market power to enjoy full flexibility in their choice of business models. Rather, it demonstrates that firms use such alliances to more effectively innovate in a competitive market. To categorically exclude such conduct would target all such efficiencies.

D. Apple's App Store

The robust rivalry in smart phones, arguably triggered by Apple iPhone but pioneered by Research In Motion's Blackberry, features highly integrated applications platforms that crucially depend on vertical coordination across multiple layers. The capital deployments of networks, the innovations of device makers, the efficiency of operating systems, and the ingenuity of application providers all tie together in a "wireless ecosystem" that consumers enter by subscribing. There are varying degrees of proprietary control exercised in this cross-platform rivalry: RIM and Apple tend towards more proprietary solutions, while Google's Android OS tends to leave more for third parties to engineer.[192] That is not to say that third parties will predictably generate more sales under one model or the other; to date the RIM and Apple forms of integration are proving most successful in attracting customers and, in Apple's case, application developers.[193] The future may reveal new winners and new models, structures chosen out of competitive confrontation in the market. Network neutrality rules seek to truncate that selection process by foreordaining that less vertical coordination is categorically preferred to more. The economics of that assertion are wrong. Often, more integrated business models outperform more neutral (less integrated) rivals, and so produce social gains.

Indeed, "walled gardens" have contributed materially to the evolution of the Internet. One important example is the business model deployed by America Online (AOL) in the mid-1990s.[194] While the World Wide Web was just beginning to feature content appealing to mass-market consumers, AOL sought to dramatically expand subscribership by offering custom features and proprietary websites.[195] Paying brand name media companies, including *TIME* and the *New York Times*, and investing in new services like the Motley Fool, a financial website, it offered its members what they could not find elsewhere.[196] This not only provided competition to rival ISPs, it gave AOL added incentives to market its services to new customers, "carpet-bombing . . . America with free AOL disks," in a campaign that would eventually distribute "more than 250 million disks bearing AOL software to the mass market."[197] Spreading easy to use access was enormously important. "Webheads would sneer that using AOL was 'the Internet on training wheels,'" writes Ken Auletta.[198] "Yet it was AOL's

---

192. Hazlett, *Modular Confines*, *supra* note 162, at 91-93.
193. *See id.* at 89-91.
194. SWISHER, *supra* note 106, at 99-100.
195. *Id.*
196. *Id.*
197. *Id.* at 99.
198. AULETTA, *supra* note 107, at 94.

user-friendliness that helped popularize the Web—and which attracted thirty-four million paid subscribers in 2002."[199] As Michael Katz summarizes,

> There is no evidence that any particular model of an "open" platform with one-sided pricing and limited network management is the only or best way to facilitate innovation, investment, and consumer welfare. Apple's iPhone provides an excellent example of a managed system that has been extremely successful in meeting consumer demands.[200]

E. Preferential Deals Between ISPs and Content Providers

Development of innovative "edge" applications has often been advanced by rivalry among content vendors seeking to secure preferential deals with ISPs. Web browsers such as Mozilla Firefox or Opera, for instance, gain traction—entering a market in which the dominance of Microsoft's Internet Explorer has been documented in U.S. antitrust courts—by entering into exclusive contracts with both ISPs and complementary application providers.[201] Google, now the world's leading search engine, strategically achieved economies of scale via exclusive contracts with ISPs.[202] On May 1, 2002, Google's service was first featured as the default choice on AOL's start-up page—a prime locational advantage sought also by search rivals Inktomi and Overture, but won by Google's commitment to compensate the country's leading ISP "with a very large financial guarantee, running to many millions of dollars."[203]

Today, services such as ESPN3 market themselves not to end-users but to ISPs; customers of nonsubscribing ISPs do not obtain access to their content.[204] This approach may or may not run afoul of net neutrality regulation, depending on rules adopted and interpretations rendered. But the more essential point is that this business model, one that creates "walled garden" content for ISPs, is instigated by the application provider and is a business model selected to advance its interests. Market structures differentiate the content available on competing ISPs are not inherently hostile to the interests of edge innovators. Treating them as if they are does not protect such entrepreneurial activity, but suppresses it.

IV. ANTITRUST'S RULE OF REASON VERSUS NET NEUTRALITY

The core of the net neutrality debate is centered upon the desirability of a

---

199. *Id.* at 94-95.

200. Katz 2010, *supra* note 112, at 46.

201. *See* Marshall Kirkpatrick, *Yahoo! Loses Mobile Giant Opera to Google; Did Google Just Buy a Mobile Browser?*, READ WRITE WEB (Feb. 27, 2008), http://www.readwriteweb.com/archives/yahoo_looses_opera.php.

202. *See* VISE & MALSEED, *supra* note 128, at 112-13.

203. *Id.* at 113.

204. *See* David S. Joachim, *Sports' Greatest Hits at One Web Site (but There's a Catch)*, N.Y. TIMES, July 31, 2006, http://www.nytimes.com/2006/07/31/technology/31cable.html?_r=2&fta=y. Since this article was published, ESPN360 changed its name to ESPN3.

regulatory rule prohibiting network operators from entering into vertical contractual relationships.[205] Proponents of net neutrality have emphasized the possibility that broadband access providers have an incentive to disadvantage rivals and ultimately harm competition.[206] The NN Order articulates these concerns:

> [A] broadband provider may act to benefit edge providers that have paid it to exclude rivals (for example, if one online video site were to contract with a broadband provider to deny a rival video site access to the broadband provider's subscribers). End users would be harmed by the inability to access desired content, and this conduct could lead to reduced innovation and fewer new services. Consistent with these concerns, delivery networks that are vertically integrated with content providers, including some MVPDs, have incentives to favor their own affiliated content.[207]

The FCC cites the standard modern vertical foreclosure references emerging from the "raising rivals' cost" literature, which considers the conditions under which an incumbent firm might successfully disadvantage rivals, reduce competition, and harm consumers.[208]

As is well known in the industrial organization literature, while vertical contracts can occasionally give rise to competitive foreclosure concerns, they can also generate significant efficiencies and enhance consumer welfare.[209] Indeed, vertical contractual arrangements are often efficient and result from the normal competitive process. They are frequently observed between firms lacking any meaningful market power, implying that there must be efficiency justifications for the practice. Economic literature is replete with such procompetitive

---

205. *See* Gary S. Becker et al., *Net Neutrality and Consumer Welfare*, 6 J. COMPETITION L. & ECON. 497, 509 (2010); Jonathan E. Nuechterlein, *Antitrust Oversight of an Antitrust Dispute: An Institutional Perspective on the Net Neutrality Debate*, 7 J. ON TELECOMM. & HIGH TECH. L 19, 21 (2009); Howard A. Shelanksi, *Network Neutrality: Regulating with More Questions than Answers*, 6 J. ON TELECOMM. & HIGH TECH. L. 23, 24 (2007).

206. *See* FCC NN Order, *supra* note 1, para. 21, at 17,915.

207. *Id.* para. 23, at 17,918 (footnotes omitted).

208. *Id.* para. 23, at 17,918-19 (citing Steven C. Salop & Thomas Krattenmaker, *Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price*, 96 YALE L.J. 209, 214 (1986); Steven C. Salop & David T. Scheffman, *Raising Rivals' Costs*, 73 AM. ECON. REV. 267, 267-71 (1983)).

209. *See* Oliver E. Williamson et al., *Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange*, 6 BELL J. ECON. 250, 269-70 (1975); *see also* R.H. Coase, *The Nature of the Firm,* 4 ECONOMICA 386, 398-401 (1937) (indicating that integration between entrepreneurs in a given market lowers the transaction costs associated with divisions of labor, thereby increasing a firm's productive processes that directly flow to the consumer); Benjamin Klein et al., *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, 21 J.L. & ECON. 297, 300 (1978) (discussing how vertical integration allows for avoidance of postcontractual opportunistic behavior).

explanations: reducing double marginalization, preventing free riding on manufacturer-supplied investments,[210] reducing free riding and facilitating investment in promotional effort,[211] to name a few. The benefits of these efficiencies are at least partially passed on to consumers in the form of lower prices, increased output, higher quality, and greater innovation.

Vertical contractual arrangements pose a more complex and nuanced problem for designing efficient legal rules than do other forms of business conduct that are either presumptively anticompetitive (e.g., cartels) or nearly always procompetitive (e.g., an above cost price reduction). A vast theoretical literature documenting both pro- and anti-competitive uses of vertical contractual arrangements evidences this problem. Vertical contracts, without more, have theoretically ambiguous welfare effects—that is, some forms of discrimination are efficient and pro-consumer, while others raise potential competitive concerns. Antitrust jurisprudence has developed a case-by-case rule of reason approach to vertical foreclosure concerns, while net neutrality regulation would ban, as a class, certain vertical relationships. A critical question, and the one to which we now turn, is whether the antitrust approach is a superior alternative to net neutrality regulation which would ban, as a class, certain vertical relationships.[212]

Approaching this question requires a clear objective function: consumer welfare. A rigorous economic approach to designing a legal rule that would maximize consumer welfare in the context of business conduct with ambiguous welfare consequences is desirable. Such an approach requires an analytical framework that takes into account: (1) the probability that the business arrangement is anticompetitive; (2) the magnitude of the social cost of errors in assessing the competitive virtue of the business arrangement, including both false positives (procompetitive conduct is erroneously barred) and false negatives (anticompetitive conduct is falsely absolved); and (3) the administrative costs of implementing the alternative legal rules. The so-called error-cost approach to the design of legal rules, which amounts to the application of economic analysis and empirical evidence in a decision-theoretic framework,[213] has a long history in the

---

210. *See generally* Howard P. Marvel, *Exclusive Dealing*, 25 J.L. & ECON. 1 (1982); *see generally also* Ryko Mfg. Co. v. Eden Servs., 823 F.2d 1215 (8th Cir. 1987).

211. *See, e.g.*, Benjamin Klein & Andres V. Lerner, *The Expanded Economics of Free-Riding: How Exclusive Dealing Prevents Free-Riding and Creates Undivided Loyalty*, 74 ANTITRUST L.J. 473 (2007); Benjamin Klein & Kevin M. Murphy, *Vertical Restraints as Contract Enforcement Mechanisms*, 31 J.L. & ECON. 265 (1988); Benjamin Klein & Joshua D. Wright, *The Economics of Slotting Contracts*, 50 J.L. & ECON. 421 (2007).

212. *See* Howard A. Shelanksi, *Competing Legal Approaches to Network Neutrality Regulation*, 3 COMM. & CONVERGENCE REV. 26, 30 (2011) ("It is this very ambiguity in the welfare effects of price discrimination and in the incentives to discriminate inefficiently that is important. The welfare ambiguity means that any rule patently barring discrimination could have unintended, negative consequences because the conduct sought to be barred—price discrimination—is neither always bad nor always good.").

213. *See* Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1 (1984).

economic analysis of law generally,[214] and antitrust specifically.[215]

Generally, the error-cost approach allows a regulator, court, or policymaker to use new evidence to update a prior belief about the anticompetitive, or procompetitive, nature of a specific business practice, either as the theoretical and empirical understanding of the practice evolves over time or with case-specific information. The optimal decision rule is then based upon the new, updated belief about the likelihood that the practice will be anticompetitive, thus minimizing a loss function measuring the social costs of false positives, false negatives, and administrative costs.[216] The cost-minimizing rule, and thus the legal rule most likely to maximize consumer welfare, depends most critically on the likelihood that particular practices are anticompetitive and the magnitude of the losses attributable to both types of error.

One can begin to approach the design of an optimal legal rule for the set of business practices under the net neutrality umbrella—a variety of vertical contractual relationships—by focusing upon a more narrow inquiry: Under what conditions would a per se prohibition on such business practices maximize consumer welfare? Once those conditions are identified, one can examine whether they are satisfied in the present setting. From a welfare perspective, a per se rule would be appropriate only if vertical contracts were overwhelmingly, but not necessarily always, likely to cause competitive harm and the ability to engage in a more fact-intensive inquiry to absolve instances of procompetitive vertical contracts provided a sufficiently small marginal benefit to consumers.

A leading antitrust casebook describes per se rules as making the "most economic sense" when factors like those below are present:

> [(1)] if permitted, the prohibited conduct will likely harm competition severely;
> [(2)] if the conduct is reviewed for reasonableness rather than held illegal per se, defendants will frequently claim that their conduct is reasonable, it will be costly and time-consuming to evaluate those claims, and in the end, few such claims will prove to be valid; and
> [(3)] little pro-competitive conduct will be deterred by establishing a rule that denies defendants the ability to prove that their conduct was reasonable.[217]

---

214. *See, e.g.*, William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971); Richard A. Posner, *The Behavior of Administrative Agencies*, 1 J. LEGAL STUD. 305 (1972).

215. *See, e.g.*, C. Frederick Beckner III & Steven C. Salop, *Decision Theory and Antitrust Rules*, 67 ANTITRUST L.J. 41 (1999); Keith N. Hylton & Michael Salinger, *Tying Law and Policy: A Decision-Theoretic Approach*, 69 ANTITRUST L.J. 469 (2001); Joshua D. Wright, *Overshot the Mark? A Simple Explanation of the Chicago School's Influence on Antitrust*, 5 COMPETITION POL'Y INT'L 189 (2009).

216. For a more formal exposition of such a loss function in the context of the vertical contracting practices at issue in the present debate, see James C. Cooper et al., *Vertical Antitrust Policy as a Problem of Inference*, 23 INT'L J. INDUS. ORG. 639 (2005).

217. ANDREW I. GAVIL ET AL., ANTITRUST LAW IN PERSPECTIVE: CASES, CONCEPTS AND PROBLEMS IN COMPETITION POLICY 105 (2d ed. 2008); *see also* Jonathan B. Baker, *Per Se Rules*

It is simple to see that the design of a consumer-welfare maximizing legal rule for vertical contracting cannot be resolved by competing theories alone; empirical evidence is a necessary input to application of the error-cost framework. Such evidence allows a court or regulator to form sensible estimates of the key parameters: How often is the potentially prohibited conduct anticompetitive? What is the magnitude of the social losses imposed by false positives or negatives? While economists can reasonably disagree about which empirical studies should receive the most weight for purposes of policy analysis, or forming a precise probability estimate, we shall see that even a summary view of the existing literature reveals that the calls for network neutrality, a per se approach, are either indifferent or immune to Bayesian updating based upon the empirical evidence on vertical contracts.

Multiple academics review the existing theory and evidence on vertical restraints and single-firm conduct more generally, and they uniformly conclude that the practices are generally procompetitive. Furthermore, they conclude antitrust rules should "slant" towards requiring plaintiffs to demonstrate clear anticompetitive effect before courts and juries can find violations.[218] No serious antitrust scholar argues that underlying economic theory and empirical evidence warrant per se treatment for vertical contracts.[219] A comprehensive survey of the vertical contracting literature in economics is beyond the scope of our present task, though it has been done admirably by others.[220] Some highlights from those surveys paint the picture of just how divergent the economic evidence is from the conditions under which a per se rule can be justified on consumer welfare

---

in the Antitrust Analysis of Horizontal Restraints, 36 ANTITRUST BULL. 733, 740 n.29 (1991) ("From a law and economics perspective, per se rules may be preferred to a rule of reason when violations are expensive for a court to observe but are strongly correlated with observable behaviors that are cheaply observed, and when it would be expensive for a violator to break the law without engaging in the observable behavior. Under such circumstances, the judicial system would minimize enforcement costs by conditioning liability on the cheaply observable behavior, and the resulting enforcement errors, corporate compliance costs, and social costs of deterring socially beneficial actions, would not produce substantial efficiency loss.").

218. *See, e.g.*, Francine Lafontaine & Margaret Slade, *Exclusive Contracts and Vertical Restraints: Empirical Evidence and Public Policy*, *in* HANDBOOK OF ANTITRUST ECONOMICS 391 (Paolo Buccirossi ed., 2008); Daniel P. O'Brien, *The Antitrust Treatment of Vertical Restraints: Beyond the Possibility Theorems*, *in* THE PROS AND CONS OF VERTICAL RESTRAINTS 40 (2008), *available at* http://www.kkv.se/upload/Filer/Trycksaker/Rapporter/Pros&Cons/rap_pros_and_cons_vertical_restraints.pdf; Cooper et al., *supra* note 216.

219. Some have argued that the per se rule previously applied to minimum resale price maintenance, prior to *Leegin Creative Leather Products, Inc. v. PSKS, Inc.*, 551 U.S. 877 (2007), is justified on stare decisis grounds. *See* Marina Lao, *Free Riding: An Overstated, and Unconvincing, Explanation for Resale Price Maintenance*, *in* HOW THE CHICAGO SCHOOL OVERSHOT THE MARK: THE EFFECT OF CONSERVATIVE ECONOMIC ANALYSIS ON U.S. ANTITRUST 196, 201-02 (Robert Pitofsky ed., 2008).

220. *See* Lafontaine & Slade, *supra* note 218; O'Brien, *supra* note 218; Cooper et al., *supra* note 216.

grounds.

While measuring the welfare effects of vertical restraints can be especially difficult in the absence of a natural experiment, over the last twenty-five years there has been a concerted effort to add empirical knowledge to our large menu of theoretical models. Two recent empirical surveys summarize the existing empirical literature. The first, authored by a group of Federal Trade Commission and Department of Justice economists, reviews twenty-four papers, published between 1984 and 2005, providing empirical effects of vertical integration and vertical restraints.[221] The second, by Francine Lafontaine and Margaret Slade, reviews twenty-three papers with some overlap with the first survey. While the reader is referred to these surveys for methodological details concerning individual studies, a careful review, provided here, of both surveys offers a synthesis of the evidence.

Cooper et al. observe:

> Empirical analyses of vertical integration and control have failed to find compelling evidence that these practices have harmed competition, and numerous studies find otherwise[, and while] . . . [s]ome studies find evidence consistent with both pro- and anticompetitive effects, . . . virtually no studies can claim to have identified instances where vertical practices were likely to have harmed competition.[222]

Lafontaine and Slade reach a similar conclusion. Summarizing and synthesizing the evidence they reviewed, the authors conclude:

> [I]t appears that when manufacturers choose to impose such restraints, not only do they make themselves better off but they also typically allow consumers to benefit from higher quality products and better service provision . . . . The evidence thus supports the conclusion that in these markets, manufacturer and consumer interests are apt to be aligned . . . .[223]

In a more recent analysis of the vertical restraints literature, Dan O'Brien notes that three additions to the literature provide new evidence that vertical restraints mitigate double marginalization and promote retailer effort.[224] O'Brien goes on to conclude that "[w]ith few exceptions, the literature does not support the view that these practices are used for anticompetitive reasons, . . . [and supports] a fairly strong prior belief that these practices are unlikely to be anti-competitive in most cases."[225]

To be clear, our claim is not that vertical contracts can never generate

---

221. *See* Luke Froeb, Director, Bureau of Econ., Fed. Trade Comm'n, Economics and Antitrust: Enforcement R&D (Sept. 2, 2005), *available at* http://www.ftc.gov/speeches/froeb/earie.pdf.

222. Cooper et al., *supra* note 216, at 658.

223. Lafontaine & Slade, *supra* note 218, at 409.

224. O'Brien, *supra* note 218, at 73-74.

225. *Id.* at 76.

foreclosure and create competitive concerns. To the contrary, we stipulate that reasonable economists can differ in their views about the likelihood of competitive harm on a case-by-case basis; net neutrality supplants that case-by-case approach with a blanket prohibition. Recalling the conditions that render per se rules desirable from a consumer welfare perspective—that is, if vertical contracts were always or almost always anticompetitive in practice—and contrasting those theoretical conditions with the state of empirical evidence indicating that vertical contracts are overwhelmingly procompetitive, it is clear that net neutrality cannot be supported on economic grounds as evidence-based policy.

A close evaluation of the studies discussed in the economic literature will reveal that few deal with network access providers, cable, or wireless. Indeed, the NN Order anticipates the objection that it is promulgating policies that far outstretch the data, citing the Goolsbee study and a number of examples of perceived anticompetitive conduct.[226] As we discuss, the Goolsbee study neither sets forth enough evidence to justify a conclusion that vertical contracts warrant application of a per se rule nor actually demonstrates anticompetitive foreclosure.[227] With respect to the anecdotal evidence of foreclosure, even taking the FCC's descriptions of these events at face value for the moment, as Professor Gerald Faulhaber observes, "[b]y any standard, four complaints about an entire industry in over a decade would seem to be cause for a commendation, not for restrictive regulations."[228]

Not only is there substantial evidence that access regulation has deterred rather than advanced broadband network deployment, but there is a plethora of marketplace experience demonstrating that "non-neutral" business models deployed by ISPs have often proven highly efficient. Vertical integration, in which a firm expands its scope to produce complementary products, and vertical restrictions, where a firm favors one set of complements over another, can enable productive coordination leading to lower costs and better products. That such strategies may also, in certain instances, produce anticonsumer results is established in economic theory,[229] even as the empirical support for successful, anticompetitive outcomes is weak.[230] But no theory or empirical analysis supports the view that such market structures are inherently anticompetitive and should be categorically restricted. This forms the basis of the very strong arguments by Alfred Kahn,[231] Jon Nuechterlein,[232] Scott Hemphill,[233] and

226. FCC NN Order, *supra* note 1, para. 23, at 17,918-19 & nn.58-61.

227. *See infra* Part VI.

228. Gerald R. Faulhaber, *Economics of Net Neutrality: A Review*, 3 COMM. & CONVERGENCE REV. 7, 11 (2011).

229. *See generally* Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age*, 17 HARV. J.L. & TECH. 85 (2003).

230. *See generally* Cooper et al., *supra* note 216.

231. *See, e.g.*, Alfred E. Kahn, *Network Neutrality* (AEI-Brookings Joint Ctr. for Regulatory Studies, Paper No. RP07-05, 2007), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_

others[234] that network discrimination conflicts are best left to antitrust enforcement. Moreover, the history of regulatory attempts to impose vertical structures on communications carriers has not, in most instances, ended happily for consumers.[235]

The economic theory and evidence do not support the conclusion that vertical contracts generally, or those under the net neutrality umbrella specifically, are always anticompetitive. That is not our burden to bear. Quite the contrary, the burden lies with those advocating the rigid per se approach to demonstrate that consumers will benefit from a legal rule prohibiting whole classes of business arrangements, as compared to alternatives. For net neutrality to generate consumer gains relative to the leading alternative, antitrust's rule of reason, it must be the case that there is a substantial basis in economic theory and empirical evidence upon which to conclude that the vertical contracting is nearly always anticompetitive. Describing the relationship between that assertion and economic reality as "at tension" would seriously understate the state of affairs.[236]

With good reason, both the FTC and DOJ have called into question a net neutrality regime and argued that antitrust is up to the task of protecting consumers from vertical contracts that threaten competition. Former Chairman Deborah Majoras observed:

> [L]et me make clear that if broadband providers engage in anticompetitive conduct, we will not hesitate to act using our existing authority. But I have to say, thus far, proponents of net neutrality regulation have not come to us to explain where the market is failing or what anticompetitive conduct we should challenge; we are open to hearing from them.[237]

---

id=973513.

232. Jonathan E. Nuechterlein, *Antitrust Oversight of an Antitrust Dispute: An Institutional Perspective on the Net Neutrality Debate*, 7 J. ON TELECOMM. & HIGH TECH. L. 19 (2009).

233. C. Scott Hemphill, *Network Neutrality and the False Promise of Zero-Price Regulation*, 25 YALE J. ON REG. 135 (2008).

234. *See* J. Gregory Sidak, *A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet*, 2 J. COMPETITION L. & ECON. 349 (2006); Philip J. Weiser, *The Next Frontier for Network Neutrality*, 60 ADMIN. L. REV. 273 (2008); Christopher S. Yoo, *Beyond Network Neutrality*, 19 HARV. J.L. & TECH. 1 (2005).

235. *See* Gerald R. Faulhaber, *Policy-Induced Competition: The Telecommunications Experiments*, 15 INFO. ECON. & POL'Y 73 (2003); Bruce M. Owen, *Antecedents to Net Neutrality*, REG., Fall 2007, at 14, 15.

236. *See* Bruce M. Owen, *Antitrust and Vertical Integration in "New Economy" Industries with Application to Broadband Access*, 38 REV. INDUS. ORG. 363, 381 (2011) ("Empirical evidence that vertical integration or vertical restraints are harmful is weak, compared to evidence that vertical integration is beneficial—again, even in cases where market power appears to be present. Thus, it is reasonable to conclude that prophylactic regulation is not necessary, and may well reduce welfare by deterring efficient investment. Sound policy is to wait for ex post evidence of harm to justify interventions in specific cases.").

237. Deborah Platt Majoras, Chairman, FCC, Luncheon Address at The Progress & Freedom

The DOJ is also on the record criticizing network neutrality. While the FCC dismissed the DOJ critique[238] as limited to "price regulation, which we are not adopting,"[239] the FCC appears to have missed the point and economic substance of their own regulatory regime. To the contrary, it is well understood that network neutrality is indeed a form of price regulation. As Becker, Carlton, and Sider correctly observe, "[n]et neutrality, however, is properly considered a form of price regulation because it limits the form of pricing that can be practiced. Such regulations thus limit a broadband provider's revenue opportunities and its ability to differentiate itself from competitors, and thereby stifle incentives to invest and innovate."[240]

None of the above says that regulators cannot improve the performance of broadband services. Procompetitive reforms can help enormously. First among these initiatives is a push for aggressive spectrum policies that permit advanced wireless broadband networks to develop more rapidly and with far larger capacities. Indeed, competition among fixed and wireless data networks is substantially truncated by the artificial lack of bandwidth, a constraint imposed not by opportunity costs but by rigid regulatory structures that leave vast swaths of valuable airspace severely under-utilized.[241] The FCC's recent emphasis on new liberal license allocations in the National Broadband Plan constitutes a

---

Foundation's Aspen Summit (Aug. 21, 2006); *accord* FED. TRADE COMM'N, BROADBAND CONNECTIVITY COMPETITION POLICY 11 (2007), *available at* http://www.ftc.gov/reports/broadband/v070000report.pdf ("[T]o date we are unaware of any significant market failure or demonstrated consumer harm from conduct by broadband providers.").

238. DOJ 2010, *supra* note 4.

239. FCC NN Order, *supra* note 1, para. 42, at 17,931 n.143.

240. Becker et al., *supra* note 205, at 513.

241. I recently submitted a proposal to the FCC for an overlay license auction which would facilitate the efficient, expeditious reallocation of TV band frequencies to potentially higher-valued uses (including mobile broadband). Thomas W. Hazlett, *Unleashing the DTV Band: A Proposal for An Overlay Auction*, *In re* A National Broadband Plan for Our Future, No. GN Docket No. 09-51 (FCC, Dec. 18, 2009). Many previous studies have noted the social cost of the spectrum allocation rigidities that artificially constrain wireless services. *See* Gerald R. Faulhaber, *The Future of Wireless Telecommunications: Spectrum as a Critical Resource*, 18 INFO. ECON. & POL'Y 256 (2006); Thomas W. Hazlett, *Optimal Abolition of FCC Spectrum Allocation*, 22 J. ECON. PERSP. 103 (2008); Thomas W. Hazlett, *The Wireless Craze, the Unlimited Bandwidth Myth, the Spectrum Auction Faux Pas, and the Punchline to Ronald Coase's "Big Joke": An Essay on Airwave Allocation Policy*, 14 HARV. J.L. & TECH. 335 (2001); Thomas W. Hazlett, *Tragedy T.V.: Rights Fragmentation and the Junk Band Problem*, 53 ARIZ. L. REV. 83 (2011); Thomas W. Hazlett & Roberto E. Muñoz, *A Welfare Analysis of Spectrum Allocation Policies*, 40 RAND J. ON ECON. 424 (2009); Thomas W. Hazlett et al., *What Really Matters in Spectrum Allocation Design*, 10 NW. J. TECH. & INTELL. PROP. 93 (2012); Reed E. Hundt & Gregory L. Rosston, *Spectrum Flexibility Will Promote Competition and the Public Interest*, IEEE COMM. MAG., Dec. 1995, at 40; Gregory L. Rosston & Jeffrey S. Steinberg, *Using Market-Based Spectrum Policy to Promote the Public Interest*, 50 FED. COMM. L.J. 87 (1997).

major shift in regulatory focus and is a welcome development.[242]

One final argument concerning the relative merits of antitrust should be addressed. The FCC and net neutrality proponents often argue that antitrust analysis might not prohibit all use of vertical contracts is a bug, rather than a feature, of that regime.[243] However, the fact that antitrust is not a "slam dunk" can also be a feature. The economic discipline of antitrust requires the FCC to establish a real theory, garner actual evidence, and convince judges who do not depend on the regulated industry for future employment. The rule of reason, as applied to vertical contractual arrangements, represents a century-old attempt to develop a legal rule aimed at reliably distinguishing procompetitive from anticompetitive arrangements.[244] Indeed, recent antitrust enforcement efforts suggest that the FTC and DOJ have no problem bringing vertical theories.[245] Net Neutrality proponents argue that the rule of reason is too restrictive.[246] They contend it may only reach instances of foreclosure or discrimination in which harm to consumers can be demonstrated, thereby absolving discrimination and other undesirable conduct that is competitively beneficial for consumers.[247] This description of the rule of reason is correct, but these features of the rule of reason are consumer protections that stem from an incremental evolution now over a century old and are based upon increasing economic knowledge and evidence. These features are precisely why net neutrality has garnered so much support from scholars and commentators.[248]

While the affirmative case for antitrust over network neutrality on consumer welfare grounds is clear, the fact that antitrust might not "work" does not default to the position that the FCC's solution will work. If, after 121 years of trying, the antitrust regime has trouble, then it is difficult to imagine that the FCC—routinely seen as "one of the more dysfunctional [agencies] in Washington" will do better.[249] Consistent with this observation, the FCC has

---

242. *See* CONNECTING AMERICA, *supra* note 121, at 73-106; *see also* Thomas W. Hazlett, Editorial, *Putting Economics Over Ideology*, BARRON'S, July 12, 2010, at 35, *available at* http://mason.gmu.edu/~thazlett/opeds/PuttingEconomicsAboveIdeology.pdf.

243. *See* FCC NN Order, *supra* note 1, para. 42, at 17,930-31 & n.141; Jon Leibowitz, Commissioner, Fed. Trade Comm'n, Concurring Statement Regarding the Staff Report: "Broadband Connectivity Competition Policy" (June 2007), *available at* www.ftc.gov/speeches/leibowitz/V070000statement.pdf ("[T]here is little agreement over whether antitrust, with its requirements for *ex post* case by case analysis, is capable of fully and in a timely fashion *resolving* many of the concerns that have animated the net neutrality debate.").

244. *See generally* Baker, *supra* note 217.

245. *See* United States v. Dentsply Int'l, Inc., 399 F.3d 181 (3d Cir. 2005); United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001); *In re* Intel Corp., No. 9341, 2010 WL 3180281 (F.T.C., Aug. 4, 2010); *In re* Transitions Optical, Inc., No. C-4288, 2010 WL 1804580 (F.T.C., Apr. 22, 2010).

246. *See supra* notes 229-32.

247. *See supra* notes 229-32.

248. *See supra* notes 228-31.

249. JAMES B. MURRAY, JR., WIRELESS NATION: THE FRENZIED LAUNCH OF THE CELLULAR

already flopped with its initial forays; the *MetroPCS* and *Comcast* matters, to the extent they are reflective of the net neutrality regime, evince no understanding of the underlying economics in the NN Order.[250]

V. THE PLURAL OF ANECDOTE IS NOT DATA: FCC CLAIMS CONCERNING HARMFUL DISCRIMINATION ARE SPECULATIVE, INCOMPLETE, OR CONTRADICTORY

In the absence of systematic empirical evidence, the FCC turns to a number of anecdotal accounts to support its claim that vertical relationships generate "dangers to Internet openness [that] are not speculative or merely theoretical."[251] Such claims are properly evaluated in the context of the FCC's case for net neutrality generally. The FCC and net neutrality proponents offer a simple case in support of net neutrality: (1) the Internet has been a "virtuous circle of innovation," (2) the circle depends upon the openness of the Internet, and thus, (3) we need to impose new rules to protect that structure.[252]

The call for new rules is a non sequitur. The Internet most certainly has spurred remarkable innovation, diverse business models, and economic growth. However, much of this innovation has occurred without regulatory requirements that constrain ISPs or others from adopting new business models and arrangements that respond to changes in technology and market conditions.[253] The FCC concedes that vertical contractual arrangements and price discrimination increase consumer welfare.[254] Thus, the case for net neutrality is not only based upon the non sequitur described above, but it also anticipates that the new regulatory regime will prohibit at least some procompetitive business arrangements and thus make some consumers worse off.

As we have stressed, from an economic perspective, the critical question is whether the tax imposed upon consumers by restricting vertical contracts facilitating competitive price discrimination can be justified on the grounds that net neutrality will create offsetting consumer welfare gains. Here, as in Part IV, the key question is whether the FCC can satisfy the burden of persuasion with an appeal to economic theory and evidence. We have demonstrated that neither basic industrial organization economic theory regarding vertical contracting nor existing empirical evidence support the FCC's position.[255]

REVOLUTION IN AMERICA 21 (2002).

250. *See* Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010); *see generally* FCC NN Order, *supra* note 1.

251. FCC NN Order, *supra* note 1, para. 35, at 17,925.

252. *Id.* para. 14, at 17,910.

253. *See* Becker et al., *supra* note 205, at 499 ("To date, and in the absence of regulatory requirements to do so, access providers have maintained business models and network management practices that, as a general rule, do not prioritize traffic or impose congestion-based charges.").

254. *See* Preserving the Open Internet, Broadband Industry Practices, 74 Fed. Reg. 62,638, 62,651 (Nov. 30, 2009) (to be codified at 47 C.F.R. pt. 8).

