UNIVERSITY OF CHICAGO
UNDERGRADUATE LAW REVIEW

VOLUME IV, ISSUE 3

SPRING 2015

UNIVERSITY OF CHICAGO
UNDERGRADUATE LAW REVIEW

Sponsored by the Student Government Finance Committee

Affirmation Bias? Supreme Court Decision-Making in the Circuit-Riding Century

Will Giles†

This article quantitatively analyzes the effect of circuit riding on judicial decision-making, particularly affirmation rates. From 1790 to 1891, the Supreme Court Justices served as circuit judges as well as Justices, a system that allowed for the possibility of the Justices reviewing their own circuit opinions on appeal to the Supreme Court. Despite encapsulating the formative period of the Court, the "Circuit-Riding Century" is understudied in the literature, and there is a dearth of materials relating to the practice. I introduce three theoretical explanations for the effect of circuit riding on Supreme Court affirmation rates: the circuit Justice's fear of reversal by his brethren, protecting the legitimacy of the Court, and the circuit Justice as an advocate. All three explanations lead to the same hypothesis: cases heard by Justices on circuit were more likely to be affirmed by the Supreme Court than cases that do not feature a Justice. In order to test my hypothesis, I constructed two groups, the experiment group (containing cases heard by Justices on the circuit level) and the control group (consisting of cases from state courts and the D.C. Circuit, which did not have a Justice assigned) in order to do a Welch's t-test on the difference of affirmation rates between the two groups. I then used three case studies to elucidate my theoretical explanations. Though the difference between affirmation rates is not statistically significant, this thesis is the first study to provide a concrete, positive answer to the speculated effect of circuit riding on Supreme Court decision-making.

I. Introduction

A central component of modern judicial ethics is the recusal of a judge or Justice from a case in which she has a conflict of interest. For example, newly-appointed Justice Elena Kagan recused herself from the cases she helped develop in her previous role as Solicitor General in the Obama Administration. The Supreme Court's decision in *Caperton v. A. T. Massey Coal Co.*[1] held that a West Virginia state Supreme Court Justice should have recused himself,, due to the probability of bias resulting from a conflict of interest, from a case as a matter of due process. Scholars, such as Leslie Abramson,[2] Debra Bassett,[3] Jeremy Miller,[4] and Margaret McKeown[5] have argued recusal is necessary for impartial and unbiased judgment.

This modern prevalence of judicial recusal has not always been the norm. From 1790 to 1891, Supreme Court Justices had to "ride circuit."[6] Justices were each assigned to a federal circuit and were required, along with a district judge from that circuit, to sit in panel as circuit judges. If a circuit case, heard by a sitting Supreme Court Justice, was appealed to the Supreme Court, the Justice could have heard the case again, potentially affecting the outcome of the Court's decision. "[T]he structure of the original court system was flawed in conception and hence problematic from the outset…Though a district judge, when sitting on a circuit court, could not vote on an appeal from his own decision, no similar provision prevented a Supreme Court Justice from voting on an appeal from a circuit court decision in which he had participated," a practice that would be considered an anathema by judges today.[7] John Frank notes that it was an "odd system…insofar as it permitted appeals from the circuit court to the Supreme Court, it permitted appeals

---

†: Will Giles graduated from Duke University in 2015 with a B.A. in Political Science.

1: 556 U.S. 868 (2009).

2: See Leslie W. Abramson, *Deciding Recusal Motions: Who Judges the Judges,* 28 Val. U. L. Rev. 543, 543-61 (1994).

3: See Debra Lyn Basset, *Recusal and the Supreme Court,* 56 Hastings L. J. 657 (2005).

4: See Jeremy M. Miller, *Judicial Recusal and Disqualification: The Need for a Per Se Rule on Friendship (Not Acquaintance),* 33 Pepp. L. Rev. 575 (2005).

5: M. Margaret McKeown, *To Judge or Not to Judge: Transparency and Recusal in the Federal System,* 30 Rev. Litig. 653 (2011).

6: The terms "riding circuit" (verb), "circuit riding" (noun), and "circuit-riding" (adjective) refer to the same practice but are grammatically distinct.

7: Daniel J. Meltzer, *The Judiciary's Bicentennial,* 56 U. Chi. L. Rev. 423, 424 (1989).

from, among others, [Circuit Judge Philip Pendleton] Barbour to [Supreme Court Justice Philip Pendleton] Barbour."[8] Though some Justices chose to recuse themselves so on their own accord, like Chief Justice John Marshall did for *Stuart v. Laird,*[9] most Justices did not do so in appeals from lower court decisions in which they participated.

II. History

Article III of the Constitution created the judicial branch, the last to be established by the Framers. Section One mandated that"[t]he judicial Power of the United States, shall be vested in one Supreme Court," while providing Congress with the power to create "such inferior Courts as [it] may from time to time ordain and establish."[10] Immediately following ratification, Congress used this power to pass the Judiciary Act of 1789.[11] The Act, written primarily by future Chief Justice Oliver Ellsworth, set the number of Justices of the Supreme Court at six—one chief Justice and five associate Justices—while also creating 13 judicial districts in 11 states.[12] Each judicial district consisted of both a circuit court and a district court.

The Act provided congressional funding for district court judges for each of the districts but did not provide separate funding for circuit court judges. Instead, two of the recently-appointed Supreme Court Justices were also given circuit duties in each of the three circuits: the Southern, Middle, or Eastern Circuit. Chief Justice Charles Evans Hughes later reasoned, "At the onset it was expected that through their circuit work the Justices of the Supreme Court would be in close contact with the people. The Justices left their impressions upon the communities they visited, and these communities had their effect upon the Justices."[13] Other explanations for the establishment of this "odd system" include budget savings, the importance of having Supreme Court Justices adjudicate in federal cases, the spread of the tenets of federalism (as the Justices would quite possibly be the only federal officers that citizens

8: John P. Frank, Justice Daniel Dissenting: A Biography of Peter V. Daniel, 1784–1860 143 (1964).

9: 5 U.S. 299 (1803).

10: See U.S. Const. Art 3 § 1.

11: The Judiciary Act of 1789 Ch. 20, 1 § 73.

12: Districts for North Carolina and Rhode Island, which had not ratified the Constitution at the time, were added as these states entered the Union.

13: See *Remarks of Chief Justice Hughes at the Judicial Conference of the Federal Judges of the Fourth Circuit, Held at Asheville, North Carolina, on Thursday Morning, June 9, 1932* Minutes of the Second Annual Meeting of the Federal Judicial Conference of the Fourth Circuit 3 (1932).

would meet), ensuring the uniformity of federal law, and unifying the judicial branch.[14] Eventually, as Congress added new circuits, a single Justice was assigned to ride circuit within that circuit.[15]

An early case, *Stuart v. Laird*, challenged the constitutionality of circuit riding itself. The case came to the Supreme Court by a writ of error from Chief Justice Marshall's Fifth Circuit.[16] The case concerned the Judiciary Act of 1802,[17] which repealed the Judiciary Act of 1801.[18] The 1801 Act provided separately-commissioned circuit Justices, thus removing the burden of circuit riding from the Justices' plates. However, the 1801 Act was passed by the lame-duck Federalist Party on the eve of Thomas Jefferson's inauguration and signed into law by President John Adams. The new Democratic-Republican majority in Congress subsequently passed the 1802 Act, reinstating the practice of circuit riding, while also restructuring the federal judicial system into six circuit courts. The Marshall Court eventually ruled the Repeal Act to be constitutional, thus endorsing the practice of circuit riding and committing to carry out the practice for years to come.[19]

The Justices spent a considerable amount of time riding circuit, traveling thousands of miles per year. The job was arduous. Justice Thomas Todd's circuit duties forced him to be absent for five of the nineteen terms he was on the Supreme Court, and Justice William Paterson died from injuries sustained from a carriage accident *en route* to hold court on circuit.[20] In an attempt to alleviate the burden of riding circuit, Congress passed the Judiciary Act of 1869,[21] creating nine circuit judgeships, or one for each circuit. The Act significantly reduced the Justices' time on circuit, requiring each Justice to hold court within his circuit once every two years.

The 1869 Act did not completely alleviate the Justices of their circuit-riding duties, however. This finally occurred with the passage of the Judiciary

---

14: Joshua Glick, *On the Road: The Supreme Court and the History of Circuit Riding*, 24 Cardozo L. Rev. 1753, 1762 (2002), and see Jean Edward Smith, John Marshall: Definer of a Nation (1996).

15: Congress eventually established ten circuits in the 1860s, with ten Justices needed to ride those circuits. Congress reduced the number of Justices to nine by the end of the decade, and the number of Justices has remained constant to this day.

16: As mentioned previously, Chief Justice Marshall recused himself from the proceedings once the case reached the Supreme Court. 5 U.S. at 299.

17: Also known as the Repeal Act. See Judiciary Act of 1802, Ch. 8, 2 § 156.

18: See Judiciary Act of 1801, Ch. 4, 2 § 89.

19: The case will be further discussed in the Case Study section, below.

20: Glick, 24 Cardozo L. Rev at 1798 (cited in note 14).

21: See Judiciary Act of 1869, Ch. 22 16 § 44.

Act of 1891,[22] which created the Circuit Courts of Appeals. The Act shifted the appellate caseload burden from the Supreme Court and old circuit courts to the new Courts of Appeals and made the federal district courts the judicial system's primary trial courts. With two circuit judges and a district judge from that circuit sitting on these newly established courts, the Supreme Court's "Circuit-Riding Century" came to an end.[23]

III. Literature Review of Circuit Riding

The "Circuit-Riding Century" was a significant period of the Court's history. Despite its contentious nature, the effect of circuit riding has not been fully explored in the literature. As noted earlier, there have been a number of scholars who have mentioned the possibility of a conflict of interest by having Supreme Court Justices ride circuit, but these scholars never attempted to expand upon this relationship by comparing circuit court and Supreme Court opinions themselves.[24] The literature mainly emphasizes the hardships that the Justices faced while riding circuit. The physical toll of traveling a large number of miles on unpaved roads, coupled with the emotional toll of being away from home, led Bernard Schwartz to label circuit riding as "the great albatross of the early Supreme Court."[25] Joshua Glick, noting the dearth of articles focusing exclusively upon the topic, provides a detailed history of court riding, as well as a useful bibliography, but fails to extend his analysis beyond a historical framework.[26] David Stras, Steven Calabresi, and David Presser argue for the reinstatement of circuit riding, as they postulate that it will give Justices more familiarity with local laws and statutes, something not

---

22: See Circuit Court of Appeals Act 25 § 826. Commonly referred to as the "Evarts Act," after the bill's author, Senator William Evarts of New York.

23: The old circuit courts were officially abolished by the Judicial Code of 1911. Today, each Justice is still assigned to a particular circuit, but this is merely anachronistic. As Justice Sandra Day O'Connor notes, "In essence, the [current] duties of the Circuit Justice typically range from consideration of routine requests for relief from formal filing requirements to stays of a lower court's mandate in controversial civil cases and of capital sentences." See Sandray Day O'Connor, *Foreword: The Changing Role of the Circuit Justice,* 17 U. Tol. L. Rev. 521, 526 (1986).

24: See, for example, Frank, Justice Daniel Dissenting (cited in note 8); Meltzer, 56 U. Chi. L. Rev. at 423 (cited in note 7); and O'Connor, 17 U. Tol. L. Rev. at 521 (cited in note 23).

25: Bernard Schwartz, A History of the Supreme Court 30 (1993). Appendix I contains a table of circuit distances travelled in 1839.

26: Glick. 24 Cardozo L. Rev at 1798 (cited in note 14).

possible under the current system that cloisters Justices in Washington.[27]

The studies by Kevin McGuire, Richard Vining, Christopher Zorn, and Susan Smelcer are perhaps the only quantitative studies to encapsulate this early period of the Court's history.[28] As McGuire acknowledges, "[L]ongitudinal knowledge of the Supreme Court is quite limited."[29] He finds that the early Court increased its institutional power only gradually, hypothesizing that the institution would be hardly recognizable to observers today. In their study of the determinants of resignation from the Supreme Court, Vining *et al* state, "As circuit riding was expected of all Justices in this period on an annual basis, our data provide little opportunity to assess the effect of its presence or absence among Justices' duties."[30] Despite focusing on the same period as this Article, these studies do not focus on the institution of circuit riding itself; Vining *et al* use the practice as a variable that possibly contributed to resignation from the Court in their multivariate analysis, while McGuire[31] cites circuit riding as an impediment to the Court's differentiation, an important step in institutional building.

As evidenced, these authors do not attempt to analyze the affirmation rates of the Justices riding circuit. Justices produced circuit opinions that could have been appealed to their brethren on the Supreme Court, and then, in turn, affected the Supreme Court's decisions and the content of American law. What influence did Justices riding circuit have upon Supreme Court decision-making? Would the decisions reached in seminal cases such as *Marbury v. Madison*,[32] *McCulloch v. Maryland,*[33] *Gibbons v. Ogden,*[34] and others have turned out differently if the Justices had not ridden circuit? This paper, thus, will quantitatively analyze circuit riding, seeking to discern the effect that circuit riding had upon Supreme Court affirmation rates. Were cases heard by Justices when they were on circuit more likely to be affirmed

27: See David R Stras, *Why Supreme Court Justices Should Ride Circuit Again,* 91 Minn. L. Rev. 1710 (2006), and Steven G. Calabresi & David C. Presser, *Reintroducing Circuit Riding: A Timely Proposal,* 90 Minn. L. Rev. 1386 (2005).

28: See Kevin T. McGuire, *The Institutionalization of the U.S. Supreme Court,* 12 Pol. Analysis 128 (2004), and Richard L. Vining, Christopher Zorn, & Susan Navarro Smelcer, *Judicial Tenure on the U.S. Supreme Court, 1790-1868: Frustration, Resignation, and Expiration on the Bench* 20 Stud. Amer. Pol. Dev. 198 (2006).

29: McGuire, 12 Pol. Analysis at 128 (cited in note 28).

30: Vining, Zorn, & Smelcer, 20 Stud. Amer. Pol. Dev. at 206 (cited in note 28).

31: McGuire, 12 Pol. Analysis at 130 (cited in note 28).

32: 5 U.S. 137 (1803).

33: 17 U.S. 159 (1819).

34: 22 U.S. 1 (1824).

than cases in which a Justice did not participate?

IV. Theory

The possibility of Justices' circuit court opinions being affirmed or reversed by their brethren on the Supreme Court may have had important implications for the decisions made by Justices, both on the circuit courts and the Supreme Court, from the period of 1790 to 1891. The Justices could have attempted to anticipate their brethren's policy positions and augment their circuit decisions. The Justices also could have resented having to travel so many miles on circuit and produced opinions that reflected their exhaustion and resentment. To tease out the possible implications, we must delve into the literature on judicial decision-making to set up our theoretical framework. The attitudinal model, espoused by Jeffrey Segal and Harold Spaeth, posits Justices as policymakers, rather than impartial "umpires" of the law.[35] In this model, Justices decide cases based on their own personal policy preferences, not upon the letter of the law. The letter of the law is used by the Justices as a means to achieving their preferred outcome, not as an end itself.

Building upon the attitudinal model, Lee Epstein and Jack Knight model the Justices as strategic actors, stating that Justices "are not unrestrained actors who make decisions based only on their own ideological attitudes. Here, Justices are strategic actors who realize that their ability to achieve their goals depends on a consideration of the preferences of other actors, the choices that they expect others to make, and the institutional context in which they act."[36] Though originally applied to the Supreme Court, the strategic model accurately describes the relationship between the circuit courts and Supreme Court during the period of circuit riding; Justices had to consider the "preferences" of their Supreme Court brethren and take into account the "choices" they were expected to make under the "institutional framework" of riding circuit.

What might we expect the effect of circuit riding to have been on how Supreme Court Justices decided cases when on circuit? Does the fact that a Supreme Court Justice heard a case on circuit, which was later appealed, affect how the Supreme Court reviewed cases from those circuit courts? There are three possible theoretical explanations for circuit riding's influence upon judicial decision-making. The first, which will be deemed the

35: See Jeffrey A. Segal & Harold J. Spaeth, The Supreme Court and the Attitudinal Model (1993), and The Supreme Court and the Attitudinal Model Revisited (2002).

36: See Lee Epstein & Jack Knight, The Choices Justices Make 10 (1998).

"Hughes Theory," suggests that the Justices' decision-making on the circuit level was meaningfully constrained. The fear of reversal by their Supreme Court brethren led Justices to strategically align their circuit opinions to the policy preferences of their fellow Justices, despite possibly not sharing those preferences. Second, the "Jay Theory" looks to the level of the Supreme Court. Perhaps due to its tenuous position in the new federal government, the Supreme Court should have avoided reversing Justices' circuit opinions, as reversal could have brought the Court's legitimacy into question. Third, in the "Marshall Theory," a Justice, having already heard the case on circuit, became an advocate for his lower court decision on appeal. The close interaction between the Justices allowed the Justice, whose circuit decision was being debated, to use his expertise and familiarity of the case to convince his colleagues to affirm his circuit reasoning.

IVa. The Hughes Theory

There is a vast literature on "reversal aversion," a lower court judge's fear of being reversed by a higher court. David Rohde, Harold Spaeth, and Jeffrey all suggest that if the inferior federal courts did not have to fear reversal, lower court judges would be freer to make decisions based on their policy preferences.[37] Shep Melnick postulates that the primary reason lower court judges do not vote their personal preferences more often is the possibility of review by a higher court.[38] A major component of the theory of Lee Epstein, William Landes, and Richard Posner is the reviewability of Courts of Appeals' decisions by the Supreme Court.[39] Students of judicial behavior commonly argue that judges do not like to have their decisions reversed. [40] Being overturned implies that a judge, or panel of judges at a higher level, has

37: See David W. Rohde & Harold J. Spaeth, Supreme Court Decision-making (1976) and Segal & Spaeth, The Supreme Court and the Attitudinal Model (cited in note 35) and The Supreme Court and the Attitudinal Model Revisited (citd in note 35).

38: See Shep R. Melnick, Regulation and the Courts: The Case of the Clean Air Act (1983).

39: See Lee Epstein, William M. Landes, & Richard A. Posner The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice (2013).

40: See Lawrence Baum, *Court Response to Supreme Court Decisions: Reconsidering a Negative Picture* The Justice System Journal, 3 Just. Sys. J. 208 (1978); Evan H. Caminker, *Why Must Inferior Courts Obey Superior Court Precedents?,* 46 Stan. L. Rev. 817 (1994); and McNollgast, *Politics and Courts: A Positive Theory of Judicial Doctrine and the Rule of Law,* 68 S. Cal. L. Rev. 1631 (1995).

determined the lower-court judge applied the law incorrectly. Walter Murphy notes that "Judges, no more than other men, enjoy the prospect of public correction and reprimand."[41] Similarly, Posner asserts that judges are "highly sensitive to being reversed."[42]

A series of studies have indicated judges on lower courts adjust their behavior to accommodate changes in higher courts.[43] Jean Smith provides some empirical support for the hypothesis that district court judges do react to reversals by avoiding the type of decision that was reversed.[44] Donald Songer, Jeffrey Segal, and Charles Cameron find that appeals court decisions were positively related to both the ideology of the appeals court judges and to changes in the current ideology of the Supreme Court.[45] Kirk Randazzo also finds evidence that liberal judges reach more conservative decisions, while conservatives come to more liberal decisions, than would otherwise be expected without higher court review.[46]

In the case of the Court's circuit-riding period, the Justices not only had to face the fear of reversal, but this reversal could have come at the hands of their colleagues. Chief Justice Hughes made special note of this phenomenon: "Not only did Justices of the Supreme Court on circuit duty suffer the extreme hardships of travel in the early days, but they were exposed to the even greater peril of subsequent reversal at the hands of their brethren."[47] Chief Justice Hughes, though perhaps being hyperbolic, clearly thought circuit riding may have affected how Justices reached decisions on circuit. Jonathan Macey characterizes reversals as "embarrassing;"[48] it is not hard to imagine a reversal by equal peers as being the ultimate form of judicial embarrassment. An important aspect of reversal aversion was the possible damage to a Justice's reputation. Individual Justices are concerned about their reputation as great

---

41: See Walter F. Murphy, Elements of Judicial Strategy 1030 (1964).

42: See Richard A. Posner, The Problems of Jurisprudence 224 (1990).

43: See Donald R. Songer, Jeffrey A. Segal, & Charles M. Cameron, *The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions,* 38 Am. J. Pol. Sci. 673 (1994), and Charles M. Cameron, Jeffrey A. Segal, & Donald Songer, *Strategic Auditing in a Political Hierarchy: An Informational Model of the Supreme Court's Certiorari Decisions*, 94 Am. Pol. Sci. Rev. 101 (2000).

44: See Smith, John Marshall (cited in note 14).

45: See Songer, Segal, & Cameron, 38 Am. J. Pol. Sci. at 673 (cited in note 43).

46: See Kirk A. Randazzo, *Strategic Anticipation of the Hierarchy of Justice in U.S. District Courts,* 36 Amer. Pol. Research 669 (2008).

47: *Remarks of Chief Justice Hughes at the Judicial Conference of the Federal Judges of the Fourth Circuit* at 2 (cited in note 13).

48: See Jonathan R. Macey, Internal and External Costs and Benefits of *Stare Decisis* 111 (1989).

legal minds and scholars.[49] Thus, the Justices should have strategically aligned their circuit opinions with the policy preferences of their brethren on the Supreme Court in order to avoid the "great peril" of being reversed by their peers on the Court and damaging their legal reputations.

Did the district court judges, the other half of the circuit court panels during this period, not have to worry about being reversed and not have to adjust their decisions accordingly? Should the affirmation rates for both district court judges and Supreme Court Justices, thus, not have been the same? Supreme Court Justices held an advantage over their circuit court colleagues: each Justice had repeated interaction with the other Justices upon the Supreme Court. This interaction allowed the Justices to better predict and anticipate the preferences of their brethren. A similar phenomena has been found in the use of *en banc* hearings by the Circuit Courts of Appeals. Michael Giles and others suggest that "Majority-minority panels may conceal their preferences and conform their decisions to the preferences of the circuit majority, thus avoiding *en banc* rehearing."[50] That is, members of a three-judge panel, due to their familiarity with the rest of judges on their circuit, strategically write their opinions in order to avoid review by the whole circuit. The district court judges during this time did not have these repeated interactions with the body, i.e. the Supreme Court, that would ultimately review their circuit opinions and were less adept at determining the Justices' policy preferences than the circuit-riding Justices.