255. *See supra* Part IV. The Commission's contention that its concerns are "not speculative or merely theoretical" depends critically upon its misinterpretation of Professor Goolsbee's analysis

The FCC attempts to elude its evidentiary burden to prove that net neutrality's benefits exceed its costs to consumers, asserting that there is no "persuasive reason to believe that in the absence of open Internet rules broadband providers would lower charges to broadband end users, or otherwise change their practices in ways that benefit innovation, investment, competition, or end users."[256] Instead of citing to convincing empirical evidence in support of this proposition, the FCC cites to its own order.[257] Without convincing empirical evidence, the FCC's case for net neutrality collapses under the weight of a cost-benefit analysis,[258] and amounts to the naked assertion that if we do not impose new rules, there is no evidence that the broadband networks will be better.[259] For example, the FCC claims that "[w]idespread interference with the Internet's openness would likely slow or even break the virtuous cycle of innovation that the Internet enables, and would likely cause harms that may be irreversible or very costly to undo."[260] This is a radical departure from evidence-based policy.

The most serious attempt to proffer empirical support of the frequency, or social costs, of business arrangements that the net neutrality regime would prevent comes in the form of a list of examples: (1) the 2005 Madison River case, resolved with a $15,000 fine;[261] (2) the 2008 *Comcast* decision, involving Comcast's alleged interference with BitTorrent traffic, which was resolved voluntarily and without FCC authority;[262] (3) a mobile operator that "allegedly blocked" access to a payment company that was not the exclusive contractor for the network, where no anticompetitive element in the agreement was noted;[263] and (4) a claim that MetroPCS "restricted the types of lawful applications that could be accessed over its 3G mobile wireless network."[264]

The NN Order goes further to catalog a handful of "additional allegations of blocking, slowing, or degrading P2P traffic."[265] The FCC then proceeds to

---

of foreclose in cable markets. FCC NN Order, *supra* note 1, para. 36, at 17,926; *see also infra* Part VI.B.

256. FCC NN Order, *supra* note 1, para. 40, at 17,929.

257. *Id.* at 17,929 n.131 (citing *id.* para. 29, at 17,922); *id.* at 17,929 n.132 (citing *id.* para. 28, at 17,921-22).

258. The FCC concedes that the benefits of network neutrality rules should exceed their costs. *Id.* paras. 38-39, at 17,927-28.

259. *See id.*

260. *Id.* para. 38, at 17,927.

261. *In re* Madison River Commc'ns, LLC, 20 FCC Rcd. 4295 (Mar. 3, 2005), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-05-543A2.pdf.

262. *See* Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010); *see also supra* note 112 and accompanying text.

263. FCC NN Order, *supra* note 1, para. 35, at 17,925. As we have discussed, the competitive benefits of exclusive dealing arrangements have long been recognized in the economics literature. For a summary, see Alden F. Abbott & Joshua D. Wright, *Antitrust Analysis of Tying Arrangements and Exclusive Dealing*, *in* 4 ANTITRUST L. & ECON. 183, 191-201 (Keith N. Hylton ed., 2010).

264. FCC NN Order, *supra* note 1, para. 35, at 17,925.

265. *Id.* para. 36, at 17,926.

document these additional allegations with yet another list, beginning with the observation that "in May 2008 a major cable broadband provider acknowledged that it had managed the traffic of P2P services."[266] Of course, universities and small ISPs routinely do the same thing,[267] and it is hardly sufficient to demonstrate the presence of market power or a serious danger of competitive harm.

Next, the FCC notes that "[i]n July 2009, another cable broadband provider entered into a class action settlement agreement stating that it had 'ceased P2P Network Management Practices,' but allowing the provider to resume throttling P2P traffic."[268] This was RCN, an 'overbuilder' with so little market power it has already declared bankruptcy![269] While RCN "denied any wrongdoing" in settling these allegations, and the FCC acknowledges that its targeting of P2P applications was aimed at easing network congestion,[270] the more pertinent point for evaluating the desirability of network neutrality rules is that RCN does not have market power, and such discrimination is much more likely to be efficient and proconsumer than result in anticompetitive foreclosure.

The FCC's next anecdotal example of conduct the network neutrality rules will address is the claim that "other broadband providers have engaged in similar degradation."[271] As with the RCN example, the FCC's economics do not distinguish what it describes as "degradation" that makes all parties worse off from that which enhances others' services. In summing up allegations of blocking, slowing, or degrading P2P traffic, and assigning equal weight to all such allegations regardless of their competitive impact, the FCC eschews a consumer-welfare oriented approach. The FCC observes that broadband providers frequently reserve "sweeping rights to block, degrade, or favor traffic," including one provider whose terms of service reserve such rights "without limitation."[272] But evaluate those terms compared to any small Internet service provider without market power (including, for example, Virginia Broadband[273]) and one will see similar network management rules. A central tenet of industrial organization economic analysis is that if one commonly observes contractual terms and business arrangements adopted by both firms with and without market

---

266. *Id.*; *see also* Amy Schatz, *Cox About to Feel Wrath of Net Neutrality Activists*, WALL ST. J. (May 15, 2008, 5:44 PM), http://blogs.wsj.com/washwire/2008/05/15/cox-about-to-feel-wrath-of-net-neutrality-activists/.

267. Ashwin Navin, *Perspective: The P2P Mistake at Ohio University*, CNET (May 7, 2007), http://news.cnet.com/2010-1027_3-6181676.html.

268. FCC NN Order, *supra* note 1, para. 36, at 17,926 (footnotes omitted).

269. *See* Dana Cimilluca & Brett Cole, *RCN Considers Sale After Emerging from Bankruptcy (Update 5)*, BLOOMBERG (Sept. 13, 2006), http://www.bloomberg.com/apps/news?pid=newsarchive&sid=akzWsX8zY5dI.

270. FCC NN Order, *supra* note 1, para. 36, at 17,926 n.110.

271. *Id.* para. 36, at 17,926.

272. *Id.* para. 36, at 17,926 & n.113.

273. *See Terms and Conditions*, *supra* note 137.

power, the practice is likely efficient.[274]

Finally, the FCC stacks two more anecdotal exemplars of blocking allegations, observing that "a major mobile broadband provider prohibits use of its wireless service for 'downloading movies using peer-to-peer file sharing services' and VoIP applications."[275] Once again, the FCC's example involves MetroPCS; once again, the FCC's anecdotal account fails to recognize that without market power, MetroPCS's wireless network management principles do not threaten consumers. Quite the contrary, these rules enhance others' services and welfare, but those benefits are omitted from the FCC's anecdotal scorecard. The FCC ends its string of anecdotes by describing complaints concerning Comcast's allegedly "overly restrictive device approval procedures."[276] Comcast has now completed a 105 MBPS Docsis 3.0 deployment across its entire market area, about forty percent of the United States, and key to advancing American consumers' access to broadband.[277]

Three consistent themes emerge from evaluation of these anecdotes. One is that they bear little to no resemblance to the concerns about economic foreclosure described throughout the NN Order. For example, market power is a necessary condition for such foreclosure; without it, industrial organization economists recognize that the same business practices condemned in the NN Order are likely to be welfare enhancing. When the FCC shifts from economic theory to its attempt to muster empirical support for its new rules, it abandons any attempt to apply the theories with rigor, or to ensure that their conditions are satisfied. Appeals to anecdote are relied upon in support of a theoretical model which, on its own terms, would reject the relevance of the example.

The second theme is that, despite the dearth of empirical data supporting its concerns and the nearly uniform recognition that vertical contracting practices are more likely to help than harm consumers, the NN Order defiantly but perversely rejects the notion that it involves heavy-handed regulation. As discussed above, the FCC converts strenuous objection from the Department of Justice into support for network neutrality on the basis of a serious misunderstanding of basic economics.[278] The FCC claims that, while the DOJ "specifically endorsed requiring greater transparency by broadband providers," it "recognized that in concentrated markets, like the broadband market, it is appropriate for policymakers to limit 'business practices that thwart innovation.'"[279] Most boldly, the FCC claims that "although the [DOJ] cautioned that care must be taken to avoid stifling infrastructure investment, it expressed

274. *See supra* note 209.

275. FCC NN Order, *supra* note 1, para. 36, at 17,926.

276. *Id.*

277. *See* Vlad Savov, *Comcast Extreme 105 Serves Up 105Mbps Internet Speeds for Home Users with Deep Pockets*, ENGADGET (Apr. 14, 2011), http://www.engadget.com/2011/04/14/comcast-extreme-105-serves-up-105mbps-internet-speeds-for-home-u/.

278. *See* FCC NN Order, *supra* note 1, para. 42, at 17,931 & n.143.

279. *Id.* at 17,931 n.143.

particular concern about price regulation, which we are not adopting."[280] The FCC appears utterly unaware that net neutrality rules regulate prices by prohibiting certain business models.[281] Even where the FCC rules do not outright prohibit certain business models, the regulations perversely aim at Internet service providers that block content so as to appeal to their subscribers' preferences! For example, the FCC would eliminate obvious procompetitive "blocking," including:

> an Internet access service that provides access to a substantial subset of Internet endpoints based on end users preference to avoid certain content, applications, or services; Internet access services that allow some uses of the Internet (such as access to the World Wide Web) but not others (such as e-mail); or a "Best of the Web" Internet access service that provides access to 100 top websites could not be used to evade the open Internet rules applicable to "broadband Internet access service."[282]

Third, the FCC's discussion of these anecdotes reveals an important flaw in the reasoning of net neutrality proponents concerning the link between incentives for network owners to discriminate, incentives to invest, and consumer welfare. Proponents of network neutrality often conflate discrimination with welfare, arguing, as Professor Barbara van Schewick does, that "[i]f network owners do not have an incentive to discriminate [anticompetitively] against independent applications anyway, the imposition of a network neutrality regime . . . will not reduce their profits."[283] It is possible to simultaneously hold the views that network owners have no incentive to discriminate against independent applications and that regulation reduces their incentives to invest. FCC Chief Economist Marius Schwartz explains the flaw in the contrary position. Schwartz correctly explains that van Schewick and others' argument "assumes counterfactually that all discrimination and, in fact, all charging, is necessarily harmful. It also ignores inevitable regulatory errors and that the resulting uncertainty will discourage investment."[284]

As we have explained, the claim is not that ISPs have no incentives to integrate or manage networks; they quite commonly do so fruitfully. But they also do so efficiently and in all ranges of manners and methods that produce benefits for consumers and appeal to their preferences. The option value to do so in the future is quite obvious. What is clear, however, after an evaluation of the theoretical and empirical support that can be mustered in favor of the FCC's network neutrality rules, is that the anticompetitive foreclosure arguments—integrating into content, creating monopolies, and then extracting

280. *Id.*

281. Becker et al., *supra* note 205, at 513.

282. FCC NN Order, *supra* note 1, para. 47, at 17,933 (footnote omitted).

283. Barbara van Schewick, *Towards an Economic Framework for Network Neutrality Regulation*, 5 J. ON TELECOMM. & HIGH TECH. L. 329, 332 n.6 (2007).

284. Schwartz 2010, *supra* note 112, at 30 n.52.

rents by leveraging ISP market power—are so far from the real world evidence that not a single serious example (one that is faithful to the anticompetitive foreclosure theories relied upon by the FCC in justifying network neutrality) can be attempted, let alone established. The premier applications—Google Search, Twitter, Facebook, Netflix, Hulu.com—are coming onto the platform. At best, these applications purchase (or create their own) CDNs, and the ISPs are starting to play. Critically, however, no monopoly has been sighted (even if only on a far-off horizon), and the idea that integration brings new rivalry is clearly a procompetitive outcome. Quashing it would reduce both consumer welfare, broadband infrastructure investment incentives, and useful services for content developers—a policy failure trifecta.

VI. THE FCC'S FAILED SEARCH FOR ECONOMIC EVIDENCE

A. Alleged Instances of Discrimination Are Uncompelling Evidence of Anticompetitive Conduct

The FCC has been crafting a net neutrality policy for the better part of a decade. The NN Order states that the rules "we adopt today follow directly from the [FCC's] bipartisan *Internet Policy Statement*, adopted unanimously in 2005,"[285] guidelines that were announced in a famous speech by former FCC Chairman Michael Powell in early 2004.[286] The FCC opened a Notice of Inquiry in 2007,[287] flowing into a 2009 Notice of Proposed Rulemaking,[288] and then to the NN Order issued in December 2010.[289]

From the beginning, regulators focused on examples of ISP conduct that restricted network usage in ways that appeared both discriminatory and unrelated to the provision of high-quality service. Powell's initial foray was driven by theories of anticompetitive foreclosure[290] and examples of allegedly "non-neutral" conduct by broadband ISPs.[291] Nothing in this rendition of marketplace

285. FCC NN Order, *supra* note 1, para. 5, at 17,907.

286. *See* Michael K. Powell, Chairman, FCC, Remarks at the Silicon Flatirons Symposium, Preserving Internet Freedom: Guiding Principles for the Industry (Feb. 8, 2004), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-243556A1.pdf.

287. *In re* Broadband Indus. Practices, 22 FCC Rcd. 7894 (2007).

288. Preserving the Open Internet, Broadband Industry Practices, 74 Fed. Reg. 62,638 (Nov. 30, 2009) (to be codified at 47 C.F.R. pt. 8).

289. *See* FCC NN Order, *supra* note 1.

290. Powell, *supra* note 286, at 4. Powell cited research that listed possible sources of anticompetitive vertical conduct: Farrell & Weiser, *supra* note 229.

291. *See* Powell, *supra* note 286, at 4. Powell noted:

> A few troubling restrictions have appeared in broadband service plan agreements. Professor Tim Wu of the University of Virginia School of Law catalogued some of these . . . things such as cable companies' early efforts to impose restrictions on use of virtual private networks, WiFi and home networking equipment and on operation of servers in the home.

Id.; *see also* Tim Wu, *Network Neutrality, Broadband Discrimination*, 2 J. ON TELECOMM. & HIGH evolution connected the ISP actions to reduced consumer welfare or suggested that NN rules would improve options for users. Vertical restrictions can generally be connected to network efficiencies in management or pricing, potentially improving overall user experience.[292] Meanwhile, reduced regulation—avoiding new NN mandates—tends to improve investment incentives for ISPs, leading to improved infrastructure build-out.[293] Hence, the examples put forth by the FCC require some overall evaluation, balancing countervailing effects, to establish a case for market failure, let alone one for regulatory effectiveness. Even while citing instances of such conduct (or misconduct), Powell surmised that mere anecdotes failed to form a compelling argument for regulatory intervention: "[T]he case for government imposed regulations regarding the use or provision of broadband content, applications and devices is unconvincing and speculative."[294]

Economists reviewing the FCC record consistently lodge the same objection. In response to the FCC's 2007 Notice of Inquiry, for instance, the Antitrust Division of the U.S. Department of Justice declared the entire record devoid of the indicators of anticonsumer conduct: "Commenters failed to submit evidence in response to the [FCC's] request for evidence of harmful discrimination or other behavior suggesting the existence of a systematic or widespread problem."[295] Following the 2009 Notice of Proposed Rulemaking, the noted industrial organization economist and former FCC Chief Economist, Michael Katz, noted that the FCC presented "a stylized and inaccurate perception of the current state of the Internet."[296] Acting to implement net neutrality regulations on this view "would be antithetical to evidence-based policy making and would lead to policy conclusions that were not grounded in reality."[297] In a paper responding to the proposed NN rules, lawyer J. Gregory Sidak and economist David J. Teece argue that "speculative fears cannot justify the overbroad prohibition embodied in the [FCC's] proposed nondiscrimination rule."[298]

TECH. L. 141 (2003).

292. *See* Hundt & Rosston, *supra* note 241, at 41.

293. This is not only implied by economic theory, and the consensus view of financial analysts, it is a relationship assumed by FCC broadband regulators. When rejecting calls for "open access" regulation on cable modem service suppliers in 1999, FCC Chairman William Kennard said that to impose such mandates on ISPs would suppress investment flows and truncate the emerging marketplace: "[T]he fact is that we don't have a duopoly in broadband. We don't even have a monopoly in broadband. We have a NO-opoly. . . . We have to get these pipes built. But how do we do it? We let the marketplace do it." William E. Kennard, Chairman, FCC, Remarks Before the National Cable Television Association, The Road Not Taken: Building a Broadband Future for America (June 15, 1999), *available at* http://transition.fcc.gov/Speeches/Kennard/ spwek921.html. For economic research highly consistent with this view, see *infra* Part VI.

294. Powell, *supra* note 286, at 4.

295. DOJ 2007, *supra* note 131, at 9.

296. Katz 2010, *supra* note 112, at 2.

297. *Id.* at 7.

298. Sidak & Teece, *supra* note 134, at 537.

Marius Schwartz, a Georgetown professor of economics (appointed the new FCC Chief Economist in June 2011[299]) offered an even blunter assessment: "no analysis of the effectiveness of competition in the broadband marketplace is presented."[300]

B. The FCC's Citation to Economic Authority

Such criticism was heard by the FCC. In the December 2010 NN Order, the Commission laid out its slim list of alleged horribles—the Madison River and Comcast episodes again the featured examples[301]—but then confronted the lack of economic evidence head-on in Footnote 60.[302] Acknowledging that potential sources of non-neutral conduct do not show anticonsumer consequence, the NN Order references a single study—written by University of Chicago economist, Austan Goolsbee[303]—and states:

> In addition to the examples of actual misconduct that we provide . . . the Goolsbee Study provides empirical evidence that cable providers have acted in the past on anticompetitive incentives to foreclose rivals, supporting our concern that these and other broadband providers would act on analogous incentives in the future. We thus disagree that we rely on "speculative harms alone" or have failed to adduce "empirical evidence."[304]

This claim, the FCC's sole citation to supporting economic scholarship,[305] is

---

299. Press Release, FCC, FCC Names Marius Schwartz Chief Economist; Jonathan Baker and Gregory Rosston to Serve as Senior Economists for Transactions (June 6, 2011), http://www.fcc.gov/document/fcc-names-marius-schwartz-chief-economist-jonathan-baker-and-gregory-rosston-serve-senior-e.

300. Schwartz 2010, *supra* note 112, at 3. Professor Schwartz went on to show how the agency's approach to economic analysis is entirely ad hoc:

> The Notice states that "imposing a fee on content, application, and service providers could reduce total welfare more than imposing the same fee on the end users and no fee on the content, application, and service providers." But one cannot presume a systematic tendency in this direction. It is true that higher fees to content providers—unaccompanied by incremental performance or other benefits—would tend to discourage their participation; but the same is true on the consumer side if higher prices are charged to them. By themselves, these observations clearly are not sufficient to guide policy.

Id. at 21 (citation omitted).

301. FCC NN Order, *supra* note 1, para. 37, at 17,926-27.

302. *Id.* at 17,918 n.60.

303. Austan Goolsbee, Vertical Integration and the Market for Broadcast and Cable Television Programming (2007), *available at* http://fjallfoss.fcc.gov/edocs_public/attachmatch/DA-07-3470A10.pdf (revised by *Research Studies on Media Ownership*, FCC (Sept. 5, 2007), http://transition.fcc.gov/ownership/studies.html).

304. FCC NN Order, *supra* note 1, at 17,918 n.60.

305. To be clear, Footnote 60 also cites the following passage: DAVID WATERMAN & ANDREW

remarkable on a number of fronts.

First, there is the simple matter that the one study they find to inform their economic analysis regarding anticompetitive conduct by broadband ISPs does not concern broadband services.[306] Hence, the FCC reveals that it was unable to locate a single study that, by examining marketplace data, concluded that credible evidence of market failure existed.

Second, there is the equally revealing fact that the Goolsbee study, which does not evaluate regulatory rules, provides no evidence whatever on the efficacy of network neutrality, or related, policies.[307] The FCC has, hence, failed to adduce any support for its assertion that the reforms instituted would positively impact consumers.

Third, Professor Goolsbee does not claim the economic conclusions the FCC claims:

> This kind of calculation, [he writes in reference to his empirical investigation,] is obviously meant only to be suggestive. But applied with better data to more narrowly defined markets, this type of approach might be able to provide an empirical basis for the threshold-type exemptions often used by the FCC and other regulatory agencies where certain markets or firms are exempted from regulation when they have been deemed to be "competitive."[308]

The FCC skips the suggestion for further research, preferring to take the economics as decided—no matter the explicit disclaimer "obviously" offered by the author.

Fourth, the FCC elects to selectively cite only the Goolsbee results, ignoring other economic research that evaluates the effects of cable TV regulation, ignoring research published in peer reviewed journals. Before noting that research, it helps to explain the key issue being investigated in papers researching the effects of vertical integration in cable TV markets.

Bias in selecting cable TV programming—say, Time Warner is more likely to include CNN-fn, rather than Bloomberg, on channel line-ups of its cable TV systems, given that it owns CNN-fn and not Bloomberg—could be explained by efficiency as readily as by any anticompetitive factor. Indeed, Time Warner may have purchased (or launched) CNN-fn because it believed its subscribers would

---

A. WEISS, VERTICAL INTEGRATION IN CABLE TELEVISION 142–43 (1997). FCC NN Order, *supra* note 1, at 17,918 n.60. Yet, these economists (Professor Waterman is at Indiana University, Dr. Weiss at the FCC) do not claim to have found anticompetitive foreclosure in cable TV markets. *See* David Waterman & Sujin Choi, *Network Neutrality and Vertical Control: Lessons from Cable T.V.* (unpublished working paper, 2010) (on file with author) ("We are reminded both by the long history of the cable industry and the short history of the broadband Internet industry, that the fundamental policy concerns from an economic perspective are not vertical integration, but the presence of horizontal market power . . . .").

306. *See* FCC NN Order, *supra* note 1, at 17,918 n.60; Goolsbee, *supra* note 303, at 2.

307. *See* Goolsbee, *supra* note 303, at 2.

308. *Id.* at 30.

find its programming valuable. Vertical integration allows the operator the opportunity to supply its own inputs, cutting out the middleman—and the costs thereof.[309] (Prices are often reduced via integrated structure, as such enterprises avoid "double marginalization."[310]) Hence, the economic literature on vertical integration seeks to discern the difference between favoritism, which drives virtually the whole of the FCC analysis, and market inefficiency, which the FCC categorically ignores. The Goolsbee study purports to have found anticompetitive conduct.

Yet, the FCC ignores the broad swath of research that reaches sharply different conclusions. For instance, in Tasneem Chipty's 2001 paper, evidence of cable operator favoritism (for owned channels) was found, with the result that subscribers gained access to a greater number of cable networks and quality-adjusted prices fell.[311] Consumer welfare increased with integration, in the presence of "foreclosure."[312] This was largely consistent with a 1997 study by George Ford and John D. Jackson that found that vertical integration reduced costs.[313] This is entirely consistent with the consensus findings in other markets. As Stanford economist and long-time cable television expert Bruce Owen summarizes:

> While there is no shortage of theoretical models in which vertical integration may be harmful, most such models have restrictive assumptions and ambiguous welfare predictions—even when market power is assumed to be present. Empirical evidence that vertical integration or vertical restraints are harmful is weak, compared to evidence that vertical integration is beneficial—again, even in cases where market power appears to be present.[314]

A recent survey of empirical research on vertical integration in cable and other sectors, conducted by four economists at the U.S. Federal Trade Commission, stated that it is "difficult to find evidence that vertical controls reduce welfare."[315] The authors concluded:

> Because the welfare effects of vertical practices are theoretically ambiguous, optimal decisions depend heavily on prior beliefs, which should be guided by empirical evidence. Empirically, vertical restraints appear to reduce price and/or increase output. Thus, absent a good

---

309. *See generally* Thomas W. Hazlett, Vertical Integration in Cable Television: The FCC Evidence (Oct. 19, 2007) [hereinafter Hazlett, *Vertical Integration*] (unpublished paper), *available at* http://www.arlingtoneconomics.com/studies/vertical-integration-in-cable-television.pdf.

310. *Id.* at 5.

311. Tasneem Chipty, *Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry*, 91 AM. ECON. REV. 428, 450 (2001).

312. *Id.*

313. George S. Ford & John D. Jackson, *Horizontal Concentration and Vertical Integration in the Cable Television Industry*, 12 REV. INDUS. ORG. 501 (1997).

314. Owen, *supra* note 236, at 381.

315. Froeb, *supra* note 221, at 23.

> natural experiment to evaluate a particular restraint's effect, an optimal policy places a heavy burden on plaintiffs to show that a restraint is anticompetitive.[316]

This survey, consistent with other findings, represents a consensus representing the state of economic understanding. In an important 2007 paper in the Journal of Economic Literature, economists Francine Lafontaine and Margaret Slade summarize economic research thusly:

> [O]verall a fairly clear empirical picture emerges. The data appear to be telling us that efficiency considerations overwhelm anticompetitive motives in most contexts. Furthermore, even when we limit attention to natural monopolies or tight oligopolies, the evidence of anticompetitive harm is not strong.
>
> . . .
>
> As to what the data reveal in relation to public policy, . . . [w]e are . . . somewhat surprised at what the weight of the evidence is telling us. It says that, under most circumstances, profit-maximizing vertical-integration decisions are efficient, not just from the firms' but also from the consumers' points of view. Although there are isolated studies that contradict this claim, the vast majority support it. Moreover, even in industries that are highly concentrated so that horizontal considerations assume substantial importance, the net effect of vertical integration appears to be positive in many instances. We therefore conclude that, faced with a vertical arrangement, the burden of evidence should be placed on competition authorities to demonstrate that that arrangement is harmful before the practice is attacked.[317]

Fifth, and perhaps most fundamentally, the FCC's citation to the Goolsbee study is *logically incorrect*. When the evidence in the research paper is properly interpreted, it does not imply market failure. Indeed, it reveals—with the marketplace data it examines using the empirical model it develops—that cable TV operators are *not*, on average, engaging in anticompetitive conduct. To reference, as its sole economic evaluation, the Goolsbee paper as supporting not only a conclusion of endemic market failure, but evidence suggesting that regulatory intervention on the order of NN rules is proconsumer, is so egregious an error as to call for a rather detailed explanation of what the Goolsbee analysis involves and what conclusions it reaches.

---

316. Cooper et al., *supra* note 216, at 639.

317. Francine Lafontaine & Margaret Slade, *Vertical Integration and Firm Boundaries: The Evidence*, 45 J. ECON. LIT. 629, 677, 680 (2007).

C. Evaluating the Goolsbee Results[318]

1. Empirical Approach.—Goolsbee studies the question of anticompetitive conduct without directly addressing the question of whether consumers are better off due to vertical integration. Indeed, the analysis assumes that most of what influences the flow of video programming to consumers is exogenous to the key question it investigates: whether, given the existing system and a number of cable TV networks as carriage choices, cable operators tend to carry their own program networks as opposed to the networks owned by others.[319] This is part of the economic analysis of vertical integration in cable TV markets, but it is dominated by other concerns. The more important questions for consumers involve how markets create new video programming and expand platforms for delivering this programming to customers.

An illustration with the weakness of this approach is found by considering the implications of Professor Goolsbee's admission that vertical integration is decreasing over time. "Tables 8A and 8B show that of the top 15 networks as measured by the size of their prime time audience, the share of vertically integrated networks has been falling over time, from eight in 1997 to four in 2005."[320] Simultaneously, a huge increase occurs in the quantity of programming being carried by cable TV operators: "The number of networks increased by 359 in the ten years from 1996 to 2005."[321] Moreover, the overwhelming proportion of these new networks is not owned by cable TV operators: "Independent networks made up 311 of that 359, vertically integrated networks only 48. The share of networks identified by the FCC as being vertically integrated has basically been cut in half over this period—from almost 40% in 1996 to just over 20% in 2005."[322]

These facts illuminate the question of how consumers are likely impacted by market structure. While they are noted in the Goolsbee analysis, however, they are then cast aside by the FCC, finding no part of the economic analysis and being ignored in the policy conclusions reached. This arbitrarily sets aside some of the most essential data bearing on the question of anticompetitive foreclosure. The dramatic expansion in cable TV programming, and audience share versus broadcast TV, implies that markets—with or without vertical integration—are improving content quality and viewer choice.[323] That most new program channels are produced by firms not owning cable TV systems is evidence that operators are not foreclosing entry into vertical markets. That the most successful cable TV networks—such as ESPN, by far the most valuable[324]—are

---

318. This section is taken from Hazlett, *Vertical Integration*, *supra* note 309, at 17-32.

319. *See* Goolsbee, *supra* note 303, at 4.

320. *Id.* at 21.

321. *Id.*

322. *Id.*

323. *See* Hazlett, *Vertical Integration*, *supra* note 309, at 10.

324. *See According to Cable Operators . . . Fox News Channel Most Valuable Network on Lineup*, VIAMEDIA BLOG (Feb. 24, 2011), http://viamediablog.com/according-to-cable-operators-fox-news-channel-most-valuable-network-on-lineup/ ("ESPN, as might be expected, was valued

owned by broadcasters, firms which compete head-to-head with cable TV operators, strongly supports this conclusion.

Goolsbee notes, "[t]he data suggest that vertical integration has been getting less prevalent over time."[325] Indeed, not only have independent entrants become increasingly successful, but cable TV programmers with ownership interests in cable or satellite distribution have been divesting, splitting their firms into specialized, standalone parts. This raises distinct implications. Were operators to extract monopoly profits by favoring their own, less valuable programming, they would sacrifice these gains by failing to become—or stay—highly integrated. Such key market evidence is simply ignored, despite its direct bearing on the defined research topic: "[I]t is . . . worth trying to understand why vertically integrated systems tend to be more likely to carry their own channels than independent cable systems and whether this can be attributed to market power."[326]

Goolsbee conducts two sets of econometric tests using data from cable TV markets. The first inquiry seeks to determine whether multiple system cable TV operator (MSO)-owned cable TV program networks out-perform cable TV networks owned by other firms.[327] This is undertaken to discover whether there are efficiencies (as per evidence of out-performance) associated with vertical integration.[328] The second investigation focuses on cable operator carriage decisions, searching for information about the degree of favoritism exhibited for networks owned by the operator.[329] The paper claims to find no evidence of efficiency, and possible evidence of anticompetitive foreclosure, given that self-carriage bias both appears and then appears to lessen in markets with more intense retail competition.[330] Both empirical pursuits are seriously flawed, however. Properly interpreted, they bring forth *no evidence* supporting the conclusion that anticompetitive foreclosure is deterring consumer welfare.

2. The Efficiency Inquiry.—Here, cable TV program networks are examined to see if those owned by cable TV systems exhibit higher subscriber growth, revenues, and program expenditures than independent networks. Finding no systematic statistical relationship, the paper concludes that there is no evidence that economic efficiency is causing or resulting from vertical integration.[331]

The interpretation is suspect in two respects. First, the lack of observed results from vertical integration could be interpreted, just as easily, as indicating that there is no evidence of an anticompetitive outcome. Were integrated cable

---

highest at $1.16 per sub/month.").

325. Goolsbee, *supra* note 303, at 2. The disparity in the text may stem from the fact that the first comment related to both broadcasting and cable, although the passage quoted did not invoke this distinction.

326. *Id.* at 21.

327. *See id.* at 3.

328. *See id.* at 4.

329. *See id.* at 3.

330. *See id.* at 2.

331. *See id.* at 29-31.

operators to favor their own programming networks, such networks could well exhibit higher growth rates. For instance, Professors Dong Chen and David Waterman's 2005 article is cited in Goolsbee for showing that cable operators may favor their own programming on basic tiers, relegating rival cable program networks to digital tiers.[332] Such discrimination would presumably result in nonintegrated program networks exhibiting relatively poor growth in subscribers, license fees, and advertising revenues. That the lack of affiliation produces no statistically significant correlation suggests that this is inaccurate. More generally, the evidence reveals neither efficiency nor anticompetitive discrimination.[333] Hence, Chen and Waterman, who find evidence of operator favoritism in cable channel selections, conclude, "[i]t was not possible to conclude from this study whether the foreclosure patterns we observe are efficiency or anti-competitively motivated, or whether consumers are on net better off or worse off as a result."[334]

Second, the cable network indices Goolsbee examines are at least two levels removed from the actual efficiencies that we seek to understand. The first level is corporate: Vertical integration, when adopted by firms, is designed to advance the economic returns of the combined entity, not just the program network.[335] Hence, some cable TV networks could be vertically integrated with cable operators to achieve important efficiencies that result in increased revenues or subscribers not for the program network but for the MSO. Indeed, the MSO might invest in certain networks that feature low earnings, at least for some period, if they expand system subscribership—possibly by expanding content menu diversity—or promise to prove more popular in future periods at which time some of the gains will be realized by the MSO as well as the network. This dynamic would explain the Goolsbee findings, but imply the existence of economies from integration.

Consider the first basic cable TV network, C-SPAN, founded in 1979 by a consortium of MSOs.[336] While funded by cable TV operators, it is non-profit, selling no ads and realizing only modest license fees from operators that are used to cover costs, not repatriate owners.[337] In the empirical framework in Goolsbee, the zero returns constitute evidence of a lack of efficiency.[338] In fact, the integration created new video content in order to expand the universe of cable subscriptions, increasing revenues flowing to operators. These benefits are ignored in the analysis. That C-SPAN has survived for over thirty years suggests

332. *Id.* at 19.

333. *See* Dong Chen & David Waterman, Vertical Foreclosure in the U.S. Cable Television Market: An Empirical Study of Program Network Carriage and Positioning 34-35 (Oct. 2005) (unpublished manuscript), *available at* http://ssrn.com/abstract=843544.

334. *Id.* at 35.

335. *See* Goolsbee, *supra* note 303, at 3.

336. WATERMAN & WEISS, *supra* note 305, at 25.

337. *Our Company*, C-SPAN, http://www.c-span.org/About/About-C-SPAN/ (last visited July 2, 2012).