IVb. The Jay Theory

Chief Justice John Jay expressed his concern about circuit riding in a letter to Congress. He worried that the "distinction made between the Supreme Court and its Judges, and appointing the same men finally to correct in one capacity, the errors which they themselves may have committed in another, is a distinction unfriendly to impartial Justice, and to that confidence in the Supreme Court, which it is so essential to the public Interest should be reposed in it."[51] His main argument was, by having the Justices review their

---

49: See Posner, The Problems of Jurisprudence (cited in note 42); Thomas J. Miceli & Metin M. Cosgel, *Reputation and Judicial Decision-Making,* 23 J. Econ. Behavior & Organization 31 (1994); and Frederick Schauer, *Incentives, Reputation, and the Inglorious Determinants of Judicial Behavior,* 68 U. Cincinnati L. Rev. 615 (2000).

50: See Michael W. Giles, Virginia A. Hettinger, Christopher Zorn, & Todd C. Peppers, *The Etiology of the Occurrence of En Banc Review in the U.S. Court of Appeals,* 51 Am. J. Poli. Sci. 449, 452 (2007).

51: *Remarks of Chief Justice Hughes at the Judicial Conference of the Federal Judges of the*

own circuit opinions, circuit riding would be ill-received by the American public and bring the legitimacy of the Court into question. During the circuit-riding period, the Court had little to no institutional power, as Congress controlled the number of Justices on the Court as well as its jurisdiction, while the President appointed the Justices and was tasked with enforcing judicial decisions. One way to interpret Chief Justice Jay's comment is that circuit riding undermined what little legitimacy the Court held in the early years of the new federal system.

The legitimacy of the Court has always been a concern amongst the Justices. It has been shown that the Supreme Court does not operate in a political vacuum; rather, the Court must take into consideration the policy preferences of the other branches[52] and often acts to avoid institutional conflict to protect its legitimacy.[53] According to David O'Brien, circuit riding was "not merely burdensome; it also diminished the Court's prestige, for a decision by a Justice on circuit could afterward be reversed by the whole Court."[54] With O'Brien, Chief Justice Jay's main concern centered on the possibility of the Supreme Court having to correct a Justice's circuit opinion by reversal. The reversal of a

---

Fourth Circuit at 163-64 (cited in note 13).

52: For example, see Murphy, Elements of Judicial Strategy (cited in note 41); John Ferejohn & Charles Shipan, *Congressional Influence on Bureaucracy,* 6 J. L. Econ. & Organization 1 (1990); Pablo T. Spiller & Rafael Gely, *Congressional Control or Judicial Independence: The Determinants of U.S. Supreme Court Labor-Relations Decisions, 1949-1988,* 23(4) RAND J. Econ. 463 (1992); Robert Lowry Clinton, *Game Theory, Legal History, and the Origins of Judicial Review: A Revisionist Analysis of Marbury v. Madison,* 38 Am. J. of Pol. Sci. 285 (1994); and Epstein & Knight, The Choices Justices Make (cited in note 36).

53: See Gregory A. Caldeira, *Public Opinion and the U.S. Supreme Court: FDR's Court-Packing Plan,* 81 Am. Pol. Sci. Rev. 1139 (1987); William Lasser, The Limits of Judicial Power: The Supreme Court in American Politics (1988); William P. Marshall, *Be Careful What You Wish For: The Problems with Using Empirical Rankings to Select Supreme Court Justices,* 78 S. Cal. L. Rev. 119 (2004); Matthew C. Stephenson, *Court of Public Opinion: Government Accountability and Judicial Independence,* 20 J. L. Econ. & Organization 379 (2004); Jeffrey K. Staton & Georg Vanberg, *The Value of Vagueness; Delegation, Defiance, and Judicial Opinions,* 52 Am. J. of Pol. Sci. 504 (2008); Clifford James Carrubba, *A Model of the Endogenous Development of Judicial Institutions in Federal and International Systems,* 71 J. Pol. 55 (2009); Barry Friedman, The Will of the People: How Public Opinion Has Influenced the Supreme Court and Shaped the Meaning of the Constitutional (2009); and Thomas S. Clark, The Limits of Judicial Independence (2011).

54: David M. O'Brien, Storm Center: The Supreme Court in American Politics 138 (1990).

fellow Justice would have endangered the perception of judicial infallibility. If a Justice could be reversed and corrected on circuit, the whole judicial system unraveled; the decisions of the Supreme Court, considered to be the final say regarding the law of the land, were then opened to debate and the possibility of being ignored by the other branches and the American public.

As stated, reversal of circuit decisions may have hurt the legitimacy of the Court, and the literature emphasizes that the Court takes precautions to avoid hurting its legitimacy. One could argue that the Court could have refused to hear cases that question its legitimacy; however, the Court did not control its docket until 1925, 34 years after circuit riding had been abolished. The Court also could not have relied upon its prestige as the ultimate arbiter of the law at this time, as it had not yet achieved this exalted status. On the other hand, the Court could have possibly acquiesced to the other branches' every whim. This, though, would have undermined the Court's goal of enhancing its legitimacy. Thus, as the only rational option, the Court should have been more likely to affirm its brethren on circuit to avoid the hurting its own legitimacy.

IVc. The Marshall Theory

The Marshall Theory proceeds as follows: the Justice, having already heard the case on circuit, was intimately familiar with the circumstances of the case, the local laws of his circuit, and the applicable precedent, all of which would have been cited in his circuit opinion. He was an expert on the particulars of the case, a conduit of information for the other Justices. In this regard, the collegial atmosphere of the Marshall Court, created by lodging in the same boardinghouse in Washington during the Court term and eating their meals together, cannot be understated. The close living situation of the Justices allowed them to better advocate and defend their circuit opinions appealed before the Supreme Court; the circuit Justice could have explained the nuances and reasoning of his decision to the other Justices, an advantage not available to the district court judges also riding circuit.

Kevin McGuire finds that a more experienced lawyer, by which he means one that appears frequently before the Court, as opposed to infrequent participants or new members of the Supreme Court Bar, significantly raises the probability of his party winning the case.[55] Repeated appearances before the Court and interactions with the Justices, it seems, breeds familiarity and trust between the advocate and the Court, making the advocate more

55: Kevin T. McGuire, *Repeat Players in the Supreme Court: The Role of Experienced Lawyers in Litigation Success,* 57 J. Pol. 187 (1996).

effective at arguing a case. During the Court's circuit riding period, the Justices themselves served a similar role, though in a different setting. The advocacy occurred in their private chambers or their shared boardinghouse, perhaps over a glass of Madeira, rather than the Supreme Court's chamber on the first floor of the Capitol.

IVd. Implications of Theory

What do these theories predict for the likelihood of a decision being affirmed? Under the Hughes Theory, the Justice on circuit should have aligned his opinion closely to the policy position of his brethren in order to avoid reversal. Recall that by the Jay theory, Supreme Court should avoid reversing their brethren's circuit opinions in order to protect the legitimacy of the Court. The Marshall Theory posits the Justices as effective advocates of their circuit opinions, leading the Supreme Court to affirm those decisions. Each theory, through different mechanisms, led to the same conclusion: cases in which a Justice sat on the circuit level should have been more likely to be affirmed than those cases in which a Justice did not sit.

Though the three theoretical explanations lead to the same conclusion, they also, conversely, would each lead to a difference in the content of the law after the decision. Under the Hughes Theory, the content of the law should have reflected the preferences of the majority of the Justices on the Supreme Court, as Justices on circuit shaped their opinions to avoid reversal by their brethren on appeal. But then, per the Jay Theory, the content of the law should have been closer to the circuit Justice's opinion, as, in order to protect its legitimacy, the Court should have affirmed the circuit Justice's opinion. Finally, the content of law under the Marshall Theory would have been located between the two ends of the spectrum, but closer to the circuit Justice's opinion. The circuit Justice's advocacy of his opinion would create a pivoting effect. Though the other Justices' preferences would have been reflected in the Court's opinion, the majority's reasoning would have always been grounded in the circuit Justice's opinion.

V. Data on Affirmation Rates

The hypothesis being tested is: "The Supreme Court is more likely to affirm a decision appealed from a fellow Justice on circuit than a decision that did not involve a Justice on the circuit level." The court cases featured in this analysis were heard by the Supreme Court in the period of 1812-1823. This period represents the longest "natural court"—defined as a period with no changes in membership of the Court—in the history of the Supreme Court. Choosing this time period provides two distinct advantages: no changes in the Court's make-up allows us to control for a number of variables (Justices' propensity to ride circuit, a change in the median Justice, etc.), and this time period presents a large number of cases to study during a time when circuit riding was more prevalent than later periods. We selected a four-year subset (1813-1817) during the natural court to study. This mitigates a number of factors, including a changing political climate over the natural court, substantively different issues reaching the Court, and territories becoming new states (and therefore adding more circuit burdens).

The dataset includes all cases heard by the Supreme Court during this period, as published in the U.S. Reports, including those that were remanded back to the circuit court or needed further proof. In order to test the hypothesis, we constructed two different groups: one group that contains Supreme Court cases heard by a Justice on circuit and the other group containing cases not heard by a Justice on circuit. To construct the first (circuit-riding) group, cases were culled based on any indication that a Justice heard a specific case on circuit, such as a Justice recusing himself based on circuit participation, a Justice referencing his circuit opinion during a dissent, or a writ of division (which requires both the Justice and district judge to disagree on circuit). If there was no apparent indication of a Justice on circuit, the remaining cases were then cross-listed with the numerous circuit reports published during the era, which garners 41 cases. Though this method is not exhaustive, it presents the best method to categorize cases. At the time, court records were notoriously incomplete; most circuit court records were not even published. If the case did not explicitly note a Supreme Court Justice on circuit and was not featured in the circuit reports, it is nearly impossible to discern whether a Justice heard the particular case on circuit, despite knowing the circuit from which the case arose. Thus, we cannot be confident that the unaccounted cases did not have a Justice on circuit.

In order to get the second group, which served as the control group, we collected cases that were appealed from the state courts and the Circuit Court of the District of Columbia. The D.C. Circuit did not have an appointed Supreme Court Justice. Rather, the court consisted of judges appointed by

the president and confirmed by the Senate (the only circuit with this feature at the time).[56] The cases that arose from state courts did not have a Supreme Court Justice sitting in on the decision, as judges on the state-level were appointed by the states' governments or elected by the voters of each state. These cases (60 in total) naturally exclude the independent variable (Justices riding circuit).[57]

The dataset is showcased in Figure I. Using the method described above, a total of 210 cases were heard during the 1813-1817 period. In 41 of these cases, a Justice sat on the circuit case. In 60 cases, no Supreme Court Justice sat on the lower level. Finally, the author could not determine whether or not a Justice sat on the circuit level in 109 cases. Eight of the circuit cases reached the Court on certificate of division, but the circuit position of the Justice could not be determined.

Figure I Supreme Court Case Dataset (1813-1817)	
Circuit-Riding Cases	41
D.C. Circuit and State Cases	60
Unaccounted Cases	109
Total Cases	210

In order to compare the affirmation rates between the circuit-riding group and the control group, a Welch's *t*-test was used. [58] This test was chosen because this paper is aiming to show that statistical evidence of the difference in affirmation rates between the two groups is not simply due to random chance, meaning riding circuit had a statistically significant effect on Supreme Court affirmation rates. In order to calculate the affirmation rates of the

56: William Cranch, who served on the D.C. Circuit from 1801 until his death in 1855 (and chief judge from 1806 onward), was the second Reporter of Decisions of the Supreme Court (from 1801 to 1815). A nephew of Abigail Adams, he was one of the "Midnight Judges" appointed under the Judiciary Act of 1801 by President John Adams. He was also the great-grandfather of T.S. Eliot. (An interesting future study would be to test whether his D.C. Circuit opinions were affirmed at a higher rate than others, due to his position as Reporter.)

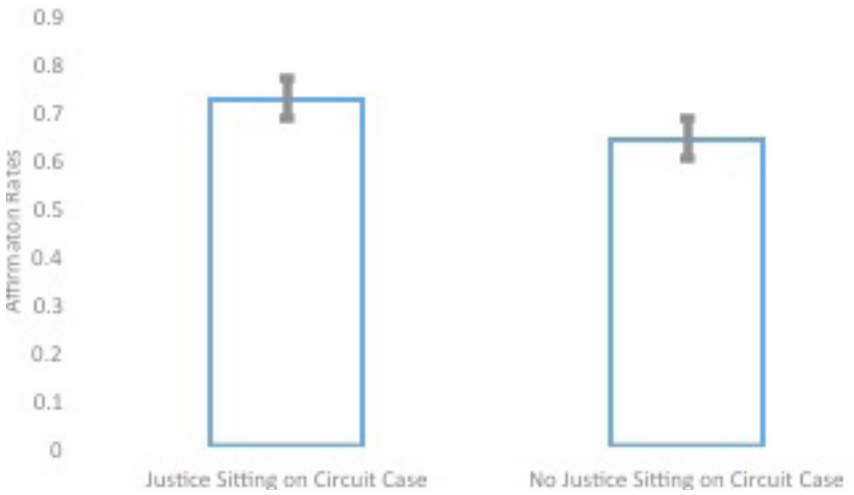
57: It should be noted that there is a heavy reliance on the D.C. Circuit cases in the control group. This occurs organically, as more cases reached the Court from that path than through the state and territorial courts.

58: A "two-sample unpooled *t*-test for unequal variances," also known as a "difference in means test."

cases, a circuit decision that was reversed by the Supreme Court was coded as "zero," while an affirmed circuit decision was coded as "one," essentially a dummy variable for Supreme Court affirmation. If the Supreme Court only reversed parts of a decision, not the entire decision, the case was coded as the proportion of affirmed clauses to overall clauses (i.e. if two clauses, out of five, of a decision were affirmed, as stated in the opinion itself, the case would be coded as ".4"); there are two of these cases in the test group and one in the control group. We then calculated the means of each group and the pooled standard error[59] and calculated a test statistic, t[60], where X_1 and X_2 are the sample means, s_1^2 and s_2^2 are the sample variances, and N_1 and N_2 are the sample sizes.

The results are featured in Figure II. Supreme Court cases featuring a Justice on circuit had an affirmation rate of 72.44% (standard error: 0.0689) while cases not featuring a Justice on circuit had an affirmation rate of 64.17% (standard error: 0.0619). The t-value associated with the Welch's t-test was -0.8937, which corresponded to a confidence interval (represented by the thick, grey lines in Figure II) of [-0.2666, 0.1012] and a p-value of 0.3738. This p-value means that we cannot reject the null hypothesis that circuit riding has no effect on Supreme Court affirmation rates.

Figure II
Comparison of Affirmation Rates



59: The formula is $SE_{pooled} = S_p[sqrt((1/N_1)+(1/N_2))]$
60: The formula is $t = (x_1 - x_2)/sqrt[(s^2_1/N_1)+(s^2_2/N_2)]$.

VI. Discussion of Results

What do the results of this test indicate for the three theoretic explanations? The explanations could simply be wrong. As it is almost impossible to discern the actual motivations behind how and why Justices make their decisions, these explanations may not factor into judicial decision-making. The aversion to reversal by other Justices (the Hughes Theory) may not have been as strong as stated. Also, the judges on the D. C. Circuit and state courts may have been equally capable as Justices of predicting the policy preferences of the Supreme Court. Though the Justices seemed to be worried about the legitimacy of the Supreme Court (the Jay Theory), this could have manifested itself in other considerations besides not wanting to reverse a fellow Justice on circuit; the Justices could have voted to reverse a circuit Justice's sound legal opinion if it was at odds with the policy preferences of the other branches, therefore lowing affirmation rates. Finally, the advocacy role (the Marshall Theory) of the circuit Justice in front of his brethren may not have been as effective as stated, as the Justices had the full case report in front of them, allowing a holistic look at the case at hand.

Or, circuit riding did matter for decision-making on the Supreme Court, but the results of our test did not translate well. While circuit riding could have had an effect on affirmation rates, it may not be the only explanatory variable. One confounding variable could be the ideological division of Supreme Court; the Court could have been deeply split along ideological lines, no matter which Justice was riding circuit. We also did not control for the composition of lower panels, whose relationships between district judges and circuit Justices could have varied greatly. Another possible confounding variable is similar issues could have been arising from different circuit courts at the same time. If the circuits differed in their proscribed outcome, the Court would have had to choose a side, meaning one Justice necessarily would have been reversed. Finally, we did not control for the total distance travelled by Justices on circuit. Spending more time on the road, traveling to each court in one's circuit, could have meant having less time to craft high quality opinions than those Justices riding smaller circuits; these lesser-quality opinions could have been reversed at a higher rate by the Supreme Court. In an ideal world, we would conduct a multivariate analysis to tease out these confounding factors, but we are limited to a simple test of comparing affirmation rates across two groups. Thus, within the confines of this Article, we were not able to test for these confounding variables.

Also, the data collected could not be very reflective of the circuit-riding phenomenon, as it is very incomplete. As mentioned earlier, court reporting was notoriously inconsistent during the nineteenth century; accordingly, in

over half of the cases in the dataset (109 cases), it could not be determined whether a Justice sat on the case on the circuit level. Furthermore, the circuit-riding group is overwhelmingly reliant upon cases from the First Circuit and decisions by Justice Joseph Story, whose records appear to be more reliable than those of other circuits (30 out of 41 cases). The cases heard in the First Circuit, with its multitude of shipping and cargo condemnation (libel) cases, could have also been substantively different from, say, cases dealing with western land sales arising from the Seventh Circuit.

As noted, the difference between Supreme Court affirmation rates in circuit cases featuring a Justice and affirmation rates for circuit cases not featuring a Justice was not statistically significant. Thus, with the possibility of having incomplete or unreflective data, several case studies will be showcased to attempt to flesh out the theoretical explanations. Due to the relatively obscure nature of these cases and the inventiveness of this study, the best approach is an in-depth reading of both the circuit and Supreme Court opinions of the following cases: *Stuart v. Laird*,[61] *Evans v. Jordan and Morehead*[62] *and Jones vs. Shore's Executors*,[63] and *Haney v. the Baltimore Steamship Company.*[64] Delving into these cases allows us to explore, examine, and illustrate the theoretical explanations in the Justices' own words.

VIa. *Stuart v. Laird*

Ironically, the constitutional test of circuit riding arose from the practice itself, as Marshall heard *Stuart v. Laird* on circuit. The case placed the Marshall Court in a quandary. Ruling against the constitutionality of the act would have removed the onerous duty of riding circuit, something that the Court would have preferred. However, this action would have incurred the wrath of the Jefferson Administration and the Democratic-Republican-controlled Congress only days after the momentous decision in *Marbury v. Madison*. In a letter, Marshall inveighed, "This is a subject not to be lightly resolved on. The consequences of refusing to carry the law into effect may be very serious… The law having been once executed will detract very much in the public estimation from the merit or opinion of the sincerity of a determination, not now to act under it."[65] Obviously, Marshall was very concerned with

61: 5. U.S. 299 (1803).

62: The appellate decision is reported at 8 F. Cas. 872 (1813) and the Supreme Court's decision is reported at 13 U.S. 199 (1815).

63: The appellate opinion can be found at 26 F. Cas. 638 (1814), and the Supreme Court's decision is reported at 14 U.S. 462 (1816).

64: 64 U.S. 287 (1859).

65: See Bruce A. Ackerman The Failure of the Founding Fathers: Jefferson,

the "public estimation," i.e. legitimacy, of the Court if it was to strike down circuit riding as unconstitutional, despite performing the practice in the past.

In another letter to Justice William Paterson, Marshall stated his policy preference as being against circuit riding and would have acted on that preference if he was not constrained. "If the question was new I should be willing to act in this character without a consultation of the Judges; but I consider it as decided & that whatever my own scruples may be I am bound by the decision."[66] However, Marshall implied he was constrained not only by the tradition of circuit riding but by the preferences of the other Justices, gleaned from "a consultation of the Judges." It is obvious that he was worried about the "survival of the institution."[67]

In the end, Justice Paterson announced the Court's decision in *Stuart*:

> [It is said] that the judges of the [S]upreme [C]ourt have no right to sit as circuit judges, not being appointed as such, or in other words, that they ought to have distinct commissions for that purpose. To this objection, which is of recent date, it is sufficient to observe, that practice and acquiescence under it for a period of several years, commencing with the organization of the judicial system, afford an irresistible answer, and have indeed fixed the construction. It is a contemporary interpretation of the most forcible nature. This practical exposition is too strong and obstinate to be shaken or controlled.[68]

Though arduous, the Court agreed to ride circuit as a conciliatory measure to the executive and legislative branches. Rather than rule on the constitutionality of the Act, the Court sidestepped the question by basing its ruling on the fact that circuit riding was already established, and reversal would have had negative effects on the Court's legitimacy. As Dean Alfange states, "The Court acted out of a fully justified fear of the political consequences of doing otherwise, not out of an overriding compulsion to reach the correct legal result at whatever sacrifice of their own political preferences."[69]

Stuart suggests an interplay of both the Hughes and Jay Theories. Knight

---

Marshall, and the Rise of the Presidential Democracy 164 (2005).

66: See George Lee Haskins & Herbert Alan Johnson, Foundations of Power: John Marshall, 1801-15, 5 (1981).

67: *Id.* at 157.

68: *Stuart,* 5 U.S. at 299.

69: See Dean Alfange, Jr., *Marbury v. Madison and Original Understandings of Judicial Review: In Defense of Traditional Wisdom,* 1993 Sup. Ct. Rev. 329, 364 (1993).

and Epstein, looking at *Stuart* through a game theoretic lens, postulate, "[P]oliticians—even those who lack an electoral connection—are strategic actors. Had this not been the case for Marshall, for example, he simply would have voted his unconstrained preferred positions…in *Stuart* (strike the Repeal Act)."[70] Despite his misgivings (and those of Justice Samuel Chase), Chief Justice Marshall ruled in favor of riding circuit in his circuit ruling. Though making different claims than both the Hughes and Jay Theories, it is reasonable that these theories can be derived from the assumptions gleaned from *Stuart*. It seems that Chief Justice Marshall acted strategically by ruling in favor of circuit riding in his circuit decision, perhaps after gauging his colleague's policy preferences by seeking the "consultation of the Judges (Justices)" (the Hughes Theory), while also being concerned by the "public estimation" of striking down the practice, which could have decreased the legitimacy of the judicial branch (Jay Theory). Striking down circuit riding would have been a double-edged sword for Marshall; not only would his circuit opinion have been reversed by his brethren on the Supreme Court level, the Court itself would have damaged its legitimacy as an institution.

VIb. *Evans v. Jordan and Morehead* and *Jones vs. Shore's Executors*

During the circuit-riding period, a unique writ existed: the writ of division. The writ of division signaled that the district judge and the Supreme Court Justice, hearing a particular case on circuit, disagreed on the outcome of the case. The certificate of division automatically appealed the case to the Supreme Court's docket. Thus, cases in which there is a writ of division can be qualitatively analyzed for evidence of the Marshall Theory, which suggests the Supreme Court was more likely to side with the Supreme Court Justice's circuit opinion, due to his advocacy role, than the district judge. The Brokenbrough volumes (1837) provide two instances of writs of division arising from Chief Justice Marshall's Fifth Circuit: *Evans v. Jordan and Morehead*[71] and *Jones vs. Shore's Executors*.[72]

In the May 1813 term of the Circuit Court of Virginia, Chief Justice Marshall and District Judge Tucker heard *Evans v. Jordan & Morehead.* Oliver Evans had been granted a 14-year patent for his invention used in the manufacture of flour and cornmeal. After the patent had expired, Jordan

---

70: See Epstein & Knight, The Choices Justices Make at 112 (cited in note 36).

71: 8 F. Cas. 872 (1813).

72: 26 F. Cas. 638. There are a number of other instances of writs of division during the 1812-1823 time period, but circuit opinions, which are needed to discern the difference of opinion, could only be found for these cases.

and Morehead started to manufacture Evans' invention. In 1808, Congress passed a law that gave Evans the exclusive right to manufacture his invention for another 14-year term. It also provided that "no person who shall have used the said [invention], or have erected the same for use, before the issuing of said patent, shall be liable for damages therefor."[73] Evans brought suit against Jordan and Morehead for the violating his exclusive patent, but Jordan and Morehead contended that they were protected by the recent Act, as they started manufacture during the interim period between the expiration of the patent and the issuance of the Act.

The question before the circuit court was whether Evans could sue Jordan and Morehead for manufacturing his invention after the expiration of the first patent but before the issuance of the patent renewal. To Chief Justice Marshall, "[Since this is] one of those subjects which is, by the Constitution of the United States, delegated entirely to the government of the Union, the question which has been made, must depend on the acts of [C]ongress."[74] The intention of the legislature, as exhibited by the wording of the Act, was to provide for the relief of Evans and grant him another patent. "The Constitution and law, taken together, give to the inventor, from the moment of invention, an inchoate property therein, which is completed by suing out a patent. This inchoate right is exclusive. It can be invaded or impaired by no person. No person can, without the consent of the inventor, acquire a property in the invention."[75] Jordan and Morehead were allowed to manufacture the invention during the period that the patent was expired (1804-1808), but once Congress passed the relief act, they had to cease manufacture. Thus, Evans was entitled to, and could sue for, damages based on the manufacture of his invention after Congress issued the new patent.

Chief Justice Marshall and District Judge Tucker split in their decision, so the case was sent to the Supreme Court on certificate of division. Justice Bushrod Washington authored the opinion of the Supreme Court in *Evans v. Jordan and Morehead*. Justice Washington's opinion read, "The language of this last proviso [that protects against damages for manufacturing the invention before the passage of the Act] is so precise, and so entirely free from all ambiguity, that it is difficult for any course of reasoning to shed light upon its meaning."[76] Clearly, Jordan and Morehead could have been liable for damages relating to the manufacture of the invention after the expiration of the first patent and before the issuance of the second. "The legislature might

---

73: *Evans,* 8 F. Cas. at 873.
74: *Id.* at 872.
75: *Id.* at 873.
76: *Evans,* 13 U.S. at 202.

have proceeded…by providing a shield for persons standing in the situation of these Defendants [and protecting them against damages]. It is [believed] that the reasonableness of such a provision could have been questioned by no one."[77] However, since Congress did not provide such a provision, the Court affirmed Chief Justice Marshall's circuit court opinion and awarded damages to Evans for the infringement of his patent by Jordan and Morehead after 1808.

An additional notable case, heard during the May 1814 term of the Circuit Court for the District of Virginia by Chief Justice Marshall and District Judge Tucker, is *United States v. Jones,*[78] which arose from the Embargo Act of 1807.[79] On November 23, 1808, Thomas Pearse took out an embargo bond on his ship, the *Sally*, to the tune of $46,300. A penalty to be assessed if the ship did not re-land its cargo in the United States, as trading with Britain and France during the Napoleonic Wars was forbidden. The case then became highly complicated, as the collector of the port of Petersburg, John Shore, died shortly after bringing the embargo bond to the district court after the *Sally* failed to re-land in the United States. His brother succeeded him in the post as collector from October 31 to December 14, 1811, during which the district court rendered its decision against George Pegram, one of the obligors of the bond. Joseph Jones then became the collector of the port. Pegram soon died, and the case was carried to the circuit court by his executors, where the district court's decision was affirmed. At the same time, Andrew Forborne, surveyor for the district of Petersburg, died in office after the suit was brought but before judgment was rendered, and John Peterson filled that position. In 1814, Pegram's executors paid the full penalty to the circuit court. A number of petitions followed, issued separately by the district attorney on behalf of the United States, by the living surveyor and collector, and the executors of the deceased surveyor and collector. In a previous case, the circuit court ruled that the United States was entitled to a moiety (half) of the penalty. The question before the court was whether the other moiety should go to the current (living) collector and commissioner or the deceased under the provisions of the Embargo Act.

After dismissing some questions not pertinent to the case, Chief Justice Marshall began, "It is first observable, that the bounty [in accordance with the law] is payable in equal proportions to the collector, naval officer, and surveyor."[80] This bounty provided motivation, in the form of money, to

77: *Id.* at 203.
78: 26 F. Cas. 638 (1814).
79: See Embargo Act of 1807 2 § 451.
80: *Jones,* 26. F. Cas. at 642.

ensure the collection laws were being followed diligently by all relevant officials. This compensation, thus, was a reward at the time of incidence for the official who caught the perpetrator, not his successor in office. "Though the office [of collector] never dies, the individuals who fill it do, and as their emoluments are considered in the light of compensation for services, the rewards of services rendered by one, ought not to be bestowed on another."[81] Though John Shore died just a few days before the judgment was rendered, he was still entitled to his compensation. Chief Justice Marshall and District Judge Tucker split in their decision when it came to the surveyor position, as Marshall felt that Forborne was entitled to his compensation, while Tucker disagreed.

On certificate of division, the Supreme Court heard *Jones v. Shore's Executors.*[82] After going over the facts pertaining to the case, Justice Joseph Story intoned, "The Court is clearly of opinion that the right of the collector to forfeitures…attaches on seizure and to personal penalties on suits brought, and in each case it is ascertained and consummated by the judgment, and it is wholly immaterial whether the collector die before or after the judgment."[83] Thus, the Court affirmed the circuit court's opinion in which the judges were not split (the collector position). Story turned to the contentious issue: "And it is further of opinion that the case of the surveyor is not in this respect distinguishable in any manner from that of the collector. It is therefore of opinion that the representatives of the deceased collector and surveyor, and not the present incumbents in office, are entitled to the distributive shares of the moiety of the money now in the registry of the circuit court."[84] The Court did not consider the positions of surveyor and collector to be fundamentally different, so the payment should not have been treated differently, either. Once again, the Court sided with one of its own, Chief Justice Marshall, over the district judge, Tucker.

In both of these cases, the Court ruled in favor of Marshall's opinion. Though some could point to the fact that this was due to Tucker being a mere district judge, this is unlikely; Tucker was one of the most revered American legal scholars of his day, having published a highly influential edition of Blackstone's *Commentaries.* Thus, the Court siding with Marshall over Tucker is not likely to be rooted in a perceived difference in legal acumen. It seems plausible that Marshall's presence on the Court gave him an advantage that Tucker, who was not a member of the Supreme Court, did not have: a venue

81: *Id.* at 643.
82: 14 U.S. 462 (1816).
83: *Id.* at 474.
84: *Id.* at 475.

to advocate for his circuit opinion by persuading the other Supreme Court Justices to his side.

VIc. *Haney v. Baltimore Steam Packet Company*

So far, we have looked at cases in which the Justice accurately anticipated the policy positions of his brethren, as the theoretical explanations predict. However, the Justices operated in an imperfect world with imperfect information. It was possible the Justices could not have known the policy preferences of their brethren, could have misapplied or misunderstood precedent, or could simply have guessed the policy preferences of their fellow Justices incorrectly.

Haney can be interpreted as an example of the last consideration. In the November 1858 term of the Circuit Court for the District of Maryland, Chief Justice Taney heard *Haney v. the Louisiana.*[85] The steamboat *Louisiana* collided with and sank the schooner *William K. Perrin* on February 26, 1858 in Chesapeake Bay. The schooner, despite sailing in an important thoroughfare, had a navigator who was inexperienced in sailing and was not familiar with the area and a pilot who was not paying attention at the helm and did not heed the warnings of his navigator. Each vessel had spotted the other from miles away, yet the chain of events that transpired resulted in the *Louisiana* ramming the *William K. Perrin* at a ninety-degree angle.

Chief Justice Taney anchored his decision on the inexperience of the schooner's crew. "[T]he weight of the testimony is decidedly on the side of those on board the *Louisiana*; for there was a steersman and a look-out, each competent and experienced, and accustomed to navigate the bay for years, each in his proper place, and each confining his attention exclusively to his appropriate duty," as opposed to the ineptitude of the crew of the *William K. Perrin*.[86] The precedent set by the Supreme Court in *The Case of the Gennesee Chief*,[87] which ruled that damages would solely be accrued by a steamboat if it was responsible for a collision, did not apply in this case, as Chief Justice Taney had established that it was the schooner's, not the steamboat's, fault. Likewise, the rule established in *St. John v. Paine,*[88] which stated, "when two vessels are meeting in opposite directions, each one shall port the helm, so as to pass each other on the larboard side," was not applicable because "[t]he rule applies only to cases where both are sailing vessels, or both are steamboats,

85: 11 F. Cas. 425 (1858).
86: *Id.* at 427.
87: 53 U.S. 443 (1851).
88: 51 U.S. 557 (1850).

not to cases where one is a steamboat and the other navigated only by sails."[89] Thus, Chief Justice Taney ruled the crew of the schooner was at fault and ordered its owners to pay for all damages.

The owners of the *Perrin* appealed to the Supreme Court in *Haney v. Baltimore Steam Packet Company.*[90] Justice Grier gave the Court's opinion, which, unlike Chief Justice Taney's circuit opinion, placed blame for the accident upon the crew of the *Louisiana*. The crew's testimony "alleges as an excuse that while the steamboat and schooner were meeting on parallel lines, the schooner suddenly changed her course and ran under the bow of the steamer. This is the stereotyped excuse usually resorted to for the purpose of justifying a careless collision. It is always improbable, and generally false."[91] Furthermore, "[t]he hypothesis set forth in the answer to excuse this collision…is not only grossly improbable in itself, but contradicted by the testimony and is a mathematical impossibility."[92] The doctrine laid out in *Chamberlain v. Ward,*[93] which required a capable watchman and helmsman on board each vessel, controlled in this case. Finding that the *Louisiana* did not adhere to this requirement, the Court overruled Chief Justice Taney's opinion and awarded damages to the owners of the *Perrin.*

In a dissent that dwarfed the length of the majority opinion,[94] Chief Justice Taney defended his circuit court opinion by saying, "I think it my duty, therefore, to state the principles of law and the evidence in the case, upon which my opinion has been formed."[95] As he did in his circuit opinion, he cited *Gennesee Chief* and *St. John* as still remaining the controlling precedent for this case. He then recounted the testimonies of the crews of both of the vessels and stated unequivocally, "I regard this as the true history of the disaster, and of the movements of the vessels by which it was produced," i.e., the collision was the fault of the inexperienced crew upon the *Perrin*, not the capable crew of the *Louisiana.* "I cannot think that the steamboat should be charged with any part of the damage which the sailing vessel brought upon itself. Those who entrust their property on the water to incompetent hands have no just right to complain of disasters, and claim indemnity for losses arising altogether from the incapacity and unfitness of those to whom they have confided it, and still less have Cory and Miles (the crew of the *Perrin*),

89: *Haney,* 11 F. Cas. at 428.
90: 64 U.S. 287 (1860).
91: *Id.* at 291.
92: *Id.* at 293.
93: 62 U.S. 548 (1858).
94: Taney comments, "This opinion…occupies more space than I anticipated." 64 U.S. at 308 (Taney, J., dissenting).
95: *Id.* at 294.

whose incapacity and misconduct were the sole cause of disaster."[96]

A closer look at *Louisiana* reveals a technologically sage opinion by Chief Justice Taney. He adroitly handles the advantages and limitations offered by the recent advent of steamboat technology and seems to be in command of the relevant law and precedent. However, he, in this case, was less adept in anticipating the policy preferences of his brethren on the Supreme Court. The Supreme Court, according to Taney, could not grasp the technological shift that steam caused in the shipping industry, but it seems that Taney did not grasp the adverse reaction of the rest of the Supreme Court to his decision. Though this is the opposite of what the Jay Theory predicts, the Court could have been worried less about its legitimacy in the middle of the nineteenth century than in the beginning. Also, Chief Justice Taney's decision could have been so far from the Court's policy preferences that the costs of affirming this decision outweighed the costs to the Court's legitimacy.

VII. Conclusion

This Article makes an important contribution to the study of the Supreme Court during its early history—the Circuit-Riding Century—a period that is understudied in the literature. It attempts to quantitatively analyze the possible effect that circuit riding had upon Supreme Court affirmation rates in an area where no quantitative work has been done before. Initial findings suggest circuit riding did not have an effect on Supreme Court affirmation rates, but further research is necessary to reach a conclusive position on either side of the issue. The findings are subject to inherent limitations: inconsistent recording and dearth of records, the difficulty of locating transcripts of circuit decisions, the lack of secondary literature on circuit riding and circuit cases in general (especially on the quantitative side), and not finding a "smoking gun," meaning we cannot discern which of the three theoretical explanations—that is, the Hughes, Jay, and Marshall Theories—if any, had an effect on affirmation rates. Although the difference between the affirmation rates of cases heard by a Justice on circuit and of circuit cases without a Justice present was not found to be statistically significant, we have found an explicit result: the thesis provides a positive counterpoint to scholars who have speculated about the detrimental effect of circuit riding on Supreme Court decision-making in strictly normative terms.

This study provides a launching pad for studying this important part of the Court's early history, and there are many paths moving forward. Locating better data or sources would allow a clearer picture of the possible effect of

96: *Id.* at 309.

circuit riding, as it would allow regression analysis to be conducted rather than a simple statistical test. The expansion of the studied time period could allow for analysis across changing political climates and the changing composition of the Supreme Court.

Appendix I		
Miles Traveled on Circuit by Justice		
Circuit	Justice	Miles Traveled
1st	Joseph Story	1896
2nd	Smith Thompson	2590
3rd	Henry Baldwin	2000
4th	Roger Brooke Taney	458
5th	Philip Pendleton Barbour	1498
6th	James Moore Wayne	2370
7th	John McLean	2500
8th	John Catron	3464
9th	John McKinley	10000
See Schwartz, History of the Supreme Court at 153 (cited in note 25).		

The European Court of Human Rights Goes to School: The "Headscarf Cases" (*Leyla Sahin v. Turkey* and *Dahlab v. Switzerland*) as Unjustified Restrictions of Religious Practice

Oren Fliegelman†

As guarantor of individual liberties in the Council of Europe's 47 member states, the European Court of Human Rights (ECtHR) is tasked with balancing the competing rights of citizens against citizens and states against citizens. In a group of controversial continent-spanning cases dealing with the right of female Muslim students and teachers to wear Islamic headscarves in public schools–known collectively as the "headscarf cases" and called by one scholar "almost a touchstone for the reflection on the presence of Islam in the public space"–the Court has found itself adjudicating between the sometimes contradictory values of state-sponsored secularism and religious freedom.

In this article, I argue that the ECtHR has failed miserably in the "headscarf cases" at fairly balancing these rights. I distill the reasoning used by the Court in *Dahlab v. Switzerland* and *Leyla Sahin v. Turkey*, the two most significant of these cases, into four major arguments: Argument from Religious Pressure, Argument from Political Symbolism, Argument from Gender Inequality, and Argument from Subjugation of Women. Using this novel argument categorization, I show that rather than protecting the rights of individuals to practice their religion freely with only the necessary restrictions, it has defended the actions of over-reaching national governments infringing on those rights under the banner of secularism. I also argue that these decisions hint that the Court does not understand the multifaceted meaning of the Islamic headscarf and has a generally negative view of Islam, both of which color its judicial decisions.

"What is lacking in this debate is the opinion of women, both those who wear the headscarf and those who choose not to."

Judge Françoise Tulkens,
Dissent in *Leyla Sahin v. Turkey*

I. Introduction

Schools are closely tied to a country's cultures and values. States rightly see public schools—ranging from primary schools all the way to universities—as spaces for the inculcation of critical national values, like good citizenship, political involvement, and tolerance. For this reason, many governments and scholars believe that it is particularly important that these educational spaces epitomize and manifest these national values.[1] For many states in Europe, secularism is a fundamental national value. While secularism has a number of localized flavors, its core demand is creating a public sphere for public activity that is separate and free from religious influence.[2] Thus, countries that believe in secularism believe that students must have the opportunity to grow in a safe environment in which they feel no pressure to submit to any specific religious beliefs. Moreover, creating this environment in European schools is more important than in traditional "adult" settings for it is setting a standard for the interaction of state and religion from which the young pupils can learn. It is not only that secularism requires a place of learning, like all public spaces, to be free from religious interference, but the students who are just beginning to learn key political concepts should see the school as a model for the type of separation that the state supports.

---

†: Oren Fliegelman is a student at Princeton University, majoring in Politics.

1: In the US, where an emphasis is placed on the First Amendment's guarantee of freedom of speech, the Supreme Court has said, "The vigilant protection of constitutional freedoms is nowhere more vital than in the community of American schools." *Shelton v. Tucker*, 364 U.S. 479, 487 (1960). In Switzerland, the highest Court has said that the country's principle of secularism "assumes particular importance in State schools, because education is compulsory for all, without any distinction being made between different faiths." *Dahlab v. Switzerland,* 42393/98 ECtHR 5 (2001). Aernout Nieuwenhuis has written that "the school system has been one of the most important battlefields between religion and state." Aernout Nieuwenhuis, *European Court of Human Rights: State and Religion, Schools and Scarves,* 1 Eur. Const. L. Rev. 495, 503 (2005).

2: See Heiner Bielefeldt, "Political Secularism and European Islam: A Challenge to Muslims and Non-Muslims," in Jamal Malik, ed., *Muslims in Europe: From the Margin to the Centre* 149 (2004), and Hilal Elver, *The Headscarf Controversy: Secularism and Freedom of Religion* 4 (2012).

The implementation of secularism in public schools, however, must not be allowed to interfere with one's right to manifest and practice one's faith. The same countries that endorse secularism also believe in the right to freedom of religion, a right that is protected by the European Convention on Human Rights (ECHR). These values—secularism and freedom of religion—although not necessarily contradictory, often find themselves on opposite sides of legal battles. Indeed, these values have been at the epicenter of a decades-long controversy that has important political and cultural dimensions: can states legitimately institute bans on the wearing of Islamic headscarves in schools or are such bans an unreasonable infringement on a woman's ability to practice her religion? Does a ban advance the protection of the rights of others at school critical to creating a modern secular society or is this an egregious violation of an individual's right to manifest her religion?

The significance of the controversy can be measured in European Court of Human Rights (ECtHR) casework: as of 2011, twelve of the sixteen cases brought to the ECtHR by individuals who had been barred from wearing religious symbols in public involved women who wanted to wear the Islamic headscarf.[3] As will be shown, the so-called ECtHR "headscarf cases" have been discussed, debated, and commented on by academics, journalists, and religious leaders.[4] Moreover, the debate is not only a legal one: it involves a number of emotional and social dimensions that are influenced by Western views on Islam. Schirin Amir-Moazami recounts that at a recent academic seminar she attended there was "a tangible sense of relief among most of the participants in the room" as a scholar expressed his disapproval of veiling, "spell[ing] out what the majority of people probably thought but were unable to voice: that veiling provokes reactions and touches on a number of embodied emotions that can hardly be addressed by legal rules."[5] Many scholars, judges, and citizens have strong antipathies toward veiling, even if their feelings are couched in legalistic terms. To emphasize the headscarf controversy's poignant social significance, one scholar has even called it "almost a touchstone for the reflection on the presence of Islam in the public space."[6] Another has said it

3: Elver, *The Headscarf Controversy* at 75 (cited in note 2).

4: Ayşe Saktanber & Gül Çorbacioğlu, *Veiling and Headscarf-Skepticism in Turkey,* 15 Soc. Pol.: Int'l Stud. Gender, St., & Soc'y 514, 521 (2008).

5: Schirin Amir-Moazami, "The Secular Embodiments of Face-Veil Controversies Across Europe" in Nilüfer Göle, ed., *Islam and public Controversy in Europe* 83 (2013).

6: Stefano Allievi, "Relations and Negotiations: Issues and Debates on Islam," in Brigitte Maréchal, ed., *Muslims in the Enlarged Europe: Religion and Society* 338 (2003).

"represents a major international challenge in this 21st century."[7]

As guarantor of individual liberties in the Council of Europe's 47 member states, the ECtHR has played a crucial role in debating this issue through its adjudication of cases. Unfortunately, its role has not been a productive one. Rather than protecting the rights of individuals to practice their religion freely with only strictly necessary restrictions, it has defended the actions of over-reaching national governments infringing on those rights. With this article, I hope to add a new perspective on the role that the ECtHR has played in this debate by distilling into four major arguments the reasoning used by the Court in two of the most significant headscarf cases.[8] I will maintain that using these four arguments, the ECtHR has sacrificed Muslim women's freedom of religious exercise in the public schools in the name of protecting a secular and sexually equal environment. I will also argue that these decisions hint that the Court does not understand the multifaceted meaning of the Islamic headscarf and has a generally negative view of Islam, both of which color its judicial decisions.