338. *See* Goolsbee, *supra* 303, at 29-31.

that it is valuable to subscribers, even as it exhibits zero returns as a standalone enterprise.

The second level of efficiency involves the question: Does vertical integration enhance consumer welfare? This would tend not to be the case were vertical foreclosure the intent and effect of cable operator-created programming. If that were the outcome, MSOs would launch program networks, give preferential carriage to these owned properties, exclude rival services, and in so doing deny the rivals the scale economies necessary to effectively compete for carriage in other systems. Yet the history of cable TV programming reveals that independent firms, and particularly those in the broadcasting business—a key competitor to cable—have owned the great majority of successful ventures.[339] From ESPN to Fox News and from Discovery to A&E, the most widely distributed cable TV program networks have little or no cable TV equity interests. And MSO ownership has been declining over time as Goolsbee notes above.[340] This market outcome is inconsistent with the foreclosure theory, which predicts that integrated operators will squeeze out independent, non-cable operator program networks over time.[341]

In 2005, just twenty-seven percent of the twenty most profitable cable TV program networks were owned by cable TV operators, weighted by equity shares of ownership and cash flows of the cable program networks.[342] This represents a substantial decline in the level of vertical integration exhibited in earlier years. In 1992, for example, cable operators owned forty-one percent of the top twenty program networks using a similar weighting system.[343]

---

339. *Id.* at 30-31.

340. *See id.* at 20.

341. *See* WATERMAN & WEISS, *supra* note 305, at 129-32.

342. *See infra* tbl.1. To clarify the weightings, suppose there are just two cable TV networks, one of which is fifty percent owned by a group of MSOs, the other having no MSO ownership. If the MSO-affiliated network accounts for sixty percent of total cash flows between the two cable program networks, the weighted-average MSO, ownership share = (0.5)*(0.6) = 0.3, or thirty percent. It should be noted that these top twenty networks accounted for 110% of all cable TV network cash flows (meaning that the networks smaller than the top twenty generated negative cash flows in aggregate).

343. *See* PAUL KAGAN ASSOCS., INC., THE ECONOMICS OF BASIC CABLE NETWORKS 37-38 (1993).

TABLE 1. OWNERSHIP SHARES OF TOP TWENTY CABLE NETWORKS BY CASH FLOW (2005)[344]

Network	*Owner*	*Share*	*Cash Flow ($ Mil)*	*MSO Share*
Nickelodeon	Viacom	100%	900.3	0%
ESPN	Disney	80%	858.9	0%
	Hearst	20%		
MTV	Viacom	100%	692.3	0%
TNT	**TW**	100%	642.6	100%
Disney	Disney	100%	441.1	0%
USA	GE	100%	416.9	0%
Discovery	**Cox**	25%	377.8	50%
	Hendricks	2%		
	Liberty	49%		
	Newhouse	25%		
Fox Sports	Fox	50%	373.6	50%
	Cablevision	50%		
CNBC	GE	100%	360.1	0%
Fox News	News Corp.	100%	350.6	0%
TLC	**Cox**	25%	338.3	50%
	Hendricks	2%		
	Liberty	49%		
	Newhouse	25%		
LIFE	Disney	50%	332.9	0%
	Hearst	50%		
CNN+HN	**TW**	100%	325.2	100%
TBS	**TW**	100%	290.0	100%
BET	Viacom	100%	285.2	0%
HGTV	Scripps	100%	193.9	0%
AMC	**Cablevision**	100%	184.8	100%
TOON	**TW**	100%	184.3	100%
VH1	Viacom	100%	184.3	0%
A&E	Disney	38%	182.8	0%
	GE	25%		
	Hearst	38%		
CF-adjusted % of Top 20 Total				27%

Note: Cable MSOs in **boldface**.

Since the Goolsbee paper was written, the trend towards *dis*integration has continued. While Comcast acquired about one-half of NBC from General

344. KAGAN RESEARCH, LLC, THE ECONOMICS OF BASIC CABLE NETWORKS 56-57, 98-426 (11th ed., 2005) [hereinafter KAGAN 2005] (using data found in each network's profile).

Electric in 2011,[345] the 2009 Time Warner spin-off of Time Warner Cable involved far more cable TV programming.[346] Time Warner's content includes CNN, HLN, HBO, TNT, TBS, TCM, truTV, and the Cartoon Network.[347] NBC, while owning a broadcast TV network, had relatively smaller cable TV programming assets, which included USA, CNBC, MSNBC, E!, and Bravo.[348] NBC also owns only about half of its joint venture with NBC Universal.[349] In later 2003, NewsCorp acquired thirty-four percent (a controlling interest) of the largest U.S. satellite TV system (and second largest video subscription service), DirecTV.[350] In 2008, however, it sold its interest, which had grown to 38.5% to Liberty Media—which does not own cable or satellite TV distribution assets.[351] By divesting, NewsCorp, owner of Fox Television, FX, Fox News Channel, Fox Business Channel, and several sports networks,[352] eliminated its integration into program distribution in the United States.

While the Goolsbee "efficiency analysis" is uncompelling due to model mis-specification, these observed trends in market structure should not be ignored. They offer important evidence counter to the hypothesis that cable operators are offering lower-quality, higher-priced programming networks that they can force upon their subscribers via foreclosure of rivals' content.

3. The Carriage Favoritism Inquiry.—The primary empirical investigation conducted in Goolsbee involves a statistical analysis that attempts to predict whether a particular cable TV network (twelve different cable TV networks are chosen for the exercise) will gain carriage on a particular cable TV system, given various characteristics of the market, the cable system, and whether or not the

345. Yinka Adegoke & Dan Levine, *Comcast Completes NBC Universal Merger*, REUTERS, Jan. 29, 2011, http://www.reuters.com/article/2011/01/29/us-comcast-nbc-idUSTRE70S2WZ20110129.

346. *See Time Warner Cable Spinoff to Finish Next Month*, N.Y. TIMES, Feb. 27, 2009, http://dealbook.nytimes.com/2009/02/27/time-warner-cable-spin-off-to-finish-next-month/.

347. *Highlights*, TIME WARNER, http://www.timewarner.com/our-content/turner-broadcasting-system/ (last visited July 2, 2012); *It's HBO*, TIME WARNER, http://www.timewarner.com/our-content/home-box-office/ (last visited July 2, 2012).

348. *NBC Universal History*, NBC UNIVERSAL, http://www.nbcuni.com/corporate/about-us/history/ (last visited July 2, 2012).

349. *Id.*

350. *Company Profile*, DirecTV, http://www.directv.com/DTVAPP/content/about_us/company-profile (last visited July 2, 2012).

351. *See* Tim Mullaney, *Liberty Completes Swap of News Corp., DirecTV Stakes (Update 3)*, BLOOMBERG (Feb. 27, 2008), http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a4JFvml25M_8; *see also Company Overview*, LIBERTY MEDIA, http://www.libertymedia.com/company-overview.aspx (last visited July 2, 2012); Jennifer LeClaire, *News Corp. Agrees to DirecTV Swap with Liberty Media*, E-COMMERCE TIMES, Dec. 22, 2006, http://www.ecommercetimes.com/story/54870.html.

352. *See Cable Network Programming*, NEWS CORP., http://www.newscorp.com/operations/cable.html (last visited July 2, 2012); *Television*, NEWS CORP., http://www.newscorp.com/operations/television.html (last visited July 2, 2012).

cable operator (or its parent company) owns the cable TV network in question.[353] Goolsbee finds that cable operators *do* tend to favor the program networks they own.[354] This evidence of favoritism for a company's own programming assets may be, as discussed, explained either by the efficiencies of vertical integration or by the strategic interests of the firm in anticompetitively foreclosing alternatives. To differentiate between these sources of bias, the Goolsbee equations include another explanatory variable: DBS penetration.[355] If higher satellite TV subscribership in the local market is associated with a reduction in favoritism (i.e., a decrease in the probability the cable operator will carry its own programming, all else equal), then the economic implication drawn by Professor Goolsbee is that enhanced competition, as proxied by the satellite penetration rate,[356] is constraining MSO carriage choices. The "evidence suggests, perhaps, an explanation rooted in competitive pressures rather than efficiencies."[357]

a. Data.—Before turning to the estimated results, the data used for analysis deserve comment. Goolsbee first considers five of the most popular cable TV networks owned, in whole or in part, by MSOs: AMC, CNN, TBS, TNT, and Discovery.[358] He notes that these program channels cannot be used in the statistical analysis due to lack of variation in cable system carriage because they have essentially ubiquitous coverage on all systems.[359] "Clearly there is little scope for strategic behavior when every system has enough capacity to carry all the major channels."[360]

That would appear to constitute evidence, however, of the fact that cable operators have created the capacity to host and carry a multitude of popular

---

353. *See generally* Goolsbee, *supra* note 303.

354. *Id.* at 29.

355. The variable is an interactive term, with a dummy for MSO ownership of the cable network whose carriage is being evaluated times the DBS penetration in the designated market area (DMA) (also known as a local television market, of which there are 210 nationally). *See id.* at 28.

356. Penetration rate = subscribers/total homes in the DMA in which the cable TV system is located. What Goolsbee calls DBS penetration is actually alternative delivery system (ADS) penetration, which includes DBS, "large dish" satellite TV, multipoint multi-channel distribution systems, and satellite master antennae systems. *See generally What Is ADS?, Alternate Delivery Systems*, TVB LOCAL MEDIA MARKETING SOLUTIONS, http://www.tvb.org/media/file/TVB_PB_Political_Bulletin_What_is_ADS.pdf (last visited July 2, 2012).

357. Goolsbee, *supra* note 303, at 29.

358. *Id.* at 27. The statement is followed by a parenthetical aside, "although the work of Chen and Waterman, 2006 does show that there may still be interesting decisions regarding what networks get carried on the digital versus the analog tier." *Id.* The reference does not plausibly explain the situation with respect to these cable TV channels because the problem with moving from the analog to the digital tier is reduced coverage (analog tiers reaching all subscribers while digital tiers reaching fewer). The first five cable networks listed in *id.* at 38 tbl.11 (the object of this discussion) achieve virtually universal coverage, overcoming whatever discrimination Goolsbee or Chen and Waterman purport to find.

359. *See infra* tbl.2 (taken from Goolsbee, *supra* note 303, at 38 tbl.11).

360. Goolsbee, *supra* note 303, at 27.

networks, not just those they own.[361] Moreover, the stated reason for excluding the evidence is that there is little self-carriage favoritism worth searching for when we already know that nonowners carry these networks just as owners do.[362] But this transmits valuable information about the *lack of foreclosure*. Excluding these data inherently biases the analysis to follow.

TABLE 2. CARRIAGE RATES FOR INTEGRATED NETWORKS BY SYSTEM TYPE[363]

	System Owns Network	*System Does Not Own Network*
Type I		
AMC	98.7%	98.4%
CNN	99.9%	99.8%
Discovery	100%	99.8%
TBS	100%	97.2%
TNT	100%	99.5%
Type II		
Boomerang	43.4%	13.2%
BBC America	89.3%	38.2%
CNN International	17.7%	5.1%
FitTV	24.5%	45.9%
FUSE	57.6%	60.8%
G4	83.7%	93.9%
PBS Kids	2.1%	8.8%
Science Channel	4.1%	15.6%
Style	5.4%	6.1%
Travel Channel	97.4%	79.7%
TV One	7.2%	9.0%
WE	97.2%	71.2%

Similarly, the study omits from the statistical analysis, and then excludes from its economic conclusions, the information yielded by the large number of widely distributed cable TV program networks featuring no MSO ownership.[364] Using industry data from 2005, there were ten networks that were at least as profitable in terms of annual cash flow as AMC (the least profitable MSO-affiliated network excluded due to its ubiquitous carriage).[365] These would appear to be extremely useful data, alone, they suggest that strategic behavior by MSOs to exclude rivals' programming is, again, not in evidence.

The MSO and non-MSO networks explicitly rejected for inclusion in the

361. *See id.* at 38 tbl.11.

362. *See id.* at 19-20.

363. *Id.* at 38 tbl.11.

364. These are: Nickelodeon, ESPN, MTV, Disney, USA, Lifetime, CNBC, Fox News, BET, and HGTV.

365. *See supra* tbl.1.

econometric investigation of anticompetitive foreclosure constitute the overwhelming share of cable program network revenues and profits. In 2005, the five MSO-affiliated program networks accounted for seventeen percent of total network cash flows, while the ten non-MSO program networks accounted for another forty-five percent.[366] If strategic moves by MSOs could block entry by rivals, the incentives would presumably be strongest here. For these reasons, tossing this evidence aside seriously biases the test conducted towards foreclosure and away from efficiency.

Having dropped the most popular and economically important networks from consideration, the paper then examines twelve cable TV program networks that are "wholly or partially vertically integrated basic cable TV networks . . . [with] carriage rates between 5% and 90%."[367] This results in the analysis of the twelve networks are identified in Table 2 as Type II. In contrast to the economic importance of Type I networks and comparable networks not owned by MSOs—which together account for about sixty-two percent of industry profits—the Type II networks accounted for just 1.3 percent of cash flows in 2005.[368]

Noteworthy among this group is that seven of the twelve networks exhibit higher carriage rates among cable TV systems that do not own them than among those which do.[369] This is remarkable in that the transactional advantages of ownership would seem to be pronounced among networks that are new and growing, with carriage obtained early on via cable TV systems owned by the parent company, to the extent that such carriage would occur at all. However, this information is, again, overlooked in favor of the estimation of a marginal favoritism metric.

b. Empirical model.—That analysis focuses on how the probability of carriage changes when (a) the cable TV network is owned in part or wholly by the cable system's parent company, as indicated by the estimated coefficient on *Vertical Integration* ["*VI*"]; and (b) satellite TV penetration changes when the cable TV network is owned in part or in whole by the cable system's parent, indicated by the coefficient on the interactive term *Vertical Integration * DBS Penetration* ["*VI*DP*"].[370] A probit regression—a prediction of the probability the cable channel is carried on a particular cable TV system, given various factors adjusted for by the independent variables[371]—is run for each of the twelve

366. *See supra* tbl.1 (using data from KAGAN 2005, *supra* note 344).

367. Goolsbee, *supra* note 303, at 27.

368. The ratio was calculated using the cash flow of the twelve networks studied by Goolsbee over the total program channel revenues.

369. *See supra* tbl.2; *see also* Goolsbee, *supra* note 303, at 38 tbl.11.

370. *See* Goolsbee, *supra* note 303, at 39.

371. The independent variables are: a dummy variable equal to one when the cable network is owned by the cable system; an interactive variable equal to the ownership dummy (equal to one when the cable operator owns the program network) times satellite TV penetration in the local TV market (DMA); the satellite TV penetration (DMA); fiber's share of system plant miles; a dummy equal to one if the system is analog only; a dummy equal to one if the system is two-way;

Type II cable networks selected.[372]

Multiple problems emerge with this model. First, *DBS penetration* does not measure the existence of competition; rather, it measures the subscribership of satellite television in the local television market in which each observed cable system exists. The data do not reflect DBS subscribership in the specific area covered by a given cable TV system, but across a far larger area in which the cable TV system happens to exist. But beyond this data mismatch, the far more fundamental problem is that the DBS penetration rate does not measure the *level of competition* because the satellite video offering—the substitute product which is (correctly) seen to potentially constrain cable TV system pricing and carriage decisions in the Goolsbee model—does not change from market to market. What varies, and what the Goolsbee equations likely measure, is the build-out of cable TV systems in the DMA. As a 2005 GAO report found, DBS penetration averages fifteen percent among households where subscribers also have access to cable TV (i.e., they live in homes already passed by cable), but achieves sixty-five percent penetration where there is no cable available.[373] But the variation in DMA cable saturation (homes passed/total homes) does not represent variation in competition *in the areas served by cable*.

Second, while neither *DBS penetration* nor *VI*DP* proxy "competition," other factors that presumably impact carriage decisions by cable TV operators are excluded. Most obviously, channel capacity is of key importance theoretically; cable systems allocate scarce channel slots to different programming choices, and, as Professor Goolsbee notes, expanded channel capacity accommodates more programming from all ownership sources.[374] Yet Goolsbee argues that including channel capacity loses too many observations (due to missing data) and that results are, in any event, unchanged.[375] When running the same model but including cable system channel capacity, however, results change substantially.[376] It may also be useful to include an explanatory variable for "cable homes passed" per local DMA, in that this could arguably help distinguish the effect of DBS penetration from the effect of cable build-out.

c. Results.—Two key estimated relationships are obtained in Goolsbee,

---

population density; population growth rate; percent of residents of Hispanic origin in local area; percent of residents under eighteen years of age; percent of residents over sixty-five years of age; percent of residents who are black; population per household; natural log of income; percent of local residents who are homeowners. *See id.*

372. The model appears in each of the reported results tables in *id.* at 39-49 tbls.12A-K.

373. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-05-257, DIRECT BROADCAST SATELLITE SUBSCRIBERSHIP HAS GROWN RAPIDLY, BUT VARIES ACROSS DIFFERENT TYPES OF MARKETS 9 (2005) (finding approximately nine percent of U.S. households were to have no access to cable).

374. *See* Goolsbee, *supra* note 303, at 18.

375. *Id.* at 28 ("Adding channel capacity did not change the results but is missing from a large number of the system level observations and thus dramatically reduced the sample.").

376. Channel capacity is defined for analog tiers. *See generally* 76 TELEVISION AND CABLE FACTBOOK (Daniel Y. Warren ed., 2008).

summarized here in Table 3.[377] The first is that, in eight of twelve estimated equations, the *Vertical Integration* dummy coefficient is positive and statistically significant at the ninety-five percent confidence level.[378] The second is that the estimated coefficient on the interactive term, *Vertical Integration * DBS penetration*, is negative and statistically significant in the same eight equations.[379] Professor Goolsbee takes this evidence to suggest that integrated cable TV firms do, as seen in other studies, favor their own programming over that owned by other firms.[380] Moreover, this favoritism diminishes in markets where the cable operator faces more competition from DBS, as measured by the estimated coefficient on *Vertical Integration * DBS penetration*.[381] This, concludes Professor Goolsbee, excludes efficiency as an explanation of self-carriage favoritism, leaving anticompetitive foreclosure.[382] In short, Professor Goolsbee finds that cable operators are more likely to carry their own program networks, and they are most likely to do it where they can "get away with" it due to a lack of competition.[383]

TABLE 3. KEY COEFFICIENT ESTIMATES IN GOOLSBEE[384]

	Vertical Integration	*VI * DBS*	*"Neutral" DBS Penetration*	*MSO Owner*
BBC America	2.733**	-0.066*	41.3%	Newhouse
Boomerang	1.407**	-0.039*	36.0%	Time Warner
CNN Int'l	2.279**	-0.066**	34.6%	Time Warner
Fit TV	1.414**	-0.074*	19.1%	Newhouse
G4	0.151	-0.121+	7.2%	Comcast
PBS Kids	2.043**	-0.156**	13.2%	Comcast
Science	3.533**	0.194**	18.2%	Comcast
Style	-0.035	-0.009	0%	Newhouse
Travel	0.752+	0.002	N.A.	Newhouse
TV One	1.583**	-0.061**	26.0%	Comcast
WE	7.317**	-0.412**	17.8%	Cablevision
FUSE	not reported due to "strange probit result"			Cablevision

** significant @ 1%; * significant @ 5%; + significant @ 10%.

377. The results for the FUSE regression are not reported, in that "the probit showed a significant positive coefficient on vertical integration and a significant negative on the interaction with DBS but something in the data lead the standard errors to be absurdly small and the coefficients absurdly large." Goolsbee, *supra* note 303, at 29 n.5; *see also infra* tbl.3.

378. *See infra* tbl.3.

379. *See* Goolsbee, *supra* note 303, at 29 n.5; *see also infra* tbl.3.

380. *See* Goolsbee, *supra* note 303, at 17.

381. *See id.* at 26.

382. *See id.* at 30-31.

383. *See id.* at 26.

384. *See id.* at 39-49 tbls.12A-K.

On their own terms, these statistical results do *not* yield evidence of anticompetitive self-carriage bias. Before explaining this, however, a number of comments on the strength of the econometric evidence are appropriate.

First, while the paper reports eleven regressions, twelve were estimated, the results from one (involving FUSE) being so "absurd" as to go unreported.[385] The statistical difficulties encountered in this estimation suggest that the data may not fit the model well. These problems are likely to be an issue in estimating the other equations. Indeed, the results obtained for WE, the other Cablevision-owned network (like FUSE) in the sample, appear economically absurd, as shown below. Both sources of information strongly undercut the validity of the estimated coefficients in explaining market behavior.

Second, only eight of twelve regressions suggest that, at standard confidence levels, there exists a statistically significant relationship between vertical integration and cable carriage choices. The evidence, even accepting the underlying economic model, becomes even weaker when it is recalled that the twelve channels chosen for analysis were selected because the very widely distributed channels owned by cable operators—such as AMC and CNN—were omitted.[386] It was argued that anticompetitive foreclosure was not a factor for these channels: "it is important to note that the historic literature on vertical integration and the carriage decision no longer applies to most of the major vertically integrated networks because all of them are carried on virtually all major cable systems."[387] The very popular channels not owned by cable operators—such as MTV and ESPN—could also have been examined (reversing the favoritism hypothesis) but were not. Presumably, the same result would obtain for the nonintegrated channels: "the historical literature on vertical integration and the carriage decision" would not apply to these ubiquitously available networks.

Further, there are numerous problems with the data, including deficiencies in the Warren cable TV database and the geographical mismatch between cable TV franchise areas and the DMAs in which they operate. Beyond these issues, the regressions do not adjust for share ownership of cable networks by MSOs; Time Warner's incentives, when owning all of Boomerang, are treated the same as Comcast's, with its fifty percent share of G4.[388] And relevant information about vertical integration in the twelve selected cable TV networks is discarded: CNN International, wholly owned by Time Warner, all but exited the U.S. market.[389] Launched in 1985 and backed by the second-largest U.S. cable

---

385. *See id.* at 29 n.5 (stating that "something in the data lead the standard errors to be absurdly small and the coefficients absurdly large").

386. *See id.* at 27.

387. *Id.* at 26-27.

388. *See* KAGAN 2005, *supra* note 344, at 56; Leo Sun, *Comcast's (CMCSA) Purchases 51% of NBC Universal from General Electric (GE), and Competitors Cry Foul*, INVESTOR GUIDE.COM (Nov. 24, 2010), http://www.investorguide.com/article/7128/comcasts-cmcsa-purchases-51-of-nbc-universal-from-general-electric-ge-and-competitors-cry-foul/.

389. *See CNN International North America*, CNN NEWS GROUP (2005), http://edition.cnn.

operator, it proved unsuccessful in the United States—much as did CNNfn, which went dark in December 2004.[390]

As shown in Table 2, CNN International had carriage in but eighteen percent of Time Warner's U.S. cable households and just five percent of other firms'.[391] The Goolsbee regressions show evidence of anticompetitive foreclosure in Time Warner carriage decisions.[392] This is because the firm is found more likely (than other MSOs) to carry CNN International, but the tendency is found to be reduced in areas (DMAs) where DBS penetration is higher.[393] The interpretation is that the intensified competition in such high-DBS areas forces Time Warner to move CNN International off its line-ups, making way for more competitive fare.[394]

That is uncompelling and even accepting the DBS penetration rate as a metric for competitiveness, it does not show what the argument implies. That would require evidence that high DBS penetration drove the Time Warner system to omit the network it owned *and* move something more valuable into its place. Dumping CNN International in such markets suggests, by itself, that fewer channels are presented to customers in such markets. Associating the dropping of an owned cable channel is then correlated, wrongly, with competitiveness.

The weakness of the results shown in Table 3 can perhaps be understood by considering the largest and most significant empirical estimates. In the WE ("Women's Entertainment") regression, the coefficient on *Vertical Integration* equals 7.3, twice the magnitude in any other estimated equation. Similarly, the coefficient on *VI*DP* equals -0.41, more than twice the magnitude (in absolute value) obtained elsewhere. Both estimated parameters are significant at the one percent level.

So this is what we can learn: Cablevision Systems, which owns WE, highly favors its own programming relative to other program networks, except when it faces a lot of competition, as measured by the DBS penetration rate for the DMA. But consider further how the DBS penetration rate varies in the case of Cablevision's systems.

The company's subscribers were located in New York, New Jersey, Connecticut, and Pennsylvania, clustering around New York City.[395] DBS penetration variance by DMA is irrelevant to Cablevision's program choices;

---

com/services/opk/cnn25/cnn_ newsgroup.htm#then (listing CNN International as being available in just over 3.5 million North American homes).

390. *See* Linda Moss, *CNNfn's Loss, Others' Gain*, MULTICHANNEL NEWS (Dec. 12, 2004), http://www.multichannel.com/article/116930-CNNfn_s_Loss_Others_Gain.php.

391. *See* Goolsbee, *supra* note 303, at 38 tbl.11.

392. *See id.* at 27-28.

393. *See id.* at 29-30, 39-49 tbls.12A-K.

394. *See id.*

395. The database Goolsbee used features thirty-three systems owed by Cablevision, all located in three DMAs: New York (DMA 1), Philadelphia (DMA 4), and Hartford/New Haven (DMA 28). *Research Studies on Media Ownership*, FCC (Sept. 5, 2007), http://transition.fcc.gov/ownership/studies.html/studydata/2007/study09data.zip (referring to Study 9: Vertical Integration and the Market for Broadcast and Cable Television Programming).

their customers face the same substitute products across the metropolitan area that it serves.[396] It is interesting that the WE regression produced manageable coefficient estimates; the other Cablevision-owned network (FUSE) did not. Since the model found a link between *VI*DP* and WE carriage, this strongly suggests spurious correlation, not strategic behavior.

Even if the results obtained in the cable program network regressions were plausible and no problems existed with data or economic interpretation, the regressions would not constitute evidence suggesting vertical integration in cable as anticompetitive. This is because the estimated parameters allow calculation of "break even" DBS penetration levels where the estimated favoritism ends for a given MSO-owned cable channel. This statistic is reported in Goolsbee as "DBS share for VI neutrality," and presented in Table 12.[397]

Goolsbee's estimates show that, given the model, MSOs are discriminating in favor of their own programming up until a DBS penetration rate of the critical value.[398] After that value, the model suggests that the operator *discriminates against its own programming*. Given existing levels of DBS penetration, it turns out that the eight estimated regressions imply that there is more likely to be this latter discrimination against self-carriage. The results break down this way:

- Twelve channels are investigated in separate regressions.
- Eight of the regressions produce statistically significant coefficients (at standard confidence levels), in the proper direction, for both *VI* variables: BBC America, Boomerang, CNN International, Fit TV, PBS Kids, Science, TV One, and WE.
- The most recent data as of the Goolsbee analysis (July 2007) showed that the national average DBS penetration, with DMAs weighted by households, equaled 26.7 percent.[399]
- Five of the eight equations that find a pattern between *VI* and carriage choice in the model (those for Fit TV, PBS Kids, Science,TV One, and WE) exhibit a DBS "neutrality share" below the current national average level of "competition."
- Three of the eight equations (those for BBC America, Boomerang, and CNN International) exhibit a DBS "neutrality share" above the national average.
- Hence, the empirical model in Goolsbee suggests that *cable systems owned by operators are at least as likely (five times in eight) to suffer negative bias*

---

396. *See* Katy Bachman, *FCC Orders Cablevision to Hand HD Sports Programming Over to Competitors*, ADWEEK, Sept. 22, 2011, http://www.adweek.com/news/television/fcc-orders-cablevision-hand-hd-sports-programming-over-competitors-135051; Kelly Riddell, *Cablevision Blackout Shows News Corp., Broadcasters Gain Edge in Fee Spats*, BLOOMBERG (Oct. 19, 2010), http://www.bloomberg.com/news/2010-10-19/cablevision-blackout-shows-news-crop-broadcasters-gain-edge-in-fee-spats.html.

397. *See* Goolsbee, *supra* note 303, at 39-49 tbls.12A-K.

398. *See id.* at 26-27.

399. The national DBS penetration rate (=26.7%) is based on authors' calculations using the database employed in the Goolsbee study and posted on the FCC's website. *See generally Research Studies on Media Ownership*, *supra* note 395.

from their parent companies as they are to enjoy favoritism.

These results, interpreted according to the model that produced them, offer no support for the conclusion that anticompetitive vertical foreclosure has been found. The evidence presented leads to implausible implications, namely that cable operators discriminate against their own programming. But those results, however explained, do not imply vertical foreclosure.

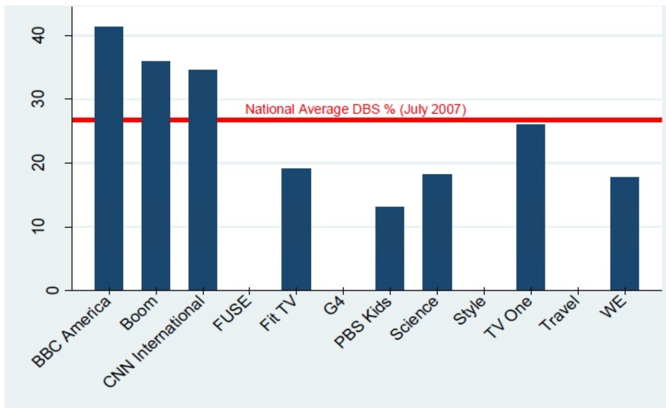

Figure 2. DBS Penetration & VI Neutrality: Goolsbee's Results[400]

It is also possible to see how adding (analog) *Channel Capacity*[401] (*CC*) and *Cable Saturation* by DMA as independent variables alters econometric results. This offers a robustness check, helping to discern whether the coefficient estimates produced in the Goolsbee model are stable across alternative specifications that include theoretically important causative factors. In fact, statistical outcomes substantially vary.

In particular, simply adding one additional variable—*CC*—eliminates the results obtained for the only three regressions, suggesting, given the assumptions of the model, that cable operators discriminate against program networks owned by rivals and do so more in areas where DBS penetration is higher.[402] In two of

400. The data from this chart comes from the combined tables 12A through 12K in Goolsbee, *supra* note 303, at 39-49. *See* Goolsbee, *supra* note 303, at 39-49 tbls.12A-K.

401. *See generally* TELEVISION AND CABLE FACTBOOK, *supra* note 376 (listing channel capacity per cable system).

402. A substantial number of observations are lost when *Channel Capacity* is included, reducing observations (each denoting a different cable TV system) from about 1400 to about 800, depending on the regression. *See Industry Data*, NAT'L CABLE & TELECOMM. ASS'N, http://www.ncta.com/Statistics.aspx (last visited July 2, 2012) (reporting data from TELEVISION AND CABLE FACTBOOK, *supra* note 376). Data unavailability already excludes the overwhelming share of cable TV systems from the analysis, however. *See id.* Goolsbee's regressions incorporate approximately 1400 observations from a universe that, in 2006, was comprised of 7090 systems. *See* NAT'L CABLE & TELECOMMC'NS ASS'N, 2007 INDUSTRY OVERVIEW 7 (2007), *available at* http://i.ncta.com/ncta_com/PDFs/NCTA_Annual_Report_04.24.07.pdf.

the equations (for Boomerang and CNN International), the estimated coefficients for the two VI variables are insignificant.[403] In the third (for BBC America), the model will not compute due to colinearity.[404] Including *Cable Saturation* as an explanatory variable produces additional instability in results.

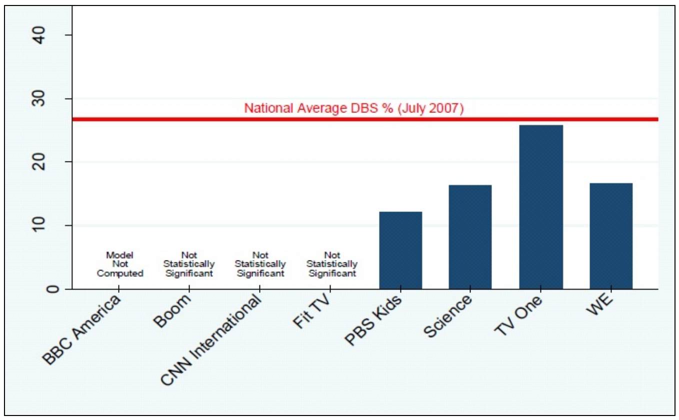

Figure 3. DBS Penetration & VI Neutrality: Adding Channel Capacity[405]

In sum, even accepting the estimates at face value, only three of the twelve regressions suggest that cable operators, on average, discriminate in favor of their own programming; five of the twelve predict that operators discriminate against their own networks; and four of the twelve offer no statistical relationship between carriage choices and vertical integration.[406] Even the results obtained for only the three regressions suggesting discrimination vanish when *Channel Capacity* is added as an explanatory variable.[407] Hence, none of the regressions produce robust results consistent with the hypothesis that vertical integration leads to anticompetitive foreclosure in the marketplace. These results are summarized in Figures 3 and 4, and Table 4, the latter of which also notes the relative economic importance of widely distributed networks (which form the lion's share of industry profits), which were excluded from this analysis because data was not expected to yield variance in carriage choices that could be associated with self-carriage bias.

403. *See* Goolsbee, *supra* note 303, at 40-41 tbls.12B-C.
404. *See infra* fig.3.
405. Hazlett, *Vertical Integration*, *supra* note 309, at 30 fig.6.
406. *See supra* note 399.
407. *See supra* fig.3.