I have selected the two ECtHR cases because they are particularly good examples of the Court's jurisprudence on this topic. I will focus on *Dahlab v. Switzerland* (2001), a case about a Zurich primary school teacher who was fired because she refused to refrain from wearing the Islamic headscarf. Next, I will discuss *Leyla Sahin v. Turkey* (2005), a Grand Chamber decision about an Istanbul University student who was suspended for not following a University-wide ban on headscarves. *Dahlab*'s reasoning has been used as precedent in almost every subsequent headscarf case and *Sahin* was the first Grand Chamber judgment on the topic of religious clothing.[9]

To advance my argument, I will first discuss the relevant article of the ECHR, explain the Court's jurisprudence on the article, and outline the details of *Leyla Sahin* and *Dahlab.*[10] Next, I will discuss and critique the four main arguments that the Court uses in the two cases to defend the

---

7: Malika Ghamidi, "The Islamic Veil: a Focal Point for Social and Political Debate," in Theodore Gabriel, ed., *Islam and the Veil: Theoretical and Regional Contexts* 143 (2011).

8: While other scholars have also attempted to distill and distinguish certain specific arguments from one or both of these decisions, I have not read any that come to the same conclusion as my own.

9: Carolyn Evans, *The Islamic Scarf in the European Court of Human Rights,* 7 Melb. J. Int'l. L. 52, 53 (2006).

10: I will only deal with the cases here from the perspective of the ECHR's Article 9. While both Dahlab and Sahin appealed to the Court on the basis of multiple infringements (including, among others, the right to an education), freedom of religion was the primary and strongest of their arguments.

headscarf bans: religious pressure, political symbolism, gender inequality, and subjugation of women. Finally, in the conclusion I will describe four of the judgments' overarching issues which highlight the Court's problematic approach to Islam, and then provide a number of general recommendations for future Court decisions in similar cases.

II. Background: The ECHR's Article 9 and the Headscarf Cases

IIa. Article 9 of the ECHR

The European Convention on Human Rights (ECHR) was established by the Council of Europe's member states in 1950 and came into force in 1953.[11] In the wake of World War II, the Convention was aimed at ensuring the rule of law by protecting the individual rights and liberties of citizens from overbearing state government. The Convention is meant to be a "living document" that adapts to the exigencies of the times and has been amended with 14 subsequent protocols, protecting additional rights such as the right to education (Protocol 1) and a ban on the death penalty (Protocol 6). Since 1959, the ECtHR has been the primary enforcer of the Convention and the protocols' guarantees of individual rights. While any citizen in a member state can bring their case to the ECtHR, it is a court of last resort; the citizen must first exhaust all the domestic legal avenues of appeal.

Article 9 of the Convention guarantees to citizens the right to freedom of religion, both in belief and practice. The ECHR provides:

> 1. Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance;
>
> 2. Freedom to manifest one's religion or beliefs shall be subject only to such limitations as are prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others.[12]

11: European Convention on Human Rights.
12: *Id.* at Art 9.

The article follows a frequent double structure of articles in the ECHR: guarantee of a right and then explicit restrictions. In the first paragraph, the right is asserted and detailed. For Article 9, the right is freedom of thought, conscience, and religion, which includes belief, manifestation of belief, religious practice, and so on. In the second paragraph, the types of legitimate interferences on the right are explained. By developing the right in this way, the Convention is saying that while in general freedom of religion and its subsidiary rights are absolute, there will be certain circumstances when, for the sake of public interest or the rights of others, the state may reasonably need to infringe upon these rights.

When handling applications that accuse states of infringements on freedom of religion, the Court first determines whether or not the measure actually interfered with the applicant's freedom of religion. In both *Leyla Sahin* and *Dahlab,* the ECtHR found the law to be an interference on the right.[13] If the Court determines that there has been an interference, it then uses a three-part test to judge whether the interference was a legitimate one within the scope of the second paragraph of Article 9.

To do this, the Court first determines whether the restriction was "prescribed by law." In *Leyla Sahin*, the Court explains that this requirement means that the measure must have "a basis in domestic law" and be expressed with enough precision so that that a regular citizen would have been able to foresee the restriction.[14] Second, the Court judges whether the restriction pursues "legitimate aims." These first two tests are generally not very difficult burdens to meet and the bans in *Leyla Sahin* and *Dahlab* passed both.

The true test of the legitimacy of the interference lies in the third component of the test, the "most complex and open-ended" of the requirements.[15] The second paragraph requires that the interference be "necessary in a democratic society," by forwarding the interests of public safety, health, or morals, or the protection of the rights of others. When examining a measure in light of this test, the Court looks at whether the law is effective in accomplishing its aims, uses the least intrusive means, and its interference is proportional to the significance of the goals of the measure. While the Court determined the "legitimate aim" test in one paragraph of *Leyla Sahin,* the explanation of the

13: In *Sahin,* the Turkish government did not dispute that the law was an interference. *Sahin v. Turkey,* 44774/98 ECtHR Grand Chamber 19, ¶ 76-78 (2005). In *Dahlab,* the Swiss government did dispute the interference. *Dahlab,* 42393/98 ECtHR at 8.

14: *Sahin*, 44774/98 ECtHR Grand Chamber at 20, ¶ 84.

15: Douwe Korff, *The Standard Approach under Articles 8-11 ECHR and Article 2 ECHR,* Euro. Commission 3, online at http://ec.europa.eu/justice/news/events/conference_dp_2009/presentations_speeches/KORFF_Douwe_a.pdf.

"necessary in a democratic society" test took up eight pages.[16] In both *Leyla Sahin* and *Dahlab,* the ECtHR found that the headscarf bans were "necessary in a democratic society."

IIb. *Dahlab v. Switzerland*

Dahlab v. Switzerland (2001) was the first petition that came before the Court that involved a ban on Islamic headscarves in public schools. Switzerland is a secular state and Geneva, the canton in which the case occurred, places an emphasis on secularism in its school system. The Geneva cantonal laws state that "the public education system shall ensure that the political and religious beliefs of pupils and parents are respected."[17] To ensure this respect for all political and religious beliefs in schools, section 6 of the Geneva Public Education Act requires that there be no "obvious means of identification imposed by a teacher on her pupils, especially in a public, secular education system."[18] While Switzerland is an overwhelmingly Christian country, the federal government—and the cantonal government of Geneva—is committed to creating a public sphere free from religious influences.[19] For this reason, Geneva places primary importance on ensuring that teachers in public schools act in religiously neutral ways in front of their students.[20] Students, however, are free to wear religious symbols as they please.[21]

Lucia Dahlab was an elementary school teacher in Geneva, and converted to Islam in 1991 after a period of "spiritual soul-searching" and in preparation for her marriage to a Muslim man.[22] At the end of that year, she began wearing the Islamic headscarf in public at all times, including during the school day in her role as a teacher. She continued to do so for the next four years with no complaints from teachers, students, or parents. So as not to religiously influence her students, when they asked about the headscarf she would tell them that it was to keep her ears warm.[23]

In January 1996, an inspector informed the Director General for Primary

16: *Sahin,* 44774/98 ECtHR Grand Chamber at 23-28 (2005).
17: *Dahlab,* 42393/98 ECtHR at 4.
18: *Id.* at 2.
19: Rene Pahud de Mortanges, *Religion and the Secular State in Switzerland,* Int'l. Ctr. L. & Religion Stud. 690, online at http://www.iclrs.org/content/blurb/files/Switzerland.1.pdf.
20: In fact, most schools in Switzerland are public. *Id.* at 692.
21: Allievi, "Relations and Negotiations: Issues and Debates on Islam" at 342 (cited in note 6).
22: *Dahlab,* 42393/98 ECtHR at 10.
23: Evans, 7 Melb. J. Int'l. L. at 59 (cited in note 9).

Education in Geneva that Dahlab wore a headscarf while she was carrying out her teaching duties. After a meeting, the Director General ordered that Dahlab stop wearing the headscarf while teaching, informing her that she was in violation of section 6 of the Geneva Public Education Act.[24] In August 1996, she appealed the decision as a violation of her Article 9 right to manifestation of religion to the Geneva cantonal government, which dismissed it. She then appealed to the Federal Swiss Court, which dismissed it, saying that it is "especially important that [teachers] should discharge their duties—that is to say, imparting knowledge and developing skills—while remaining denominationally neutral."[25] The Court added that "it must also be acknowledged that it is difficult to reconcile the wearing of a headscarf with the principle of gender equality."[26]

After exhausting all of her domestic avenues of appeal, Dahlab submitted her case to the ECtHR. In 2001, the Court declared her application for the case to be heard as inadmissible, stating that her complaint was "manifestly ill-founded" partially because of the wide margin of appreciation granted to Switzerland in this type of case. In other words, the Court believed that Dahlab's complaints were insignificant or unconvincing enough that they did not merit a full examination before a chamber of the Court. The Court's decision to not even fully hear the case provoked a significant wave of criticism from scholars.[27]

IIc. *Leyla Sahin v. Turkey*

Leyla Sahin v. Turkey (2005) was the first decision issued by the ECtHR Grand Chamber on religious clothing in the public sphere. It has been the most influential and widely debated of the headscarf decisions and has been cited as precedent in all subsequent related ECtHR decisions. Secularism has been a primary value of Turkey since the creation of the Republic of Turkey

24: *Dahlab,,* 42393/98 ECtHR at 2.

25: *Id.* at 6.

26: *Id.*

27: Carolyn Evans has written of *Dahlab*, "A woman with an otherwise spotless employment record who had spent years wearing Islamic clothing to which no-one objected had been effectively sacked because of her religion. But the issue was so clear that it did not even deserve a full and proper consideration by the Court" Evans, 7 Melb. J. Int'l. L. at 60 (cited in note 9). Eva Brems has written that this "one-sided opinion about the negative signification of the head-scarf is in my view not befitting to a body of the stature of the European Court of Human Rights" Eva Brems, *Diversity in the Classroom: The Headscarf Controversy in European Schools*, 31 Peace & Change 117, 127 (2006).

in 1923, following the breakdown of the Ottoman Empire after World War I. Article II of Turkey's current constitution declares that the country is "a democratic, secular and social state," a section of the constitution that is marked as "unchangeable."[28] The president's oath includes the promise to safeguard and abide by "the principles of the secular republic."[29] The commitment to secularism has special significance in Turkey, given that the country's founder deliberately set up a non-religious republic in a country in which approximately 99 percent of the country's population identifies as Muslim.[30] The state is particularly sensitive to religious clothing because of the close connection that existed between clothing and religious and social status in the Ottoman Empire.[31] Against this background, in 1982 Turkey became the first (and only) European state to ban Islamic headscarves at institutions of higher education.[32] This ban was upheld in 1984 by the Turkish Supreme Administrative Court, writing that "the headscarf is in the process of becoming the symbol of a vision that is contrary to the freedoms of women and the fundamental principles of the Republic."[33]

Leyla Sahin was a fifth-year medical student at the Faculty of Medicine at the University of Bursa when she chose to transfer to the Faculty of Medicine at the University of Istanbul in August 1997. Sahin was 24 years old and wore a headscarf during her time at Bursa and continued to wear it at the University of Istanbul until February 1998.[34] At that time, the vice-chancellor of the University issued a statement that said students with headscarves and beards should be barred from lectures, courses, and tutorials.[35] Throughout the next few months, Sahin was refused entrance to written exams and lectures, making it impossible for her to participate in school. After receiving a warning from the dean of the faculty, she filed a case to end the ban with the Istanbul Administrative Court. She argued that the ban violated her Article 9 rights to practice her religion, in addition to her Protocol 1, Article 2 rights to an education, among others. She was suspended from the University in April 1999 and her appeal of the headscarf ban was dismissed by the Court in November and by the Supreme Administrative Court the next year.[36]

28: Turkish Const., Art. 2, 4.
29: *Id.* at Art. 103.
30: *International Religious Freedom Report 2004: Turkey*, U.S. Dep't. State (2004), online at http://www.state.gov/j/drl/rls/irf/2004/35489.htm.
31: *Sahin,* 44774/98 ECtHR at 8-9, ¶ 36.
32: *Id.* at 13, ¶ 55.
33: *Id.* at 7, ¶ 34.
34: *Id.* at 3, ¶ 11.
35: *Id.* at 3, ¶ 12.
36: *Id.* at 4-5, ¶ 24-27.

Sahin appealed the Turkish courts' decisions to the ECtHR. Unlike *Dahlab,* the Court allowed the case to proceed to a discussion of its merits. The fourth section issued its Chamber ruling in June 2004, saying that the ban was "justified in principle and proportionate to the aims pursued, and therefore, could be regarded as 'necessary in a democratic society.'"[37] Sahin then requested a referral to the Grand Chamber, which the Court granted. In November 2005, the Grand Chamber handed down a 51-page judgment, including a short concurrence and a 9-page dissent, which declared that the ban's interference with her religious practice was legitimate and within the scope of Article 9's second paragraph.[38] The decision cited *Dahlab* six separate times. By then, Sahin had already left Turkey to continue her studies at the University of Vienna.

III. The Four Arguments

The specifics of *Leyla Sahin* and *Dahlab* are quite different. One was a Grand Chamber judgment regarding an adult medical student and the other was a case dismissed for lack of merit about a primary school teacher. Moreover, one case dealt with Turkey, a country with an overwhelmingly Muslim population, while the other took place in Switzerland, a country in which 26 of 28 cantons have established Christian churches.[39] Surprisingly, despite these differences, three of the four main arguments used by the EctHR in explaining its decisions were similar in the two cases. Indeed, the similarity between their arguments is a significant indicator to the generalized and stereotype-influenced manner in which the Court approaches the issue of Islamic headscarves. A probable cause of this approach is the Court's handling of the margin of appreciation in both cases: by granting a wide margin to Switzerland and Turkey, the EctHR has given itself legal wiggle-room to step back from the specifics and leave the details—as important as they are—to the national governments.

In her scathing dissent in *Leyla Sahin,* Judge Françoise Tulkens asserts that the Court's arguments for the legality of the university headscarf ban are in "general and abstract terms."[40] She goes on to say that the majority's arguments are based entirely on the apparent protection of two main principles of Turkish democracy—secularism and equality—from the headscarf. In its

37: *Leyla Sahin*, 44774/98 ECtHR 26, ¶ 114 (2004).

38: *Sahin,* 44774/98 ECtHR Grand Chamber at 30, ¶ 123.

39: Mortanges, *Religion and the Secular State in Switzerland* at 689 (cited in note 19).

40: *Sahin,* 44774/98 ECtHR Grand Chamber at 44, ¶ 4 (Tulkens, J., dissenting).

jurisprudence, the Court approves and even encourages the notion that the headscarf is incompatible with these principles. I would like to expand and adjust Judge Tulkens' statement, because I believe that her dichotomy does not only hold true for the arguments used in *Leyla Sahin,* but also for the arguments used by the Court in *Dahlab.*

To show this, I have categorized the Court's arguments into four types. Two of the arguments—religious pressure and political symbolism—fall under the "secularism" category. While religious pressure appears in both decisions, political symbolism appears in only *Leyla Sahin* because of the unique nature of Turkey's majority Muslim population. Two of the arguments—gender inequality and subjugation of women—fall under the "equality" category and also appear in both decisions. I believe that the Court places more of an emphasis on the "secularism" category and gender inequality arguments, while the subjugation of women argument is secondary.[41] I will detail the Court's reasoning in each of the arguments, while critiquing its approach and offering potential ways forward.

IIIa. Argument from Religious Pressure

The argument from religious pressure is the most clear-cut and fully reasoned of the four arguments given by the Court and it appears prominently in both cases. In *Dahlab*, the Court expresses concern that a teacher wearing a headscarf will inappropriately pressure or influence his or her students to become interested in Islam or, even more distressingly, want to convert. The Court makes clear that this pressure does not need to be spoken or even intentional: the headscarf is a "powerful external symbol" and merely wearing it potentially proselytizes and exerts pressure on the students.[42] This is because the ECtHR agrees with the Swiss court in that the Islamic headscarf constitutes "a 'powerful' religious symbol" that "clearly identified [the wearer] as a member of a particular faith."[43] Dahlab maintained that she told her pupils that she wore the headscarf to keep her ears warm so as to abide by the Geneva rules for school teachers, but the Swiss Court (quoted by the ECtHR) said that the students "will realise that she is evading the issue. It is therefore

41: While Judge Tulkens does not say this explicitly in her dissent, I believe she also thinks the "secularism" category of arguments is more significant than the "equality" category of arguments. Although admittedly a rough barometer of importance, the division of pages in her dissent is telling: arguments from secularism merit three full pages of explanation, while the arguments from equality merit one.

42: *Dahlab,* 42393/98 ECtHR at 13.

43: *Id.* at 2.

difficult for her to reply without stating her beliefs."[44]

From the government's perspective, this type of pressure from an authority figure in a public school would be a clear violation of the country's principle of secularism and the laws passed to enforce it. This pressure was an even graver concern to the Court because of the young age of Dahlab's students. Teachers are "important role models for their pupils, especially when, as in the applicant's case, the pupils were very young children attending compulsory primary school," said the Court.[45] The Court quotes the Swiss court in saying that by wearing the headscarf "the appellant may have interfered with the religious beliefs of her pupils, other pupils at the school and the pupils' parents."[46]

In *Leyla Sahin,* the Court applies a similar argument, saying that Sahin's wearing of the headscarf makes other non-observant Muslim students uncomfortable and pressures them into wearing headscarves. The Chamber explains that wearing the headscarf "may have an [impact] on those who choose not to wear it" because it is a symbol that is "presented or perceived as a compulsory religious duty."[47] The Court is saying that by the mere act of following her religious beliefs and wearing the headscarf, Sahin could be intimidating or pressuring her peers to do the same. Since the majority of Turkey's population "profess[] a strong attachment to the rights of women and a secular way of life" but still "adhere to the Islamic faith," the Court views the wearing of the headscarf as disrespecting the rights of other female citizens who choose not to wear it.[48] By wearing the headscarf, Sahin would be interfering with the right of other students to not wear it.

It seems like a difficult challenge to prove that one can pressure others to dress like one's self just by wearing a certain article of clothing. Indeed, despite these assertions, the Court fails to prove anything of the sort, whether it be pressuring students in the case of *Dahlab* or pressuring peers in the case of *Leyla Sahin*. The Court does not cite any evidence that Dahlab or Sahin's headscarves had made any student or peer uncomfortable and even says explicitly that neither verbally proselytized. In Dahlab's case, there was no evidence from other cases or studies that indicated that students felt uncomfortable with teachers who wore headscarves. While it is reasonable to worry somewhat about the way a civil servant represents the state, the Court simply trusted the Swiss court's assumptions of pressure, saying that "it

---

44: *Id.* at 6.
45: *Id.* at 9.
46: *Id.* at 4.
47: *Sahin*, 44774/98 ECtHR Grand Chamber at 28, ¶ 115.
48: *Id.*

cannot be denied outright that the wearing of a headscarf might have some kind of proselytizing effect."[49] But certainly, not being able to deny something outright is not nearly the same as being able to prove it. As Carolyn Evans notes, the Court in *Dahlab* tries to blur the effects on Dahlab's students of her wearing the headscarf by declaring them unknowable rather than admitting that there most likely was just no harm.[50] She adds that "even if the Court's assumption—that there might be some proselytizing effect on the children—is true, it is not clear why this effect is sufficient" to discharge the high burden of "necessary in a democratic society."[51] Surely, one teacher in one year at school would not be enough to challenge a child's complete religious outlook.

The Court's reasoning in *Leyla Sahin* is even less convincing. It is not surprising that in a generally non-observant student body, a woman wearing a headscarf might make some feel uncomfortable. But just like the ECHR guarantees the majority the right *not* to wear the headscarf, naturally it should also guarantee Sahin the right *to wear* the headscarf. Nowhere does the Court explain how the mere wearing of the headscarf would make other students feel uncomfortable in such a way that it justifies the state's interference with Sahin's right. Like in *Dahlab,* the Court speaks solely in generalities. It says that democracy is a "constant search for a balance between the fundamental rights of each individual" and that it "entail[s] various concessions" from individuals, but does not explain why headscarf-wearing women should make the concessions and not the non-headscarf-wearing.[52] Is it more difficult for most women to ignore the headscarf-wearing women than it is for a woman to ignore what she believes to be a religious commandment? In her dissent, Judge Tulkens compares the argument to freedom of expression jurisprudence, saying that "the Court has never accepted that interference with the exercise of the right to freedom of expression can be justified by the fact that ideas or views concerned are not shared by everyone and may even offend some people."[53]

While the Court's generalities are useful in contextualizing the problem, they are not helpful in finding a solid justification for why the rights of non-observant Muslims trump those of observant ones. Nowhere does the Court offer such a reason or justification. Moreover, the Court relies on the "powerful external symbol" rationale from *Dahlab*, but does not note the key difference between the two cases: that Dahlab was in a position of authority,

---

49: *Dahlab,* 42393/98 ECtHR at 13.
50: Evans, 7 Melb. J. Int'l. L. at 63 (cited in note 9).
51: *Ibid.*
52: *Sahin,* 44774/98 ECtHR Grand Chamber at 26, ¶ 108.
53: *Id.* at 47, ¶ *9 (*Tulkens, J., dissenting).

while Sahin was in a peer position. In her dissent, Judge Tulkens also takes issue with the Court attempting to attribute to Sahin the same proselytizing potential as Dahlab.

In truth, the Court appears to be confusing who is pressuring who: the non-religious state governments are pressuring, through the bans, observant Muslim woman into abandoning the headscarf. Indeed, *Leyla Sahin* seems like a manifestation of what Aernout Niewuwenhis describes as the "special" component of Turkish secularism: Turkey is "militant against the pressure put on the secularity of the state by Islam, the religion of the vast majority of the people."[54] The state is so afraid of agitation against secularism that it is proactively infringing upon the rights of the religious whether or not they threaten the secular state. An easy way to accomplish this is to turn religious practices into a public expression of faith. Ayse Saktanber and Gul Corbacioglu argue that the Turkish government, with help from the ECtHR and the observant women themselves, have done exactly that: transformed headscarves into a "public question of religious expression" in Turkey rather than "a private question of piety."[55] Choosing to wear the headscarf is no longer seen as simply a personal religious choice that happens to be publicly visible, but an intentionally public religious-political expression that exerts pressure on others. In essence, the wearing of the headscarf is no longer directed inward—to emphasize one's own spirituality—but directed outwards at others, as if a challenge or threat. Once this transformation in the public eye has taken place, it is much easier from the human rights perspective to restrict public and potentially pressuring expressions of faith than private commitments to practice a religion. While Switzerland does not seem to have this same fear about the overthrow of the secular system, it remains aggressive in protecting its citizens from any religious pressure, as represented in *Dahlab.*

IIIb. Argument from Political Symbolism

In *Leyla Sahin,* the ECtHR invokes the political symbolism of the headscarf as a primary reason for barring it from educational institutions. This unusual argument originated from the Turkish government and courts and is prominent in Turkey because of its majority Muslim population and history. Since the mid-1980s, Turkey has seen a rise of Islamist leaders and parties who openly advocate the overthrow of the secular nature of government or are believed by secularists to be secretly pursuing it. While the

54: Nieuwenhuis, 1 Eur. Const. L. Rev. at 501 (cited in note 1).

55: Saktanber & Çorbacioğlu, 15 Soc. Pol.: Int'l Stud. Gender, St., & Soc'y. at 518 (cited in note 4).

most well-known of these parties is the now-banned Welfare (Refah) party, others include the National Salvation party in the early 1980s, some members of the Motherland party, and the current ruling Justice and Development party.[56] The female supporters of many of these parties wear headscarves, and by the 1980s, many headscarf-wearing women had begun entering higher education.[57] Opponents of the Islamists began identifying the headscarf as a rhetorical symbol used by the fundamentalists, and some fundamentalists indeed began using it as such. As Özlem Denli explains, "Turkish legislators and judges portray veiling as a politicized symbol of systematic rejection targeting the laicist [secularist] organizing principle of the state."[58] Many Turkish officials feel that if a woman wears a headscarf, she is automatically opposed to the secular nature of the state. They believe that it is a symbol of these fundamentalist Islamist groups and assume that any woman wearing it is a supporter of fundamentalism.

The Turkish government's linkage of headscarf-wearing women to fundamentalist Islamic parties was given a prominent role at the ECtHR. In the "History and Background" section of the Grand Chamber decision, the Court explains that while those in favor of the headscarf view wearing it as "a duty and/or a form of expression linked to religious identity," "supporters of secularism" believe the headscarf is "a symbol of political Islam."[59] The Court goes on to essentially endorse this attitude and linkage. It references the linkage multiple times in the Court's 2003 decision in *Refah v. Turkey*, in which the Court affirmed that the Turkish Constitutional Court's dissolution of the fundamentalist Islamist Refah party for attempting to overthrow the secular nature of the government was in compliance with Article 11 of the ECHR. In the decision, the Court agrees that it is legitimate for a democracy to ban a political party that is aimed at eliminating a core concept of the state, including secularism. In relation to headscarves, the Court in *Refah* specifically noted how the party was planning on using headscarves in universities to display the power and popularity of the party.[60] As will be discussed later, Patrick Macklem has pointed to *Refah* and *Leyla Sahin* as examples of what he calls "militant secularism," the state's proactive attack of any organization or movement that appears to threaten a state's secularism.[61]

56: Özlem Denli, "Between Laicist State Ideology and Modern Public Religion: The Head-cover Controversy in Contemporary Turkey," in Tore Lindholm, ed., *Facilitating Freedom of Religion or Belief: A Deskbook* 506-507 (2004).

57: *Id.* at 507.

58: *Id.* at 504.

59: *Sahin,* 44774/98 ECtHR Grand Chamber at 8, ¶ 35.

60: *Refah Party v. Turkey,* 41340/98 ECtHR Grand Chamber 8 (2003).

61: Patrick Macklem, *Guarding the Perimeter: Militant Democracy and Religious*

The ECtHR then applies some of its reasoning for banning Refah to why it is understandable that the Turkish government would want to bar headscarves at universities. In *Leyla Sahin,* the Grand Chamber quotes the Chamber's decision saying that "The Court does not lose sight of the fact that there are extremist political movements in Turkey which seek to impose on society as a whole their religious symbols and conception of a society founded on religious precepts."[62] The Chamber also added that "it should be noted" that the Turkish Constitutional Court believed that the opinions expressed by the leaders of fundamentalist Islamist parties "on the question whether the Islamic headscarf should be worn in the public sector and/or schools demonstrated an intention to set up a regime based on the Sharia."[63] Without explicitly saying it, the Court marks Leyla Sahin's wearing of the headscarf as a symbol of fundamentalist political Islam like the banned Refah party.

The ECtHR's linkage of these two issues is highly troubling. Indeed, it seems like the Court is painting all Muslim women who choose to wear the headscarf as antagonists of the secular system with one broad brush stroke. It divides the population between "supporters of secularism" and all others, presumably opponents of secularism, a group that includes all women who choose to wear headscarves. While it is legitimate to protect the secular system from people who would like to overthrow it, it is not legitimate to try to do this by restricting the rights of those who happen to dress similarly or follow similar religious beliefs to those who would like to overthrow the system. Judge Tulkens expresses as much in her dissent, when she writes that "merely wearing the headscarf cannot be associated with fundamentalism and it is vital to distinguish between those who wear the headscarf and 'extremists' who seek to impose the headscarf as they do other religious symbols."[64] The Court does not only fail to distinguish these political and non-political individuals, but it also does not take into account the number of different reasons that Muslim women choose to wear the headscarf or what types of political affiliations they might have. There was nothing that Leyla Sahin said or did to indicate that she held anti-secular beliefs, aside from the fact that she did not believe that the principle of secularism barred her from being able to wear the headscarf as a student at a public university.[65]

In her dissent, Judge Tulkens takes issue with the Court's assumption of the political nature of Leyla Sahin's headscarf. In particular, she believes that

---

Freedom in Europe, 19 Constellations 575, 579-81 (2012).

62: *Sahin,* 44774/98 ECtHR Grand Chamber at 28, ¶ 115.

63: *Sahin,* 44774/98 ECtHR at 7, ¶ 32.

64: *Sahin,* 44774/98 ECtHR Grand Chamber at 47, ¶ 10 (Tulkens, J., dissenting).

65: *Id.* at 20, ¶ 85.

the connection that the Court emphasizes between Leyla Sahin's headscarf and fundamentalist Islamic hopes to overthrow the secular government is incorrect and inappropriate. She writes, "Not all women who wear the headscarf are fundamentalists and there is nothing to suggest that the applicant held fundamentalist views."[66] Indeed, it seems that the *Refah* decision and discussion of the political nature of the headscarf is entirely inappropriate for this decision since it is clear that Leyla Sahin was not wearing the headscarf for political reasons.[67] Patrick Macklem argues that this component of *Leyla Sahin* is a quintessential example of the ECtHR approving Turkey's use of "militant secularism." Like Niewuwenhis, Macklem believes that Turkey has taken a proactive stance in protecting its secular nature by trying to ban any acts or symbols that it even remotely sees as a threat. Secularism is so important to Turkey's government, Macklem argues, that it views a threat to secularism as an existential threat to the democracy. From the Turkish government's handling of the *Leyla Sahin* case—and of headscarf issues, more generally— Macklem's opinion and its logic certainly appears correct. He says that "characterizing the headscarf ban" as a threat to democracy "means that it is not only a measure designed to promote secularism, but also a preemptive measure taken by the state designed to combat a threat to a democratic Turkey."[68] While one can understand why the Turkish government might bring this "political symbol" argument before the Court, the Court should have explicitly refused it, explaining that a ban due to political symbolism would not outweigh the restrictions that the ban would place on women who wore the headscarf because of religious (and not political) convictions.

IIIc. Argument from Gender Inequality

In both *Leyla Sahin* and *Dahlab,* the ECtHR accepts the premise offered by the national governments that the Islamic requirement of only women wearing headscarves is inherently sexually unequal. The Court seems to believe that since the Muslim tradition requires that only women wear the headscarf—a clothing regulation thought of as restrictive—headscarves are by definition opposed to the ideological underpinning of democracy, that

66: *Id,* at 47, ¶ 10 (Tulkens, J., dissenting).

67: However, this is not to deny that some women do use the veil as a political challenge of acceptance in a number of countries. Seyla Benhabib writes about the *l'affaire foulard* ("the Scarf Affair") in France, which began in 1989 when three French school girls wore the scarf to school so as to "defy[] the state." Seyla Benhabib, *The Claims of Culture* 94-102 (2002).

68: Macklem, 19 Constellations at 581 (cited in note 61).

every individual, both male and female, merits an equal presence in the public state. By forcing only women to wear the veil, the judges believe that Islam is marking women as separate and inferior. In *Dahlab*, the Court writes that it is "difficult to reconcile the wearing of an Islamic headscarf with the message of tolerance, respect for others, and, above all, equality and non-discrimination."[69] Not only does the headscarf contradict the value of gender equality, but it also advocates discrimination and a lack of respect for others. The Court emphasizes that these are not values that a teacher should be transmitting to his or her pupils in the classroom, thus implying that the mere wearing of the headscarf symbolizes these values. The Court also describes the headscarf as a requirement that "appears to be imposed on women by a precept which is laid down in the Koran."[70] Even the word choice of the Court expresses dissatisfaction with the headscarf. As Carolyn Evans points out, describing the headscarf as what "appears" to be a requirement that is "imposed" by the Koran is using a clearly "loaded" manner of describing the religious rule.[71] She goes on to say that it is unusual for the Court to refer to religious duties in such negative terms.[72]

In asserting the gender inequality of the Islamic headscarf requirement, *Sahin* references the above statements from *Dahlab.* The Grand Chamber also stood by the Chamber's comments about the inherent inequality of the headscarf. The Chamber wrote that "it is understandable" that the Turkish authorities "would consider that it ran counter to the furtherance of such values [of tolerance and respect for others] to accept the wearing of religious insignia, including as in the present case, that women students cover their heads with a headscarf while on university premises."[73] While the Court presents this rule as general—applying to all religious insignia from any faith—

69: *Dahlab,* 42393/98 ECtHR at 13.

70: *Id*. Navigating the Islamic law of headscarves is difficult territory, with multiple expert opinions developed over more than a millennium. Since examining this is out of this paper's scope, I will go by the majority of Islamic jurisprudence which believes that women are required to cover their hair in some way; of course, this does not mean that all Muslim women follow (or should follow) these religious rules. However, since this paper deals with Turkey, it is interesting to note that even in the heavily secular country 67% of men and women prefer that women at least partially cover their hair. See Jacob Poushter, *How People in Muslim Countries Prefer Women to Dress in Public,* Pew Research Ctr. (Jan. 8, 2014), online at http://www.pewresearch.org/fact-tank/2014/01/08/what-is-appropriate-attire-for-women-in-muslim-countries/.

71: Evans, 7 Melb. J. Int'l. L. at 65 (cited in note 9).

72: *Id.*

73: *Sahin,* 44774/98 ECtHR Grand Chamber at 26, ¶ 110.

it is clear that that the Court and those representing the Turkish government are emphasizing the rule against Islamic headscarves. In fact, Sahin explains that the Turkish authorities only apply the restriction of religious symbols to Muslims, whereas Jews are allowed to wear *kippot* (skullcaps) and Christians are generally allowed to wear crucifixes.[74] Moreover, the circular sent out by the University of Istanbul that Sahin was appealing specifically targeted Muslim students. It only barred from classes and examinations students "whose 'heads are covered' (who wear the Islamic headscarf) and students… with beards," a Muslim religious requirement for men.[75] It was not until months later that the university issued another circular that mentioned other religious symbols.

The Court does little in either case to prove that duty to wear the headscarf is an inherently misogynistic and sexually unequal rule. They simply state it as fact with little to no explanation. In fact, it seems from the opinions that the judges did little research into the Koran or other Islamic texts to examine the reasoning for the headscarf, or spoke with Muslim women about the reasons why they might choose to wear or not to wear it.[76] Jill Marshall phrases the issue nicely when she writes that "the assumed conflict between the Islamic faith and the rights of women goes uninvestigated" by the Court.[77] She goes on to argue that by declaring the headscarf as inherently unequal without deep analysis of the religious reasoning behind it displays an obvious disrespect for Islam as a whole and the Muslim women who choose to wear it. Carolyn Evans agrees, emphasizing that the Court states that the headscarf is a symbol of gender inequality with hardly a line of explanation. She says that the ECtHR's assumptions about the inherent gender inequality embodied by the headscarf seem to be based off a popular western view of Islam as oppressive to women. For the Court, there is no reason to go into further analysis or detail about the headscarf because "it is a self-evident, shared understanding of Islam" that the religion is oppressive to women.[78] Cindy Skach adds that the "Grand Chamber's analysis in this regard was notably thin and unsatisfying," simply repeating what the ECtHR had said in earlier decisions.[79] In fact, much of the Grand Chamber decision is quotations from

74: *Id.* at 21, ¶ 88.

75: *Id.* at 3, ¶ 16.

76: See Saktanber & Çorbacioğlu, 15 Soc. Pol.: Int'l Stud. Gender, St., & Soc'y. at 530 (cited in note 4); Jill Marshall, *Conditions for Freedom? European Human Rights Law and the Islamic Headscarf Debate,* 30 Hum. Rts. Q. 631 (2008); and Evans, 7 Melb. J. Int'l. L. at 65 (cited in note 9).

77: Marshall, 30 Hum. Rts. Q. at 633 (cited in note 76).

78: Evans, 7 Melb. J. Int'l. L. at 65 (cited in note 9).

79: Cindy Skach, Sahin v. Turkey *& 'Teacher Headscarf,'* 100 Am. J. Int'l. L. 186, 192 (2006).

the earlier Chamber decisions, *Dahlab,* and *Refah.* As Judge Tulkens phrases it, the Court views the headscarf as the alienation of women, but "what, in fact, is the connection between the ban and sexual equality? The judgment does not say."[80]

If the Court would have examined the meanings of the headscarf more thoroughly, the judges would have discovered a much more nuanced picture. Many women find wearing the headscarf liberating. After interviewing Canadian Muslim women who chose to wear the headscarf in 1998 and following up with them ten years later, Katherine Bullock explains that the women find wearing the headscarf freeing and enjoyable. She says that the headscarf has two dimensions: one that represents piety and religiosity, and another that is a "healthy way of desexualizing women in public space." This last dimension is key because it allows women to feel liberated from the fashion industry and the Western cultural ideal of skinny beautiful femininity.[81] In Bullock's words, the veil returns to the women "personhood, dignity and respect" that is lost in the sex-heavy Western culture.[82] Rather than viewing the veil as a marker of inequality, these women voluntarily choose to wear it because they believe that it makes them more equal in the public sphere. While Bullock's work deals with only Canadian women, it is but one study of many that show the multifaceted way in which Muslim women view the veil—a feature of veiling that has been noted by a significant number of scholars.

It is for this reason that Seyla Benhabib calls veiling a "*complex* institution that exhibits great variety." [83] She laments that the West has "reduced" this complicated cultural tradition to just a few items of clothing that play oversized symbolic roles in Westerners' judgment of Islam. Rather than simply assuming that the headscarf marks women as inferior, the ECtHR has a duty to challenge its initial negative reactions to the veil and study its true meaning to many Muslim women. While, of course, not everyone will agree with this picture of the headscarf, a response like that would miss the point. The views of the headscarf as expressed in Bullock and Benhabib's research are legitimate and cannot simply be brushed aside by states that view the headscarves as restrictive. Moreover, I agree with Judge Tulkens that it is not the Court's role to judge the meaning or legitimacy of religious practices. As

---

80: *Sahin,* 44774/98 ECtHR Grand Chamber at 48, ¶ 11 (Tulkens, J., dissenting).

81: Katherine Bullock, "Hijab and Belonging: Canadian Muslim Women," in Theodore Gabriel, ed., *Islam and the Veil: Theoretical and Regional Contexts* 168 (2011).

82: *Id.*

83: Benhabib, *The Claims of Culture* at 94 (cited in note 67).

long as the mandatory wearing of the headscarf is not so terrible a practice that it corrupts "democratic society"—which seems highly unlikely for mere headscarves—then the government has little right to judge the legitimacy or morality of the practice.[84]

IIId. Argument from Subjugation of Women

The Court does not only believe that the headscarf is inherently unequal, but it also indicates that it thinks that the headscarf leads to inequality in another way: it permits the communal and familial subjugation of women. Although the Court does not say so as explicitly, much of the Court's discussion of the headscarf's gender inequality is based on the idea that the state must protect Muslim women from family or community pressure to wear the headscarf against their wishes. Nowhere is this clearer than when the Court in *Leyla Sahin* compares the headscarves worn in Turkey to the full-body coverings forced on Afghani and Iranian women by the Taliban and Iranian government.[85] Like the judges did with the argument from political symbolism and gender inequality, they paint headscarf-wearing women and Islam in broad brush strokes. In the Court's pronouncement that banning the headscarf for the sake of protecting "the rights and freedoms of others, public order and public safety," there is an implicit indication that women must be protected from their own ideas about what is good for them. Rather than dealing with the details, the Court generalizes. By banning the headscarf, the Court believes that it can prevent this pressure on women. Certainly, the subjugation of women by those who pressure them into donning the headscarf is problematic and the state should work to stop such pressure. But there is no evidence that either Dahlab or Sahin were forced to wear the veil and, more generally, there is little evidence that these are widespread problems in Turkey or Switzerland.[86]

In fact, by denying women the right to voluntarily don the headscarf, the

---

84: Judge Tulkens wrote in her dissent, "It is not the Court's role to make an appraisal of this type—in this instance a unilateral and negative one—of a religion or religious practice, just as it is not its role to determine in a general and abstract way the signification of wearing the headscarf or to impose its viewpoint on the applicant." *Sahin*, 44774/98 ECtHR Grand Chamber at 48, ¶ 12 (Tulkens, J., dissenting).

85: *Id.* at 22, ¶ 92.

86: Marshall, 30 Hum. Rts. Q. at 649 (cited in note 76). However, the same may not be true about France. The Stasi Committee reported in 2003 that many young French Muslim women felt pressured to wear headscarves. Nieuwenhuis, 1 Eur. Const. L. Rev. at 506 (cited in note 1).

EctHR and the governments are the ones pressuring women. Now observant Muslim women must choose between following one of their religious edicts and studying at a university, a choice that, at least ostensibly, should be barred by the right to education established in the ECHR's Protocol 1. Jill Marshall makes a strong case that the EctHR is taking away from Muslim women the exact right that it promises to protect: the personal autonomy of women. She believes that the Court is making assumptions about women who wear the headscarf, because wearing it "may not be a symbol of humiliation, subordination, or oppression by a patriarchal religion or community" as the Court believes.[87] As mentioned earlier, many women embrace the headscarf, believing that it liberates them and gives them more control over their lives. Rather than protecting women's rights by saving them from a misogynistic religious rule, she argues that "in the Islamic headscarf cases the Court tells people how to behave."[88] If anything, making the assumption that the Court is implicitly making—that most women wear headscarves because they are forced to— is "controversial, some may even say insulting."[89] To make such an argument, the Court needs evidence which it does not attempt to provide.

There is another issue with this "paternalistic" argument, as Judge Tulkens calls it: the argument projects an image of women as submissive that contradicts the image that is presented by EctHR's argument from pressure.[90] This contradiction is highlighted by Carolyn Evans, who explains that the Court views women as both "victims" and "aggressors," and is an argument that fits nicely into my division of the EctHR's arguments. On the one hand, the argument from subjugation of women displays women as victims of a "gender oppressive religion" who "need protection from abusive, violent male relatives."[91] On the other hand, the argument from pressure displays women as aggressors who "force[] values onto the unwilling and undefended" and who are "inherently and unavoidably engaged in ruthlessly propagating [their] views."[92] This dual image of women as both victims and aggressors is littered throughout both *Leyla Sahin* and *Dahlab* and begs questions about the feasibility of using both the argument from pressure and the argument from subjugation of women: if veiled women only wear the headscarf because they are forced to by male relatives, why would other women feel that veiled women were forcefully encouraging them to follow suit? Yet, if veiled women

87: Marshall, 30 Hum. Rts. Q. at 649 (cited in note 76).
88: *Id.* at 642.
89: *Id.* at 649.
90: *Sahin*, 44774/98 ECtHR Grand Chamber at 48-49, ¶ 11-13 (Tulkens, J., dissenting).
91: Evans, 7 Melb. J. Int'l. L. at 71 (cited in note 9).
92: *Id.* at 72.

are encouraging other women to wear the headscarf against their wishes, then surely the veiled women would not need to be forced to wear the veil in the first place?[93] This dual image accounts for the contradictory way in which the Court approaches the relationship between headscarves, women, and Islam. As Evans concludes, "In the course of a single sentence in *Dahlab*, Ms Dahlab transforms from a woman who needs rescuing from Islam to an Islamic woman from whom everyone else needs rescuing."[94] Indeed, it shows that there is little reasoning behind the Court's arguments and it seems that there is not much that headscarf-wearing women can do that would disprove the Court's dual vision of them. For the Court to legitimate its reasoning, it must come to terms with this contradiction.

IV. Conclusion

In *Dahlab* and *Leyla Sahin*, the ECtHR employs four main arguments to defend bans against wearing Islamic headscarves in school environments. The argument from pressure states that by wearing the headscarf a Muslim woman is inherently pressuring other Muslim women to wear it, and proselytizing to non-Muslims. The argument from political symbolism states that in countries where there are strong Islamic fundamentalist movements, the headscarf acts as a political symbol of defiance to secularism and marks its wearer as a supporter of overthrowing the secular system. In the argument from gender inequality, the headscarf marks women as inferior to men and wearing it is antithetical to the values systems of modern democracies. In the argument from the subjugation of women, the Court views itself in a paternalistic role protecting women from families and communities who pressure them to wear the headscarf against their will.

I believe that the Court's usage of these arguments exposes four overarching issues with the Court's reasoning. Together, these issues do not only indicate that these two judgments were badly reasoned and not fully investigated; they also provide hints that the entire line of ECtHR jurisprudence on this topic faces difficulties and must be re-examined. While not all of the significant wave of criticism that this line of jurisprudence has received has

93: To put this point another way, the pressure that non-veiled Muslim women feel from veiled women is one of condescension for not properly practicing their religion. This pressure is necessarily premised on the voluntary basis by which observant women veil themselves or else they would be in the same "voluntary observance" category as the non-veiled women. If veiled women do not wear the headscarves voluntarily, the condescension-based pressure would not exist.

94: Evans, 7 Melb. J. Int'l. L. at 73 (cited in note 9).

been well-argued, most of the legitimate objections to the Court's decisions are contained within these four issues. After discussing them, I will outline three general recommendations for the Court that it should use to solve these problems in its jurisprudence by taking advantage of the next case on this topic that appears on its docket.