TABLE 4. EVIDENCE ON EFFECT OF VERTICAL INTEGRATION IN GOOLSBEE (2007)[408]

Network Category	*% of 2005 Cable Program Net Cash Flow*	*Included in Regressions*	*Implications of Evidence*
Top MSO-owned program networks (*AMC, CNN, Discovery, TBS, TNT*)*	17	No	Widely carried MSO-owned channels implies lack of "self-carriage" favoritism or, therefore, anticompetitive foreclosure
Top non-MSO program networks (*Nickelodeon, ESPN, MTV, Disney, USA, Lifetime, CNBC, Fox News, BET, HGTV*)**	45	No	Widely carried non-MSO channels implies lack of "self-carriage" favoritism or, therefore, anticompetitive foreclosure
12 smaller cable networks (*BBC America, Boomerang, CNN International, Fit TV, FUSE, PBS Kids, Science Channel, Style, Travel Channel, TV One, WE*)	1.3***	Yes	8 of 12 regressions report two statistically significant coefficients which, jointly, are consistent with a "self-carriage" bias that lessens as DBS penetration rises. Five of the eight biases become neutral with DBS penetration at or below the national average, meaning no anticompetitive foreclosure is generally in evidence. The 3 regressions consistent with foreclosure at typical DBS penetration are not robust to the inclusion of *Channel Capacity*.[409] Robust results consistent with the observation of anticompetitive vertical foreclosure in the marketplace: 0 for 12.

* Networks rejected in Goolsbee for inclusion in foreclosure tests due to widespread coverage.

** Networks not owned by MSOs which had 2005 cash flows exceeding those for AMC, the least profitable network excluded from foreclosure estimates due to widespread coverage.

*** Data from Kagan (2005); CNN International and PBS Kids not listed or included in totals.

4. Public Policy.—The Goolsbee analysis does not offer even the beginnings of an economic case for further regulation. To make that case, two substantial elements would have to be established. The first is a showing that vertical

408. Hazlett, *Vertical Integration*, *supra* note 309, at 31 tbl.6; *see generally* Goolsbee, *supra* note 303.

409. *See supra* fig.3.

integration threatens consumer welfare. The second is a cost-benefit analysis suggesting that proposed regulatory changes will reduce quality-adjusted prices for consumers. The first showing is attempted, but the results wither upon scrutiny. The second showing is not even begun. The empirical predicate for the FCC's net neutrality rules is, by Footnote 60, nonexistent.

VII. Economic Evidence from U.S. Broadband Markets

While citing sources that do not support the FCC's regulatory initiative, the NN Order ignores empirical evidence that—while precisely on point—strongly rejects the economic presumptions on which the NN Order is based.[410] Indeed, when considering the question of market failure, and considering regulatory fixes to remedy the observed inadequacies, virtually the first investigation would seem to focus on market growth. If the marketplace is not performing well, and can be improved by regulation, there ought to be evidence of that in the output patterns evident.

Hence, when the U.S. Department of Justice Antitrust Division commented on the first regulatory foray by the FCC in this area, it immediately questioned the existence of market failure by citing robust growth statistics:

> On the empirical side, despite the Commission's request for evidence of harmful discrimination or behavior . . . commenters failed to present evidence suggesting that a problem exists. To the contrary, it appears that the Internet is flourishing without the proposed sectoral regulation. Statistics evidence an explosion in Internet usage in recent years due to new applications and increased broadband subscribership. . . .
>
> Between June 2005 and June 2006, the Commission found that high-speed lines increased by 52 percent (or 22.2 million lines).[411]

Similarly, three years later, two prominent economists objected to the FCC's move to regulate net neutrality due, in part, to market performance: "The rapid growth in recent years in broadband subscribers, Internet usage, service quality and reductions in price indicate that consumers have derived significant benefits from competition and innovation."[412]

Of course, overall trends may not be helped by alleged discriminatory practices; the possibility remains that negative effects are masked by macroeconomic effects and, moreover, that such growth impediments could be reduced by regulation of broadband ISPs to counter anticompetitive practices.

---

410. *See* FCC NN Order, *supra* note 1, para. 23, at 17,918 & n.60.

411. DOJ 2007, *supra* note 131, at 5-6.

412. Attachment A: Declaration of Gary S. Becker and Dennis W. Carlton, at 5, *In re* Preserving the Open Internet Broadband Indus. Practices, GN Docket No. 09- 191 (FCC, Jan. 14, 2010) [hereinafter Becker & Carlton 2010]. Both of the authors are economists at the University of Chicago. Gary Becker was awarded the Nobel Prize in Economics in 1983, while Dennis Carlton, one of the world's leading industrial organization experts, formerly served as Deputy Assistant Attorney General for Economic Analysis, Antitrust Division, U.S. Department of Justice. *Id.* at 2-3.

Fortunately, a series of natural experiments have been conducted in the U.S. broadband marketplace that casts light on just this set of questions. Research has evaluated how both networks and subscribers change their behavior in response to changes in "open access" rules: regulations that aim to achieve similarly "open" or "neutral" traffic flows across broadband networks, and such research has been published in peer-reviewed scholarly journals.[413] The FCC ignores such evidence, which—unlike studies of vertical integration that deal only with the cable TV market, and not regulatory, data—cast direct light on the effectiveness of the *broadband industry rules* such as they seek to implement.

Cable TV operators have never been subject to "open access" rules, regulations that would force a network owner to share its facilities with rival service suppliers at regulated terms and conditions.[414] Legally, this unregulated environment stems from cable broadband's categorization as an "information service" rather than a "telecommunications service."[415] In terms of regulatory decision-making, it also reflects the fact that policy makers have believed that imposing such mandates on cable broadband providers would discourage investment in infrastructure and stymie the growth of new services.[416]

However, telephone carriers offering competing broadband services, primarily digital subscriber line (DSL) services, were originally subject to such network sharing rules before being deregulated.[417] This regulatory structure, varying across time and technologies, allows us to observe how markets change following regime switches. Not only can cable modem (CM) service growth be compared to DSL, but the regime governing DSL was abruptly changed when the FCC voted to end "line sharing" in February 2003.[418] This policy had allowed third parties to lease the high-frequency portion of incumbent telephone companies' voice lines, using the shared link to provide data services to retail customers at relatively low (wholesale) rates.[419] Ending "line-sharing" dramatically raised the prices charged competitive local exchange carriers (CLECs), undercutting their business models and effectively driving them from the market.[420] Then, in August 2005, after the U.S. Supreme Court had refused to overturn the FCC's deregulatory policy for cable TV systems,[421] the FCC further deregulated telephone carriers, eliminating remaining network access rules by designating DSL—like cable modems—an "information service."[422]

This regulatory pattern allows three windows with which to view the

---

413. *See generally* Thomas W. Hazlett & Anil Caliskan, *Natural Experiments in U.S. Broadband Regulation*, 7 REV. NETWORK ECON. 460 (2008).

414. *See id.* at 462.

415. *See id.* at 473.

416. *See, e.g.*, Kennard, *supra* note 293.

417. Hazlett & Caliskan, *supra* note 413, at 464-65.

418. *Id.* at 464.

419. *See id.* at 462-63.

420. *See id.* at 463-64.

421. *See* Nat'l Cable & Telecomms. Ass'n. v. Brand X Internet Servs., 545 U.S. 967 (2005).

422. Hazlett & Caliskan, *supra* note 413, at 465.

competition between cable modem and DSL services. Of interest is the relative success of CM services versus DSL in terms of subscribership. Where the implementation of "open access" rules stimulates innovation sufficient to dominate any potential investment disincentives, subscriber growth should reflect this. Specifically, cable growth should be disadvantaged relative to DSL growth during the time that "open access" regulation is in effect.

A. Period I (DSL Regulated)

Until the first quarter of 2003, DSL was regulated under relatively tight wholesale price controls. During this period, cable operators emerged as leaders in the "broadband race."[423] Through 2002, CM households held nearly a two-to-one advantage over DSL households.[424]

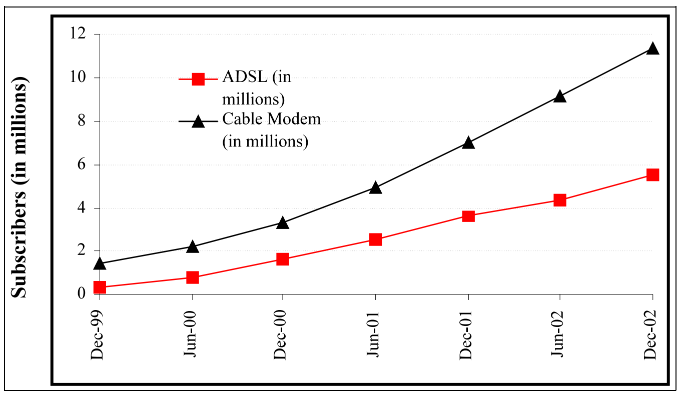

Figure 4. Cable Modem v. DSL Subscribership, 1999-2002[425]

Was this decisive CM edge caused by regulatory differences? Later evidence, gleaned from regime change, would soon suggest that it was. But even at the time, there was important knowledge weighing on this issue. GTE, a large local exchange carrier that joined with AOL, the leading (dial-up) Internet Service Provider, sought to promote "open access" rules on cable modem suppliers.[426] Given that GTE was a direct cable rival in the emerging market for broadband services, it is clear the firm believed that regulation would retard, rather than stimulate, CM deployment.[427]

---

423. *See id.* at 466-68.

424. *See infra* fig.4.

425. Hazlett & Caliskan, *supra* note 413, at 467.

426. *See* George Bittlingmayer & Thomas W. Hazlett, *"Open Access:" The Ideal and the Real*, 26 TELECOMM. POL'Y 295, 296 (2002).

427. That firms often lobby for regulatory rules that will handicap rivals is logically compelling and a widely understand paradigm in economic analysis. *See, e.g.*, Thomas G. Krattenmaker & Steven C. Salop, *Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power Over Price*, 96 YALE L.J. 209, 215 (1986); Bruce Yandle, *Bootleggers and Baptists—The Education of a Regulatory Economist*, REG., May/June 1983, at 12.

B. Period II (DSL "Line Sharing" Ended)

The elimination of line-sharing raised wholesale rates, undermining "open access." In line with this, data-CLEC growth was adversely impacted. But the key issue is what happened to overall broadband growth, and DSL in particular?

DSL lines spurted in the post-deregulation period, sending the total broadband trend much higher. Quarterly subscriber growth, which had trailed cable nearly two-to-one under line-sharing, matched cable modem subscriber growth within just a few quarters.

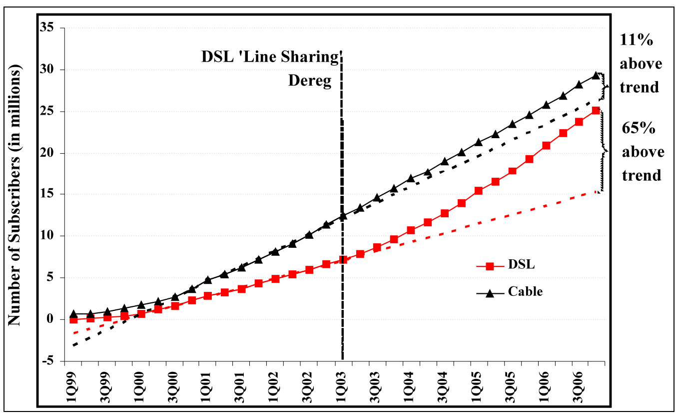

Figure 5. CM & DSL Subscriber Growth Pre- and Post-Deregulation[428]

C. Period III (DSL an "Information Service")

With further deregulation in mid-2005, DSL exhibited another increase from the trend.[429] The larger impact was apparently associated with the policy reform of first quarter 2003, however. At that pivot point, the projected year-end 2006 DSL universe is projected to be about fifteen million households.[430] The actual year-end DSL subscribership was over twenty-five million.[431] The sixty-five percent increase from trend did not come at the expense of cable modems, which maintained their growth profile.[432] And the results cannot be attributed to marketplace changes unconnected to U.S. policy shifts, a possibility adjusted for in regressions that used Canadian cable modem and DSL subscribership as control variables.[433] The implication of the evidence is that U.S. consumers

---

428. Hazlett & Caliskan, *supra* note 413, at 469 fig.2.

429. That incremental effect is not shown here, as the 2003 and 2005 DSL deregulations are combined in Figure 4. The effect is quantified, and found statistically significant, in *id.* at 474-76.

430. *Id.* at 475.

431. *Id.*

432. *See id.*

433. For the econometric results, see *id.* at 476. The empirical model explains U.S. DSL

responded very positively to policy choices that refrained from imposing "open access" or eliminated such rules once in place. The entire three-stage regulatory path is consistent with this interpretation.

D. Fiber Unbundling

Network sharing mandates have also been imposed, and then eliminated, on broadband networks using fiber-to-the-home (FTTH) technology. This opens a fourth policy window to consider. In simple terms, there was virtually no FTTH deployed in the United States prior to the October 2004 decision by the FCC to preempt fiber network "open access" rules (also known as "unbundling").[434] At that point, substantial investments commenced such that today, more than twenty million households are able to subscribe to the Internet through ultra-fast fiber connections, with over six million subscribing.[435]

These data are consistent with the hypothesis that broadband network regulation is inversely related to deployment, and experts have asserted causality to the relationship. Fiber industry sales forecasts, for example, were projected by industry consultancy Gartner.[436] Their initial forecast for 2004-08 was undertaken prior to the FCC deregulation.[437] When the decision to bar unbundling obligations on new fiber networks was made, Gartner explicitly upped its sales estimates for fiber optic inputs.[438]

E. Summary

Across all regime windows, the less regulated broadband platform out-performs. Cable modem service, unburdened by "open access" mandates, spurted out to an early, and quite substantial, advantage in terms of build-out and subscribership. When network sharing mandates on telephone networks were dramatically (and unexpectedly) reduced, DSL—while predicted to shrink, due to the inability of third party providers to continue to compete for retail customers[439]—surged. The further deregulation of DSL services in 2005

---

subscriber growth as a function of contemporaneous U.S. CM growth, as well as Canadian CM and DSL growth (separately). *See id.* at 477. This controls for changes in technology or equipment pricing that would be reflected in a market such as Canada (with similar CM v. DSL rivalry) but does not operate under the same regulatory regime switches as in the U.S. The pronounced upswing in post-regulation DSL growth is not explained by such other factors. *See id.*

434. *Growth of Fiber to the Home*, FTTH COUNCIL, http://www.ftthcouncil.org/en/content/the-growth-of-fiber-to-the-home (last visited July 2, 2012) (depicting graph regarding FTTH homes passed and subscribers).

435. *Id.*

436. *See* Gartner Consulting, *One Gigabit or Bust Roundtable* 12 (Nov. 15, 2004), http://www.cenic.org/events/archives/1gob/112004/mgilbertpres.pdf.

437. Gartner Consulting prepared their projections in November 2004. Deregulation was not implemented until August 2005. *See* Hazlett & Caliskan, *supra* note 413, at 465.

438. *See* Gartner Consulting, *supra* note 436, at 15 (graphing fiber forecasts, 2004-2008).

439. *See* Saul Hansell, *High-Speed Service May Cost More*, N.Y. TIMES, Feb. 21, 2003, at C4.

reinforced this trend, effectively giving DSL (de)regulation with CM services. By year-end 2006, DSL had increased its growth rate so sharply that it accounted for some twenty-five million households—some sixty-five percent, or ten million homes, more than the trend up through 2002 would have predicted.[440]

These results do not directly examine how vertical integration or restrictions levied by broadband ISPs are impacting customers; rather, they focus on the policy margin regulators should be most interested in: Will rules limiting the power of ISPs to exclude rival services, content, or applications be likely to expand network growth and serve consumer interests? The historical data, given the FCC's varying regulatory broadband policies, is rich. It suggests that broadband deployment is furthered when "open access" regulations are eliminated. That is information that a policy analysis, unless arbitrary and capricious, must consider in its decision-making.

CONCLUSION

Consumers may take small solace in the fact that the FCC's NN Order, as presently constituted, will likely meet a swift end before the D.C. Circuit for the same or similar reasons as its order in *Comcast* a few short months ago. As broader regulation of network management appears beyond the scope of the Communications Act, the enduring imposition of net neutrality, as with most other regulatory decisions, will lay with Congress. It seems unlikely Congress will offer up net neutrality for the foreseeable future; activists met substantial opposition in even the Democratic-controlled Congress of 2008-10.[441] An element of the 2008 DNC Party Platform,[442] the newly Republican House will almost certainly refuse to condone through legislation what net neutrality proponents have thus far sought primarily through regulation. Sound economic analysis, supported by robust and meaningful empirical data, must animate any upcoming debate over net neutrality—as it must with any consumer-welfare-oriented legislation. Agencies often overlook the long-run costs of both their proposed policies as well as the regulatory forms those policies take. Net neutrality-like rules are far from novel, and their imposition is anything but a partisan affair. From C-Block licenses to cable broadcasts, promulgating regulations with obvious protectionist implications and dubious consumer welfare benefits has proven a truly bipartisan affair. Net neutrality follows in this pedigree, privileging certain market participants at the expense of others. Both Congress's and the courts' rejection of net neutrality is therefore welcome news for consumers at large.

Yet such solace is indeed small, as net neutrality's short-run welfare costs begin to mount in earnest. The FCC has received its first complaint under the NN order, against MetroPCS, a small cellular carrier primarily serving low-income

---

440. *See supra* fig.5.

441. *See, e.g.*, Ira Teinowitz, *Congress' Takeover of Net Neutrality Is Dead*, WRAP (Sept. 29, 2010), http://www.thewrap.com/media/column-post/congress-net-neutrality-takeover-dead-21304.

442. *See 2008 Democratic Party Platform*, AM. PRESIDENCY PROJECT (Aug. 25, 2008), http://www.presidency.ucsb.edu/ws/index.php?pid=78283#axzz1zsKF9mQl.

consumers. The complaint alleges MetroPCS favors certain websites (namely YouTube, owned by Google) at the expense of others, harming consumers by excluding other websites.[443] The complaint ignores hard-won observations elementary to antitrust in its application of a categorical regulatory prohibition. MetroPCS enjoys no financial stake in increasing Google's traffic, it holds an incredibly small market share, and it provides massive savings to consumers.[444] The MetroPCS complaint demonstrates the failures of net neutrality unto itself: a firm, providing concrete and known benefits to consumers, using inexpensive technology through a mutually beneficial arrangement with a third party, increases consumer welfare in untold ways. With its lack of an intellectually and economically coherent method of recognizing a simple distinction between procompetitive and anticompetitive network discrimination, the FCC prepares to condemn the cellular carrier of choice amongst many low-income phone users. This systemic failing in categorical prohibitions might be an unexpected consequence were it new; unfortunately, it is a problem antitrust has spent a hundred years correcting. One can hope net neutrality takes nowhere near as long.

443. *See supra* notes 12-16 and accompanying text.

444. *See* text accompanying *supra* note 16.

Indiana Law Review

Volume 45 2012 Number 3

NOTES

NET INCOME WITH MAKE-UP CHARITABLE REMAINDER UNITRUSTS AND THE TRUSTEE'S POWER TO ADJUST UNDER INDIANA'S UNIFORM PRINCIPAL AND INCOME ACT

TED R. BATSON, JR.*

INTRODUCTION

In 1997, the National Conference of Commissioners on Uniform State Laws (Commissioners) promulgated a new version of the Uniform Principal and Income Act (UPAIA 1997).[1] Indiana adopted the UPAIA 1997 as the Indiana Uniform Principal and Income Act (the Indiana Act) in 2002.[2] Both the Indiana Act and the UPAIA 1997 provide in part that "[a] trustee may adjust between principal and income to the extent the trustee considers necessary" if certain conditions are met.[3] An adjustment under this provision is commonly referred to as the trustee's "power to adjust." One of the necessary conditions to an exercise of the power to adjust is that the trust's governing instrument "describe the amount that may or must be distributed to a beneficiary by referring to the trust's income."[4] A net income with make-up charitable remainder unitrust (NIMCRUT) is a type of split-interest trust defined in the Internal Revenue Code that, by definition, requires the distribution of the trust's income subject to a unitrust

---

* J.D. Candidate, 2012, Indiana University Robert H. McKinney School of Law; M.B.A., 1990, Indiana University; B.A., 1986, Asbury University, Wilmore, Kentucky. I am grateful to the following colleagues and friends who unselfishly gave of their time and expertise during the preparation of this Note: Professor Peter A. Prescott, Gregory W. Baker, Steven R. Bone, Emanuel J. Kallina, Jonathan D. Ackerman, David W. Newman, and Frances D. Schafer. As with most significant endeavors, I am also deeply indebted to my wife Ruth and my two daughters, Kelsey and Mackenzie.

1. UNIF. PRINCIPAL & INCOME ACT (amended 2008), 7A pt. III U.L.A. 363 (Supp. 2010), *available at* http://www.law.upenn.edu/bll/archives/ulc/upaia/2000final.pdf. Previous versions were promulgated in 1931, UNIF. PRINCIPAL & INCOME ACT, 7A pt. III U.L.A. 589 (2006), and 1962, UNIF. PRINCIPAL & INCOME ACT, REVISED 1962 ACT, 7A pt. III U.L.A. 547 (2006).

2. Uniform Principal and Income Act, Pub. L. No. 84-2002, § 2, 2002 Ind. Acts 987 (codified at IND. CODE §§ 30-2-14-0.1 to -44 (2011), *available at* http://www.in.gov/legislative/ic/code/title30/ar2/ch14.html).

3. IND. CODE § 30-2-14-15(a) (2011); UNIF. PRINCIPAL & INCOME ACT § 104(a).

4. IND. CODE § 30-2-14-15(a)(2); UNIF. PRINCIPAL & INCOME ACT § 104(a).

limitation.[5]

The trustee's power to adjust is not a power of unfettered discretion. Both the Indiana Act and the UPAIA 1997 impose limitations on a trustee's ability to exercise the power to adjust. These limitations have special implications in the context of a NIMCRUT. In addition, the Internal Revenue Code grants certain federal tax benefits to the settlor of a NIMCRUT and exempts a NIMCRUT from income tax. Similar benefits are often conferred under state law. Therefore, an Indiana NIMCRUT trustee desiring to exercise the power to adjust must be aware of the limitations imposed by the Indiana Act, and federal and state tax authorities.

Part I of this Note provides contextual background describing the NIMCRUT and the policy rationale behind the power to adjust. Part II examines the power to adjust as found in the Indiana Act, focusing on the threshold conditions for exercising the power and the internal limitations the Indiana Act places on a trustee's exercise of the power. Part III advocates an amendment to the Indiana Act to clarify the availability of the power to adjust to Indiana NIMCRUT trustees.

I. Setting the Stage

A. The Net Income with Make-Up Charitable Remainder Unitrust

The term "charitable remainder trust" (CRT) is used to denote a trust arrangement that creates two distinct classes of beneficiaries: (1) income beneficiaries[6] who are generally entitled to receive an income interest for life,[7] and (2) charitable remainder beneficiaries who are entitled to receive the remainder interest upon the expiration of the income interest.[8] In 1969, Congress enacted the Tax Reform Act of 1969[9] and prescribed rules such that a properly drafted, funded, and administered CRT (a qualified CRT) provides a number of tax benefits to the creator of the trust. The NIMCRUT is a special case of the CRT that is distinguished by the manner in which the income beneficiary's right to income is defined.

1. Charitable Remainder Trusts Generally.—CRTs are defined at I.R.C. § 664 and its supporting Treasury regulations. A qualified CRT is a trust arrangement that provides for the payment of a distribution at least annually to

---

5. *See* I.R.C. § 664(d)(3) (2006).

6. *Id.* § 664(d)(1)(A) (describing a charitable remainder annuity trust); *Id.* § 664(d)(2)(A) (describing a charitable remainder unitrust).

7. The term of a charitable remainder trust may be "for a term of years (not in excess of 20 years) or for the life or lives of such individual or individuals." *Id.* §§ 664(d)(1)(A), 664(d)(2)(A).

8. *Id.* § 664(d)(1)(C) (describing a charitable remainder annuity trust); *id.* § 664(d)(2)(C) (describing a charitable remainder unitrust).

9. Tax Reform Act of 1969, Pub. L. No. 91-172, § 201(e)(1), 83 Stat. 487, 562-64 (1969), (codified as amended at I.R.C. § 664).

one or more non-charitable income beneficiaries[10] with the trust remainder irrevocably designated to, or for the use of, one or more charitable organizations described in I.R.C. § 170(c).[11] CRTs are generally classified into two categories based upon the method used to compute the required payout to the income beneficiary. In the first category, the charitable remainder *annuity trust* (CRAT), the trust pays the income beneficiary a *fixed amount* of not less than five percent (and no more than fifty percent) of the trust's initial value (the annuity amount).[12] In the second category, the charitable remainder *unitrust* (CRUT), the trust pays the income beneficiary a *fixed percentage* of not less than five percent (and no more than fifty percent) of the trust's fair market value as redetermined annually (the annual unitrust amount).[13]

In either a CRAT or a CRUT, the trustee may pay the annuity amount or the annual unitrust amount from income or principal but must pay the income beneficiary the required distribution from principal if the trust's income is insufficient.[14] Trust income in excess of the stated annuity amount or the annual unitrust amount must be added to the trust's principal.[15]

A CRT created in conformity with I.R.C. § 664 is exempt from tax[16] unless it has unrelated business taxable income.[17] A transferor to an inter vivos CRT receives an income tax charitable deduction,[18] a gift tax charitable deduction,[19] and, if applicable, an estate tax charitable deduction[20] equal to the present value

---

10. I.R.C. §§ 664(d)(1)(A), 664(d)(2)(A); *see also* Treas. Reg. § 1.664-1(a)(1)(I) (1972). The Internal Revenue Code requires that the current income beneficiary(ies) of a charitable remainder trust be "one or more persons (at least one of which is not [a charitable organization])." I.R.C. §§ 664(d)(1)(A), 664(d)(2)(A). Accordingly, charitable income beneficiaries are permitted so long as there is at least one non-charitable beneficiary.

11. I.R.C. § 664(d)(2)(C). Provision is also made for the remainder interest to remain in trust for the use of an organization described in section 170(c). *Id.*; *see also* Treas. Reg. § 1.664-1(a)(1)(I).

12. I.R.C. § 664(d)(1)(A); *see also* Treas. Reg. § 1.664-1(a)(1)(I).

13. I.R.C. § 664(d)(2)(A); *see also* Treas. Reg. § 1.664-1(a)(1)(I).

14. Rev. Rul. 72-395 § 4.01, 1972-2 C.B. 340, 342 (describing the requirement for CRATs); *id.* § 6.01 (describing the requirement for CRUTs).

15. *Id.* § 7.01 (The sample trust provision illustrated states, "Any income of the trust in excess of such payments *shall* be added to principal." (emphasis added)). *But see id.* § 6.01(2) ("Any income of the trust in excess of the unitrust amount *may, but need not*, be added to principal. Care should be taken, however, *to assure that*, under applicable local law, *such excess income is retained by the trust*." (emphasis added)).

16. I.R.C. § 664(c)(1) (providing that a CRT is exempt from income taxation); *see also* Treas. Reg. § 1.664-1(a)(1)(I).

17. I.R.C. § 664(c)(2) (imposing an excise tax of one hundred percent on a CRT's unrelated business taxable income).

18. *Id.* § 170(f)(2)(A) (2006).

19. *Id.* § 2522(c)(2)(A) (2006).

20. *Id.* § 2055(e)(2)(A) (2006). The estate tax charitable deduction is only available if the settlor retains an income interest. If the settlor names someone other than himself or herself as an

of the charitable remainder interest.[21] Similarly, a transferor to a testamentary CRT receives an estate tax charitable deduction.[22]

2. The NIMCRUT, a Special Case.—A NIMCRUT is a special case of a CRUT. Internal Revenue Code section 664(d)(3)(A) adds a twist to the standard CRUT definition by providing that a CRUT agreement may define the payment to the income beneficiary to be the *lesser* of the CRUT's annual unitrust amount *and* the trust's accounting income.[23] Internal Revenue Code section 664(d)(3)(B) adds a further twist by permitting the payment of trust accounting income in excess of the annual unitrust amount to the extent of any accumulated prior year deficiencies to "make-up" for those prior year shortfalls.[24] Prior year deficiencies are measured as the excess of the CRUT's annual unitrust amount over the CRUT's trust accounting income each year.[25] The combination of the trust income provision and the deficiency "make-up" provision gives rise to the name commonly associated with this type of CRUT, the *n*et *i*ncome with *m*ake-up *c*haritable *r*emainder *u*ni*t*rust, or NIMCRUT.[26] Despite the difference in payout structure between a standard CRUT and a NIMCRUT, the computation of the present value of the charitable remainder interest (i.e., the deductible amount) is exactly the same for both types of CRUT.[27]

3. Use of NIMCRUTs in Charitable Planning.—Donors wishing to make charitable contributions of illiquid assets in exchange for an income stream commonly use the NIMCRUT.[28] The NIMCRUT is particularly suited for this

---

income beneficiary, then there is no estate inclusion of the CRT corpus and, as a result, no charitable estate tax deduction. *See id.*

21. Treas. Reg. § 1.664-3(d) (1972); *see also* Treas. Reg. § 1.664-4(a) (1972).

22. I.R.C. § 2055(e)(2)(A).

23. "Trust accounting income" is the term used to describe that portion of a trust portfolio's total return realized in the form of interest, dividends, rents, and other statutorily defined categories. UNIF. PRINCIPAL & INCOME ACT (amended 2008) § 104 cmt., 7A pt. III U.L.A. 363 (Supp. 2010). Trust accounting income is the amount to which an income beneficiary is entitled when a trust agreement "express[es] the income beneficiary's distribution rights in terms of the right to receive 'income.'" *Id.*

24. I.R.C. § 664(d)(3)(B).

25. *Id.*

26. The make-up provision is optional. I.R.C. § 664(d)(3) ("[T]he trust instrument *may* provide that the trustee shall pay the income beneficiary for any year – . . . (B) any amount of the trust income which is in excess of [the annual unitrust amount]." (emphasis added)); *see also* Rev. Rul. 72-395 § 7.01 cmt. 1, 1972-2 C.B. 340, 350 (explaining that "[m]akeup of deficiencies [is] not required" (emphasis omitted)). In rare cases this provision is omitted and the resulting trust is typically called a "net income charitable remainder unitrust," or NICRUT.

27. *See* Rev. Rul. 72-395 § 7.01 cmt. 2, 1972-2 C.B. 340, 350 ("[N]otwithstanding the [net income limitation], the computation of the charitable deduction will be determined on the basis that the regular unitrust amount will be distributed in each taxable year of the trust.").

28. *See* B. Howard Pearson, *Charitable Remainder Trusts and Other Charitable Techniques*, *in* 30TH ANNUAL ESTATE PLANNING INSTITUTE, at 967, 976-77 (PLI Tax Law & Est. Plan., Course Handbook Series No. D-283, 1999).

type of contribution because, unlike a CRAT or a standard CRUT, a NIMCRUT is not required to make a distribution to an income beneficiary unless the trust receives trust accounting income. Thus, for example, the NIMCRUT trustee in possession of an illiquid asset, such as raw land, is not required to make a distribution until the land is sold and the proceeds are invested in an income-producing portfolio. Were this property contributed to a CRAT or a standard CRUT, the trustee would be required to make a distribution according to the distribution schedule defined in the trust agreement. Failure to make distributions when required will generally cause a charitable remainder trust to lose its tax-qualified status, potentially resulting in catastrophic results to the donor.[29]

NIMCRUTs are used much less frequently than CRATs or standard CRUTs. This is most likely because income distributions from a NIMCRUT are limited to the trust's net income instead of a more predictable (and, in many cases, larger) annuity or unitrust amount. In 2007, for example, 21,475 NIMCRUT returns were filed compared with 74,773 returns filed by standard CRUTs.[30] Based on these statistics, NIMCRUTs represented 22.3% of all charitable remainder unitrusts filing returns in 2007.[31]

4. The Intersection of Federal Income Tax Law and State Principal and Income Law.—Treasury Regulation section 1.664-3(a)(1)(i)(b)(3) states that "trust income generally means income as defined under [I.R.C. §] 643(b) and the applicable regulations."[32] Internal Revenue Code section 643(b) specifies that "the term 'income,' when not preceded by the words 'taxable,' 'distributable net,' 'undistributed net,' or 'gross,' means the amount of income of the estate or trust for the taxable year *determined under the terms of the governing instrument and applicable local law*."[33] The applicable local law in Indiana is the Indiana Act.[34]

B. The Modernization of Principal and Income Accounting: A Play in Two Acts

In administering a trust, a trustee is bound by a duty of impartiality to balance

---

29. *See, e.g.*, Estate of Atkinson v. Comm'r, 309 F.3d 1290, 1295-96 (11th Cir. 2002) (denying an estate tax charitable deduction where the trustee of a CRAT failed to make the required annual annuity payments to the income beneficiary).

30. *See* Lisa Schreiber Rosenmerkel, *Changing Times: An Analysis of the 2007 Revision of the Split-Interest Trust Information Return*, 29 I.R.S. STAT. INCOME BULL., Mar. 2010, at 130, 135, *available at* http://www.irs.gov/pub/irs-soi/10winbul.pdf. For purposes of this analysis, the number of NIMCRUT returns filed was computed as the sum of "Net Income CRUTs," 4,628, and "Net Income with makeup CRUTs," 16,847. *See id.*

31. *See id.* For purposes of this analysis, the percentage of NIMCRUT returns filed was computed by dividing the sum of "Net Income CRUTs," 4,628, and "Net Income with makeup CRUTs," 16,847, or 21,475, by the total number of returns filed, 96,248. *See id.*

32. Treas. Reg. § 1.664-3(a)(1)(i)(b)(3) (as amended in 2004).

33. I.R.C. § 643(b) (2006) (emphasis added).

34. IND. CODE § 30-2-14-14(a) (2011) ("The following applies to a fiduciary in allocating receipts and disbursements to or between principal and income . . . .").

the competing interests of various classes of trust beneficiaries.[35] In the context of principal and income accounting, these competing interests are illustrated, on the one hand, by the interest of income beneficiaries that the trust's investment portfolio be productive of income for distribution, and, on the other hand, by the interest of remainder beneficiaries in the preservation and growth of the trust's principal. To strike this balance, trustees have traditionally invested trust portfolios according to a prudence standard and allocated the resulting investment receipts and disbursements between the income account (payable to the income beneficiaries) and principal account (payable to the remainder beneficiaries at the expiration of the trust's term).[36]

In the thirty-five years between the promulgation of the Revised Uniform Principal and Income Act[37] and the promulgation of UPAIA 1997, both investment practice and scholarship about the behavior of the investment markets changed significantly.[38] The Commissioners first responded to these changes in 1994 by promulgating the Uniform Prudent Investor Act (UPIA).[39] The UPIA set aside the "prudent man rule" of *Harvard College v. Amory*[40] in favor of the "prudent investor rule" of the *Restatement (Third) of Trusts*.[41] The Commissioners followed the UPIA with the UPAIA 1997 to harmonize the adoption of the prudent investor rule in the UPIA with the rules relating to the allocation of receipts and disbursements to the income and principal accounts.[42]

1. Act 1: The Uniform Prudent Investor Act.—The UPIA modernized trust investment law in light of the widespread acceptance of modern portfolio theory.[43] Under modern portfolio theory, a trustee invests a trust's portfolio as a whole to maximize total return given an acceptable level of risk.[44] In the language of the UPIA, "[a] trustee's investment and management decisions respecting individual assets must be evaluated not in isolation but in the context of the trust portfolio as a whole and as a part of an overall investment strategy

---

35. *See* RESTATEMENT (THIRD) OF TRUSTS § 79 (2007).

36. *See* Robert B. Wolf, *Defeating the Duty to Disappoint Equally—The Total Return Trust*, 32 REAL PROP. PROB. & TR. J. 45, 50-51 (1997).

37. UNIF. PRINCIPAL & INCOME ACT, REVISED 1962 ACT, 7A pt. III U.L.A. 547 (2006).

38. *See* UNIF. PRINCIPAL & INCOME ACT (amended 2008) prefatory n., 7A pt. III U.L.A. 363, 366-67 (Supp. 2010); UNIF. PRUDENT INVESTOR ACT, prefatory n., 7B U.L.A. 1, 3 (2006); RESTATEMENT (THIRD) OF TRUSTS, ch. 17 intro. n. (2007).

39. UNIF. PRUDENT INVESTOR ACT, 7B U.L.A. 1.

40. 26 Mass. (9 Pick.) 446, 469 (1830) ("[Trustees should] observe how men of prudence, discretion and intelligence manage their own affairs, not in regard to speculation, but in regard to the permanent disposition of their funds, considering the probable income, as well as the probable safety of the capital to be invested.").

41. RESTATEMENT (THIRD) OF TRUSTS § 90.

42. *See* UNIF. PRINCIPAL & INCOME ACT prefatory n.

43. *See* UNIF. PRUDENT INVESTOR ACT prefatory n.

44. *See* S. Alan Medlin, *Limitations on the Trustee's Power to Adjust*, 42 REAL PROP. PROB. & TR. J. 717, 719, 736 (2008).

having *risk and return objectives reasonably suited to the trust*."[45] This is distinguishable from the prior rule of *Amory*, which over time came to treat certain classes of assets as per se imprudent as well as to focus on the prudence or imprudence of the selection of individual assets in a trust's portfolio.[46]

Inherent in the UPIA's approach is the absence of any concern about the *categories* of returns generated by an investment portfolio.[47] However, under traditional notions of trust accounting income, the categories of returns generated by an investment portfolio are important because only investment returns from specific categories (e.g., interest, dividends, rents, and royalties) are considered allocable to income[48] and, therefore, payable to income beneficiaries. Similarly, absent language to the contrary in the trust agreement, capital appreciation, realized or unrealized, has been traditionally allocable to the trust principal[49] and, therefore, retained for the benefit of the trust's remainder beneficiaries. As a result, in states that adopted the UPIA, trustees were potentially placed in a bind. Those trustees were duty-bound to invest according to the modern portfolio theory principles of the UPIA.[50] At the same time, those trustees were duty-bound to impartially balance the interests of both income and remainder beneficiaries with respect to the production of income under the traditional income categories while adding to trust principal through capital appreciation. This balance was complicated by the fact that, in general, the greater the proportion of a trust portfolio invested in equity securities, the smaller the proportion of the overall expected investment return that will come from one of the traditional categories of income distributable to a trust's income beneficiaries.[51] Conversely, the greater the proportion of a portfolio invested in fixed-income securities, the smaller the proportion of the overall expected investment return that will result from capital appreciation that increases the trust's principal.[52] In other words, under the traditional model, to achieve a targeted amount of income to distribute to the trust's income beneficiaries, the trustee was necessarily required to sacrifice some overall investment return potential in order to generate returns from categories that were distributable to the trust's income beneficiaries.

This dilemma can be illustrated as follows[53]:

---

45. UNIF. PRUDENT INVESTOR ACT § 2(b) (emphasis added).

46. *See* RESTATEMENT (THIRD) OF TRUSTS, Ch. 17 intro. n.

47. *See* UNIF. PRINCIPAL & INCOME ACT prefatory n.; Medlin, *supra* note 44, at 719.

48. *See* UNIF. PRINCIPAL & INCOME ACT prefatory n.; Christopher P. Cline, *The Uniform Prudent Investor and Principal and Income Acts: Changing the Trust Landscape*, 42 REAL PROP. PROB. & TR. J. 611, 647 (2008).

49. *See* Cline, *supra* note 48, at 647.

50. *See, e.g.*, IND. CODE § 30-4-3.5-1(a) (2011) ("Except as otherwise provided [by the terms of the trust agreement], a trustee who invests and manages trust assets owes a duty to the beneficiaries of the trust to comply with the prudent investor rule set forth in this chapter.").

51. *See* Wolf, *supra* note 36, at 59.

52. *See id.*

53. This example is patterned after an example found in Medlin, *supra* note 44, at 721.

Assume that for the current year, a trustee can choose between Portfolio A invested solely in the stock market (a one hundred percent equity allocation) and Portfolio B invested solely in fixed-income securities[54] (a one hundred percent fixed-income allocation). Further assume Portfolio A will earn a fifteen percent return, all in the form of unrealized capital appreciation, while Portfolio B will earn a four percent return, all in the form of interest income. Assuming (1) an initial value of $1,000,000 and (2) that all trust accounting income is distributed to the trust's income beneficiaries, then the trust's investment and distribution activity can be summarized as follows:

	Portfolio A 100% Equity	Portfolio B 100% Fixed-Income
Beginning of Year Value	$1,000,000	$1,000,000
Capital Appreciation	$150,000	$0
Trust Accounting Income	$0	$40,000
Distribution to Income Beneficiaries	$0	($40,000)
End of Year Value	$1,150,000	$1,000,000

Assuming that the returns illustrated represent the maximum possible returns for each portfolio given an equally acceptable degree of risk,[55] one can see that Portfolio A maximizes the trust's investment performance but generates no income to pay to the income beneficiaries; all of the investment return inures to the benefit of the remainder beneficiaries. But, if the trustee selects Portfolio B to generate income to distribute to the income beneficiaries, then the trustee will clearly favor the interests of the income beneficiaries over the remainder beneficiaries because the portfolio generates no growth in the trust portfolio. Thus, the trustee is placed in a position of conflict between the two classes of beneficiaries, each desiring that the trustee invest with their best interests in mind.

Trustees have traditionally managed this conflict in Solomon-like fashion by devising portfolios, composed in part of equity securities and in part of fixed-income securities, expected to generate blended returns comprised of interest, dividends, and capital appreciation. Returning to the previous example, if the trustee invested instead in Portfolio C, which invests forty percent in equity

54. Fixed-income securities, generally corporate bonds and other similar corporate debt instruments, produce fixed rates of return in the form of interest. *See* Medlin, *supra* note 44 at 719. While the rate of return from fixed-income securities is subject to adjustment for inflation, deflation, and other economic conditions, the rate of return on fixed-income securities is generally more stable than the rate of return on equity securities (e.g., corporate stocks). *See id.* at 720.

55. Because equity securities are generally more risky than fixed income securities, this assumption is unlikely to hold in a real-world investment scenario. *See* Wolf, *supra* note 36, at 52-60.

securities and sixty percent in fixed-income securities, the resulting activity can be summarized as follows[56]:

	Portfolio A 100% Equity	Portfolio B 100% Fixed-Income	Portfolio C 40% Equity/ 60% Fixed-Income
Beginning of Year Value	$1,000,000	$1,000,000	$1,000,000
Capital Appreciation	$150,000	$0	$60,000
Trust Accounting Income	$0	$40,000	$24,000
Distribution to Income Beneficiaries	$0	($40,000)	($24,000)
End of Year Value	$1,150,000	$1,000,000	$1,060,000

While Portfolio C does generate income that is distributable to the trust's income beneficiaries, it is only sixty percent of the amount they would receive under Portfolio B. Similarly, the remainder beneficiaries would receive some capital appreciation, but only forty percent of the amount they would receive under Portfolio A. One commentator has characterized this result as the trustee's duty to disappoint equally.[57]

2. Act 2: The Uniform Principal and Income Act.—The Commissioners' response to the potential conflict between an investment standard that is indifferent to the categories of returns generated and a trust accounting income regime principally based on categories of returns, was the introduction of an equitable power of adjustment in the UPAIA 1997. UPAIA 1997 section 104(a) states:

> A trustee may adjust between principal and income to the extent the trustee considers necessary if the trustee invests and manages trust assets as a prudent investor, the terms of the trust describe the amount that may or must be distributed to a beneficiary by referring to the trust's income, and the trustee determines, after applying the rules in Section 103(a), that the trustee is unable to comply with Section 103(b).[58]

The comments to section 104 (Comments) explicitly state that "[t]he purpose of [the power to adjust] is to enable a trustee to select investments using the standards of a prudent investor without having to realize a particular portion of

56. This example is an extension of an example found in Medlin, *supra* note 44, at 721.

57. *See* Wolf, *supra* note 36, at 72 ("All the trustee can do is to try to fulfill their duty of impartiality by disappointing income and remainder beneficiaries equally!").

58. UNIF. PRINCIPAL & INCOME ACT § 104(a) (amended 2008), 7A pt. III U.L.A. 363, 434 (Supp. 2010).

the portfolio's total return in the form of traditional trust accounting income such as interest, dividends, and rents."[59]

To guide trustees in determining whether and to what extent to exercise the power to adjust, UPAIA 1997 section 104(b) provides the following nine factors for consideration:

(1) the nature, purpose, and expected duration of the trust;
(2) the intent of the settlor;
(3) the identity and circumstances of the beneficiaries;
(4) the needs for liquidity, regularity of income, and preservation and appreciation of capital;
(5) the assets held in the trust; the extent to which they consist of financial assets, interests in closely held enterprises, tangible and intangible personal property, or real property; the extent to which an asset is used by a beneficiary; and whether an asset was purchased by the trustee or received from the settlor;
(6) the net amount allocated to income under the other sections of this [Act] and the increase or decrease in the value of the principal assets, which the trustee may estimate as to assets for which market values are not readily available;
(7) whether and to what extent the terms of the trust give the trustee the power to invade principal or accumulate income or prohibit the trustee from invading principal or accumulating income, and the extent to which the trustee has exercised a power from time to time to invade principal or accumulate income;
(8) the actual and anticipated effect of economic conditions on principal and income and effects of inflation and deflation; and
(9) the anticipated tax consequences of an adjustment.[60]

The Comments state that these factors are intended to be consistent with a similar list of factors in the UPIA "so that, to the extent possible, comparable factors will apply to investment decisions and decisions involving the power to adjust."[61]

The Comments provide several examples illustrating the interplay between the trustee's investment decisions and the exercise of the power to adjust. For example, in Example 1 from the Comments, a newly appointed trustee determines that a shift in the trust's investment allocation from a portfolio invested twenty percent in stocks (i.e., equity securities) and eighty percent in bonds (i.e., fixed-income securities) to a portfolio equally weighted between the two asset classes is appropriate because this alternative portfolio "has risk and return objectives that are reasonably suited to the trust."[62] However, having made this decision, the

59. *Id.* § 104 cmt.

60. *Id.* § 104(b) (alteration in original).

61. *Id.* § 104 cmt.

62. *Id.* Note that the phrase "risk and return objectives that are reasonably suited to the trust" is from section 2(b) of the UPIA. UNIF. PRUDENT INVESTOR ACT § 2(b).

trustee also determines that this new portfolio allocation will result in the receipt of less dividend and interest income.[63] After considering the factors listed in UPAIA 1997 section 104(b), the "[trustee] may transfer cash from principal to income to the extent [the trustee] considers it necessary to increase the amount distributed to the income beneficiary."[64]

II. Indiana's Power to Adjust

The power to adjust is codified in the Indiana Act as follows:

> (a) A trustee may adjust between principal and income to the extent the trustee considers necessary if:
>
> (1) the trustee invests and manages trust assets as a prudent investor;
>
> (2) the terms of the trust describe the amount that may or must be distributed to a beneficiary by referring to the trust's income; and
>
> (3) the trustee determines:
>
> (A) after applying the rules in section 14(a) of this chapter; and
>
> (B) considering any power the trustee may have under the trust or the will to invade principal or accumulate income;
>
> that the trustee is unable to comply with section 14(b) of this chapter.[65]

As adopted, Indiana's power to adjust statute is substantially the same as UPAIA 1997 section 104. The only substantive change was the insertion of the phrase "considering any power the trustee may have under the trust or the will to invade principal or accumulate income" in Indiana Code section 30-2-14-15(a)(3).[66] This phrase does not substantively alter the meaning of the power to adjust as embodied in the UPAIA 1997, but rather serves to highlight the requirement that the trustee first administer a trust according to its terms as required by Indiana Code section 30-2-14-14(a)(1).

Under the Indiana Act, a number of threshold conditions must be satisfied prior to an exercise of the power to adjust.[67] These threshold conditions can be divided into those that are external to the power to adjust (embodied in the references to section 14(a) and section 14(b) in the text of the statute) and those that are internal to the power to adjust's statutory text. Assuming the threshold conditions are met, the trustee must then determine whether any of the statutory prohibitions limiting the exercise of the power to adjust found in Indiana Code section 30-2-14-15(c) are applicable. If, after reviewing the prohibitions on the exercise of the power to adjust found in Indiana Code section 30-2-14-15(c), the trustee is still not prevented from pursuing the power to adjust, then the trustee

---

63. *See* UNIF. PRINCIPAL & INCOME ACT § 104 cmt.

64. *Id*.

65. IND. CODE § 30-2-14-15(a) (2011).

66. *Id.* § 30-2-14-15(a)(3).

67. *See* Medlin, *supra* note 44, at 726-36 (discussing these threshold conditions in the context of the UPAIA 1997).

should assess the factors found at Indiana Code section 30-2-14-15(b)[68] to determine whether and to what extent to exercise the power to adjust.

A. External Threshold Conditions

As an initial matter, a trustee *must* administer a trust according to the terms of the trust agreement, even if an applicable trust provision differs from the Indiana Act.[69] The Comments[70] explain that this rule "emphasize[s] . . . that provisions in the terms of the trust are paramount"; the UPAIA 1997 is a set of default rules that only have effect when the document is silent.[71] Second, a trustee *may* exercise any discretionary powers of administration given the trustee under the terms of the trust agreement.[72] Third, a trustee *must* administer the trust in accordance with the applicable principal and income act in matters that are not addressed in the trust agreement or covered by a discretionary power of administration granted the trustee by the terms of the trust agreement.[73] Finally, "[i]n exercising the power to adjust . . . a fiduciary shall administer a trust . . . impartially . . . ."[74] The Comments describe this requirement of impartiality as the "paramount consideration in applying [the power to adjust]."[75]

In sum, before a trustee considers the appropriateness of an exercise of the power to adjust, the trustee must first administer the trust using all principal and income rules contained in the trust agreement and all the discretionary and non-discretionary powers granted to the trustee under the terms of the trust agreement. Next, the trustee must apply all of the provisions of the Indiana Act *other than the power to adjust*. Having taken into account powers granted by the trust agreement and having applied the Indiana Act, the trustee must then determine whether the trustee has impartially administered the trust as between the income and remainder beneficiaries before proceeding to the power to adjust. If, in the trustee's assessment, the trustee has impartially administered the trust as between

---

68. The list of factors in IND. CODE § 30-2-14-15(b) corresponds to the list of factors found at UNIF. PRINCIPAL & INCOME ACT § 104(b). *See supra* note 60 and accompanying text.

69. IND. CODE § 30-2-14-14(a)(1) (2011).

70. Indiana did not create its own explanatory comments when it enacted the Indiana Act. However, where a statute is based on a uniform act, "[t]he comments to [the] uniform act are indicative of the Legislature's intent." Basileh v. Alghusain, 912 N.E.2d 814, 821 (Ind. 2009) (addressing the Uniform Interstate Family Support Act).

71. UNIF. PRINCIPAL & INCOME ACT § 103 cmt.

72. IND. CODE § 30-2-14-14(a)(2).

73. *Id.* § 30-2-14-14(a)(3).

74. *Id.* § 30-2-14-14(b).

75. UNIF. PRINCIPAL & INCOME ACT § 104 cmt.; *see also* RESTATEMENT (THIRD) OF TRUSTS § 79 gen. cmt. a (2007) (A trustee has a duty to impartially administer a trust "in balancing the naturally conflicting concerns of life and remainder beneficiaries . . . . The duty of impartiality is applicable to . . . . the allocation of receipts and expenditures between principal and income accounts . . . , especially as fiduciary discretion, or the making of adjustments . . . , may be involved . . . .").

the income and remainder beneficiaries after satisfying these external threshold conditions, then the inquiry stops. However, if, after satisfying the external conditions, the trustee determines that partiality is being shown to one class of beneficiary over another, then the trustee must determine whether the internal threshold conditions of the Indiana Act are met.

In general, a NIMCRUT trustee should be able to satisfy the external threshold conditions. A trustee prudently overseeing the every day administration of a NIMCRUT will, as a matter of course, apply any discretionary and non-discretionary powers in the administration of the trust. Similarly, where the trust agreement is silent as to the proper allocation of receipts and disbursements between principal and income, the Indiana Act is the standard that guides the decisions of a prudent trustee with regard to the administration of principal and income matters.[76] The more difficult question is whether the trustee has, in fact, impartially administered the trust. What is "fair and reasonable" is a highly fact-sensitive determination that is likely to be colored by one's perspective. An income beneficiary whose trust distributions are less than expected will almost certainly have a different perspective from that of a charitable remainder beneficiary with its own expectancy of what it should receive when the trust's term expires.

B. Internal Threshold Conditions

In addition to the external threshold conditions described above, Indiana Code section 30-2-14-15(a) imposes two internal threshold conditions that must be met before an exercise of the power to adjust is proper. First, the trustee must "invest[] and manage[] [the] trust assets as a prudent investor."[77] The Comments state that this "condition will be met [where] the prudent investor rule applies because the Uniform [Prudent Investor] Act . . . has been enacted."[78] For a NIMCRUT administered under Indiana law, this condition is met by virtue of Indiana's adoption of the Indiana Uniform Prudent Investor Act.[79]

Second, by its terms, the trust agreement must "describe the amount that may or must be distributed to a beneficiary by referring to the trust's income."[80] The Comments state that this requirement is met

> when the terms of the trust require all of the "income" to be distributed at regular intervals; or when the terms of the trust require a trustee to distribute all of the income, but permit the trustee to decide how much to distribute to each member of a class of beneficiaries; or when the terms

---

76. *See* IND. CODE § 30-2-14-14(a)(1).

77. *Id.* § 30-2-14-15(a)(1).

78. UNIF. PRINCIPAL & INCOME ACT § 104 cmt.

79. Ind. Unif. Prudent Investor Act, Pub. L. No. 137-1999, § 3, 1999 Ind. Acts 756 (codified at IND. CODE §§ 30-4-3.5-1 to -13). Note that Indiana's adoption of UPIA may satisfy this condition in an objective sense, but whether the trustee is actually investing as a prudent investor is an individualized determination based on the trustee's actions in a particular case.

80. IND. CODE § 30-2-14-15(a)(2).

> of a trust provide that the beneficiary shall receive the greater of the trust accounting income and a fixed dollar amount (an annuity), or of trust accounting income and a fractional share of the value of the trust assets (a unitrust amount). If the trust authorizes the trustee in its discretion to distribute the trust's income to the beneficiary or to accumulate some or all of the income, the condition will be met *because the terms of the trust do not permit the trustee to distribute more than the trust accounting income*.[81]

While this Comment does not specifically describe a NIMCRUT, it does not preclude a NIMCRUT from meeting the condition. The Comment clearly states that the condition is satisfied where a beneficiary is entitled to the *greater* of trust accounting income or an annuity or unitrust amount. But, a NIMCRUT income beneficiary is entitled to the *lesser* of trust accounting income *and* the annual unitrust amount. The resolution of this distinction is found in the last clause of the last sentence of the Comment: "because the terms of the trust do not permit the trustee to distribute more than the trust accounting income."[82] The implication of this clause is that a trust term permitting the distribution of more than trust accounting income does not "describe the amount that may or must be distributed to a beneficiary by referring to the trust's income"[83] and, therefore, the power to adjust is not available. Conversely, a trust term that "describes the amount that may or must be distributed" in terms of trust income, but does not permit the distribution of trust income in excess of a stated unitrust amount, still meets the requirement of the statute because it does not permit the distribution of *more* than the trust's accounting income. This latter condition is clearly the case with a NIMCRUT where trust accounting income, up to the annual unitrust amount, must be distributed.[84] Even if there is a make-up amount, no more than trust accounting income may be distributed.[85]

Furthermore, in the case of a NIMCRUT, if trust accounting income exceeds the annual unitrust amount, such excess must be retained and added to trust principal.[86] The Comment does not address the case where such retention is mandatory, but the last sentence of the Comment does speak approvingly of a discretionary power to retain income so long as "the terms of the trust do not permit the trustee to distribute more than the trust accounting income."[87] Following the same rationale stated in the last clause of the Comment, there is little reason to distinguish between a NIMCRUT's mandatory retention of income and a discretionary retention of income coupled with a requirement that no more than trust accounting income be distributed. In neither case may a trustee

---

81. UNIF. PRINCIPAL & INCOME ACT § 104 cmt. (emphasis added).
82. *Id.*
83. IND. CODE § 30-2-14-15(a)(2).
84. *See supra* Part I.A.2.
85. *See supra* Part I.A.2.
86. *See supra* Part I.A.2. and note 15.
87. UNIF. PRINCIPAL & INCOME ACT § 104 cmt.

distribute more than the trust accounting income. Therefore, a qualified NIMCRUT should satisfy the requirement that the trust agreement "describe the amount that may or must be distributed to a beneficiary by referring to the trust's income."[88]

C. Statutory Prohibitions on the Exercise of the Power to Adjust

Indiana Code section 30-2-14-15(c) contains seven statutory prohibitions that limit a fiduciary's exercise of the power to adjust. Of these seven, four are relevant to a NIMCRUT.[89]

1. Prohibition on Exercise Where Trustee Is a Beneficiary.—Indiana Code section 30-2-14-15(c)(7) states that the power to adjust may not be exercised "if the trustee is a beneficiary of the trust."[90] Thus, if a NIMCRUT trustee is also a beneficiary of the trust (including a charitable remainder beneficiary) then that beneficiary-trustee is prohibited from exercising the power to adjust. However, the Indiana Act provides that a co-trustee who is not a beneficiary may exercise the power unless prohibited from doing so by the terms of the trust.[91] Thus, if the limitation of Indiana Code section 30-2-14-15(c)(7) applies, it may be possible to appoint a co-trustee who is not a beneficiary to exercise the power to adjust.[92]

2. Grantor Trust Prohibition.—Indiana Code section 30-2-14-15(c)(5) prohibits a trustee from

> possessing or exercising the power to make an adjustment [if this] causes an individual to be treated as the owner of all or part of the trust for income tax purposes, and the individual would not be treated as the owner if the trustee did not possess the power to make an adjustment.[93]

This prohibition is intended to prevent the availability or exercise of the power to adjust from causing a trust that is otherwise a non-grantor trust from becoming

---

88. IND. CODE § 30-2-14-15(a)(2).

89. The three that are not relevant are IND. CODE § 30-2-14-15(c)(1) (prohibiting an exercise of the power to adjust that would adversely impact a claimed marital deduction); IND. CODE § 30-2-14-15(c)(2) (prohibiting an exercise of the power to adjust that would adversely impact a claimed gift tax exclusion); and IND. CODE § 30-2-14-15(c)(6) (prohibiting an exercise of the power to adjust that would cause the corpus of the trust to be included in an individual's estate).

90. IND. CODE § 30-2-14-15(c)(7).

91. *Id.* § 30-2-14-15(d).

92. Indiana did not adopt the provision in the UPAIA 1997 that prohibited an exercise of the power to adjust by a non-beneficiary trustee where the adjustment would benefit the trustee indirectly. This raises the question of whether a non-beneficiary co-trustee could be the spouse or child of a beneficiary. If such a person is appointed co-trustee, the trustee's duty of loyalty, described in the *Restatement (Third) of Trusts* § 78 as the "duty to administer the trust solely in the interest of the beneficiaries," probably precludes such a co-trustee from acting improperly to his or her own benefit. RESTATEMENT (THIRD) OF TRUSTS § 78.

93. IND. CODE § 30-2-14-15(c)(5).

a grantor trust.[94]

A qualified NIMCRUT is not a grantor trust.[95] For example, notwithstanding the general rule that a retained right to income from a trust will cause the grantor to be treated as the owner of the trust,[96] by rule a NIMCRUT is not a grantor trust simply because the grantor or the grantor's spouse are income beneficiaries of the trust.[97] However, a trust, including a NIMCRUT, is a grantor trust where "the beneficial enjoyment of the corpus or the income therefrom is subject to a power of disposition, exercisable by the grantor or a nonadverse party, or both, without the approval or consent of any adverse party."[98]

Where the power to adjust is exercisable by the grantor or a nonadverse party, or both, it is likely that the power to adjust is a power of disposition over trust income that would cause the trust to be a grantor trust. The power to adjust can only be exercised where the amount to be distributed to income beneficiaries is described by reference to trust income. Moreover, any adjustment between income and principal necessarily has a direct effect on the amount of income to be distributed. As a power that necessarily affects the disposition of trust income, it is probable that the IRS would characterize the power to adjust as a "power of disposition" controlling the "beneficial enjoyment" of the income of the trust. Therefore, even if the terms of the trust agreement do not limit the grantor's exercise of the power to adjust, a grantor who is the trustee of a NIMCRUT should not exercise the power to adjust, but rather should defer such an exercise to a co-trustee who is neither a grantor nor a beneficiary.

3. Prohibition on Changing a Fixed Annuity or Unitrust Amount.—Indiana Code section 30-2-14-15(c)(3) prohibits an exercise of the power to adjust "that changes the amount payable to a beneficiary as a fixed annuity or a fixed fraction of the value of the trust assets."[99] Referring to this provision, the Comments state that "a beneficiary's right to receive a fixed annuity or a fixed fraction of the value of a trust's assets is not subject to [the power to adjust]."[100] One commentator suggests that this Comment "seem[s] to indicate that [UPAIA 1997

---

94. The term "grantor trust" is derived from I.R.C. § 671 which provides that "[w]here . . . the *grantor* . . . [is] treated as the owner of any portion of a trust, there shall then be included in computing the taxable income and credits of the grantor . . . those items of income, deductions, and credits against tax of the trust which are attributable to that portion of the trust . . . ." (emphasis added). I.R.C. § 671 (2006).

95. *See* Treas. Reg. § 1.664-1(a)(4) (1972).

96. I.R.C. § 677(a) (2006) ("The grantor shall be treated as the owner of any portion of a trust . . . whose income . . . is, or . . . may be-- (1) distributed to the grantor or the grantor's spouse . . . .").

97. Treas. Reg. § 1.664-1(a)(4) ("[N]either the grantor nor his spouse shall be treated as the owner of the trust . . . merely because the grantor or his spouse is named as a recipient.").

98. I.R.C. § 674(a) (2006).

99. IND. CODE § 30-2-14-15(c)(3) (2011).

100. UNIF. PRINCIPAL & INCOME ACT § 104 cmt. (amended 2008), 7A pt. III U.L.A. 363, 438 (Supp. 2010).

section] 104(c)(3)[[101]] governs charitable remainder trusts, *including the net-income-with-makeup charitable remainder unitrust*."[102]

It seems clear that this section works to prohibit an exercise of the power to adjust in the case of a CRAT or standard CRUT which describe the amount due to the income beneficiary as a fixed annuity or fixed fraction (or percentage) of the value of the trust assets, respectively. However, it is less clear whether this prohibition should apply to a NIMCRUT. A NIMCRUT requires that a trustee distribute to the income beneficiary the *lesser* of the trust's accounting income *and* an annual unitrust amount.[103] An exercise of the power to adjust by a NIMCRUT trustee is not a change to the annual unitrust amount, but rather a change to the amount of the trust accounting income. For this reason, it seems unlikely that the prohibition applies to NIMCRUTs.

4. Prohibition on Adjustments from Amounts Permanently Set Aside for Charitable Purposes.—It is the fourth prohibition that poses the greatest test to the propriety of the exercise of the power to adjust in the case of a NIMCRUT. Indiana Code section 30-2-14-15(c)(4) prohibits an adjustment "from any amount that is permanently set aside for charitable purposes under a will or the terms of a trust unless both income and principal are so set aside."[104] Of specific concern is whether an exercise of the power to adjust in the context of a NIMCRUT is "an adjustment . . . from [an] amount . . . permanently set aside for charitable purposes."[105]

a. In a NIMCRUT, is there an amount permanently set aside for charitable purposes?—As previously noted, a transferor who transfers property to a qualified NIMCRUT receives a charitable income tax deduction and charitable transfer tax deduction equal to the present value of the remainder interest in the trust.[106] To qualify for the income and transfer tax deductions, the remainder interest must be irrevocably designated to, or for the use of, one or more charitable organizations.[107] Therefore, at least facially, a NIMCRUT's charitable remainder interest is an amount permanently set aside for charitable purposes. Further, because a NIMCRUT defines the interest of the income beneficiaries in terms of income and the only other available interest is the right to the trust principal, at any given point in time, a NIMCRUT's charitable remainder is, at least theoretically, equivalent to the NIMCRUT's principal. Under this analytical framework, an adjustment from principal to income would therefore be an adjustment from an amount permanently set aside for charitable purposes.

However, this conclusion is arguably incorrect for at least three reasons. First, the computation of the present value of the remainder interest (the

101. UPAIA 1997 § 104(c)(3) corresponds to IND. CODE § 30-2-14-15(c)(3).

102. CHRISTOPHER P. CLINE, THE LAW OF TRUSTEE INVESTMENTS 82 (2009) (emphasis added).

103. *See supra* Part I.A.2.

104. IND. CODE § 30-2-14-15(c)(4).

105. *Id.*

106. *See supra* notes 18-21 and accompanying text.

107. *See* I.R.C. § 664(d)(1)(C) (2006) (describing a charitable remainder annuity trust); I.R.C. § 664(d)(2)(C) (describing a charitable remainder unitrust).

deductible amount) casts doubt on the assumption that the amount a NIMCRUT donor intended to permanently set aside for charitable purposes is always equal to the NIMCRUT's principal. Second, an appropriate exercise of the power to adjust probably will not cause a NIMCRUT to lose tax benefits obtained when it was created. Finally, denying NIMCRUT trustees access to the power to adjust is contrary to the policy underlying the power.

b. The computation of the present value of the remainder interest and its relationship to the amount permanently set aside for charitable purposes.—The prescribed method for computing the present value of the remainder interest of a standard CRUT and a NIMCRUT are the same.[108] The computation requires seven inputs: (1) the date of transfer;[109] (2) the age of each beneficiary;[110] (3) the unitrust percentage;[111] (4) the applicable I.R.C. § 7520 rate;[112] (5) the frequency of the required unitrust distributions (e.g., monthly, quarterly, semi-annual, or annual distribution periods);[113] (6) the timing of distributions within the distribution period (e.g., whether distributions occur at the beginning or end of the distribution period);[114] and (7) the date on which trust property is to be valued relative to the distribution year (e.g., at the beginning or the end of the period).[115]

As discussed above, the computation of the present value of the remainder interest for a standard CRUT and a NIMCRUT is the same.[116] The computation assumes that the annual unitrust amount is distributed in full each year; it makes no adjustment for the possibility that, in the case of a NIMCRUT, trust accounting income may be less than the annual unitrust amount. The computed amount is the basis of the charitable income tax and transfer tax deductions. As such, it is arguable that this amount reflects the amount permanently set aside for charitable purposes.

For example, assume that Fred, age 68, and Jane, age 62, transfer $1,000,000 to a NIMCRUT on January 15, 2011. Further assume that the trust's stated unitrust percentage is five percent. Based on a transfer date of January 15, 2011,

108. *See supra* Part I.A.2. and note 27 and accompanying text; *see also* Treas. Reg. § 1.664-4 (2009) (describing the computation of the remainder interest in a charitable remainder unitrust).

109. The date of transfer establishes the "valuation date" used to determine the applicable life table as required by Treas. Reg. § 1.664-4(a)(1) (2011) and the I.R.C. § 7520 rate to be used as required by Treas. Reg. § 1.664-4(a)(2) (2009).

110. The age of each beneficiary is an indirect input in computation. A beneficiary's age is relevant to the selection of the remainder factor from the appropriate table as described in IRS Publ'n 1457, Actuarial Valuations (5-2009), *available at* http://www.irs.gov/pub/irs-pdf/p1457.pdf.