IVa. Overarching Issues

The first overarching issue with the arguments is that they make clear that the Court does not understand Islam and that the judges allow the West's often negative picture of the religion to color their decisions. In the arguments from "secularism" the Court places meaning on the headscarf that most Muslim women never intend to give it. They become political enemies of the modern state, opposed to its secular and democratic form. Of course some fundamentalist Muslim women might use the headscarf as a political tool, but that is no reason to restrict all women from following their religious duties. Many of these women, like Dahlab and Sahin, are dedicated citizens who hope to give back to their communities through teaching and medicine.

In the arguments from "equality," the Court treats headscarf-wearing women as individuals who do not know what is best for themselves. It strips them of their personal autonomy to make decisions regarding their beliefs and religion. Ironically, when the ECtHR claims that it is "necessary in a democratic society" to treat headscarf-wearing women in this way, it seems reasonable for the women to question what modern secularism and pluralism is all about. From these decisions it seems that the judges allow the popular negative Western attitude toward Islam to color their adjudication of the cases. By labeling Islam and headscarves as anti-secularism and anti-gender equality without more supporting evidence, the ECtHR seriously limits the ability of observant Muslim women to participate in the public sphere without sacrificing their religious practices. Unfortunately, the Court shows little interest in investigating Islam or its beliefs more deeply, instead trusting that the state governments can be relied upon to balance freedom of religion with protecting secularism.

The second issue that the arguments highlight is that the Court often ignores nuances in the headscarf cases, instead treating disparate cases in similar ways. Despite the highly different situations from which the *Leyla Sahin* and *Dahlab* cases arose—one involving a teacher in a primary school in a Christian-majority country, the other involving a student at a university in a Muslim-majority country—the Court utilizes strikingly similar arguments for both. In fact, all but one of these four main arguments appears prominently in both cases. At first glance, this continuity might seem admirable: despite

the inevitably different details that will arise in cases before a supra-national court, the ECtHR is attempting to establish a uniform jurisprudence that is coherent. A uniform jurisprudence is often important so that countries and citizens understand what is allowed and not allowed by the ECHR and can act accordingly.

However, when examined more closely, the similarity in the arguments is troubling. The Court is not establishing a uniform jurisprudence on a specific type of case; it is, in fact, applying a uniform jurisprudence to two quite *different* cases. This is especially problematic for a human rights court, because of the importance placed on understanding the types of pressures that minority applicants feel. In cases involving religious liberties, judges can only understand these pressures by grasping the importance of certain rituals or practices to those individuals who believe in the faith—regardless of how strange or unusual they may seem—and placing them in the case's context. However, the Court chose not to allow these detailed issues to influence its arguments. For example, the Court cites *Dahlab* extensively in *Leyla Sahin* in making the argument from pressure, failing to make the distinctions that Dahlab was in a much better position to "pressure" as a teacher and that Dahlab was dealing with more impressionable primary school students while Leyla Sahin was dealing with much older university students. Meanwhile, the Court used the very stark difference in the minority communities to which the two women belonged—Dahlab as a Muslim in a Christian-majority country and Sahin as an observant Muslim in a secular-majority country—only to highlight the feelings of intimidation felt by the Turkish majority, not Sahin.

Much of the decisions' dearth of details or nuanced arguments mostly likely stems from the third overarching issue: the Court's granting a wide margin of appreciation to the states in dealing with the headscarf cases. In both cases, the Court announced that the sensitivity of the topic and the lack of European-wide standards on headscarves in public schools indicated that Turkey and Switzerland deserved a significant margin of appreciation in determining their domestic laws. Because of the significant margin of appreciation, the Court explained that the countries had leeway in how to handle headscarves and it was not appropriate for the Court to get too involved.

As with the second issue, this significant margin of appreciation appears positive on its face for it allows countries to handle the headscarf issue in ways appropriate to their specific cultural tradition, something that is often seen as an advantageous component of the ECHR system. However, both countries chose to ban the headscarf in similar ways, as do a significant number of other countries. When looking at the similarity of argument in this way, it seems that the Court's grant of a significant margin of appreciation is actually

allowing the states to run free in unjustly restricting their citizens' religious rights. Instead of the wide margin of appreciation permitting countries to deal reasonably with the headscarf issues, it is creating a situation in which there is just no "European supervision" at all.[95] More specifically, since the wide margin means it is unlikely that the Court will intervene in national decisions, it creates an excuse for the lack of care and diligence in examining the details and nuances of the cases on the topic, as demonstrated by these two decisions.

The fourth issue with the decisions is tied in with the first problem: the Court does not seem to examine the full extent of the ramifications that the bans—and thus its own decisionswill have on observant female Muslims in Europe. The evolving line of ECtHR jurisprudence that is represented in *Leyla Sahin* and *Dahlab* will have a significant negative effect on the roles that observant Muslim women can play in the European public spheres. As the European-wide guarantor of individual liberties, the Court is looked to as a defender of political, cultural, and religious minorities. Instead, the Court has approved the essential barring of thousands of observant Muslim women from the teaching profession in Switzerland and from pursuing higher education in Turkey. Outside of just the specific details of these two cases, by allowing—and even advocating—for the relegation of headscarf-wearing women to the sidelines, the Court is marking as acceptable similar attitudes and behavior across Europe. It is deeply troubling that the majority in *Leyla Sahin* did not grapple with the ways that this line of jurisprudence will affect Muslim women that were brought up by Judge Tulkens in her dissent.

IVb. General Recommendations

When adjudicating decisions as delicate as the headscarf cases, the ECtHR must balance a number of divergent interests and opposing rights. Managing these is a complicated task, but it is crucial that in future cases the Court avoids these four overarching issues with its current jurisprudence. While it is difficult to provide particular standards to the Court without knowing the specifics of the future cases, there are three general recommendations that the Court should begin following.

The ECtHR should decide future headscarf cases with a narrow margin of appreciation. The main jurisprudence of the ECtHR has been wise to recognize the margin of appreciation because it has played a critical role in ensuring that national governments do not feel as if the Court is hoisting a strict set of standards upon countries. However, the Court's belief that

95: *Sahin*, 44774/98 ECtHR Grand Chamber at 44, ¶ 3 (Tulkens, J., dissenting).

these cases necessitate a particularly wide margin because the "relationship between State and religions are at stake" and there is significant "diversity of the approaches taken by national authorities on the issue" is misguided.[96] This is precisely the type of issue that the Court must supervise closely. In countries with as significant religious majorities as Switzerland and Turkey, it is especially likely that the rights of minority religious groups will need protection. The Court cannot just trust that the national governments—by definition, the majorities—will defend these rights, especially given the great impact these decisions will have on observant Muslim women. Moreover, the margin of appreciation should be especially narrow in cases dealing with universities because European opinion on the subject of headscarves is, in fact, not diverse like the Court says: Turkey is the only country in the Council of Europe to ban headscarves at universities.[97]

Second, the ECtHR should make clear that the argument from pressure can only be a valid reason for an interference with a woman's right when there is clear evidence that the woman actually pressured others. In both *Dahlab* and *Leyla Sahin*, the Court made assumptions about the effects of the women's headscarves on others without examining statistical or anecdotal evidence. Without finding evidence of the pressure, the Court has little justification for interfering with a woman's religious right. Looking at the *Dahlab* case, the Court, responding to Dahlab's assertion that she told students who asked about her headscarf that she wore it because it was cold outside, assumed without speaking to the students or their parents that they felt pressured. Because of this, the Court unjustly denied her the right to wear the headscarf. While this unfairly placed blame for religious coercion on Dahlab, the case would have been different if Dahlab had instead given her students a long religious explanation for why she wore the headscarf, or if the students had complained that she acted in such a way that they felt pressured by her wearing it. In these situations, there would have been evidence that Dahlab either actually promoted Islam in comparison to other religions or the young children felt as if she did. Thus, it might be reasonable for the Court to say her behavior—but not merely the wearing of the headscarf—placed unreasonable pressure on the students. However, when dealing with students old, mature, and intelligent enough to be in university as in *Leyla Sahin*, it seems unreasonable to mark actions short of actual proselytizing as placing pressure on peers significant enough to justify interference with a religious right.

Lastly, the ECtHR should make clear in its judgments that it understands

96: *Id.* at 26, ¶ 109.
97: *Id.* at 13, ¶ 55, and 44, ¶ 3.

the complexity behind the decision of Muslim women to wear or not to wear the headscarf and challenge states' negative generalizations about the practice. While it is unlikely that secularist national governments such as Turkey and Switzerland will stop viewing the headscarf in a negative light, the Court has an opportunity to defend the reputation of a significant European minority. In *Dahlab* and *Leyla Sahin*, the Court did little to call into question the assertions of the state governments regarding headscarves as symbols of subjugation and gender inequality, even though the two applicants clearly did not see them as such. In the future, the Court should more actively investigate the states' arguments on this topic, challenging stereotypes and assumptions. A primary way to do this would be to replace the one-line statement with veiling as a tradition that "appears to be imposed on women," with an examination of the diverse Muslim views of the topic and the importance the tradition plays in Islamic culture.[98] The Court should also take care to only use arguments that are applicable to the specific case that it is deciding, such as not using the argument from subjugation to defend a ban in a country where there is no evidence that this is a predominant problem. Rather than inappropriately judging the practices and adherents of a religion, these changes would show respect to the diversity of opinions on the topic and indicate that the Court is speaking about the religion in the terms of its adherents.

With the Muslim population of Europe continuing to grow, the legal issue debated in the headscarf cases will not disappear. In upcoming cases, the ECtHR must strive to adjudicate the headscarf fairly and un-prejudicially, grant a narrower margin of appreciation, grapple with the details of the situations rather than make generalized assumptions, and make clear that it understands the significant ramifications of its judgments. If the Court does not alter its jurisprudence on this topic, it risks ostracizing a significant European minority and failing to fulfill its role as a defender of religious liberties for all.

98: *Dahlab,* 42393/98 ECtHR at 13.

Tax Avoidance Law: Finding the Supreme Court's Role in Uncertainty

Frank Yan†

I. Introduction

In 2014, a practice where U.S.-based corporations enter complex income shifting transactions to avoid U.S. income tax took center stage in tax avoidance policy.[1] This practice known as corporate inversions is just one attempt of many by corporations to continually find new strategies to avoid tax liabilities while still remaining within the letter of the law. The struggle for tax law is that it must continually adapt to these strategies or it will become outdated and ineffective.[2] Given this challenge, the judiciary has served as a policymaker by providing means beyond the written law for tax enforcement.

In this article, I will use the landmark case *Gregory v. Helvering*[3] to explain the principles of extralegal tax adjudication central to federal tax enforcement. The term "extralegal" refers to approaches of statutory interpretation where the Court fills a gap in the law by interpreting a statute's intent or purpose. In reaction to these principles, I will raise concerns about the blurring of the separation of powers between the Court and Congress and reduced certainty in tax avoidance law. I will attempt to clarify the answers to these contentions by examining the development of modern tax avoidance doctrine, highlighting the inconsistent applications of *Gregory* principles by lower federal courts and the clarification of the separation of powers through Congress' codification of tax avoidance doctrine. Finally, I argue that it is justified and viable for modern extralegal tax avoidance doctrine to sacrifice some degree of democracy and some certainty in tax law because such extralegal principles help maintain the long-term efficacy of American government while posing little threat to the integrity of democratic government. At the same time, the Supreme Court must be more active in maintaining national uniformity of court doctrine

---

†: Frank Yan is a third-year in the College, majoring in Economics and Political Science.

1: Shayndi Raice, *How Tax Inversions Became the Hottest Trend in M&A*, Wall St. J. (Aug. 5, 2014).

2: A distinction should be raised between tax *avoidance* and tax *evasion*. While the two behaviors are not so different with regard to the ultimate goal of reducing one's tax liability, tax avoidance practices are formally legal whereas tax evasion practices are a direct rebuke of the law.

3: 295 U.S. 465 (1935).

over lower federal courts in order to prevent a degradation of federal power.

A final preliminary matter needs addressed. At first blush, this article may seem to rest on the assumption that the practice of tax mitigation is wrong in a moral or patriotic sense. Responses against attacks on the tax mitigation and avoidance include Justice Learned Hand's lower court opinion in *Gregory*, where he states that, "Anyone may so arrange his affairs that his taxes shall be as low as possible; he is not bound to choose that pattern which will best pay the Treasury; there is not even a patriotic duty to increase one's taxes."[4] This essay does not censure the idea that individuals and firms should be allowed to lower their tax liabilities. There is also no assumption that government taxation and spending occurs in the most efficient and socially beneficial way. However, taxation is here to stay for the long (*long*) term. The focus of this essay instead is on those abusive tax strategies that, if allowed to proliferate, would mean that select individuals and firms with the ability to purchase professional tax services can significantly lower their tax liabilities, while other segments of the American population disproportionately bear the burden of federal taxation.

II. Tax Avoidance Politics, Policy, and Law

Before delving into the Supreme Court's role in tax avoidance adjudication, it is important to first situate the judicial system's role in tax enforcement relative to the legislative and executive branches. In a world of constantly evolving tax avoidance strategies, tax law is to some degree destined to be obsolete once it is passed, but Congress severely lags behind in revising the tax code to address these strategies.[5] For example, corporate inversions began in the U.S. during the 1990s and Congress reacted with legislation in 2004, but the legislation has been weak and still allows for the proliferation of inversions transactions a decade later.[6] Major tax reform that can address the fundamental issues incentivizing tax avoidance takes even longer; the most recent major tax reform was President Ronald Reagan's 1986 reform.[7] The cause of Congress's slow reaction can be found in the fundamental structure of American lawmaking between the executive and legislative branch. Deliberations over

4: *Helvering v. Gregory*, 69 F. 2d 809, 810 (1934).

5: Lee Hamilton, *Why Isn't Congress More Efficient?*, Center on Congress at Indiana University (2007), online at http://www.centeroncongress.org/why-isnt-congress-more-efficient.

6: Mindy Herzeld, *News Analysis: What's Next in Inversion Land?*, Tax Analysts (2014), online at http://www.taxanalysts.com/www/features.nsf/Features/F817995A255AFD2485257CF900427B06?OpenDocument.

7: See, e.g., Jeffrey Birnbaum & Alan Murray, *Showdown at Gucci Gulch* (1988).

bills can persist over several years and election cycles, because 535 lawmakers in Congress each have the desire and individual resources to voice their opinions and their versions of a given bill, Congressional procedure from committee hearings to a floor vote is a tedious process, and because the executive branch has veto power to check Congressional action.[8] Under these circumstances, Congress is almost always ill equipped to react quickly to new tax avoidance strategies.

When Congress does react in a timely manner, it does not always create effective laws and rules to curb tax avoidance. On the one hand, restrictions targeting specific avoidance strategies only provide more text for tax professionals to operate on and even more incentive for tax avoidance. The oft-cited allusion in federal taxation literature to Moses' rod sums up the dilemma: "every stick crafted to beat on the head of a taxpayer will metamorphose sooner or later into a large green snake and bite the commissioner on the hind part."[9] On the other hand, Congress may draft broad statutes, such as passive activity loss rules that aim to curb general categories of tax avoidance practices by restricting groups of outcomes rather than specific processes of tax avoidance.[10] Yet, even with these broad "outcomes-oriented" statutes, the fact that they still unavoidably reference certain types of transactions albeit on a broader level have permitted cunning tax professionals to plan around these statutes.[11] Overall, then, Congress can try to fight tax avoidance, but it cannot effectively play a direct role.

Beyond Congress, actions by the Internal Revenue Service (IRS) and the federal courts against tax avoidance are intertwined, with arguably more weight given towards the federal courts. The IRS can be effective at interpreting the tax Code and using its rulemaking authority to curb tax avoidance strategy, but even so, the IRS is not the final authority; its rules and interpretations can be resisted through litigation in federal courts. The prevalence of litigation is amplified in tax avoidance, because when the rewards of avoidance can be in the hundreds of millions of dollars, corporations have a strong incentive to spend the several million dollars to litigate.[12] Consequently, the IRS must work side-by-side with federal courts to maintain interpretations of the Code

8: See, e.g., Woodrow Wilson *Congressional Government* (1885).

9: Martin Ginsburg, *Making Tax Law through the Judicial Process*, 70 A.B.A. J. 74, 76 (1984).

10: Erik Jensen, *Legislative and Regulatory Responses to Tax Avoidance: Explicating and Evaluating Alternatives*, 57 St. Louis U. L.J. 1 (2012).

11: *Id.* at 19.

12: In the recent case *WFC v. U.S.*, Wells Fargo sought a $426 million tax deduction through a tax avoidance strategy developed by accounting firm KPMG. See *WFC Holdings Corp. v. U.S,* 728 F. 3d 736 (8th Cir. 2013*).*

favorable to the government.

Given these considerations on Congress and the Internal Revenue Service, both are unable to definitively stop tax avoidance strategies. Federal courts, then, are important in tax avoidance because they fill a vacuum in federal policy where the Congress and the IRS stand slow or incompetent to react.

IIa. *Gregory v. Helvering*

Under these circumstances, the Supreme Court has historically moved beyond the inadequacies and limitations of analyzing the plain meaning of a statute to interpreting the legislative intent and purpose of tax laws in order to fill in gaps in the tax code. Over the decades since the 1930s, tax avoidance adjudication has derived its foundational principles from the early case of *Gregory v. Helvering*.[13] In *Gregory*, petitioner Evelyn Gregory owned United Mortgage Corporation (UMC), which owned the subsidiary Monitor Corporation (MC).[14] Gregory wanted to sell MC without incurring taxes, so she set up a new corporation called Averill Corp solely to acquire MC, liquidate MC to herself, and personally sell the stocks of MC to avoid income tax.[15] At issue in the case was whether Gregory's transactions qualified as a corporate reorganization under section 112 of the Revenue Act of 1928.[16]

IIb. Principles of *Gregory*

The Supreme Court's decision in *Gregory* is important because it embodies three of the arguably most important principles in federal tax avoidance adjudication: examining legislative intent, business purpose, and substance over form. Writing for the majority, Justice Sutherland struck down the tax benefits by arguing that Gregory's actions did not qualify under the definition of a reorganization, because Gregory's actions had "no business or corporate *purpose*—a mere device which put on the form of a corporate reorganization as a disguise for concealing its *real character*" and that "the transaction upon

---

13: 293 U.S. 465 (1935).

14: *Id.* at 467.

15: *Id.*

16: *Id.* at 467-68. Section 112 of the 1928 Revenue Act reads, "The term `reorganization' means . . . (B) a transfer by a corporation of all or a part of its assets to another corporation if immediately after the transfer the transferor or its stockholders or both are in control of the corporation to which the assets are transferred." egislateuthority to tuionally npowers and uncertainty in lawf the extent it creates uncertainty and entre doctre is an encroach

its face lies outside the *plain intent* of the statute." [17] When Justice Sutherland referred to the "plain intent," he had to personally evaluate Congressional reports on the bill version of the law.[18] Justice Sutherland found that the section's explicit goal was to protect normal business transactions aimed at reorganization from being taxed, and consequently Justice Sutherland concluded that the application of the reorganization statute required that Gregory's transaction have a business purpose in order to be tax exempt.[19] Finally, Justice Sutherland broke from the literal text of the 1928 statute to argue that even though Gregory's transaction followed the statutory definition of reorganization in *form*, the "real character" or *substance* of the transaction in reality had no business purpose.[20] Gregory's transaction thus could not qualify as a tax-exempt reorganization.

IIc. Merits and Flaws: Legislative Intent, Business Purpose, and Substance over Form

Practically speaking, the principles from *Gregory* have served as a highly useful foundation for combatting tax avoidance strategies. As explained earlier, Congress' laws inevitably fail to account for all possible situations for a given issue. By looking at legislative intent and purpose instead, the Court can act on a broader base beyond the literal statute, in order to help fill the gaps of existing law as they appear.[21] These extralegal tools bring significant practical results, allowing the federal government to consistently win tax avoidance cases involving business purpose and substance over form.

On a more fundamental level, however, the Court's extralegal approaches in tax avoidance adjudication raise important issues over statutory interpretation, the influence of courts over policy, and the rule of law. In the first place, Justice Sutherland's opinion is based on an intentionalist theory of statutory interpretation, where a judge or justice ascertains what lawmakers

---

17: *Id.* at 469-70 (emphasis added). It is interesting that Justice Sutherland, a highly conservative justice, voted against the taxpayer and wrote an opinion expanding the government's taxing ability. Tel Aviv Law School Lecturer Assaf Likhovski notes the political context of the decision within the Great Depression and of the potential indictment of millionaire Andrew Mellon for tax evasion at the time. Assaf Likhovski, *The Duke and the Lady:* Helvering v. Gregory *and the History of Tax Avoidance Adjudication*. 25 Cardozo L. Rev. 953 (2003).

18: Jasper Cummings, *The Supreme Court's Federal Tax Jurisprudence* 90 (2010).

19: *Gregory,* 293 U.S. at 469.

20: *Id.*

21: Ray Knight & Lee G. Knight, *Substance over Form: The Cornerstone of Our Tax System or a Lethal Weapon in the IRS's Arsenal?*, 8 Akron Tax J. 91 (1991).

intended in passing a given statute.[22] Justice Sutherland's interpretation of the Congressional report in *Gregory* was quite clear-cut because the author of the statute's reorganization section specified in Congressional testimony that his bill aimed to protect ordinary business transactions.[23] However, legislative intent is not always so easily ascertained, partly because of the multitude of Congressional reports, floor debate transcripts, and other documents that may indicate different or even contradictory intentions.[24] This example in *Gregory* of an intentionalist approach yields two sides of the coin. On the one side, interpreting Congressional intent helps prevent abuse of the law and reinforces the efficacy of federal legislation. In this sense, the federal courts help maintain Congress as the most powerful branch of government. On the other side, a troubling caveat is that if judges and justices are allowed to proceed freely with interpreting legislative intent, any judge may read his or her personal values into a decision by selectively choosing how to interpret a statute's intent.[25]

Furthermore, while the merits of legislative intent are ambiguous, the Court's development of tax avoidance doctrine over time significantly transgresses Congressional intent towards a common law approach. With regard to *Gregory,* although Justice Sutherland had reasoned that business purpose was necessary because the intent of the statute seemed to be geared towards protecting normal business transactions, the business purpose requirement in modern court doctrine has become a requirement independent of a statute's legislative intent.[26] By making the business purpose doctrine a standalone evaluation of a contended tax transaction, the Court has presupposed an assumption about the IRS code as a whole: that the purpose of the tax code is only to tax regular business transactions rather than extremely technical strategies involved in avoiding tax liabilities.[27] The Court has essentially developed doctrine on an more extreme philosophy of intentionalist interpretation, where the courts may still interpret legislative intent but also rely on mental exercises of thinking broadly about a statute's purpose, not necessarily directly relying on any Congressional records or

22: See, e.g., William Eskridge, Philip P. Frickey, & Elizabeth Garrett, *Legislation and Statutory Interpretation* (2000).