111. *See* Treas. Reg. § 1.664-4(e)(3) (2009).

112. Each month the IRS publishes an interest rate used to determine "the value of any annuity, any interest for life or a term of years, or any remainder or reversionary interest." I.R.C. § 7520(a) (2006); *see also* Rev. Rul. 2011-6, 2011-10 I.R.B. 537, 538 tbl. 4 (providing an example of the monthly publication of the I.R.C. § 7520 rate).

113. *See* Treas. Reg. § 1.664-4(e)(3).

114. *See id.* § 1.664-4(a)(3).

115. *See id.* § 1.664-4(e)(3).

116. *See supra* Part I.A.2. and note 27 and accompanying text.

the best[117] section 7520 rate allowed is 2.4%. Under these facts, the present value of the remainder interest (and the putative amount permanently set aside for charitable purposes) is $365,700.[118] As previously explained, this computed amount is the same regardless of whether Fred and Jane's trust is a standard CRUT or a NIMCRUT.

What is the effect if the NIMCRUT's investment portfolio fails to produce trust accounting income greater than or equal to the annual unitrust amount but the overall investment return is greater than or equal to the annual unitrust amount?[119] Under this scenario, each year the income beneficiaries will receive less than the annual unitrust amount while, at the expiration of the trust's term, the charitable remainder beneficiaries will receive a larger charitable remainder—more than the putative amount permanently set aside for charitable purposes. This increase in the charitable remainder would be reflected in an increase in the value of the trust's principal over time. Assuming the amount originally computed as the present value of the remainder interest represents the amount permanently set aside for charitable purposes, the difference between the trust principal and the computed present value represents an amount *not* permanently set aside for charitable purposes, an amount from which adjustments could be made.

It is possible to illustrate this difference by substituting in the present value computation, in place of the unitrust percentage, an estimate of the real rate of return the trust will receive in the form of trust accounting income while holding all other factors constant. Assuming that Fred and Jane's NIMCRUT will receive a real rate of return of trust accounting income of three percent, the resulting estimate of the present value of the remainder interest is $541,824, a difference of $176,124. It is from this difference that an adjustment between principal and income would be made under the power to adjust. Assuming this difference is over and above the amount permanently set aside for charitable purposes (as measured by the computed present value of the remainder interest), an adjustment

---

117. Internal Revenue Code section 7520(a) provides in part that "[i]f an income, estate, or gift tax charitable contribution is allowable for any part of the property transferred, the taxpayer may elect to use such Federal midterm rate *for either of the [two] months preceding the month in which the valuation date falls* . . . ." I.R.C. § 7520(a) (emphasis added). When computing the present value of the remainder interest for unitrusts requiring distributions more frequently than annually, a higher I.R.C. § 7520 rate will produce a slightly larger result.

118. Further assume that income beneficiary distributions are paid at the end of each calendar quarter and that the trust property is valued on the first day of each year.

119. If the trust accounting income rate of return exceeds the stated unitrust percentage, then there is no distinction between the expected results of a standard CRUT and a NIMCRUT. This is because, in either case, the required distributions are exactly equal to the annual unitrust amount and the income in excess of the annual unitrust amount is retained in the trust for the benefit of the charitable remainder beneficiary. Conversely, if the trust accounting income rate of return *and* the overall rate of return are less than the stated unitrust percentage, or perhaps even reflect a negative return, there is little or no amount from which to make an adjustment and the question of the propriety of an adjustment is in all probability moot.

from this amount would not be subject to the limitation of Indiana Code section 30-2-14-15(c)(4).

In short, the computed present value of the remainder interest used to substantiate the charitable income tax and transfer tax deductions is a reasonable estimate of the amount the donor intended to "permanently set aside for charitable purposes." In Indiana, "[t]he primary goal in construing a trust document is to ascertain and effectuate the intent of the settlor."[120] The charitable tax deductions a donor receives upon the transfer of property to a charitable remainder trust suggests that it was the donor's intent to permanently set aside for charitable purposes a specific amount that, depending on the performance of the trust's investment portfolio, may be less than the trust principal. Therefore, an adjustment from principal amounts in excess of the amount permanently set aside for charitable purposes should not be treated as an adjustment from an amount permanently set aside for charitable purposes. A contrary interpretation is that altruistic donors create NIMCRUTs with the expectation that the charitable remainder beneficiary will receive the excess described above.

An additional consideration in this analysis is the role of the unitrust amount in limiting the potential for adjustment. Because a NIMCRUT trustee can never distribute more than the annual unitrust amount plus the makeup amount,[121] there is an upper bound to the amount of an adjustment from principal to income that will be payable to the NIMCRUT's income beneficiaries. This upper bound serves to limit a trustee's ability to adjust from an amount permanently set aside for charitable purposes. Consequently the potential for abuse is significantly diminished.

c. An appropriate exercise of the power to adjust will not cause the loss of federal or state tax benefits obtained when the trust was created.—The Comments provide that "[t]he purpose of [UPAIA 1997 sections 104(c)(1)] through (4) is to preserve tax benefits that may have been an important purpose for creating the trust."[122] The implication is that, if upon enactment of the UPAIA 1997 by a state, the availability of the power to adjust could cause a trust to lose a tax benefit obtained when the trust was created, then, to preserve that tax benefit, the trustee of that trust is denied access to the power to adjust. In the case of a NIMCRUT, three such tax benefits are relevant: (1) the donor's charitable income tax deduction, (2) the charitable transfer tax deduction, and (3) tax-exempt status during the administration of the trust.

With respect to federal tax benefits, the IRS has approved the use of the power to adjust. As noted above, the starting point in computing trust accounting income is the I.R.C. § 643(b) and its applicable regulations.[123] In response to the publication of the UPAIA 1997 and its subsequent adoption by several states, on January 2, 2004 the IRS released final regulations (the 2004 Regulations),

---

120. Malachowski v. Bank One, Indianapolis, 590 N.E.2d 559, 565 (Ind. 1992).

121. *See supra* Part I.A.2.

122. UNIF. PRINCIPAL & INCOME ACT § 104 cmt. (amended 2008), 7A pt. III U.L.A. 363, 438 (Supp. 2010).

123. *See supra* Part I.A.4.

amending the existing regulations under I.R.C. § 643(b), designed to

> recognize that state statutes are in the process of changing traditional concepts of income and principal in response to investment strategies that seek total positive return on trust assets. These statutes are designed to ensure that, when a trust invests in assets that may generate little traditional income (including dividends, interest, and rents), the income and remainder beneficiaries are allocated reasonable amounts of the total return of the trust (including both traditional income and capital appreciation of trust assets) so that both classes of beneficiaries are treated impartially.[124]

Consistent with this purpose, Treasury Regulation section 1.643(b)-1 was amended by adding the following language:

> Allocations pursuant to methods prescribed by [state principal and income] statutes for apportioning the total return of a trust between income and principal *will be respected* regardless of whether the trust provides that the income must be distributed to one or more beneficiaries or may be accumulated in whole or in part . . . .[125]

However, the 2004 Regulations included an express provision permitting a NIMCRUT trustee to possess a discretionary power to apportion realized capital gains between income and principal "under the terms of the governing instrument but only to the extent that the state statute permits the trustee to make adjustments between income and principal to treat beneficiaries impartially."[126] Because this language appears in the regulation, is couched in language similar to the UPAIA 1997's power to adjust, and is the only such language that appears in the charitable remainder trust regulations, it is tempting to interpret this regulation as the extent of the IRS's permissive attitude toward the power to adjust.[127] That is, to assume that the IRS recognizes the valid existence of such a power *only* for realized capital gains and *only* if the power is granted in the governing instrument and *only* if there is a state statute permitting adjustments between income and principal under a duty of impartiality. This narrow view is contrary to the express language of Treasury Regulation section 1.643(b)-1 recited above, which imposes no limitation on the use of the power to adjust.

Furthermore, a careful reading of the preamble to the 2004 Regulations indicates that it is unlikely that this language from the charitable remainder trust regulations circumscribes the availability of the power to adjust. Speaking in the context of a discretionary power to apportion realized capital gains between

124. T.D. 9102, 2004-1 C.B. 366, 366-67.

125. *Id.* at 373-74 (emphasis added).

126. Treas. Reg. § 1.664-3(a)(1)(i)(b)(3) (as amended in 2004); *see also* T.D. 9102, 2004-1 C.B. 366, 374.

127. *See, e.g.*, Medlin, *supra* note 44, at 743-44 ("The revised Treasury regulations allow the allocation of capital gains, traditionally principal, to income for charitable remainder unitrusts, but only if both the trust instrument and applicable state law allow such an equitable adjustment.").

income and principal, the IRS stated, "The provision in the . . . regulations [that describes the apportionment of capital gains between income and principal] *has no effect on the determination of trust accounting income under applicable state law that grants the trustee a power to reasonably apportion the total return of the trust*."[128] This statement from the preamble, coupled with the express language of Treasury Regulation section 1.643(b)-1, clearly indicates that, as long as a NIMCRUT trustee observes the formalities imposed by the applicable state's principal and income statute, the IRS will respect the trustee's exercise of the power to adjust. As a result, a reasonable exercise of the power to adjust under a state-law-imposed duty of impartiality should not cause a NIMCRUT to lose any federal tax benefits obtained when the trust was created or to lose the trust's exempt status during the trust term.

The same result is obtained when considering the impact on state tax benefits. The starting point for computing the tax benefits conferred by the State of Indiana on a qualified NIMCRUT is federal adjusted gross income as modified by state statute.[129] Indiana does not allow a taxpayer to claim a charitable deduction in arriving at taxable income.[130] Therefore, there is no direct Indiana tax benefit to a donor who creates and transfers property to a NIMCRUT.[131] Because a qualified charitable remainder trust is exempt from federal income tax,[132] its federal adjusted gross income is zero dollars. As a result, a charitable remainder trust is effectively exempt from income tax in Indiana so long as its federal tax exemption is preserved. Furthermore, the charitable remainder interest in a qualifying charitable remainder trust is exempt from Indiana's inheritance tax.[133] Therefore, an exercise of the power to adjust that does not cause an otherwise qualified charitable remainder trust to lose its exemption from federal income tax granted by I.R.C. § 664(a), or to lose the estate tax deduction granted by I.R.C. § 2055(a), will not cause the trust to lose any state tax benefits obtained when the trust was created or to lose the trust's exempt status during the trust term.

One commentator has noted a potential problem with applying this tax-benefit-preservation rationale to circumvent the limitation of UPAIA 1997 section 104(c) (and its Indiana counterpart Indiana Code section 30-2-14-15(c)(4)): The plain language of the statute fails to incorporate the tax-benefit-preservation rationale.[134] Absent a reference to the tax benefits to be preserved in the statute,

128. T.D. 9102, 2004-1 C.B. 366, 369 (emphasis added).

129. *See* IND. CODE §§ 6-3-2-1, 6-3-1-3.5 (2011); *see also id.* § 6-3-1-14 ("The term 'person' means an individual, trust, or estate.").

130. *See id.* §§ 6-3-2-1 to -21.7 (describing amounts includible in taxable income and amounts deductible in arriving at taxable income).

131. However, because the excess of the value over the adjusted tax basis (i.e., the unrealized capital gain) of property transferred to a NIMCRUT is excluded from federal adjusted gross income and consequently excluded from Indiana taxable income, there is an indirect tax avoidance benefit.

132. *See supra* note 16.

133. IND. CODE § 6-4.1-3-1 (2011) ("Each transfer described in section 2055(a) of the Internal Revenue Code is exempt from the inheritance tax."); *see also supra* note 20.

134. *See* Medlin, *supra* note 44, at 744 ("[I]t is arguable that state law, based on the UPAIA

it is uncertain that a court interpreting the statute would look beyond the plain language of the statute to the Comments.[135] Had the drafters of the UPAIA 1997 had the benefit of the IRS's commentary accompanying the 2004 Regulations when the UPAIA 1997 was promulgated, perhaps the prohibition on the exercise of the power to adjust in the case of amounts permanently set aside for charitable purposes would have been tempered to make it clear that the power to adjust should be exercisable by NIMCRUT trustees.

d. Denying NIMCRUT trustees access to the power to adjust is contrary to the policy underlying the power.—As discussed above, the UPIA and the UPAIA 1997 are two halves of a whole.[136] The UPIA creates a duty to invest as a prudent investor, seeking total return without regard to the categories of income that make up the return, while the UPAIA 1997, through the mechanism of the power to adjust, empowers a trustee to ensure impartial treatment of income and remainder beneficiaries should inequities arise in the pursuit of optimal returns.[137] If the prohibition of Indiana Code section 30-2-14-15(c)(4) includes NIMCRUTs, then the statutory scheme arguably creates a direct conflict between the trustee's duty to invest as a prudent investor and the trustee's duty of impartiality; a conflict that is not diminished by the fact that one class of beneficiary is made up of one or more charitable organizations.

An Indiana NIMCRUT trustee has a duty to invest as a prudent investor.[138] Further, a trustee is also required to invest impartially, "taking into account any differing interest of the beneficiaries."[139] The trustee's investment decisions are "evaluated . . . in the context of the trust portfolio as a whole and as a part of an overall investment strategy having risk and return objectives reasonably suited to

---

[1997], . . . prohibits a reallocation from principal to income because section 104(c)(4) does not use tax terms.").

135. In Indiana, statutes that are "clear and unambiguous . . . [are construed by taking] words and phrases . . . in their plain, ordinary, and usual sense." Basileh v. Alghusain, 912 N.E.2d 814, 821 (Ind. 2009). If a "statute is susceptible to more than one interpretation it is deemed ambiguous and thus open to judicial construction." *Id*. When construing an ambiguous statute, a primary rule of statutory construction "is to determine, give effect to, and implement the intent of the Legislature." *Id*. Where the statute in question is based on a uniform act, "[t]he comments to [the] uniform act are indicative of the Legislature's intent." *Id*. However, "[i]t is always presumed, in regard to a statute, that no absurd or unreasonable result was intended by the legislature." Simmons v. State, 129 N.E.2d 121, 125 (Ind. 1955) (quoting Marks v. State, 40 N.E.2d 108, 111 (Ind. 1942)). Finally, when "applying and construing [the Indiana Act], consideration must be given to the need to promote uniformity of the law . . . among states that enact [the UPAIA 1997]." IND. CODE § 30-2-14-44 (2011); *see also* 2A NORMAN J. SINGER & J.D. SHAMBIE SINGER, STATUTES AND STATUTORY CONSTRUCTION § 48:11 (7th ed. 2007) ("Legislation adopted from a Uniform Act should be construed in a manner which makes it operate uniformly with other states that have adopted the Uniform Act.").

136. *See* discussion *supra* Part I.B.

137. *See supra* Part I.B.

138. IND. CODE § 30-4-3.5-1(a) (2011); *see also id.* § 30-4-3-6(b)(1) (2011).

139. *Id.* § 30-4-3.5-6.

the trust."[140] However, as previously noted, such an investment strategy is indifferent to the categories of income that make up the returns produced.[141] In fact, the pursuit of total return using "an overall investment strategy having risk and return objectives reasonably suited to the trust"[142] may result in a portfolio that produces no trust accounting income and all capital appreciation (or loss), or vice versa.

Under a scenario where the trustee's compliance with the duty to invest as a prudent investor produces minimal trust accounting income under the traditional categories of income while producing significant capital appreciation, the result is a windfall to one class of beneficiary at the expense of the other. In the alternative, if a NIMCRUT trustee, considering the prohibition on the exercise of the power to adjust imposed by Indiana Code section 30-2-14-15(c)(4), determines that "risk and return objectives reasonably suited to" a NIMCRUT requires investing in a fashion that sacrifices total return in favor of an allocation which necessarily disappoints equally,[143] then neither class of beneficiary is served and the amount that ultimately passes to the charitable remainder beneficiaries would, in all likelihood, be diminished. This diminishment may even be greater than the effect of an adjustment between principal and income under the power to adjust.

In summary, a NIMCRUT trustee's adjustment from principal to income is not barred per se by Indiana Code section 30-2-14-15(c)(4). If the present value of the charitable remainder interest when a NIMCRUT is funded defines the amount permanently set aside for charitable purposes, then, if the total return from the trust's investment portfolio exceeds the annual unitrust amount, and the trust accounting income is less than the annual unitrust amount, the value of the trust principal will exceed the amount permanently set aside for charitable purposes. An adjustment from this excess, up to the annual unitrust amount, is an adjustment from an amount that is *not* permanently set aside from charitable purposes. In addition, an exercise of the power to adjust will not result in the loss of federal or state tax benefits. Finally, a NIMCRUT trustee denied access to the power to adjust may be forced to invest in less productive investment portfolios to the mutual detriment of both the income and remainder beneficiaries.

III. WHAT IS NEEDED TO BREATHE LIFE INTO THE POWER TO ADJUST FOR NIMCRUTS?: LEGISLATIVE INTERVENTION

The foregoing analysis notwithstanding, the lack of explicit statutory authority or case law (in Indiana or any other state) supporting an exercise of the power to adjust by a NIMCRUT trustee will give many trustees pause. Therefore, to breathe life into the power to adjust, a legislative change may be necessary to give Indiana NIMCRUT trustees the confidence to exercise the power to adjust.

140. *Id.* § 30-4-3.5-2(b).

141. *See supra* Part I.B.1.

142. UNIF. PRUDENT INVESTOR ACT § 2(b), 7B U.L.A. 1, 20 (2006).

143. *See supra* note 57 and accompanying text.

To that end, the author proposes that the Indiana legislature amend Indiana Code section 30-2-14-15(c)(4) to read:

> (c) A trustee may not make an adjustment: . . . (4) from any amount that is permanently set aside for charitable purposes under a will or the terms of a trust *(A) for which a deduction or exemption from tax would be allowed under Title 6 of the Indiana Code or sections 170, 664, 2055, or 2522 of the Internal Revenue Code if the trustee did not have the power to make the adjustment; or (B)* unless both the income and principal are so set aside. (Proposed change appears in italicized text)

The proposed change is appropriate for four reasons. First, it is consistent with the tax-benefit-preservation rationale articulated by the Commissioners in the Comments. Second, it is consistent with the IRS's current position with respect to the power to adjust. Third, it resolves the conflict between a trustee's duty to administer a NIMCRUT impartially between income beneficiaries and remainder beneficiaries and the trustee's duty to invest the trust prudently. Finally, it is consistent with the favorable stance Indiana has taken toward NIMCRUTs in other modifications made when the Indiana Act was adopted.

A. The Proposed Change Is Consistent with the Tax-Benefit-Preservation Rationale Found in the Comments

As discussed above, the Comments state that "[t]he purpose of subsection[] [104(c)](4) is to preserve tax benefits that may have been an important purpose for creating the trust."[144] The drafters of the UPAIA 1997 explicitly included language incorporating this tax-benefit rationale into two other subsections under UPAIA 1997 section 104(c). UPAIA 1997 section 104(c)(1) states that

> (c) A trustee may not make an adjustment:
> (1) that diminishes the income interest in a trust that requires all of the income to be paid at least annually to a spouse *and for which an estate tax or gift tax marital deduction would be allowed, in whole or in part, if the trustee did not have the power to make the adjustment*.[145]

The highlighted text precludes the use of the power to adjust only if a specific tax benefit, e.g., the estate tax or gift tax marital deduction, would be lost if the trustee possesses the power to adjust.[146] By drafting the statute in this manner, the drafters left open the possibility that the power to adjust would be available in the context of a marital deduction trust if the IRS ruled that such an exercise would not result in the loss of the marital deduction.

Similarly, UPAIA 1997 section 104(c)(2) states that "(c) A trustee may not

144. UNIF. PRINCIPAL & INCOME ACT § 104 cmt. (amended 2008), 7A pt. III U.L.A. 363, 438 (Supp. 2010); *see* discussion *supra* at Part II.C.4.(a)(2).

145. UNIF. PRINCIPAL & INCOME ACT § 104(c)(1) (emphasis added); *see also* IND. CODE § 30-2-14-15(c)(1) (2011).

146. *See* Medlin, *supra* note 44, at 744.

make an adjustment: . . . (2) that reduces the actuarial value of the income interest in a trust to which a person transfers property with *the intent to qualify for a gift tax exclusion*."[147] In this case, the statute only precludes the exercise of the power to adjust if the creator of the trust intended for the transfer of property to qualify for a gift tax exclusion. Once again the drafters of the UPAIA 1997 anchored the prohibition on the exercise of the power to adjust to the preservation of an expected tax benefit. For example, in the case where a settlor transferred property to a trust with no intent to qualify for a gift tax exclusion (perhaps because the settlor had already exhausted all available gift tax exclusions), the use of the power to adjust would not be precluded.

The proposed amendment explicitly imports the tax-benefit-preservation rationale that underlies UPAIA 1997 sections 104(c)(1) and (2) into Indiana Code section 30-2-14-15(c)(4). The proposed amendment precludes the exercise of the power to adjust only if the authorities that grant a tax benefit to be preserved determine that the availability of the power to adjust would result in the loss of the tax benefit.

B. The Proposed Change Is Consistent with the IRS's Current Position with Respect to the Power to Adjust

As discussed above, the IRS has approved the exercise of the power to adjust by NIMCRUT trustees.[148] However, some commentators have suggested that this permissive stance is limited to situations involving a discretionary power to apportion realized capital gains between income and principal, and then only when the discretionary power is "granted to the trustee under the terms of the governing instrument but only to the extent that the state statute permits the trustee to make adjustments between income and principal to treat beneficiaries impartially."[149] But this narrow view is at odds with the language of Treasury Regulation section 1.643(b)-1 and the broad pronouncement in the preamble to the 2004 Regulations that the provision limiting the apportionment of realized capital gains "has no effect on the determination of trust accounting income under applicable state law that grants the trustee a power to reasonably apportion the total return of the trust."[150]

Both the narrow interpretation of Treasury Regulation section 1.664-3(a)(1)(i)(b)(3) and the broader statement of Treasury Regulation section 1.643(b)-1 are currently at odds with a facial reading of the limitation imposed by Indiana Code section 30-2-14-15(c)(4). By tying the availability of the power to adjust to the IRS's determination of the effect of the power to adjust on tax benefits, the proposed amendment to Indiana Code section 30-2-14-15(c)(4)

---

147. UNIF. PRINCIPAL & INCOME ACT § 104(c)(2) (emphasis added); *see also* IND. CODE § 30-2-14-15(c)(2).

148. *See supra* Part II.C.4.(a)(2).

149. Treas. Reg. § 1.664-3(a)(1)(i)(b)(3) (as amended in 2004); *see supra* note 127 and the accompanying text.

150. T.D. 9102, 2004-1 C.B. 366, 369; *see also supra* notes 124-26 and accompanying text.

would permit the use of the power to adjust within the boundaries described by either the narrow interpretation or the broad interpretation.

C. The Proposed Change Resolves the Conflict Between a Trustee's Duty of Impartiality and the Trustee's Duty to Invest the Trust Prudently

As discussed above, Indiana Code section 30-2-14-15(c)(4) as currently written creates a facial conflict between a trustee's duty to treat the income and remainder beneficiaries of a NIMCRUT impartially and the trustee's duty to invest the trust prudently.[151] The proposed amendment eliminates this conflict by making the power to adjust available to NIMCRUT trustees until federal or state tax authorities (or legislatures) assert the preeminence of a tax policy favoring the protection of all of a NIMCRUT's principal.

D. The Proposed Change Is Consistent with Indiana's Favorable Stance Toward NIMCRUTs

The proposed amendment to Indiana Code section 30-2-14-15(c)(4) is consistent with the beneficial treatment Indiana has afforded NIMCRUTs historically. Indiana made two NIMCRUT-friendly changes to UPAIA 1997 when it enacted the Indiana Act in 2002.[152] The Indiana General Assembly amended UPAIA 1997 section 406 (relating to debt instruments) by adding the following subsection:

> (c) Notwithstanding any other provision of this section, when an obligation described in this section is held as an asset of a *charitable remainder trust*, an increase in the value of the obligation over the value of the obligation at the time of acquisition by the trust is distributable as income. For purposes of this subsection, the increase in value is available for distribution only when the trustee receives cash on account of the obligation.[153]

This provision ensures that the increase in value of an obligation, such as an original issue discount bond (e.g., a zero-coupon bond), is allocable to income, but only when the trustee receives cash on account of the obligation, typically as a result of the sale or maturity of the bond.

Similarly, the Indiana General Assembly amended UPAIA 1997 section 409 by adding the following subsection:

---

151. *See supra* Part II.C.4.(a)(3).

152. The author's employer, Renaissance Administration LLC (but not the author) played a role in recommending these changes to the Indiana General Assembly through the participation of its in-house counsel as a member of the Probate, Trust, and Real Property Section of the Indiana State Bar Association.

153. IND. CODE § 30-2-14-28(c) (emphasis added). The statute refers to "charitable remainder trusts" and not specifically to NIMCRUTs. However, because NIMCRUTs are the only variety of charitable remainder trust that determines income beneficiary payments by reference to trust income, this provision is effectively targeted at NIMCRUTs.

> (g) Notwithstanding any other provision of this section, when a private or commercial deferred annuity is held as an asset of a *charitable remainder trust*, an increase in the value of the obligation over the value of the obligation at the time of the acquisition by the trust is distributable as income. For purposes of this subsection, the increase in value is available for distribution only when the trustee exercises a right of withdrawal or otherwise receives cash on account of the obligation.[154]

This provision ensures that the increase in value of an obligation, such as a commercial deferred annuity, is allocable to income, but only when the trustee receives cash on account of the obligation. The proposed amendment is in line with the type of incremental benefits accorded charitable remainder trusts by these two enacted amendments.

In summary, the proposed amendment to Indiana Code section 30-2-14-15(c)(4) is consistent with the policies underlying the prohibition on adjustments from amounts permanently set aside for charitable purposes. Moreover the proposed amendment is in harmony with the IRS's position on an exercise of the power to adjust by a NIMCRUT trustee. Furthermore, the proposed amendment resolves a conflict in the duties a NIMCRUT trustee owes trust beneficiaries. Finally, the proposed amendment is consistent with Indiana's historically favorable treatment of NIMCRUTs.

CONCLUSION

Based on the foregoing analysis, an Indiana NIMCRUT trustee can exercise the power to adjust. However, the lack of explicit statutory authority or case law that explicitly addresses the use of the power to adjust by NIMCRUT trustees is likely to cause trustees to hesitate to exercise the power to adjust. The Indiana General Assembly should amend Indiana Code section 30-2-14-15(c)(4) to explicitly authorize the exercise of the power to adjust by NIMCRUT trustees. Indiana has historically shown leadership in effecting positive change in principal and income law as it relates to NIMCRUTs. The proposed amendment furthers Indiana's leadership in this arena. It is now up to the Indiana legislature to make it clear that the power to adjust is alive and well for Indiana's NIMCRUT trustees.

---

154. *Id.* § 30-2-14-31(g) (emphasis added).

NEW FAME IN A NEW BALLGAME: RIGHT OF PUBLICITY IN THE ERA OF INSTANT CELEBRITY

MICHAEL MULLINS*

INTRODUCTION

The right of publicity emerged as a distinct, named doctrine only as recently as 1953.[1] A philosophical cousin of the right to privacy, right of publicity doctrine asserts that an individual whose likeness has commercial value should have the exclusive, but assignable, right to control the commercial use of his likeness and reap the profits from its use.[2] All but necessarily, the right applies only to the well-known—only to celebrities.

The 1953 court decision that gave the right of publicity its name, *Haelan Laboratories, Inc. v. Topps Chewing Gum, Inc.*, involved a baseball player's conflicting contracts with rival baseball card manufacturers; the defendant had induced the player to allow it to use his likeness on its baseball cards when it was aware that the player had already signed a contract granting the plaintiff the exclusive use of his likeness on the plaintiff's baseball cards.[3] The court recognized that the injury suffered by the plaintiff—which was not loss of the use of the player's likeness, but rather loss of the value of the *exclusive* use of the player's likeness—was cognizable only if the right to use a person's likeness were assignable.[4] And for the right to be assignable, it had to be recognized as property, or at least something in the nature of property.[5] The court held that an individual has just such a quasi-property right to control the commercial use of his likeness, and dubbed the right the "right of publicity."[6]

In the decades since *Haelan Laboratories*, adoption of the right of publicity

---

* J.D. Candidate, 2012, Indiana University Robert H. McKinney School of Law; B.S., 1997, Interdisciplinary Engineering, Purdue University, West Lafayette, Indiana. I would like to thank Professor Jonathan Faber, whose course in this subject piqued my interest in it, and Mr. Robert Roach, my seventh and eighth grade English teacher, for making me a writer.

1. *See* Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866 (2d Cir. 1953).

2. Black's Law Dictionary defines the right of publicity as: "The right to control the use of one's own name, picture, or likeness and to prevent another from using it for commercial benefit without one's consent." BLACK'S LAW DICTIONARY 1439 (9th ed. 2009). In one of the most influential treatises on the subject, J. Thomas McCarthy describes the right more philosophically as: "[T]he inherent right of every human being to control the commercial use of his or her identity." 1 J. THOMAS MCCARTHY, THE RIGHTS OF PUBLICITY AND PRIVACY § 1.3 (2d ed. 2008).

3. *Haelan Labs., Inc.*, 202 F.2d at 867.

4. *Id*. at 868 ("This right of publicity would usually yield [prominent persons] no money unless it could be made the subject of an exclusive grant which barred any other advertiser from using their pictures.").

5. *Id*. ("Whether it be labelled [sic] a 'property' right is immaterial; for here, as often elsewhere, the tag 'property' simply symbolizes the fact that courts enforce a claim which has pecuniary worth.").

6. *Id*.

has been confined to state law—statutory law in some states, common law in others[7]—that is substantively different from state to state. This lack of uniformity has led some legal commentators to call for a federal right of publicity commensurate with federal protection for copyrights.[8] However, the right of publicity doctrine itself has not been uniformly accepted by legal theorists. The right of publicity has been attacked as needless—duplicative of the protections of copyright, trademark, and unfair competition law and of the false endorsement protections of the Lanham Act[9]—as constitutionally suspect,[10] or even as fundamentally adverse to society's interests.[11] Some commentators have called

---

7. *See Statutes*, RIGHT PUBLICITY, http://rightofpublicity.com/statutes (last visited March 1, 2011). Jonathan Faber, an attorney who specializes in right of publicity cases and teaches a course on the subject at Indiana University Robert H. McKinney School of Law, cites at least twelve states that statutorily confer some form of a right of publicity, but estimates that when states whose courts have recognized a common law right of publicity are included, only two states can be said to have no right of publicity at all. E-mail from Jonathan Faber, attorney, McNeely Stephenson Thopy & Harrold, to author (Dec. 9, 2010, 9:04 EDT) (on file with author).

8. *See, e.g.*, Eric J. Goodman, *A National Identity Crisis: The Need for a Federal Right of Publicity Statute*, 9 DEPAUL-LCA J. ART & ENT. L. & POL'Y 227 (1999); Richard S. Robinson, *Preemption, the Right of Publicity, and a New Federal Statute*, 16 CARDOZO ARTS & ENT. L.J. 183 (1998); Sean D. Whaley, *"I'm a Highway Star": An Outline for a Federal Right of Publicity*, 31 HASTINGS COMM. & ENT. L.J. 257 (2009). State enactment of a uniform law that is substantively identical in every state—like the Uniform Commercial Code—has been suggested as an alternative to a federal right of publicity statute. *Compare* Brittany A. Adkins, Comment, *Crying Out for Uniformity: Eliminating State Inconsistencies in Right of Publicity Protection Through a Uniform Right of Publicity Act*, 40 CUMB. L. REV. 499, 526-28 (2010) (advocating a uniform act as preferable to federal legislation, for reasons including flexibility in responding to changes in this still-evolving area of the law, respect for federalism, and the promise of states learning from one another's experiences), *with* Symposium, *Rights of Publicity: An In-Depth Analysis of the New Legislative Proposals to Congress*, 16 CARDOZO ARTS & ENT. L.J. 209, 210 (1998) (American Bar Association joint task force on federalizing the right of publicity concluded that a federal right of publicity law is preferable to a uniform act, due to unlikelihood that such an act would pass in identical form in every state and unlikelihood that states would repeal their own, existing right of publicity laws in favor of such an act).

9. *See, e.g.*, Lee Goldman, *Elvis Is Alive, But He Shouldn't Be: The Right of Publicity Revisited*, 1992 BYU L. REV. 597, 606, 610-11, 621-24.

10. *See, e.g.*, *id.* at 616-21; Jeremy T. Marr, Note, *Constitutional Restraints on State Right of Publicity Laws*, 44 B.C. L. REV. 863, 878-99 (2003); Eugene Volokh, *Freedom of Speech and the Right of Publicity*, 40 HOUS. L. REV. 903, 908-29 (2003); Diane Leenheer Zimmerman, *Who Put the Right in the Right of Publicity?*, 9 DEPAUL-LCA J. ART & ENT. L. & POL'Y 35, 53-82 (1998).

11. *See, e.g.*, Goldman, *supra* note 9, at 604-05 (questioning rhetorically whether "we want a society in which fame has economic value apart from the activity that creates celebrity"), 613-17 (right of publicity creates monopolies that result in higher costs for consumers), 620 (right of publicity curtails consumers' ability to express themselves through choices in celebrity imagery); Michael Madow, *Private Ownership of Public Image: Popular Culture and Publicity Rights*, 81

for a halt to expansion of the right of publicity, in terms of both the right's statutory adoption in additional states[12] and the scope of its application in those states that have already adopted it,[13] and others have gone so far as to recommend its outright abolition.[14] In some respects, the hostility directed at the right of publicity reflects discomfiture with the reality that while the right applies to all, it provides a remedy only insofar as one's persona has marketable value. Thus, in a very literal sense, the measure of one's right of publicity is one's fame.

However, the very nature of fame in our society is undergoing a seismic change, to the extent that fame today is a very different creature than it was in 1993, much less 1953. In the past decade, the Internet—representing the latest development in an exponential change in mass media heralded by the growth of cable television and specialized, niche marketing—has very nearly brought to manifest reality Andy Warhol's putative prediction that "[i]n the future, everyone will be world-famous for fifteen minutes."[15] In all likelihood, the judge in *Haelan Laboratories* could not have imagined the ever-expanding constituency to whom the doctrine he begat might someday apply. This changing nature of fame in our society will necessarily have ramifications for right of publicity doctrine, rendering many of the justifications proffered for the right since *Haelan Laboratories* infirm—but at the same time, rendering many of the arguments against the right moot. In the place of the old conversation regarding the right of publicity, two surprising developments will emerge: right of publicity will be reborn as nothing so much as a populist doctrine, and its existence will find singular justification in the reasoning that gave rise to it in the first place, in 1953 in *Haelan Laboratories*.

Part I of this Note briefly reviews how the right of publicity emerged, how it has developed in the decades since its formal christening in *Haelan Laboratories*, and how those post-*Haelan Laboratories* developments have diluted the formative conception of the right in ways that have made it a target of

---

CALIF. L. REV. 125, 136-37 (right of publicity allocates more wealth to already-wealthy individuals), 137-46 (right of publicity "facilitate[s] private censorship of popular culture"), 211-12 (right of publicity may decrease cultural creation by allowing individuals to subsist on the value of their fame alone, without further creative endeavor), 216-19 (right of publicity may result in disproportionate pursuit of fame-generating careers at the expense of pursuit of more socially beneficial fields of endeavor, especially amongst economically disadvantaged segments of the population); 226-28 (right of publicity further elevates disproportionate power of celebrities as a cultural and political influence) (1993); Sudakshina Sen, Comment, *Fluency of the Flesh: Perils of an Expanding Right of Publicity*, 59 ALB. L. REV. 739, 752-61 (1995) (right of publicity stifles cultural discourse).

12. *See* Goldman, *supra* note 9, at 628.

13. *See id.*; Sen, *supra* note 11, at 754-55.

14. *See* Goldman, *supra* note 9, at 602.

15. This quote is most often attributed to the catalog that accompanied Warhol's 1968 exhibition at the Moderna Museet gallery in Stockholm, Sweden. *See, e.g.*, RALPH KEYES, THE QUOTE VERIFIER: WHO SAID WHAT, WHERE, AND WHEN 288 (2006).

skepticism and criticism.[16] Part II describes the tremendous, unprecedented changes in mass media wrought by the Internet in the last decade, which may already be reshaping the very character of "fame" in our society and, consequently, the application of the right of publicity in our law.[17] Even before the dawning of this new age of Internet celebrity, the "raging debate . . . about whether [publicity] rights should exist at all"[18] had already seemed to have reached an academic stalemate.[19] Part III reviews the arguments that have long been made in this debate, both for and against the right of publicity.[20] These arguments, almost without exception, concern post-*Haelan Laboratories* developments in right of publicity doctrine and serve to neither strengthen nor weaken the case for the right of publicity as articulated by the court in *Haelan Laboratories*; rather they serve simply to highlight the doctrinal confusion evinced by these developments. Part III concludes that nearly every argument in this theoretical debate over the justification for a right of publicity will be rendered moot by the new paradigm of fame ushered in by the Internet's revolutionizing of mass media. Part IV posits that what remains in the wake of the old conversation regarding the right of publicity is the right as originally articulated in *Haelan Laboratories*—as cogent and relevant today as it was when *Haelan Laboratories* was decided.[21]

I. History of the Right of Publicity

A. The Right of Publicity Distinguished from the Right to Privacy

To understand the right of publicity, how it has changed in the decades since it was first recognized, and how it will apply in the changing culture of today's media personalities, it is necessary to recognize how the right evolved from the right to privacy. While the two doctrines ultimately serve fundamentally different purposes, as explicitly noted by the court in *Haelan Laboratories*,[22] one important effect of the changing nature of fame will be to highlight not the differences between the two doctrines, but rather their common foundations. In particular, both doctrines are grounded in a humanist concern for personal sanctity and autonomy that serves as the most compelling justification for both.

---

16. *See infra* pp. 872-79.

17. *See infra* pp. 879-80.

18. David Westfall & David Landau, *Publicity Rights as Property Rights*, 23 CARDOZO ARTS & ENT. L.J. 71, 118 (2005).

19. *See id.* at 122 (declaring the debate over the justifiability of a right of publicity to have become "pointless," inasmuch as "the right of publicity is both hard to object to and hard to support"); *see also id*. at 121 (stating, "neither those favoring publicity rights nor those opposed to publicity rights have managed to articulate particularly strong policy rationales supporting their positions").

20. *See infra* pp. 880-93.

21. *See infra* pp. 893-95.

22. Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 868 (2d Cir. 1953).

The humanist concerns undergirding a right of privacy were recognized when the right was still little more than academic theory. In their seminal article on the subject in 1890, Samuel Warren and Louis Brandeis described privacy as "a part of the more general right to the immunity of the person,—the right to one's personality."[23] Warren and Brandeis described this right as akin to a property right, though not to any recognized at the time:

> [S]ince the latest advances in photographic art have rendered it possible to take pictures surreptitiously, the doctrines of contract and of trust are inadequate to support the required protection, and the law of tort must be resorted to. The right of property in its widest sense, including all possession, including all rights and privileges, and hence embracing the right to an inviolate personality, affords alone that broad basis upon which the protection which the individual demands can be rested.
>
> . . . .
>
> We must therefore conclude that the rights, so protected, whatever their exact nature, are not rights arising from contract or from special trust, but are rights as against the world; and, as above stated, the principle which has been applied to protect these rights is in reality not the principle of private property, unless that word be used in an extended and unusual sense.[24]

Some seventy years later, William Prosser built upon the Warren and Brandeis article with his own influential treatise on the right of privacy.[25] Prosser included as one of his four enumerated privacy torts the tort of appropriation, which he characterized as "[a]ppropriation, for the defendant's advantage, of the plaintiff's name or likeness."[26] In doing so, Prosser, like Warren and Brandeis, recognized that this brand of privacy tort derived from one's intrinsic rights in one's own identity:

> It is the plaintiff's name as a symbol of his identity that is involved here, and not his name as a mere name.
>
> . . . .
>
> The interest protected is not so much a mental as a proprietary one, in the exclusive use of the plaintiff's name and likeness as an aspect of his identity. It seems quite pointless to dispute over whether such a right is to be classified as "property." If it is not, it is at least, once it is protected by the law, a right of value upon which the plaintiff can capitalize by selling licenses.[27]

Haelan Laboratories had already been decided, but not yet followed,

23. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 207 (1890).

24. *Id.* at 211, 213.

25. William L. Prosser, *Privacy*, 48 CAL. L. REV. 383 (1960).

26. *Id.* at 389.

27. *Id.* at 403, 406 (citation omitted).

when Prosser's privacy article was published.[28] While Prosser's acknowledgement of an inherent right in one's identity was an important academic validation of the nascent right of publicity, he missed a crucial distinction that the court in *Haelan Laboratories* did not: in defining the right of publicity as simply a variant of the right to privacy, Prosser failed to recognize that the interest protected by the right of publicity is fundamentally distinct from the interest protected by the right of privacy. The right of privacy protects an individual from the emotional and reputational harm caused by unauthorized distribution of the individual's likeness; the kinds of cases cited by Prosser as exemplary of his appropriation tort involved unauthorized use of the likeness of a person who felt embarrassed or aggrieved by the use.[29] This kind of tort had, in fact, been recognized in some states long before Prosser published his privacy treatise. In 1909, the New York state legislature, in response to the state's lack of statutory authority to grant relief to a private person whose likeness had been used in advertisements without her permission,[30] passed the nation's first law proscribing the unauthorized use of a private individual's likeness.[31] The Georgia Supreme Court, also in the absence of such statutory authority, chose to recognize such a tort judicially in 1905.[32]

But as the court in *Haelan Laboratories* rightly noted, not only may the famous not object to having their likenesses disseminated, they may actually desire it—provided they can exercise control of the dissemination and any monetary gain earned from it.[33] Hence, the right of publicity protects an individual's ability to control the use of, and to profit from, the use of his identity in commerce, allowing an individual whose likeness carries a merchantable value to recoup the putative market value of an unauthorized commercial use of his likeness.[34] So while the right of publicity shares as a common foundation with

28. *Id*. at 406-07.

29. *Id*. at 401-02 nn.156-61. In one of the cases cited by Prosser, the plaintiff in fact brought a claim for the presumptive loss in value in any future use he might make of the photograph of him that was used by the defendant—a claim akin to a right of publicity claim. *See* Cont'l Optical Co. v. Reed, 86 N.E.2d 306 (Ind. App. 1949). This claim was, however, rejected by the court, which held that the plaintiff had no more right to the particular photograph in question than did the defendant. *Id.* at 310.

30. *See* Roberson v. Rochester Folding Box Co., 64 N.E. 442 (N.Y. 1901).

31. N.Y. CIV. RIGHTS LAW §§ 50, 51 (McKinney 2009) (enacted 1909).

32. *See* Pavesich v. New England Life Ins. Co., 50 S.E. 68 (Ga. 1905).

33. Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 868 (2d Cir. 1953).

34. *See, e.g.*, Ventura v. Titan Sports, Inc., 65 F.3d 725, 733-35 (8th Cir. 1995) (holding that lower court did not abuse its discretion in accepting, for purposes of assessing damages, expert testimony as to a speculative royalty rate for plaintiff, a professional wrestler, based on royalty rates that had been negotiated for comparable sports and entertainment figures); Hoffman v. Capital Cities/ABC, Inc., 33 F. Supp. 2d 867, 872-73 (C.D. Cal. 1999), *rev'd on other grounds*, 255 F.3d 1180 (9th Cir. 2001) (awarding actor $1,500,000 in compensatory damages for unauthorized use of his image, based on actor's stature in the motion picture industry and long-time eschewal of commercial endorsements, and the unique promotional opportunity represented by use of the image

the right of privacy a recognition of the intrinsic right one has in one's own identity—what Warren and Brandeis called "the right to an inviolate personality"[35]—the right of privacy recognizes an abstract emotional and social interest in one's personality, while the right of publicity recognizes a commercial one.

B. Haelan Laboratories

Haelan Laboratories merits close inspection not only because it was the court decision that first put a name to the right of publicity, but also because it defined the right with such startling clarity and prescience. The plaintiff in *Haelan Laboratories*, chewing gum manufacturer Haelan, signed a contract with a famous baseball player granting Haelan the exclusive use of the player's image in marketing chewing gum; the defendant, Topps, a rival chewing gum manufacturer, and aware of the player's contract with Haelan, induced the player to sign a contract granting Topps the right to use the player's image in marketing its chewing gum.[36] Topps claimed that the player's contract with Haelan constituted nothing more than a release of Haelan from liability under New York's landmark 1909 law proscribing unauthorized use of a person's likeness, and that the player's rights under that law were personal and thus non-assignable.[37] The court disagreed, holding in two paragraphs that constituted nearly the entirety of the substance of its decision:

> We think that, in addition to and independent of that right of privacy (which in New York derives from statute), a man has a right in the publicity value of his photograph, i.e., the right to grant the exclusive privilege of publishing his picture, and that such a grant may validly be made 'in gross,' i.e., without an accompanying transfer of a business or of anything else. Whether it be labelled [sic] a 'property' right is immaterial; for here, as often elsewhere, the tag 'property' simply symbolizes the fact that courts enforce a claim which has pecuniary worth.
>
> This right might be called a 'right of publicity.' For it is common knowledge that many prominent persons (especially actors and ball-players), far from having their feelings bruised through public exposure of their likenesses, would feel sorely deprived if they no longer received money for authorizing advertisements, popularizing their countenances, displayed in newspapers, magazines, busses, trains and subways. This right of publicity would usually yield them no money unless it could be

---

in question); Apple Corps, Ltd. v. Leber, 229 U.S.P.Q. 1015, 1018 (Cal. Sup. Ct. 1986) (holding that because the Beatles could have "named [their] own price" had they reunited for a tour, it was reasonable to assess as damages against a Beatles imitator group a royalty rate of 12.5% of the gross receipts of the imitator group's tour and $2,000,000 for the film made from the tour).

35. Warren & Brandeis, *supra* note 23, at 211.

36. *Haelan Labs., Inc.*, 202 F.2d at 867.

37. *Id*.

> made the subject of an exclusive grant which barred any other advertiser from using their pictures.[38]

In the course of two compact paragraphs, the court not only christened the right of publicity, but illuminated its nature and scope with a precision belying further elaboration: the right is personal; is distinct from the right to privacy; is akin to a property right (in the manner suggested by Warren and Brandeis with regard to privacy); and is assignable. The right would never again be so exactly delineated.

C. Adoption and Expansion; Dilution and Dissent

In the decades that followed *Haelan Laboratories*, the right of publicity saw sporadic adoption in state law.[39] But as the right was invoked and interpreted in various courts, the neat boundaries of the right drawn by the court in *Haelan Laboratories* also became blurred, as different courts, interpreting different state laws, arrived at different conclusions as to the scope and the precise contours of the right.[40] In the process, the right of publicity became increasingly vulnerable to criticism: already a legal doctrine that could easily be painted as providing

---

38. *Id.* at 868.

39. *See, e.g.*, CAL. CIV. CODE § 3344.1 (2011) (enacted 1984); FLA. STAT. § 540.08 (West 2012) (enacted 1967); 765 ILL. COMP. STAT. ANN. 1075/1 to 1075/60 (West 2011) (enacted 1999); IND. CODE §§ 32-36-1-1 to -20 (2011) (enacted 1994 as IND. CODE §§ 32-13-1-1 to -20); KY. REV. STAT. ANN. § 391.170 (West 2011) (enacted 1984); NEV. REV. STAT. ANN. §§ 597.790 to-.810 (West 2011) (enacted 1989); OHIO REV. CODE ANN. § 2741.01 to -.09 (West 2011) (enacted 1999); OKLA. STAT. tit. 12 § 1448-1449 (2011) (enacted 1986); 42 PA. CONS. STAT. § 8316 (West 2011) (enacted 2002); TENN. CODE ANN. §§ 47-25-1101 to -1107 (2012) (enacted 1984); TEX. PROP. CODE ANN. §§ 26.001 to -.015 (West 2011) (enacted 1987); WASH. REV. CODE ANN. §§ 63.60.010 to -.080 (West 2011) (enacted 1998), *declared unconstitutional in part by* Experience Hendrix, L.L.C. v. HendrixLicensing.com, Ltd., 766 F. Supp. 2d 1122 (W.D. Wash. 2011); *see also* Martin Luther King, Jr. Ctr. for Soc. Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697 (Ga. 1982) (recognizing common-law right of publicity in Georgia).

40. *Compare, e.g.*, Winter v. DC Comics, 69 P.3d 473 (Cal. 2003) (using the artistic transformation test to determine whether right of publicity of professional musicians under California right of publicity statute was superseded by First Amendment right of comic book artist who used the musicians as the basis for comic book characters), *with* Doe v. McFarlane, 207 S.W.3d 52 (Mo. Ct. App. 2006) (using the predominant use test to determine whether right of publicity of professional hockey player under common law right of publicity was superseded by First Amendment right of comic book artist who used the hockey player as the basis for a comic book character). *Compare, e.g.*, Hicks v. Casablanca Records, 464 F. Supp. 426 (S.D.N.Y. 1978) (holding that in order for an individual's common law right of publicity to be exercised by his heirs or assigns after his death, the individual must have commercially exploited his right of publicity during his lifetime), *with Martin Luther King, Jr. Ctr. for Soc. Change, Inc.*, 296 S.E.2d at 706 (holding that the right of an individual's heirs or assigns to exercise the individual's common-law right of publicity after his death is not dependent on whether the individual commercially exploited his right of publicity during his lifetime).

economic relief for those who need it least, the occasional, anomalous extensions of the right's scope—frequently by the Ninth Circuit, whose jurisdiction encompasses the heart of the entertainment industry—allowed naysayers further justification to impugn the right as governmental pandering to the rich and famous.

Perhaps no other court decision did more to simultaneously validate and undermine the right of publicity than the Supreme Court's sole consideration of the right in *Zacchini v. Scripps-Howard Broadcasting Co.*[41] Importantly, the Court in *Zacchini* affirmatively recognized the right (at least insofar as it is implemented by state law).[42] However, the Court's decision construed the right in a way that obscured it more than illuminated it, and has provided more grist for condemnation of the right than for support of it. Zacchini performed a human cannonball act, and brought suit when a local newscast, reporting on the state fair at which he was performing, broadcast his act in its entirety.[43] The Supreme Court of Ohio recognized that Zacchini had a "right to the publicity value of his performance" and that the network had violated this right by broadcasting his entire act, thus depriving him entirely of its economic value.[44] However, the court held that Zacchini's right of publicity was trumped by the station's First Amendment right to "report in its newscasts matters of legitimate public interest . . . unless the actual intent of the TV station was to appropriate the benefit of the publicity for some non-privileged private use, or unless the actual intent was to injure the individual."[45] The Supreme Court reversed, holding that the news network's First Amendment rights did not entitle it to completely deprive Zacchini of the economic value of his performance.[46]

Aside from the fact that the case hardly presented a model of the kind of interest the right of publicity is generally thought to protect (after all, it was not appropriation of Zacchini's likeness that was at issue, but rather the appropriation of his act), or perhaps because of that fact, the Court justified the right of publicity in terms of protecting and promoting creative endeavor, the traditional rationale behind copyright and patent law:

> [T]he State's interest in permitting a 'right of publicity' is in protecting the proprietary interest of the individual in his act in part to encourage such entertainment. . . . [T]he State's interest is closely analogous to the goals of patent and copyright law, focusing on the right of the individual to reap the reward of his endeavors and having little to do with protecting

41. 433 U.S. 562 (1977).

42. *Id.* at 566 ("There is no doubt that petitioner's complaint was grounded in state law and that the right of publicity which petitioner was held to possess was a right arising under Ohio law.").

43. *Id*. at 563-64.

44. Zacchini v. Scripps-Howard Broad. Co., 351 N.E.2d 454, 455 (Ohio 1976), *rev'd*, 433 U.S. 562 (1977).

45. *Id*.

46. *Zacchini*, 433 U.S. at 574-76.

> feelings or reputation.[47]

This rationale for the right of publicity is completely at odds with the right as articulated in *Haelan Laboratories*, which recognized the right not in terms of a state or societal interest, but as a right protecting a personal interest.[48] The Court gave passing acknowledgment to the right of publicity as recognized in *Haelan Laboratories*, but compounded its earlier mischaracterization by explicitly devaluing the *Haelan Laboratories* court's formative model of the right of publicity as a protection of the promotional value of an individual's likeness:

> [I]n this case, Ohio has recognized what may be the *strongest case* for a 'right of publicity' involving, not the appropriation of an entertainer's reputation to enhance the attractiveness of a commercial product, but the appropriation of the very activity by which the entertainer acquired his reputation in the first place.[49]

This attempt to analogize the right of publicity to copyright and patent law is doctrinally untenable. The right of publicity protects not an individual's right to a tangible result of his labors, but a more abstract right in the sanctity of his person, making it an ill-fitting conceptual stable-mate with copyright and patent doctrine. As such, attempts to lash right of publicity doctrine to existing intellectual property rationales—rather than to rationales for the right to privacy, to which the right of publicity is more closely related—serve only to obscure the humanist concerns that drive this outwardly commercial doctrine. Regardless, the Supreme Court's doctrinally confused characterization of the right to publicity has provided critics of the right with some of the most perennially resonant aspersions against the right, particularly to the extent that the Court justified the right in terms of rewarding one's endeavors—a justification that is unpersuasive if one considers the right as protecting only the already rich and famous.[50] In short, the Supreme Court's recognition of the right of publicity in *Zacchini* may have been more damaging to the subsequent vitality of the right than the Court's recognition was worth.[51]

Similarly, the degree to which courts have expanded the scope of the right of publicity since *Haelan Laboratories* has made the right an easier target for critics. The chief void in the right that has led to the creeping outward advance of its

---

47. *Id*. at 573 (citation omitted).

48. Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 867 (2d Cir. 1953).

49. *Zacchini*, 433 U.S. at 576 (emphasis added).

50. *See, e.g.*, Madow, *supra* note 11, at 189 ("Publicity rights operate to channel additional dollars to the very people—Einstein rather than Bohr, Vanilla Ice rather than Too Short—who happen to draw first-prize tickets in the fame lottery.").

51. Advocates of the right of publicity have seemed no happier with the *Zacchini* decision than are opponents of the right. *See, e.g.*, Alice Haemmerli, *Whose Who? The Case for a Kantian Right of Publicity*, 49 DUKE L.J. 383, 402 (1999) ("Despite its affirmation of the right of publicity . . . *Zacchini*'s distortion of state law and its fixation on act rather than identity has had a profound, and negative, effect on publicity rights doctrine.").

scope is ambiguity over what constitutes a person's "likeness" for purposes of protection under the right. Courts have ruled that because a singer's voice and idiosyncratic, identifiable singing style are a part of the singer's persona, a singer's right of publicity can be violated when an advertiser employs a sound-alike to imitate the singer,[52] even when the sound-alike is not employed to perform a song associated with the singer.[53] In even more striking extensions of the right, courts have held that a party may violate an individual's right of publicity though commercial use of the image of a robot that broadly invokes the individual,[54] and even of a robot that broadly invokes not the individual himself, but the character played by the individual on a television show (a character to which the individual does not own the copyright).[55] These decisions can be seen as the product of courts struggling to define the scope of a legal doctrine still very much in its infancy, but they may also contribute to a perception of courts employing the right of publicity in fawning deference to the rarified needs of the rich and famous.

II. THE EMERGING MEDIA: A WHOLE NEW BALLGAME

In July 2010, Antoine Dodson was interviewed by a local newscast after an intruder allegedly broke into his Huntsville, Alabama home and attempted to rape his sister.[56] Dodson's videotaped diatribe was posted to YouTube, where it garnered over a million views by the next day.[57] Only a few days later, the comedy troupe behind the YouTube series "Auto-Tune the News" transformed Dodson's rant into a song.[58] Three weeks later, the song was number sixteen on the pop sales chart of music download retailer iTunes, and Dodson claimed he had made enough money from sales of the song, sales of his merchandise, and donations through his Internet site to buy his family a new home in a better neighborhood.[59]

Dodson provides a startling example of fame in a new media age. The Internet—in addition to, and even more so than, preceding developments in niche programming and niche marketing driven by cable television and its ever-proliferating array of specialty channels—has changed the media landscape profoundly, to an extent that likely could not have been imagined only a decade before Dodson's fateful interview, much less when *Haelan Laboratories* was

---

52. *See* Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1988).

53. *See* Waits v. Frito-Lay, Inc., 978 F.2d 1093 (9th Cir. 1992).

54. *See* White v. Samsung Elecs. Am., Inc., 971 F.2d 1395 (9th Cir. 1992).

55. *See* Wendt v. Host Int'l, Inc., 197 F.3d 1284 (9th Cir. 1999).

56. *See* Ki Mae Heussner, *From Sex Assault Fury to Internet-Fueled Fortune*, ABC NEWS, Aug. 17, 2010, http://abcnews.go.com/Technology/sex-assault-fury-internet-fueled-fortune/story?id=11413687.

57. *Id*.

58. *Auto-Tune the News: Bed Intruder Song!!! (now on iTunes)*, YOUTUBE.COM, http://www.youtube.com/watc?v=/hMtZfW2z9dw (last accessed April 25, 2012).

59. Heussner, *supra* note 56.

decided in 1953. This new media age may fundamentally and permanently alter the very nature of fame in our society: in terms of fame, what the baseball player at the center of the *Haelan Laboratories* controversy took years, or perhaps even the better part of his lifetime, to accomplish through practice, work, and perseverance, Antoine Dodson arguably accomplished in little more than three weeks with very little effort at all.

The hallmark of the new, Internet-driven media age is accessibility. No longer are the channels of wide-spread commercial distribution controlled by a small oligarchy of corporate gatekeepers: television networks, radio stations, record labels, and the like. The Internet provides anyone with a computer and an Internet service provider direct access to not only nation-wide distribution, but world-wide distribution. The "Auto-Tune the News" comedy group that immortalized Antoine Dodson in song epitomizes the drastically reduced economic barriers to entry engendered by the Internet: where even a decade earlier, the group's access to nation-wide media distribution of its works would have entailed laborious, highly competitive, and (at best) uncertain campaigning for patronage, capital, and access to television networks or record labels, the group was able, through the Internet, to market a profitable song in a few weeks at nominal cost. This accessibility extends not only to the channels of distribution, but also to the means of production: increasingly advanced home video cameras, inexpensive audio samplers and synthesizers, and digital video editing and digital audio recording software now allow one to produce, on a relatively inexpensive home computer, media of a quality scarcely distinguishable from that of similar works produced by traditional, corporate media providers.

The drastically lower barriers to the production and dissemination of media in the Internet age necessarily result in correspondingly reduced barriers to fame. Antoine Dodson's three-week rise to nation-wide prominence may have been a particularly exceptional example of just how far the barriers to fame have fallen—or it may represent the dawning of an emerging, unprecedented new paradigm of fame in our society. Either way, the right of publicity applies to the prominent persons of the Internet age just as it did to star baseball players in the 1950s, and just as it has to prominent persons of all stripes in the decades in between. Invocation of the right of publicity by this new breed of "instant" celebrity will, however, reshape the debate over the right of publicity, and may well reshape the doctrine itself—ironically, bringing it right back to where it started in 1953.

III. THE DYING DEBATE

It is the perception of the right of publicity as a legal gratuity granted exclusively to the rich and famous that is perhaps most damaging to the right's credibility in the legal and academic world. Criticism of the right of publicity is almost invariably shot through with explicit or implicit indignity over the notion that the right serves as a legal remedy to reimburse individuals who may be seen as already financially well-off, as living lives of leisure, or as having not done any "real" work to achieve their rarified positions.[60] What is most striking about this criticism, however, is that it is, in the main, criticism of the post-*Haelan Laboratories* developments in right of publicity doctrine. These developments, like the Supreme Court's reshaping of the doctrine to conform it to the mold of existing intellectual property principles,[61] have distanced the right of publicity from its initial conception in *Haelan Laboratories*, spawning both justifications for and arguments against the doctrine that are suspect at their root. As such, the debate over the right of publicity can be seen, if viewed from the vantage of *Haelan Laboratories*, as little more than rhetorical tail-chasing. And for whatever vitality remains of this debate, the emergent changes in the nature of fame wrought by the Internet will render the majority of the contentions on both sides of the argument moot. A survey of these arguments, on the cusp of becoming a retrospective exercise, is instructive in understanding just how sweepingly the landscape defining the purview of the right of publicity is about to change.

A. Right of Publicity as a Reward and Incentive for Creative Endeavor

The right of publicity has been justified as a reward and an incentive for creative endeavor, in the same way that copyright and patent protection are justified. That this justification for the right of publicity was consecrated by the Supreme Court[62] does not compulsorily render it persuasive, and critics of the right of publicity have been quick to assail its logic.[63] And its logic—at least as applied within the traditional purview of the right of publicity—is flimsy at best, and very nearly absurd on its face. The Internet age of celebrity may bifurcate the traditionally-related intellectual property rationales of reward and incentive as applied to the right of publicity, leaving the former as unpersuasive as it has ever been and providing the latter with a potency it never previously had.

1. The Right of Publicity as a Reward for Creation.—The "Lockean"[64] rationale for intellectual property protection as an assurance that one's creative

60. *See, e.g.*, Goldman, *supra* note 9, at 604 ("It is the Madonnas and Michael Jordans who reap the greatest rewards from the right of publicity, not the struggling actor or author."); Madow, *supra* note 11, at 137 ("Why, we may properly ask, should the law confer a source of *additional* wealth on athletes and entertainers who are already very handsomely compensated for the primary activities to which they owe their fame?").

61. *See* Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 573 (1977).

62. *Id*. at 576.

63. *See, e.g.*, Goldman, *supra* note 9, at 603-05; Madow, *supra* note 11, at 182-96; Zimmerman, *supra* note 10, at 77-78.

64. *See* Madow, *supra* note 11, at 175 n.239 ("It was John Locke who first elaborated a philosophical argument for private property based on individual 'labor.' He purported to derive the right of private property from logically prior property rights in one's body and its labor."); *see also* Melville B. Nimmer, *The Right of Publicity*, 19 L. & CONTEMP. PROBS. 203, 216 (1954) ("It would seem to be a first principle of Anglo-American jurisprudence, an axiom of the most fundamental nature, that every person is entitled to the fruit of his labors unless there are important countervailing public policy considerations.").

effort is properly rewarded has never been cogent as a justification for the right of publicity. This impotency stems from the rationale's necessary assumption that an individual who has achieved fame has earned it through hard work deserving of reward—an assumption of dubious credibility.[65] We as a culture may debate whether achievement in sports or the arts sufficiently benefits society to merit rewarding effort in these areas; the even more subjective question as to whether acting, singing in a popular band, or playing baseball truly constitutes hard work—or even "work" at all—that merits supplementary, commendatory compensation is more a debate for social scholars than for the law. Even so, this debate serves to highlight another way in which the work-reward rationale of copyright and patent law fails as a justification for the right of publicity.

Copyrights and patents protect and reward the tangible product of creative endeavor. While reasonable people might debate the relative usefulness of a given invention or a given work of art, the law, through copyrights and patents, recognizes a presumptive societal value in the creation of more inventions and more art. The "end product" protected by the right of publicity, on the other hand, is the persona of the individual. The benefit that the right of publicity seeks to protect is not a broad societal benefit in this end product, but rather—like the right of privacy—an individual "benefit" in one's personal sanctity and autonomy.[66] This distinction between the interest protected by the right of publicity and the interests protected by copyrights and trademarks becomes transparently obvious when one attempts to quantify the right of publicity in terms of the societal value that copyrights and trademarks seek to serve: How much does society benefit from the fame Madonna has worked so hard to achieve? And given that assessment, to what degree should we reward Madonna

---

65. *See, e.g.*, Goldman, *supra* note 9, at 611 ("One might even question whether the commercial value of a celebrity's identity should be regarded as the fruits of the celebrity's labor. Fame frequently is fortuitous."); Madow, *supra* note 11, at 188 ("[A]n elementary, but occasionally overlooked, sociological truth about fame [is that] fame is a 'relational' phenomenon, something that is *conferred by others*. A person can, within the limits of his natural talents, *make himself* strong or swift or learned. But he cannot, in this same sense, make himself famous, any more than he can make himself loved." (citation omitted)); Mark P. McKenna, *The Right of Publicity and Autonomous Self-Definition*, 67 U. PITT. L. REV. 225, 252 (2005) ("Fame is complex and paradoxical, a phenomenon for which labor is neither a necessary nor sufficient condition. In some instances, an individual with a public persona has indeed worked to create that persona, making professional and lifestyle choices aimed at developing a particular look or image. But that is certainly not true of every celebrity.").

66. *See generally* Roberta Rosenthal Kwall, *A Perspective on Human Dignity, the First Amendment, and the Right of Publicity*, 50 B.C. L. Rev. 1345 (2009)*; see also* Rex D. Glensy, *The Right to Dignity*, 43 COLUM. HUM. RTS. L. REV. 65, 115 (2011) (describing right of publicity as a "[cause] of action relating to a violation of one's dignity"); Haemmerli, *supra* note 51, at 421 ("The right to control the use of one's image or other objectification of identity is a property right based directly on freedom, autonomy, or personality."); Ira J. Kaplan, *They Can't Take That Away from Me: Protecting Free Trade in Public Images from Right of Publicity Claims*, 18 LOY. L.A. ENT. L. REV. 37, 46-49 (1997) (describing unauthorized use of one's persona as a "dignitary harm").

for the effort she expended in achieving fame? To ask these questions is to realize the futility of attempting to contort the right of publicity into the work-reward model of copyrights and patents.[67]

Further, the right of publicity affords individuals an exclusionary right: An individual, however prominent, may choose not to license his likeness for commercial gain.[68] Notable examples of celebrities who eschewed licensure of their likenesses, and whose right of publicity became the subject of legal wrangling as a result, range from civil rights leader Martin Luther King, Jr.,[69] to singer and songwriter Tom Waits.[70] The incongruence of the right of publicity as a right to exclude with the conception of the right as a reward for one's labors highlights the fundamental level at which this rationale for the right is inapposite, because it fails to recognize the humanist underpinnings of the right: While the right of publicity protects an interest that is economic, it also protects an interest in the sanctity and autonomy of the person.[71] To recognize the right of publicity as a right to exclude is to recognize again that the right of publicity is more closely related, philosophically and doctrinally, to the right of privacy than it is to traditional intellectual property doctrines like copyright and patents.

Regardless of how persuasive or unpersuasive legal academics might have traditionally found the work-reward justification for the right of publicity, the degree to which the emergent mass media have lowered the barriers to production and distribution—making it faster, easier, and less expensive than ever to achieve fame—serves to moot the debate regarding the work-reward rationale altogether. As Lee Goldman and Michael Madow observe, fame is not something "earned"

---

67. The related rationale for the right of publicity as a remedy for the unjust enrichment of the appropriator of an individual's identity for commercial gain—the rationale that free-riders should not be able to reap what others have sown, so to speak—is incumbent upon this same assumption that fame deserves protection because it is the product of one's labors, and is therefore untenable on the same grounds. *See, e.g.*, McKenna, *supra* note 65, at 248 ("Unjust enrichment is an empty concept . . . . A defendant is only 'free riding' (and his enrichment therefore 'unjust') if one assumes that the value of identity belongs to the plaintiff in the first instance.").