23: Cummings, *The Supreme Court's Federal Tax Jurisprudence* at 90 (cited in note 18).

24: Eskridge, *Legislation and Statutory Interpretation* at 214 (cited in note 22).

25: *Id*. at 222.

26: See the business purpose prong of the economic substance doctrine, as described below.

27: Michael Livingston, *Practical Reason, 'Purpovism,' and the Interpretation of Tax Statutes*, 51 Tax. L. Rev. 677 (1996).

other concrete evidence.

The discussion over statutory interpretation and independent construction of court doctrine boils down to important implications for the separation of powers between the Court and Congress. The slippery slope of abusing legislative intent or purpose leads the Court to the point of "legislating" substantial rules, even though it is officially, within the federal balance of powers, not the role of the courts to legislate. At the root of the issue is the fact that the legitimacy of American government is based in popular sovereignty, or in other words the ability of citizens to participate in government and vote for their leaders in the legislative and executive branches.[28] In contrast to the power of popular sovereignty, Supreme Court justices and federal court judges are politically appointed, serve for life tenure, and have developed their official duty of maintaining of the rule of law and the Constitution over the whims of popular sovereignty.[29] Furthermore, it may be argued that Congress itself does not have the authority to delegate its own constitutional powers away to any other branch of government. Case law on the non-delegation doctrine is especially notable during the Great Depression, when the Court struck down provisions of the National Industrial Recovery Act for delegating legislative power to President Franklin Roosevelt during a time of extremely aggressive executive power.[30] The suggestion, then, is that extralegal Court doctrines are not sustainable and contradictory of American democratic values.

Finally, there are philosophical issues against Court flexibility with regard to the rule of law. There is an issue with the uncertainty of tax law under extralegal doctrine, which stands in contrast to the conventional sense of concrete, written law. On an abstract level, the written law is ideally absolute and consistent throughout its application because the resulting certainty provides legitimacy to the laws of Congress and upholds the structure of government based upon statutory law.[31] While realistically, most law is not absolutely clear in the ideal sense, tax avoidance law has more ambiguous statutes than other areas of law because of the highly fluid nature of its subject.[32] At this greater extreme, it may be questioned whether Court doctrine on tax avoidance has become so unpredictable that it is counterproductive against the rule of law.

This section has summarized the range of issues involving statutory interpretation, the separation of powers, and the rule of law that result when

28: Robert McCloskey, *The American Supreme Court* (2010).

29: *Id.* at 7.

30: See *Panama Refining Co. v. Ryan*, 293 U.S. 288 (1935).

31: Rebecca Prebble & John Prebble, *Does The Use Of General Anti-Avoidance Rules To Combat Tax Avoidance Breach Principles Of The Rule Of Law? A Comparative Study*, 55 St. Louis U. L.J. 21 (2010).

32: *Id.*

the federal courts deviate from the plain text of the law in tax avoidance cases. As Chief Justice John Marshall noted in *Marbury v. Madison*[33] and Justice Antonin Scalia alluded to in his treatise on textualism, this country is "a government of laws, and not of men."[34] The common problem among all these issues is how the roles of the three branches of government should be distributed in tax avoidance issues such that the issues are dealt with effectively while maintaining the structure and stability of American government.

III. Modern Economic Substance Doctrine: Reviewing the Merits and Flaws of Court Doctrine

The answer to these contentions in tax avoidance adjudication can be clarified by examining the developments of Court doctrine in modern adjudication. Based upon the principles found in *Gregory,* modern tax avoidance doctrine since 1978 is primarily embodied in what is known as the economic substance doctrine (ESD). Under the ESD, contested transactions face a two-prong test: the economic substance and business purpose prongs.[35] The economic substance prong is an objective test, where the taxpayer must show that the transaction in question led to an economic gain other than a reduction of tax liability.[36] The second business purpose prong is a subjective test identical to the one found in *Gregory*, based upon the ability of the Court to identify the substance of a transaction from the legal form.[37]

The significance of ESD's development is that while the Supreme Court first introduced the prongs of ESD, the Court eventually receded and left the development of the ESD largely to the lower federal courts.[38] By 1978, the Court had decided two major cases, *Knetsch v. U.S.*[39] and *Frank Lyon v. U.S.*[40] where it applied the two prongs of the ESD but did not explicitly describe them as Court doctrine. After *Frank Lyon*, however, the Court has not decided *any* cases on the issues of substance over form and business purpose

33: 5 U.S. 137 (1803).
34: *Id.* at 163. Also see, e.g., Antonin Scalia, *A Matter of Interpretation* (1997).
35: Philip Sancilio, *Clarifying (Or is It Codifying?) The 'Notably Abstruse': Step Transactions, Economic Substance, and the Tax Code*, 113 Colum. L. Rev. 138 (2013).
36: Senate Finance Committee, *Economic Substance Doctrine* (2007)
37: *Id.*
38: Cummings, *The Supreme Court's Federal Tax Jurisprudence* at 205 (cited in note 18).
39: 364 U.S. 361 (1960).
40: 435 U.S. 561 (1978).

in tax avoidance.[41] The result has been that the lower federal courts have been free to develop their own variations of the Court's two-pronged approach in *Knetsch* and *Frank Lyon*, resulting in different versions of ESD across the circuits. In other words, the freedom of the lower courts has resulted in significant uncertainty about the outcomes of tax avoidance adjudication. And while Congress did codify ESD in 2010 in an attempt to increase clarity, the codification only firmly delineated a small part of the ESD.

IIIa. Inconsistency in the Lower Federal Courts

In one decision from the 9th Circuit Court of Appeals, Judge Owen Panner complained about the lower courts' applications of ESD. He noted, "The casebooks are already glutted with tests. Many such tests proliferate because they give the comforting illusion of consistency and precision."[42] Judge Panner's remarks reflect the inconsistent development and application of the economic substance doctrine across the lower federal courts since the Supreme Court's self-restraint from the issue. From 2000 to 2010, lower courts decided upon versions of the economic substance doctrine that differed on whether the doctrine was a conjunctive (requiring both prongs) or disjunctive (requiring one of two prongs) of the economic substance and business purpose prongs.[43] Furthermore, the lower courts have disagreed on whether either prong of the ESD has been satisfied.[44] While some courts have accepted arguments about whether there was a potential for profit other than a lower tax burden to satisfy the economic substance prong, other courts have compared the given transaction to other transactions that usually occur in the similar transactions, and still other courts make a cost-benefit analysis of the transaction.[45]

41: Cummings, *The Supreme Court's Federal Tax Jurisprudence* at 205 (cited in note 18).

42: *Collins v. Commissioner of Internal Revenue,* 857 F. 2d 1383, 1386 (9th Cir. 1988).

43: See Yeoram Keinan, *It is Time for the Supreme Court to Voice its Opinion on Economic Substance*, 7 Hous. Bus. & Tax L.J. 93 (2006).

44: *Id.* at 134-37.

45: Yeoram Keinan, *The Many Faces of the Economist Substance's Two-Prong Test: Time for Reconcilitation?,* 1 N.Y.U. J. L. & Bus. 371 (2005). One specific example comes from the Court of Appeals for the Fourth Circuit. In *Black & Decker Corporation* v. *U.S.*, 436 F. 3d 431 (4th Cir. 2006), the court reversed and criticized the lower district court's analysis of Black & Decker's general business activities in evaluating the economic substance prong, rather than using the profit potential test.

These patterns of lower court interpretation point back to the first caveat coming out of *Gregory*, that the extralegal approach bears a risk of excessive uncertainty detrimental to the rule of federal law. In Federalist 80, Alexander Hamilton argued one of the major advantages to having a national Supreme Court was that it could uniformly interpret and enforce federal law and the Constitution across the States.[46] In a sense the lack of uniformity in applying ESD is a result of the uncertainty as to what requirements (conjunctive or disjunctive) the ESD entails. The existing uncertainty in federal court doctrine is detrimental to the nation because corporations in different circuits are subject to different tax treatment.[47] Without national uniformity in the application of court doctrine, corporations under the jurisdiction of circuits that apply a disjunctive ESD will be more likely to enjoy tax benefits, whereas those under the jurisdiction of circuits that apply a conjunctive test will be less likely to enjoy benefits.[48] This result could put corporations in some areas of the country at a competitive advantage against states in other regions, or even affect the residency choices of corporations for the sake of tax benefits. As seen in very recent cases like *WFC v. U.S.* on this issue, the tax deduction benefits numbering in the hundreds of millions of dollars are a substantial impetus for such possibilities.[49] The greater implication is that the federal governments' actions are distorting economic relationships between states, as well as weakening the rule of federal law and the strength of the federalist system.[50] Due to this weakness from the uncertainty in the tax code, then, either the Supreme Court or Congress needs to take action and reign in the lack of uniformity in avoidance doctrine.

IIIb. Codification of the Economic Substance Doctrine

In reaction to the lack of uniformity in ESD application, Congress has in fact attempted to remedy the situation by codifying ESD. As a part of the Health Care and Education Reconciliation Act of 2010,[51] Congress codified the doctrine in Section 7701(o) of the tax code, stating:

46: Publius, "The Powers of the Judiciary," in *The Federalist Papers* (1788).

47: The Second, Fourth, Eighth, and D.C. Circuits apply a disjunctive test, and the First, Seventh, Eleventh, and Federal Circuits apply a conjunctive test. See Tracy Kaye, "United States," in *A Comparative Look at Regulation of Corporate Tax Avoidance* 335-79 (2012).

48: "Amicus Curiae Brief of Atlantic Legal Foundation in Support of Petitioner," *WFC Holdings Corporation v. U.S.*, WL 1309320.

49: *WFC Holdings Corp. v. U.S.* 728 F. 3d 736, 741-42 (8th Cir. 2013*).*

50: Publius, "The Powers of the Judiciary" (cited in note 46).

51: Pub. L. No. 111-152.

> In the case of any transaction to which the economic substance doctrine is relevant, such transaction shall be treated as having economic substance only if—
>
> **(A)** the transaction changes in a meaningful way (apart from Federal income tax effects) the taxpayer's economic position, and
>
> **(B)** the taxpayer has a substantial purpose (apart from Federal income tax effects) for entering into such transaction.[52]

The above section summarizes the objective economic substance and subjective business purpose prongs of the doctrine, establishing that the doctrine is in fact a conjunctive test.

With regard to the lack of uniformity resulting from uncertainty in Court doctrine, the 400-word codification still leaves a lot that needs to be improved upon. The only clear improvement for uniformity is the specification that a conjunctive test must be used, thereby barring several of the circuits from using a disjunctive test. Besides this specification, there remains only additional ambiguity. First, by defining the satisfaction of the economic substance prong as any transaction that changes the taxpayer's economic position in a "meaningful way," the codification is ambiguous as to what is "meaningful."[53] Next, the statute's failure to define a "transaction" is problematic because it becomes uncertain which transactions the ESD applies to, and more importantly, which portions of the transaction will be evaluated under ESD.[54] When even the codification by Congress is this ambiguous, the Supreme Court must step in to introduce some certainty into ESD in providing guidance and imposing uniformity of doctrine application across the lower courts.

As long as the Supreme Court offers guidance to maintain uniformity, however, it seems that broader uncertainty in the reading of the ESD is still a viable option. First, Congress' codification of the ESD inherently endorses the Court's extralegal approach by putting a legislative stamp on the resulting doctrine from that approach. Furthermore, the practical aspects of an ambiguous ESD seem to outweigh the social costs of the lack of concrete rules in tax

---

52: 26 U.S.C. §7701(o) (2010).

53: "Amici Curiae Brief of the Cato Institute, The Chamber of Commerce of the United States of America, and the Financial Services Roundtable in Support of Petitioner," *WFC Holdings Corporation v. U.S,* WL 1285828.

54: Jasper Cummings, *Economic Substance Doctrine Confusion*, Alston & Bird (2014), online at http://www.alston.com/advisories/economic-substance-doctrine/.

avoidance law.[55] When unpredictable tax avoidance strategies can ultimately hinder the ability of the government to collect tax revenue and function, some consistency in tax law must be sacrificed to ensure the survival of government and the rule of law in the long run.[56] In other words, the weakening of the rule of law in tax avoidance adjudication is a necessary evil to combat equally detrimental tax avoidance strategies.

In another vein, Congress' codification of ESD is also important for the caveat on the separation of powers raised earlier because the text of the codification eliminates some of the issues with the separation of powers and leaves the Court with an implicit authorization to read into the legislative intent of statutes. In explicitly codifying and backing the purely court-made business purpose prong of the ESD with legislative authority, Congress eliminates the problem that the embodied business purpose and substance over form principles were previously "legislated" by the Court without authority. What is left for the courts now is to decide the legislative intent of a statute, in order to determine whether the pursuit of tax benefits can be a justifiable reason to satisfy the definition of a certain statute.[57] This ability seems to be implicitly endorsed by Congress in the provision stating that the "determination of whether the economic substance doctrine is relevant to a transaction shall be made in the same manner as if this subsection had never been enacted."[58] Determining the relevancy of ESD is a way of saying that the courts need to decide whether the statute in a given case intended to allow for tax benefits, and if not, then ESD is relevant. Ambiguous codification, then, is not a blank check for the courts to create their own tax policies. Rather, it eliminates the problems with the separation of powers by officially providing the Court with the authority to examine business purpose, and furthermore only reinforces the Court's ability to interpret statutes and legislative intent.

55: Rebecca Prebble & John Prebble, 55 St. Louis U. L.J at 38 (cited in note 31).

56: Judith Freedman, *Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle*, 4 Brit. Tax Rev. 332 (2004).

57: Brian Galle, *Interpretative Theory and Tax Shelter Regulation,* 26 Virginia Tax Review 357 (2006).

58: 26 U.S.C. §7701(o)(5)(C) (2010).

IV. Strength in Unpredictability: Unpredictable methods to combat unpredictable strategies

In summary, the previous section's analysis yields four main conclusions about the tolerance of the federal government structure towards the federal courts' application of extralegal doctrines against tax avoidance. First, uncertainty in federal court tax avoidance doctrine should be permitted only to the extent that it does not lead to inconsistency in the application of the doctrine across circuit courts. Next, due to the lack of uniformity in the application of ESD, the Court must step in to impose nationally uniform guidelines. Third, the presence of uncertainty in Supreme Court doctrine is viable as supported by Congress' implicit approval and as a practical barrier against the unpredictable nature of tax avoidance strategies, in contrast to the certainty of the rule of law. Finally, the separation of powers issue in the Court's extralegal approaches to tax avoidance is remedied by the fact that Congress has provided official legislative authority behind the Court's now formerly extralegal methods. The Court is now left primarily with the responsibility of interpreting both the codification of ESD and the legislative intent of the relevant tax statute in a given case.

IVa. Implications for Democratic Government

On the last point that Congress has sanctioned the ESD, there remains pushback against whether the Court's use of the doctrine is justifiable in democratic government, even with Congressional approval. Based on Article I, Section 8 of the Constitution, there is no justification for the Court's tax policymaking under ESD. Congress has the sole power founded in popular sovereignty to legislate tax law, and therefore Congress rather than the Court should be taking action to stem tax avoidance strategies. However, a wider perspective regarding the trade-off between the effects of tax avoidance strategies and extralegal court doctrine on American government must be considered.

If the Supreme Court has no ability to develop extralegal doctrine such as the ESD, then the federal courts would be forced to shift towards plain text readings of tax statutes. Even if Congress or executive agencies took on more responsibility to monitor and combat tax avoidance, tax advisory professionals would have an even greater ability to manipulate the plain text readings of tax statutes, and corporations would have even greater incentives to litigate than they do now because the chance of winning a court case would increase based on plain text readings. The result is that corporations and individuals with the resources to hire tax professionals will do so to lower their tax liabilities as

much as possible. When the wealthiest 5% of individuals pay about 57% of total individual income tax revenue[59] and when corporate income taxes make up 10% of all federal tax revenue,[60] it is disconcerting to think about the state of the federal government if corporations and individuals had so much latitude to avoid federal taxes. The federal government would potentially lose hundreds of billions of dollars in tax revenue, and would have to pursue significantly more deficit spending than it does now or significantly cut federal programs and funding. Not to mention, the tax burden would shift to poorer individuals and smaller businesses that do not have the resources to hire tax professionals. Under these circumstances, democratic values would be harmed through the deterioration of federal government administration and through the disproportionate economic power of the wealthy.

On the other hand, if the Supreme Court is permitted the latitude to flexibly interpret tax statutes under the ESD, the threat could be that the Court attempts to dramatically suppress the ability of corporations and individuals to mitigate their tax liabilities. It can be imagined that at some point in the future, justices could decide to apply extralegal tax doctrines so aggressively as to prohibit tax benefits on regular business transactions, hindering and deterring economic activity within the United States. Just as in the previous case of allowing tax avoidance to proliferate, American citizens and companies would suffer financially, tax revenues would decrease, and the government would see setbacks in administrative efficacy.

Between these two cases, the threat of Supreme Court overreach in tax avoidance adjudication is highly preferential over the threat of aggressive tax avoidance strategies. The crux of the entire issue lies on the incentives of each party. On the one hand, the Supreme Court does not have incentives, the resources, or the influence to aggressively prosecute tax avoidance. Moreover, Supreme Court justices are always cautious of overstepping into the authority of the other branches of government for fear of political retaliation, and President Franklin Roosevelt's threat of court packing during the New Deal is a historical lesson always in the back of the justices' minds.[61] On the other

59: Kyle Pomerleau, *Summary of Latest Federal Income Tax Data*, Tax Foundation (2013), online at http://taxfoundation.org/article/summary-latest-federal-income-tax-data.

60: *Historical Amount of Revenue by Source*, Tax Policy Center (2015), online at http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=203.

61: In modern cases such as *National Federation of Independent Business v. Sebelius* the Court implicitly notes its historical transgressions of legislative authority. In *Sebelius*, Justice Ruth Ginsburg writes, "Since 1937, our precedent has recognized Congress' large authority to set the Nation's course in the economic and social welfare realm. The Chief Justice's crabbed reading of the Commerce

hand, there is an entire industry of tax avoidance professionals driven by the incentive of corporate profits. Between these two incentives, to allow the Court to act on extralegal doctrine that leads to minimal threats towards democracy in the short term is not so bad a trade-off in order to maintain the long-term viability of American government.

V. Comparing Tax Law to Other Areas of Law

To further bolster the argument for broad flexibility in the Supreme Court's tax avoidance doctrines, it is useful to discuss the broader issues of statutory interpretation and construction common to other areas of law. Viewing these issues in other areas such as contract law, we see similar problems in these areas of law with similar solutions that maintain long-term social benefits at the short-term expense of providing judicial flexibility.

In contract law, the Uniform Commercial Code (UCC) could be viewed as an analog to the Internal Revenue Code. Like the complex and wide-ranging tax code, the UCC is a complex set of laws identically codified in each state that governs a wide range of areas including transactions in borrowing and lending money, leasing equipment, contracts, and bank deposits. Like the constantly evolving nature of tax avoidance strategies, new types of business transactions are constantly being invented and old ones developing alongside new technology.[62] Curiously, unlike the *ad hoc* development of court doctrine leading up to the codification of the economic substance doctrine, a primary drafter of the UCC, Karl Llewellyn, incorporated a broad section in the first place meant to serve as a guide to statutory interpretation of the complex code.[63] Similar to general guidelines of the codified ESD, section 1-102(1) of the Uniform Commercial Code Act reads, "This Act shall be liberally construed and applied to promote its underlying purposes and policies."[64] On the state Supreme Court level, these provisions have allowed for the flexibility

---

Clause harks back to the era in which the Court routinely thwarted Congress' efforts to regulate the national economy in the interest of those who labor to sustain it." *National Federation of Independent Business v. Sebelius*, 132 S. Ct. 2566, 2609 (2012) (Ginsburg, J., dissenting).

62: John Gedid, *U.C.C. Methodology: Taking a Realistic Look at the Code,* 29 William and Mary L. Rev. 341 (1988).

63: *Id*. at 385.

64: U.C.C., §1-102(1). The purposes are, "(a) to simplify, clarify and modernize the law governing commercial transactions; (b) to permit the continued expansion of commercial practices through custom, usage and agreement of the parties; (c) to make uniform the law among the various jurisdictions."

of the UCC to adapt to new transaction situations.[65] The lesson is that like the ESD's use against tax avoidance, it is only by flexible judicial interpretation of the UCC beyond the literal text can the code keep up with the constantly evolving nature of business transactions and efficiently facilitate economic activity. And while these actions do require some overreach of courts into "lawmaking," the ultimate end of keeping the statute relevant and maintaining the legal infrastructure for economic activity is a worthy long-term benefit for the trade-off of a nominal threat of judicial overreach.

Extending this conclusion to a more general discussion of law, Rutgers Law School Professor Michael Livingston suggests that the UCC and the tax code are on one end of a spectrum of laws with complex and broad structure, while on the other end of the spectrum are standalone statutes.[66] The same tools for statutory interpretation and construction can be used on all statutes on this spectrum, but different weights are given to the range of tools including analyzing legislative intent and legislative purpose.[67] Livingston's suggestion is quite logical and supportable by modern adjudication. When looking at complex systems of law such as the UCC or the tax code, the overall structure of and patterns among statutes within the code can provide a strong sense the code's purpose, which in turn can help judges and justices interpret and apply a specific statute. This approach is even supported by Justice Scalia, who espouses a "new textualist" approach to statutory interpretation but who authored the majority opinion in *United States v. Woods*,[68] which supported an analysis of the "structure of [the Tax Equity and Fiscal Responsibility Act] [TEFRA] and its other provisions" to interpret a specific section of the 400-page statute.[69] In contrast, when looking at shorter statutes such as the Civil Rights Act of 1964 in *United Steelworkers v. Weber*,[70] the Court suggested that the relative brevity of that statute does not permit as much insight into its general purpose as would be the case with a 400-page statute or a 70,000-page tax code.[71] Consequently, interpretations of the general purpose of these statutes are much more contentious, and in *Weber* Justice Brennan faced pushback from Justice Rehnquist on the wide stretch of Justice Brennan's interpretation of legislative purpose against the plain text of the statute[72] Of course, this argument about the spectrum of statutory interpretation and the

65: Gedid, 29 William and Mary L. Rev at 361 (cited in note 62).
66: Livingston, 51 Tax. L. Rev. at 687 (cited in note 27).
67: *Id.*
68: 134 S. Ct. 557 (2013).
69: *Id.* at 563 (citing *Maracich v. Spears*, 133 S. Ct. 2191, 2193 (2013)).
70: 443 U.S. 193 (1979).
71: *Id.* at 200-08.
72: *Id.* at 219-220 (Rehnquist, J., dissenting).

prevalence of more flexible interpretative tools on one end does not necessarily assume anything about the coherence of longer and more structured laws that permits for the realization of overall purpose. No one would admit that the tax code is coherent at 70,000 pages. However, the volume of this code provides for relatively more evidence for interpreting overall purpose and code patterns, allowing for more weight given to flexibly interpreting purpose beyond the text of the statute.