68. A salient schism in both state law and academic opinion regarding the right of publicity is in regard to whether the right should be descendible post-mortem; a constituent branch of that debate is whether the descendibility of the right should be dependent upon whether the decedent used his right of publicity for commercial gain in his lifetime. *See, e.g.*, Martin Luther King, Jr. Ctr. for Soc. Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697, 705-96 (Ga. 1982) (examining, but declining to follow, prior case law holding that an individual's right of publicity is descendible only if the individual commercially exploited his right of publicity during his lifetime).

69. *Id*.

70. Waits v. Frito-Lay, Inc., 978 F.2d 1093 (9th Cir. 1992).

71. See Kwall, *supra* note 66, and McKenna, *supra* note 65, for a more comprehensive examination of this aspect of the right of publicity. *See also* Haemmerli, *supra* note 51 (making the same argument regarding the right and buttressing it philosophically in terms of Immanuel Kant's conception of humanity as a function of individual self-determination).

at all.[72] It is more accurate to characterize it as something one simply inherits, or as a windfall.[73] To tie one's right of publicity to the effort one expended in achieving fame would be no more justified, legally or morally, than depriving an heir of money or property gained through testamentary inheritance simply because he did not "work for it." Antoine Dodson did little or nothing to achieve his fame.[74] Should this mean that Dodson does not deserve the right to control the use of his likeness for profit? To concede that he is no less deserving of that right than an athlete or actor who worked at his profession for decades is to drive the final nail in the coffin of the work-reward justification for the right of publicity.

2. The Right of Publicity as an Incentive for Creation.—The incentive rationale of intellectual property rights holds that assuring a financial benefit in one's societally-beneficial creation encourages such efforts by the creator and by others, with society benefitting from the additional creation thus encouraged.[75] This same incentive rationale has been cited as a justification for the right of publicity: assuring a person's right to the financial benefits of the marketing opportunities attendant to fame encourages individuals to pursue careers in the arts and in sports, to the benefit of society.[76]

In the context of the traditional celebrity to which the right of publicity has customarily applied, this incentive rationale of intellectual property is singularly unpersuasive: It strains reason to presume that an individual would choose to

---

72. *See supra* note 65.

73. Madow's characterization of fame as "conferred" by the public, *see supra* note 65, on the other hand, overstates the consciousness and deliberateness of the role played by the public in making a person famous.

74. *See supra* notes 56-59 and accompanying text.

75. *See* U.S. CONST. art. I, § 8, cl. 8 ("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 . . . ."); Mazer v. Stein, 347 U.S. 201, 219 (1954) ("The economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors in 'Science and useful Arts.'").

76. *See* Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 576 (1977) ("Ohio's decision to protect petitioner's right of publicity here rests on more than a desire to compensate the performer for the time and effort invested in his act; the protection provides an economic incentive for him to make the investment required to produce a performance of interest to the public. This same consideration underlies the patent and copyright laws long enforced by this Court."); *see also, e.g.*, D. Scott Gurney, *Celebrities and the First Amendment: Broader Protection Against the Unauthorized Publication of Photographs*, 61 IND. L.J. 697, 707 (1986) ("By protecting the right of celebrities to commercially exploit their fame generally rather than solely in the context of performances or commercial endorsements, the law would help maximize incentive to develop and maintain skills and talents that society finds appealing."); David E. Shipley, *Publicity Never Dies; It Just Fades Away: The Right of Publicity and Federal Preemption*, 66 CORNELL L. REV. 673, 681 (1981) ("Protecting the right of publicity provides incentive for performers to make the economic investments required to produce performances appealing to the public.").

pursue a career in the arts or in sports, or would choose to continue pursuing it, based on the promise of earnings from his right of publicity. The court in *Haelan Laboratories* made no assertion that the baseball player whose image was at issue would not have played baseball, or would not continue to play baseball, were he not allowed to earn money by licensing his image to baseball card manufacturers; such an assertion would be almost laughable. And that assertion would be equally laughable when applied to the gamut of famous persons to whom the right of publicity has traditionally applied. A much more credible proposition is that most people who pursue such careers do so out of reverence for the career itself (keeping in mind that careers in sports and the arts are, in and of themselves, fairly enviable and inherently rewarding fields of endeavor), in pursuit of fame as its own reward, or because the work that the career inherently entails is traditionally well-compensated. Further, this argument treats the right of publicity as a purely economic protection, which it is not: Again, the right of publicity affords one an exclusionary right in one's persona, allowing an individual—like Martin Luther King, Jr.,[77] and Tom Waits[78]—to deny the world the use of his persona and forego the financial opportunities such use may offer.

However, to the new breed of media creators plying their craft on the Internet, the ability to profit from merchandising may indeed be a strong incentive to create, inasmuch as the income generating possibilities afforded by their creations themselves may be slim or, particularly in the case of individuals employing free-access web sites like YouTube to disseminate their works, none. Antoine Dodson was quick to parlay his Internet notoriety into shirts, coffee mugs, and other trinkets bearing his image or excerpts from his infamous rant;[79] it is not unreasonable to think that this merchandise will continue to provide steady income for Dodson even after sales of his song no longer are. For Internet celebrities who do not have a creation as marketable as a song, merchandise may provide the only income their creative endeavors will yield. As such, in the dawning age of Internet celebrity, the incentive function of intellectual property doctrine, once logically and doctrinally bootless in the debate over the justifiability of the right of publicity, suddenly seems relevant (if still subject to disagreement over the value of the kinds of works whose creation is being incentivized).

B. Right of Publicity as a Cultural Thief and Enemy of Free Speech

The idea that fame is granted by the public, and that the public thus retains a certain interest in it, dovetails into two related sociological criticisms of the

---

77. *See* Martin Luther King, Jr. Ctr. for Soc. Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697 (Ga. 1982).

78. *See* Waits v. Frito-Lay, Inc., 978 F.2d 1093 (9th Cir. 1992).

79. *See Antoine Dodson*, FACEBOOK, http://www.facebook.com/pages/Antoine-Dodson/102461723145137 (last visited April 3, 2012); Antoine Dodson Official Merchandise, DISTRICT LINES, http://www.districtlines.com/antoine-Dodson-bed-intruder (last visited April 3, 2012).

right of publicity: that the right deprives society of the free use of celebrity personae, which provide some of modern society's most vibrant and powerful cultural symbols and articles of common cultural currency;[80] and that the right allows celebrities to stifle free speech, possibly to an extent sufficient to raise First Amendment concerns.[81]

The famous can, of course, take on a cultural meaning beyond their control (or that of their heirs or assigns). Michael Madow offers as an example John Wayne, whose persona came to symbolize virile masculinity; in this context, Wayne's persona has been both invoked by a United States president on the eve of war and subverted by a greeting card manufacturer catering to the gay and lesbian community.[82] In another example of how a celebrity persona might be used as a signifier in cultural debate, Lee Goldman opines that one could paint a red "X" through a picture of Madonna to express a great many things, from a simple dislike of Madonna to a more complex statement about sexual politics.[83] Madow argues that to grant celebrities monopolistic control over the use of their personae is to rob our society—and in particular, culturally under-represented segments of our society—of one of its most powerful tools for defining itself:

> What it comes down to, then, is that the power to license is the power to suppress. When the law gives a celebrity a right of publicity, it does more than funnel additional income her way. It gives her (or her assignee) a substantial measure of power over the production and circulation of meaning and identity in our society: power, if she so chooses, to suppress readings or appropriations of her persona that depart from, challenge, or subvert the meaning she prefers; power to deny to others the use of her persona in the construction and communication of alternative or oppositional identities and social relations; power, ultimately, to limit the expressive and communicative opportunities of the rest of us.[84]

But the cultural censorship accusation ignores the limitation of the right of publicity to *commercial* use of a celebrity's persona. The right of publicity—correctly construed—cannot prevent an artist from creating an image of John Wayne superscribed with lipstick, nor one from wearing a shirt on which one has drawn an large, red "X" through a picture of Madonna; it serves only to prevent one from profiting from the sale of such an image without the consent of the

---

80. *See* Goldman, *supra* note 9, at 616-21 (arguing that First Amendment rights of consumers to make cultural statements through their choices of which celebrity representations to endorse via their purchases may outweigh celebrities' economic interest in their personae); Madow, *supra* note 11, at 137-46 (right of publicity "facilitate[s] private censorship of popular culture"); Sen, *supra* note 11, at 752-55 (1995) (arguing that expansion of right of publicity to protect all aspects of identity may stifle cultural dialogue).

81. *See generally* Marr, *supra* note 10; Volokh, *supra* note 10; Zimmerman, *supra* note 10.

82. Madow, *supra* note 11, at 128, 144-45.

83. Goldman, *supra* note 9, at 620.

84. Madow, *supra* note 11, at 145-46.

pictured celebrity.[85]

Arguments that the right of publicity may infringe the First Amendment are similarly unavailing. The Supreme Court has held that commercial speech is entitled to a lesser degree of protection under the First Amendment than is expressive speech.[86] However, a print of the aforementioned image of John Wayne superscribed with lipstick is arguably expressive speech, accorded stringent protection under the First Amendment.[87] Displayed in an art gallery,

---

85. Courts have held that cultural criticism regarding celebrities is expression that should be afforded generous First Amendment protection. *See, e.g.*, Cardtoons, L.C. v. Major League Baseball Players Ass'n, 95 F.3d 959, 969 (10th Cir. 1996) ("While not core political speech," social commentary on public figures engaged in "an important social institution constitutes protected expression."); Comedy III Prods., Inc. v. Gary Saderup, Inc., 21 P.3d 797, 807 (Cal. 2001) ("Once the celebrity thrusts himself or herself forward into the limelight, the First Amendment dictates that the right to comment on, parody, lampoon, and make other expressive uses of the celebrity image must be given broad scope."); Edison v. Edison Polyform Mfg. Co., 67 A. 392, 394 (N.J. Ch. 1907) ("It is certain that a man in public life may not claim the same immunity from publicity that a private citizen may."). Likewise, the Supreme Court has held that factual, informational reporting on matters of public interest constitutes fully protected First Amendment expression, *see, e.g.*, N.Y. Times Co. v. Sullivan, 376 U.S. 254, 269 (1964) ("The general proposition that freedom of expression upon public questions is secured by the First Amendment has long been settled by our decisions."), and that information regarding even private individuals may be such a matter of public interest, *see, e.g.*, Time, Inc. v. Hill, 385 U.S. 374, 388-89 (1967) (First Amendment protections for freedom of speech and press are not limited to reporting on public officials), Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 574 (1977) ("It is evident, and there is no claim here to the contrary, that petitioner's state-law right of publicity would not serve to prevent respondent from reporting the newsworthy facts about petitioner's act."). Thus, the invocation of a celebrity's persona in the context of factual, informational reporting on the celebrity should not give rise to injuries cognizable under the right of publicity.

86. *See* Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 561-66 (1980); *see also* Zauderer v. Office of Disciplinary Counsel of Supreme Court, 471 U.S. 626, 637 (1985) ("There is no longer any room to doubt that what has come to be known as 'commercial speech' is entitled to the protection of the First Amendment, albeit to protection somewhat less extensive than that afforded 'noncommercial speech.'").

87. *See, e.g.*, *Cardtoons, L.C.*, 95 F.3d at 969-70 (holding that reproduction of the image of a public figure in the context of parody and social commentary is fully-protected, expressive speech). Additionally, factual, informational reporting on celebrities is not necessarily commercial speech, even when the reporting is in the context of commercial, for-profit publishing, and even when it is disseminated strictly for the purposes of entertainment. *See, e.g.*, *Zacchini*, 433 U.S. at 578 ("There is no doubt that entertainment, as well as news, enjoys First Amendment protection. It is also true that entertainment itself can be important news."). If the reporting is not in service of the public interest or the information reported is false, the legal regimes of privacy, defamation, and libel provide proper avenues of relief. *See, e.g.*, *Time, Inc.*, 385 U.S. at 387-88 (holding that the First Amendment precludes redress for publication of false information regarding matters of public interest unless it can be shown that the publisher disseminated the information "with knowledge of its falsity or in reckless disregard of the truth"); Jenkins v. Dell Publ'g. Co., 251 F.2d

sold as an original work, or plastered on a wall in public, such an image is entitled to rigorous First Amendment protection that would almost inarguably trump Wayne's right of publicity.[88] Additionally, courts have held that even mechanical reproductions of a work of expression, sold for profit, are not necessarily commercial speech, and may be due the same stringent level of First Amendment protection as is the original work.[89]

447, 450 (3d Cir. 1958) (noting that "the interest of the public in the free dissemination of the truth and unimpeded access to news is so broad, so difficult to define and so dangerous to circumscribe that courts have been reluctant to make such factually accurate public disclosures tortious, except where the lack of any meritorious public interest in the disclosure is very clear and its offensiveness to ordinary sensibilities is equally clear"); C.B.C. Distrib. & Mktg., Inc. v. Major League Baseball Advanced Media, L.P., 443 F. Supp. 2d 1077, 1092 (E.D. Mo. 2006), *aff'd*, 505 F.3d 818 (8th Cir. 2007) ("Courts have found that First Amendment freedom of expression is applicable in cases where the subject matter at issue involved *factual data and historical facts*.").

88. *See, e.g.*, *Cardtoons, L.C.*, 95 F.3d at 970 ("The fact that expressive materials are sold neither renders the speech unprotected, nor alters the level of protection under the First Amendment.") (internal citations omitted); *Comedy III Prods., Inc.*, 21 P.3d at 810 ("[B]ecause single original works of fine art are not forms of merchandising, the state has little if any interest in preventing the exhibition and sale of such works, and the First Amendment rights of the artist should therefore prevail."). The Supreme Court of California in *Comedy III* held that the element separating an artistic portrait that is expressive speech from one that is commercial speech is the degree to which the artist has transformed the image of the individual portrayed. *Comedy III Prods., Inc.*, 21 P.3d at 808-10. In its broadest outlines, this "artistic transformation" doctrine holds that there is a point at which an artist has sufficiently imbued the image of a given person with the artist's own expression that the image becomes more definable as the work of the artist than as an image of the person pictured; conceptually, this can also be thought of as the point as which it can be assumed that a purchaser is purchasing the image as a work of the artist, and not as an image of the person depicted, and that the artist is thus not competing with the depicted person in the market for images of the person. *Id.* Discerning where this line is crossed is a somewhat subjective judgment. For contrasting examples of courts locating this boundary, compare *Comedy III Prods., Inc.*, 21 P.3d 797 (finding defendant's charcoal drawings of the Three Stooges were not sufficiently artistically transformative to merit judgment that consumers were likely buying his work as examples of his art, and not as Three Stooges merchandise), with *Winter v. DC Comics*, 69 P.3d 473 (Cal. 2003) (finding comic book characters clearly based on musicians Edgar and Johnny Winter did not violate the musicians' right of publicity, because the comic's invocation of the musicians was not intended to attract buyers based on the musicians' fame, and because the comic book depictions did not compete with real images of the musicians). *Cf.* Doe v. TCI Cablevision, 110 S.W.3d 363 (Mo. 2003) (en banc) (rejecting artistic transformation test in favor of predominant use test in determining whether artistic representation of individual was more accurately classified as commercial or expressive in nature).

89. *See* ETW Corp. v. Jireh Publ'g, Inc., 332 F.3d 915, 937-38 (6th Cir. 2003) (holding that artist's painting of golfer Tiger Woods was expressive speech under the *Comedy III* transformative effects test, that even the sale of mechanically reproduced copies of the painting constituted "expression which is entitled to the full protection of the First Amendment and not the more limited protection afforded to commercial speech," and that this First Amendment protection outweighed

In short, the right of publicity cannot prevent people from using celebrity images as symbols or signifiers, or from incorporating celebrity images into their own artistic expressions. The right constrains only unauthorized commercial exploitation and profit-taking from the commoditization of such expression, activities philosophically divorced from the cultural concerns expressed by critics like Goldman, Madow, and Sen[90] and addressed by the protections of the First Amendment.

Further, even at the acme of celebrity's power as a cultural touchstone, its importance to consumers cannot supersede the intrinsic personal interest an individual has in the sanctity of his identity. To cast the argument in more human terms, even if Madonna were the most powerful symbol of female sexuality in our culture, she remains first and foremost a human being, whose intrinsic authority over her person should be given stronger recognition than the desire of consumers—even of every other man, woman, and child in the United States—to use her persona to express their beliefs. To hold otherwise is not only to demean Madonna's humanity, but to devalue the expressive abilities of those wishing to make a cultural statement: Surely, even if we were to be completely deprived of the use of Madonna as a common cultural signifier, society would not be wholly deprived of its ability to express its feelings regarding female sexuality.[91]

That the cultural deprivation argument against the right of publicity neglects the much more compelling countervailing consideration of individual autonomy is brought into sharper relief when the right is considered in the context of its application to a celebrity like Antoine Dodson. While it is perhaps easy to lose sight of the fact that a celebrity of Madonna's or Michael Jordan's stature is a human being—especially when the celebrity himself is also a brand in commerce, and may have even encouraged or actively participated in commoditizing himself as such—human beings they are. Even the biggest stars, seemingly remote and abstract, seen only through carefully composed and vetted pictures, are individuals with an intrinsic right to the sanctity of their individual humanity—of their "inviolate personality," as Warren and Brandeis christened it.[92] However

---

Woods's right of publicity); *Comedy III Prods, Inc.*, 21 P.3d at 810 ("[A] reproduction of a celebrity image that . . . contains significant creative elements is entitled to as much First Amendment protection as an original work of art."); *see also* Diane Leenheer Zimmerman, *Fitting Publicity Rights into Intellectual Property and Free Speech Theory: Sam, You Made the Pants Too Long!*, 10 DEPAUL-LCA J. ART & ENT. L. & POL'Y 283, 299-301 (2000) (questioning whether the First Amendment should recognize a difference between sale of an original work and copies of that work).

90. *See supra* note 80.

91. And yet Michael Madow, ostensibly advocating the position of the cultural proletariat, demeans the public's ability to express itself by making this very argument: "Individuals and groups must do this symbolic work with centrally produced and distributed commodities. They must make their culture out of these commodities, for there are no other material or discursive resources available to them." Madow, *supra* note 11, at 140.

92. Warren & Brandeis, *supra* note 23, at 211.

unsympathetic one might be toward the super-rich and the super-famous,[93] it is much harder to disclaim the right of someone like Antoine Dodson—an unknown private person one day, and a national celebrity through happenstance literally the next—to be secure in his authority over his own persona. It is also much harder to demand that he surrender this right for the good of a richer cultural commons.

Dodson may represent an extreme—and extremely sympathetic—example of an Internet celebrity to whom the right of publicity applies; after all, the "fifteen minutes of fame" on which he is financially capitalizing were the result of circumstances in which he was largely a passive player, making his case one much closer to those that precipitated early developments in privacy law.[94] But even for individuals who deliberately use the Internet for the purposes of self-expression and who achieve a degree of marketable recognizability as a result, the demand that they sacrifice the ability to profit from that recognizability, or the right to not allow it to be commercialized at all, for the good of simplifying[95] cultural discourse for the rest of us is much harder to rationalize.

C. Right of Publicity as a Redundant Legal Protection

The right of publicity has been attacked as legally redundant, its protections duplicative of those already provided by existing intellectual property law. But this conclusion is a product of doctrinal mischaracterization of the right of publicity. As noted in Part I with regard to the Supreme Court's characterization of the right of publicity in *Zacchini*, copyright and trademark law protect the end product of creative endeavor, while the personal autonomy protected by the right of publicity is an intrinsic human entitlement that is not necessarily the end product of any endeavor.[96] And nowhere is that more clear than in the case of a

---

93. And to be sure, Michael Madow explicitly impugns the right of publicity in part because it "redistributes wealth *upwards*," suggesting that the right of publicity is inherently suspect because, "governmental actions that make the rich richer surely demand very compelling justification." Madow, *supra* note 11, at 137.

94. *See, e.g.*, Pavesich v. New England Life Ins. Co., 50 S.E. 68 (Ga. 1905) (recognizing right of action for invasion of privacy, arising out of natural law, for plaintiff whose name was used without his permission in advertisements); Roberson v. Rochester Folding Box Co., 64 N.E. 442 (N.Y. 1902) (holding that neither common law nor state law provided a cause of action for plaintiff whose image was used without her permission in advertisements).

95. Notwithstanding critics' couching of the cultural-commons argument against the right of publicity in terms of "democratizing" the elements of popular culture, *see* Madow, *supra* note 11, at 239, or protecting the "emotive and communicative value" of celebrity for consumers of popular culture, Goldman, *supra* note 9, at 620, the gravamen of this argument is a decidedly more prosaic—and decidedly less compelling—concern over the ease and convenience with which cultural dialogue may be conducted.

96. The court in *White v. Samsung Elecs. Am., Inc.*, 971 F.2d 1395 (9th Cir. 1992), made a game effort at adapting for use in right of publicity claims the eight-prong test enumerated in *AMF, Inc. v. Sleekcraft Boats*, 599 F.2d 341 (9th Cir. 1979), for evaluating the likelihood of trademark confusion. The results, however, provide more than anything a demonstrative illustration of the

celebrity like Antoine Dodson[97]: Dodson did not endeavor to create anything, and yet he should be no less entitled to control the exploitation of his persona than should be a famous athlete who worked for years honing his craft.

In part, the redundancy accusation leveled at the right of publicity is a response to another suspect, post-*Haelan Laboratories* justification advanced for the right. Advocates of the right of publicity assert that the right performs a consumer advocacy function, protecting consumers from being duped by unlicensed merchandise into believing that a particular celebrity has endorsed a product he has not.[98] But as critics rightly point out, this very protection is already provided by the federal Lanham Act,[99] which provides a civil remedy for an individual who feels he has been, or is likely to be, injured by another's false implication that the individual is associated with a product.[100]

It is notable that the debate over the legal redundancy of the right of publicity circles entirely around post-*Haelan Laboratories* justifications espoused for the right. Nowhere did the *Haelan Laboratories* court contend that the harm suffered by the ball-player was deprivation of earnings from a commodity he had created. Instead—in what could be characterized as a pedagogical omission or as a thoroughly forward-looking understanding of the intersection of fame and commerce—the court simply accepted that the ball-player's likeness had commercial value, however the value came about, and moved directly to the question of where the rights in that value should vest.[101] Nor did the court express concern over consumers of baseball cards being deceived. The right of publicity as articulated by the *Haelan Laboratories* court was a unique and thoroughly modern response to an emergent, unaddressed issue in the law. So prescient was the court's conception of the right that it is as apposite a remedy for a celebrity like Antoine Dodson as it was for a famous baseball player of the 1950s.

D. Right of Publicity as Agent of Economic Inefficiency

A purely economic argument against the right of publicity notes that the right grants a monopoly, and that monopolies result in economic inefficiency. In the context of the right of publicity, what this means is that if celebrities are granted monopolistic control over the use of their personae, the cost of using celebrity personae in commerce increases. The countervailing argument is that if society at large is allowed untrammeled use of celebrity personae, they will be rendered

---

degree to which the two doctrines are incongruent. *See White*, 971 F.2d at 1399-1401.

97. *See supra* notes 56-59 and accompanying text.

98. *See, e.g.*, James M. Treece, *Commercial Exploitation of Names, Likenesses, and Personal Histories*, 51 TEX. L. REV. 637, 647 (1973).

99. Lanham Act, 15 U.S.C. § 1125 (2006).

100. At least one advocate of a federal right of publicity statute has suggested simply amending the Lanham Act, under Congress's commerce clause authority, to incorporate the right of publicity. *See* Haemmerli, *supra* note 51, at 477-91.

101. Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 868 (2d Cir. 1953).

valueless for all.

The latter argument is a variation of the "tragedy of the commons" theory, which posits that if all comers are allowed to exploit a resource in the absence of the restraint of property rights in the resource, their unrestrained use will completely deplete the resource, to the detriment of all.[102] However, as both right of publicity critic Michael Madow and right of publicity advocate Mark McKenna point out, the hypothetical pasture of the tragedy of the commons and the value of celebrities' fame are neither logically analogous nor economically fungible.[103] While the tragedy of the commons theory relies on the exhaustibility of a resource, the exhaustibility of a famous individual's marketability is much more questionable. Even if one can conceive of a point at which a particular celebrity becomes "over-exposed" to the point at which his marketable value drops to zero, this point is highly speculative, and likely inextricably dependent on the celebrity in question, on the degree of the celebrity's fame, and on any number of other, external factors. Further, it is equally likely that up to that point, additional exposure *increases* the marketability of the celebrity: all publicity is good publicity, as they say.[104]

But the tragedy of the commons argument for the right of publicity becomes particularly infirm when applied to a celebrity like Antoine Dodson. Only time will tell as to the longevity of celebrity in the new media. Quite possibly, the expected "shelf life" of an Internet celebrity will conform to the same model that celebrity has always followed: Some personalities will be enduring in their popularity, while others will be "flashes in the pan." But for an individual like Dodson, elevated to instant fame by a moment of happenstance given worldwide attention via the Internet,[105] it is not unlikely that fame will be fleeting indeed. Thus, for Dodson and celebrities like him, there is likely no such thing as "over-exposure." By the time the hypothetical point of "face wearout"[106] has been reached for an individual who gains fame via an Internet video that becomes a brief sensation, the public will likely have moved on to the next Internet phenomenon. The blindingly quick turnover rate of Internet phenomena—a function of the massive deluge of information that continuously issues forth from the Internet—likely ensures that the tragedy of the commons is a peril to which Internet celebrity is immune: to return to the original analogy, there can be no tragedy of the commons when the grass is sprouting faster than the ability of any

102. *See, e.g.*, Richard A. Posner, *The Right of Privacy*, 12 GA. L. REV. 393, 411-13 (1978) ("[T]he multiple use of the identical photograph to advertise different products would reduce its advertising value, perhaps to zero."). The "tragedy of the commons" theory itself was given full explication in Garrett Hardin, *The Tragedy of the Commons,* 162 SCI., Dec. 13, 1968, at 1243, *available at* http://www.garretthardinsociety.org/articles/art_tragedy_of_the_commons.html.

103. *See* Madow, *supra* note 11, at 220-25; McKenna, *supra* note 65, at 268-75.

104. *See, e.g.*, Madow, *supra* note 11, at 221-22 ("A Madonna T-shirt may be worth *more*, not less, to consumers precisely because millions of her fans are already wearing them. The value of the T-shirt may be greater just because 'everybody's got one.'").

105. *See supra* notes 56-59 and accompanying text.

106. Madow, *supra* note 11, at 222.

number of grazing animals to consume it.

On the other hand, it is a fair accusation that if celebrities are allowed a monopoly on their likenesses, they can charge whatever they like in the absence of competing unofficial merchandise, with the likely result being that celebrity merchandise becomes more expensive. However, while some critics of the right of publicity cite this consequence as cause for alarm in and of itself,[107] it is fair to ask: of just what consequence is this consequence? How concerned should the law be about consumers being forced to pay more for a Madonna shirt or a Michael Jordan poster than they otherwise would in a competitive marketplace? One can ascribe the maximum importance one wishes to celebrities as cultural touchstones and still inevitably conclude that celebrity merchandise is not an indispensible commodity in our society.

Further, as applied to celebrities such as Antoine Dodson, the argument that the right of publicity fosters market imbalance very nearly functions as an argument in favor of the right. Again, only time will tell if the Internet will produce celebrities with long-lasting careers, but the ephemeral nature of the Internet phenomena that produce celebrities like Dodson would seem to augur against the Internet producing long, economically-productive media careers. For Dodson, and other celebrities born of fleeting Internet phenomena, the monopolistic power granted them by the right of publicity may be the only way for them to capitalize in any meaningful manner from their sudden notoriety. To shift the market in favor of such individuals hardly represents Lee Goldman's postulated result of giving yet more money to Madonna or Michael Jordan.[108] Dodson is scarcely someone one would begrudge the fruits of his unexpected fame; he could be any one of us—and in the age of the new media, any one of us could be him.

IV. THE NEW CONVERSATION

So what remains of the debate over the propriety of the right of publicity in the Internet age? Is the right of publicity more or less justifiable in an age in which one can become famous literally overnight, if only for a few weeks (or less), literally without trying?

The foremost change in the debate is that after weathering decades of derogation as law made only to benefit the rich and famous, the right of publicity will function in the Internet age as a surprisingly populist doctrine. This is not because of any expansion of the applicability of the right itself—the right was always universally applicable in theory. Rather, *fame* in our society has expanded.[109] Had Antoine Dodson's on-camera rant occurred ten years ago, it is

107. *See* Goldman, *supra* note 9, at 620.

108. *See supra* note 60.

109. For a consideration of the right of publicity as applied to individuals who achieve fame through reality television, see Porsche T. Farr, *What Good Is Fame if You Can't Be Famous in Your Own Right?: Publicity Right Woes of the Almost Famous*, 16 MARQ. INTELL. PROP. L. REV. 467 (2012).

conceivable that his distinctive visage may have been used without his permission in a local advertisement. If it had, Dodson could have brought a claim for a violation of his right of publicity. However, the damages to which he would have been entitled would have been nominal at best; Dodson would have been a marketable personality only in the viewing area of his local newscast, his opportunities to market himself limited to those to which his local notoriety (if any) would give rise in that geographic area. Today, in a media age in which Dodson's rant could be seen around the world mere hours after it occurred, and turned into a marketable song in only a day with little or no capital investment, Antoine Dodson's putative market value as a personality is that of a bona fide national—or even international—celebrity. It could happen to any of us—making any of us a potential claimant for substantive damages under the right of publicity.

As its applicability expands, so too will the importance and visibility of the right of publicity. One must concede this to the critics of the right of publicity: Humanist concerns aside, the right traditionally operated financially as a protection to those who least needed financial protection. As such, the right of publicity made good grist for academic debate, but the limited, rarified constituency to which it had any practical application ensured that it was less than a vital concern of wider public or judicial debate. That will likely change, as the Internet provides the opportunity for anyone with a camcorder and a phone line to potentially make his personality a product. For a celebrity like Antoine Dodson, the right of publicity may well be *the* most important business law on the books.[110]

Increasing prominence of the right of publicity and the inevitable, concomitant increase in its invocation as a legal remedy cannot help but add voices to the long-standing call for a preemptive federal right of publicity. What would such a federal right of publicity law look like? Ironically, with most of the post-*Haelan Laboratories* arguments both for and against the right rendered moot by the changed nature of fame wrought by the new media (the incentive rationale for intellectual property—elevated from baseless justification to cogent justification—being an exception), it might look very much like the right articulated in *Haelan Laboratories*: a legal protection encompassing elements of both a property right and a civil right; a unique, and uniquely modern, legal doctrine tailored to address a distinct, emergent legal concern.[111] In fact, perhaps

110. It is not mere idle speculation to consider the potential for violation of Antoine Dodson's right of publicity: rival, non-licensed merchandise bearing Dodson's image and quotes from his famous declamation can be found for sale on multiple Internet sites. *See, e.g.*, *Bed Intruder Gifts*, ZAZZLE.COM, http://www.zazzle.com/bed+intruder+gifts (last visited Apr. 6, 2012); *"Antoine Dodson" Gifts*, CAFEPRESS, http://shop.cafepress.com/antoine-dodson (last visited Apr. 6, 2012). Dodson himself complained publicly about unauthorized "Bed Intruder" costumes offered for sale in competition with his own Halloween costume issued in his image. *See* Robert Quigley, *Antoine Dodson Not Amused by Unauthorized Halloween Costumes*, GEEKOSYSTEM (Sept. 27, 2010, 10:50 AM), http://www.geekosystem.com/antoine-dodson-halloween-costumes.

111. For proposals as to the possible shape of a federal right of publicity, see Haemmerli,

the key to defining the proper scope of the right of publicity in the future is to cease to circumscribe the right under the umbrella of intellectual property, and begin to view it instead as equally, or even predominantly, a civil right, based on the sanctity of one's dominion over one's own person.

If the right of publicity is adopted as federal law, or if it gains wider recognition in state law, yawning doctrinal questions remain to be addressed. The most substantive of these unresolved issues is the descendibility of the right. Because the right is at once both a property right (typically descendible) and a civil right (typically vested in the individual, and not surviving him), the question of whether an individual's heirs or assigns should be able to continue to exercise the individual's right of publicity after his death will necessarily be a question of public policy—and in all likelihood, spirited debate. If the right is defined as descendible, related questions as to how long an individual's heirs or assigns may continue to exercise the individual's right of publicity, and whether the descendibility of an individual's right of publicity should be dependent upon whether the individual exercised the right during his lifetime, will have to be resolved. While these issues will, if anything, likely be less salient in the practical application of the right to celebrities like Antoine Dodson (given the reasonable likelihood that Internet celebrity will confer marketability of a particularly fleeting nature), these details will keep the debate over the contours of the right of publicity burning even once the right's justification in the Internet era is established beyond question.

CONCLUSION

Invocation of the right of publicity by Internet celebrities will serve as an overdue reminder that the right of publicity shares with the right of privacy an underlying concern with protecting personal autonomy and the sanctity of the self. As the barriers to fame fall ever lower in concert with the barriers to media creation and distribution in the digital age, the right will become a vital legal protection for an increasingly broad and diverse constituency. Consequently, the right will be transformed from an esoteric subject of academic debate to a legal doctrine on the forefront of legal and economic discourse.

This ensures one thing above all else: The debate over the justifiability of the right of publicity is over. The right of publicity is, and will remain, as practical and as necessary a legal doctrine in today's Internet age as it was in the age of baseball cards.

---

supra note 51, at 477-92 and Whaley, *supra* note 8, at 267-82.