VI. Conclusion

Tax avoidance is a constantly evolving practice that is difficult for all three branches of the federal government to fight. As the legislature is slow to react and the executive's measures are subject to litigation, the Supreme Court and lower federal courts have filled a vacuum in tax avoidance policy by developing extralegal means of interpreting the tax code. Over time, the extralegal means established in *Gregory* and based on legislative intent, business purpose, and substance over form have served as the foundation to tax avoidance adjudication. Deep issues, however, arise from extralegal approaches on the uncertainties undermining the rule of law and on the grey areas in the separation of powers between the Court and Congress. Over the development of the modern economic substance doctrine, it is apparent that there needs to be national uniformity in the application of extralegal court doctrine to reduce uncertainty. Upon this condition of national uniformity, the uncertainty element in the application of extralegal doctrine can remain an essential and constitutionally viable factor in tax avoidance adjudication to counter the unpredictability of tax avoidance strategies. Going forward then, federal court tax avoidance doctrine can carry on as a fluid and unpredictable area of law, but only if the Supreme Court takes a more active role and becomes the guiding force in tax avoidance adjudication that it once was.

Gang Members of Common Intelligence: The Relevance of the Vagueness Doctrine in RICO Prosecutions of Urban Street Gangs

Kevin Hasenfang†

This article will examine the constitutionality of the criminal provisions of the Racketeer Influenced and Corrupt Organizations ("RICO") Act, specifically focusing on whether or not the application of the statute to sophisticated urban street gangs is in violation of the vagueness doctrine derived from the due process clauses of the Fifth and Fourteenth Amendments.[1] The analysis will concentrate on constitutional challenges to specific elements of the statute,[2] with distinct attention paid to its controversial "pattern of racketeering" and "enterprise" requirements. Given the legal precedent set by numerous challenges to various elements of the statute and recent attempts to define those elements, I argue that the RICO statute, though broad in scope, is not unconstitutionally vague or ambiguous in the context of urban street gang prosecutions.

---

†: Kevin Hasenfang is a third-year in the College, majoring in Public Policy Studies.

1: For a brief discussion of other challenges to the constitutionality of RICO prosecutions, based on the First Amendment, the Tenth Amendment, and the Ex Post Facto Clause of Article I § 10 of the United States Constitution, see Frank Marine, *Criminal RICO: 18 U.S.C. 1961-1968, A Manual for Federal Prosecutors* 303-310 (5th ed. 2009), online at http://www.justice.gov/sites/default/files/usam/legacy/2014/10/17/rico.pdf.

2: I do not address, in the present analysis, the "interstate or foreign commerce" requirement, as the standard is so well defined in federal law that no vagueness challenge to this element can realistically be raised. See, *e.g.*, *United States v. Kramer,* 355 F.2d 891 (7th Cir. 1966), *cert. denied in part and granted in part*, 384 U.S. 100 (1966).

I. Background

To many, the term "organized crime" evokes exaggerated images of Italian-American Mafiosi operating during the Prohibition era. The concept, however, is continuously evolving. In reality, any group with malicious intent, from a drug cartel to a hedge fund, can be classified as an organized criminal enterprise. In major cities across the United States, urban street gangs are beginning to emerge as the new face of organized crime. These street gangs pose a serious and steadily increasing threat to the modern urban center.

Most street gangs function as highly sophisticated criminal enterprises, deeply embedded within their communities and increasingly engaged in non-traditional gang-related crimes such as human, drug, and weapons trafficking.[3] Additionally, street gangs mercilessly continue to be disproportionately responsible for violent crime in urban areas; roughly 48 percent of all violent crime is gang-related, and that proportion increases to nearly 90 percent when looking at urban and suburban communities alone.[4] These street gangs recruit aggressively, specifically targeting at-risk youth in low-income neighborhoods. Consequently, there has been a 40 percent increase in gang membership nationally between 2009 and 2011.[5]

Traditional gang enforcement strategies, particularly in urban and suburban areas, have commonly relied on reactive police responses to street-level criminal activity. In coordination with local law enforcement, prosecutors have then typically pursued strategic plea agreements with these street-level offenders in order to develop actionable intelligence on higher-ranking individuals within the gang's hierarchy. This, however, is inefficient; as these gangs evolve, becoming ever more structured and methodical, so must the government's enforcement mechanism. In the late 1980s, federal prosecutors began exploring how a broadened interpretation of the Racketeer Influenced and Corrupt Organizations ("RICO") Act might function as an aggressive means of dismantling sophisticated urban street gangs.

3: *2013 National Gang Report,* National Gang Intelligence Center 12 (2013), online at http://www.fbi.gov/stats-services/publications/national-gang-report-2013.

4: *2011 National Gang Threat Assessment: Emerging Trends,* National Gang Intelligence Center 9 (2011), online at http://www.fbi.gov/stats-services/publications/2011-national-gang-threat-assessment/2011-national-gang-threat-assessment-emerging-trends.

5: *Id.* at 11.

II. Overview of Criminal RICO

The RICO Act, 18 U.S.C. §§ 1961-1968, was enacted in 1970 as Title IX of the Organized Crime Control Act of 1970[6] in order to target the Mafia and white-collar crime. RICO provides for both civil remedies[7] and criminal penalties.[8] The statute's criminal provisions, with which this analysis is concerned, provide significant criminal penalties for "persons" who engage in a "pattern of racketeering activity" or "collection of an unlawful debt" while maintaining a relationship with an "enterprise" engaged in or affecting "interstate or foreign commerce."[9]

Four substantive criminal activities are specifically proscribed by the RICO statute.[10] Section 1962(a) prohibits any person from using or investing, directly or indirectly, the proceeds of a pattern of racketeering activity or collection of unlawful debt in the acquisition, as well as the establishment or operation of any enterprise affecting interstate or foreign commerce. Section 1962(b) prohibits any person from acquiring or maintaining, directly or indirectly, an interest in an enterprise affecting interstate or foreign commerce through a pattern of racketeering activity or collection of unlawful debt. Section 1962(c) prohibits any person associated with an enterprise affecting interstate or foreign commerce from conducting or participating, directly or indirectly, in the enterprise's affairs through a pattern of racketeering activity or collection of unlawful debt. Section 1962(d) prohibits any person from conspiring to commit any of the three preceding substantive criminal violations. The RICO statute provides criminal penalties, based on the underlying racketeering activity, up to and including a maximum life sentence and/or a fine under Title 18, along with guidelines for asset forfeiture and the pre-trial prevention of asset dissipation.[11]

III. Vagueness Doctrine

Criminal statutes are required, by constitutional rule, to state explicitly and definitely what conduct is punishable. This rule, the vagueness doctrine, is derived from the due process clauses of the Fifth and Fourteenth Amendments. Criminal statutes that violate this doctrine are "void for vagueness" or

6: See Pub. L. No. 91-452 (1970).
7: See 18 U.S.C. § 1964.
8: See 18 U.S.C. § 1963.
9: Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 1 (cited in note 1).
10: See 18 U.S.C. § 1962.
11: See 18 U.S.C. § 1963.

"unconstitutionally vague." By requiring overt notice of what conduct does and does not violate a criminal law, the vagueness doctrine protects the rights of citizens to fair criminal procedure and due process of law. Additionally, the vagueness doctrine protects citizens from arbitrary enforcement of the laws and arbitrary prosecutions by imposing limits on the extent of a legislature's delegation of legal authority to judges or administrators.[12]

The Supreme Court in *Connally v. General Construction Co.*,[13] wrote that:

> the terms of a penal statute must be sufficiently explicit to inform those who are subject to it what conduct on their part will render them liable to its penalties…and a statute which either forbids or requires the doing of an act in terms so vague that men of common intelligence must necessarily guess at its meaning and differ as to its application violates the first essential of due process of law.[14]

In order to satisfy the due process clauses of the Fifth and Fourteenth Amendments, then, courts have held that "individuals are entitled to understand the scope and nature of statutes which might subject them to criminal penalties."[15] Going further, the Court outlined in *Skilling v. United States*[16] that a penal statute "must define the criminal offense (1) with sufficient definiteness that ordinary people can understand what conduct is prohibited and (2) in a manner that does not encourage arbitrary and discriminatory enforcement."[17]

The Court has consistently restricted vagueness challenges to real cases, on an as-applied basis, as opposed to general considerations of hypothetical sets of facts. It has maintained that "[a] plaintiff who engages in some conduct that is clearly proscribed cannot complain of the vagueness of the law as applied to the conduct of others."[18] This is necessary so as to prevent the court from "[invalidating] a criminal statute on its face even when it could conceivably have had some valid application."[19] To this end, the Court held

---

12: *Kolender v. Lawson*, 461 U.S. 352 (1983).
13: 269 U.S. 385 (1926).
14: *Id.* at 391. See also *International Harvester Co. v. Kentucky*, 234 U.S. 216, 221 (1914).
15: *Hedges v. Obama*, 2012 U.S. Dist. LEXIS 68683 (2012).
16: 130 S. Ct. 2896 (2010).
17: *Id.* at 2928. Also see *Coates v. Cincinnati,* 402 U.S. 611 (1971).
18: *Village of Hoffman Estates v. Flipside, Hoffman Estates Inc.*, 455 U.S. 489, 495 (1982).
19: *Kolender v. Lawson*, 461 U.S. 352, 358 n. 8 (1983).

in *United States v. Mazurie*[20] that "vagueness challenges to statutes which do not involve First Amendment freedoms must be examined in the light of the facts of the case at hand"[21] without regard to the statute's facial validity.[22] Additionally, the Court definitively stated in *United States v. Salerno*[23] that a facial challenge to a statute "must establish that no set of circumstances exists under which the [statute] would be valid."[24]

Vagueness challenges to RICO directed toward either its civil or criminal provisions, must necessarily rely on "the facts of the particular case in which the claim is asserted."[25] Such attacks to the statute, therefore, have historically not seen much success in the courts. Regardless, there is still great concern about the difficulty of interpreting the plain language of the statute and arriving at a well-understood definition for each of its various elements. Note, for example, the concurring opinion written by Justice Scalia in *H.J. Inc. v. Northwestern Bell Telephone Co.*,[26] in which four Justices emphasized that:

> it is also true that RICO, since it has criminal applications as well, must, even in its civil applications, possess the degree of certainty required for criminal laws. No constitutional challenge to this law has been raised in the present case, and so that issue is not before us. That the highest Court in the land has been unable to derive from this statute anything more than today's meager guidance bodes ill for the day when that challenge is presented.[27]

This acknowledgement prompted numerous vagueness challenges, many of which were led by gang members, to the RICO statute. Between the *H. J. Inc.* decision in 1989 and 2009, all ten federal courts of appeals that addressed the RICO vagueness argument rejected it. In each case, courts found that "the defendants had adequate notice that their conduct fell within the proscriptions of RICO and that consequently their vagueness challenges…were meritless."[28] However, the constitutional concern here—

20: 419 U.S. 544 (1975).
21: *Id.* at 550.
22: Also see *United States v. Nadi*, 996 F. 2d 548 (2d Cir. 1993).
23: 481 U.S. 739 (1987).
24: *Id.* at 745.
25: Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 304 (cited in note 1).
26: 492 U.S. 229 (1989).
27: *Id.* at 255-56 (Scalia, J., concurring). Chief Justice Rehnquist and Justices O'Connor and Kennedy joined Justice Scalia's concurrence.
28: Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 304 (cited in note 1).

whether RICO is "too vague to put persons on notice that their activities are illegal and, alternately, casts too large a net and "captures" innocent actors in an enterprise"[29]—remains.

IIIa. Pattern of Racketeering Activity

Section 1961(5) defines a "pattern of racketeering activity" as "at least two acts of racketeering activity, one of which occurred after the effective date of this chapter and the last of which occurred within ten years (excluding any period of imprisonment) after the commission of a prior act of racketeering activity."[30] According to internal guidelines at the U.S. Department of Justice, these violations "may both be state offenses, federal offenses, or a combination of the two; they may be violations of the same statute, or of different statutes; and the acts need not have previously been charged."[31]

A number of cases, specifically regarding the ambiguity of the term "pattern," have held that the term itself is not unconstitutionally vague.[32] This is important, as was noted in *United States v. Campanale*,[33] because "[w]ith respect to the argument that this statute is vague and ambiguous, it is true that, if undefined, terms such as "pattern of racketeering activity" would be unmanageable."[34] To resolve this dispute, the Ninth Circuit held in *Campanale* that any ambiguity "is cured by 18 U.S.C. § 1961, which defines "racketeering activity" with reference to specific offenses, "pattern of racketeering activity" with reference to a definite number of acts of "racketeering activity" within specified time periods, and "enterprise" and "person" with standard language of established meaning."[35]

A significant vagueness challenge to RICO's "pattern of racketeering activity" requirement which focused on the term "pattern," on the other hand, was raised in *United States v. Tripp*.[36] In *Tripp*, the defendant claimed that "the RICO statute, by adopting the laws of the states in defining 'racketeering activity'…violates the Fifth Amendment Due Process Clause by creating a

---

29: *United States v. Swiderski*, 593 F. 2d 1246, 1249 (D.C. Cir. 1978), *cert. denied*, 441 U.S. 933 (1979).

30: 18 U.S.C. § 1961(5). For the substantive list of enumerated criminal offenses which qualify as predicate acts under the racketeering activity requirement, see 18 U.S.C. § 1961(1).

31: Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 89 (cited in note 1).

32: See *Swiderski,* 593 F. 2d at 1246.

33: 518 F. 2d 352 (9th Cir. 1975).

34: *Id.* at 364.

35: *Id.*

36: 782 F.2d 38 (1986).

lack of national uniformity and by not properly notifying a person that his conduct is illegal."[37] The court swiftly rejected this argument, declaring that there is no "constitutional objection to a criminal statute that incorporates state law for purposes of defining illegal conduct" and that this holds true "even if the result is that conduct that is lawful under the federal statute in one state is unlawful in another."[38] Moreover, the court suggests in *Tripp* that a more appropriate challenge would be to "allege that the state offenses involved [in 18 U.S.C. § 1961(1)(A)] themselves are 'void-for-vagueness.'"[39]

Legal precedent indicates, then, that the plain language of the statute, specifically the definition of the "pattern of racketeering activity" requirement, is sufficient to put defendants on notice. In *Beck v. Edward D. Jones & Co.*,[40] an Illinois district court ruled that "any person of average intelligence could determine what actions would make him liable for participating in an enterprise through a pattern of racketeering activity."[41] This suggests that future vagueness challenges to the element will undoubtedly fail.

IIIb. Collection of an Unlawful Debt

Section 1961(6) defines an "unlawful debt" as "a debt (a) incurred or contracted in gambling activity which was in violation of the law…or which is unenforceable under State or Federal law in whole or in part as to principal or interest because of the laws relating to usury, and (b) which was incurred in connection with the business of gambling in violation of the law…or the business of lending money or a thing of value at a rate usurious under State or Federal law, where the usurious rate is at least twice the enforceable rate."[42] This is, of course, an alternative to the "pattern of racketeering activity" requirement and not necessary to pursue RICO charges. Because the unlawful debt must be incurred through illicit gambling or loansharking activities, this is superfluous to the present analysis. Though urban street gangs may, in fact, engage in illicit gambling or loansharking activities, the "pattern of racketeering activity" is significantly more common in pursuit of RICO charges against street gang members.

37: *Id.* at 41-42.
38: *Id.* at 42.
39: *Id.*
40: 735 F. Supp. 903 (I.L.C.D. 1990).
41: *Id.* 906.
42: 18 U.S.C. § 1961(6).

IIIc. Enterprise

Section 1961(4) defines an "enterprise" as "any individual, partnership, corporation, association, or other legal entity, and any union or group of individuals associated in fact although not a legal entity."[43] This definition broadly encompasses a wide variety of types of enterprises, yet many have questioned whether wholly illegitimate organizations, and thereby street gangs, fit at all within the plain language of the element. Specifically, critics of a broadened application of the statute argue that "the sole purpose of RICO [is] to protect legitimate enterprises from infiltration by organized crime."[44]

In *United States v. Turkette*,[45] the Court ended this dispute by holding that "neither the language nor structure of RICO limits its application to legitimate 'enterprises'" and that an application "also to criminal organizations does not render any portion of the statute superfluous nor does it create any structural incongruities within the framework of the Act."[46] Moreover, the Court went as far as to suggest that "insulating the wholly criminal enterprise from prosecution under RICO is…incongruous."[47] Additionally, *Turkette* narrowed the applicability of the RICO statute by requiring "evidence of an ongoing organization, formal or informal" in which "the various associates function as a continuing unit."[48] Still, as one court has noted, the definition of "enterprise" is continuously shifting based on the "fluid nature of criminal associations."[49] Furthermore, the Court held in *Boyle v. United States*[50] that an "association-in-fact enterprise under RICO" needs not have "an ascertainable structure beyond that inherent in the pattern of racketeering activity in which it engages."[51]

In order to establish that the members of the enterprise "operated together in a coordinated manner, in furtherance" of shared objectives, prosecutors may rely on:

43: 18 U.S.C. § 1961(4).
44: Jan Neuenschwander, *RICO Extended to Apply to Wholly Illegitimate Enterprises*, 72 J. Crim. L. & Criminology 1426, 1429 (1981).
45: 452 U.S. 576 (1981).
46: *Id.* at 587.
47: *Id.*
48: *Id.* at 583.
49: *Swiderski,* 593 F. 2d at 1249.
50: 556 U.S. 938 (2009).
51: *Id.* at 941.

> a wide variety of direct and circumstantial evidence including, but not limited to, inferences from the members' commission of similar racketeering acts in furtherance of a shared objective, financial ties, coordination of activities, community interests and objectives, interlocking nature of the schemes, and overlapping nature of the wrongful conduct.[52]

This is particularly interesting, in terms of statutory ambiguity and vagueness, because it is rare that every individual who contributes to the mission of a criminal enterprise actually identifies as a member of the illegitimate organization; in many instances, individuals that engage in criminal activity that benefits the enterprise are only associated-in-fact.

It has been determined, in this regard, that it is not necessary, in proving the existence of the enterprise, to show that "every member of the enterprise participated in or knew about all its activities."[53] Rather, it is only necessary that the defendant "know the general nature of the [enterprise] and that the [enterprise] extends beyond his individual role."[54] To this end, it was determined by the District of Columbia Circuit in *United States v. Perholtz*[55] that "it is not essential that each and every person named in the indictment be proven to be a part of the enterprise," as the enterprise "may exist even if its membership changes over time or if certain defendants are found… not to have been members at any time."[56] As such, it also appears that an individual newly associated with an active urban street gang may be indicted under RICO charges predicated on any and all racketeering activities, charged against the defendant or uncharged altogether, which are carried out by the associated gang prior to the individual defendant's involvement.[57]

For some time, defense attorneys and scholars questioned whether there exists an economic motive requirement implicit in RICO's "enterprise" element. Indeed, the organized criminal enterprises at which RICO was originally targeted were profit-seeking entities and motivated by economic goals. However, the Court held in *National Organization for Women v.*

52: Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 62 (cited in note 1).
53: *United States v. Cagnina,* 697 F. 2d. 915, 922 (11th Cir. 1983).
54: *United States v. Rastelli*, 870 F. 2d. 822, 827-28 (2d Cir. 1989).
55: 842 F. 2d. 343 (D.C. Cir. 1988), *cert. denied*, 488 U.S. 821 (1988).
56: *Id.* at 364 (citations omitted).
57: For a brief discussion regarding the admission of evidence related to uncharged crimes, see Marine, *Criminal RICO: 18 U.S.C. 1961-1968* at 371-73 (cited in note 1).

Scheidler[58] that RICO does not require "proof that either the racketeering enterprise or the predicate acts of racketeering were motivated by an economic purpose."[59] Moreover, the Court went so far as to declare that "the statutory language is unambiguous" and there is no "clearly expressed intent to the contrary [in the legislative history] that would warrant a different construction."[60] This will undoubtedly prove significant in the near future, as the Department of Justice actively explores the use of purely violent criminal acts as historical predicates for RICO prosecutions of urban street gangs.

IV. Conclusion

It is true that RICO prosecutions of urban street gangs have always been met with a great deal of skepticism. In 1993, survey research demonstrated that roughly one-third of local prosecutors "would continue to rely entirely on traditional criminal statutes in prosecuting organized crime, with 27 percent citing the possibility that RICO-driven prosecutions will fail and 17 percent citing the legal complexity of the statute itself."[61] Some prosecutors, especially on the local level, still assume that judges would only allow a narrow application of the statute, thus "limiting practical applications of the statute to white-collar offenses alone."[62] This, as we have seen, is simply not the case.

It is true that Congress, when they passed the RICO Act in 1970, was not anticipating the fluid nature of organized crime or the application of the statute to urban street gangs. As the Court declared in *Sedima, S.P.R.L. v. Imrex Co.,*[63] however, "[t]he fact that RICO has been applied in situations not expressly anticipated by Congress does not demonstrate ambiguity. It demonstrates breadth."[64] Indeed, it has been shown that RICO prosecutions of urban street gangs fit well within the plain language of the statute. Given the legal precedent set by numerous challenges to elements of the statute, particularly those directed toward the "pattern of racketeering activity" and "enterprise" requirements, and the many recent efforts by various courts to

58: 510 U.S. 249 (1994).
59: *Id.* at 252.
60: *Id.* at 261 (citations omitted).
61: Donald Rebovich, Kenneth R. Coyle, & John C. Schaaf, *Local Prosecution of Organized Crime: The Use of State RICO Statutes,* National Criminal Justice Reference Service 14 (1993), online at https://www.ncjrs.gov/pdffiles1/Digitization/143502NCJRS.pdf.
62: *Id.* at 16.
63: 473 U.S. 479 (1985).
64: *Id.* at 499 (citing *Haroco, Inc. v. American National Bank & Trust Co. of Chicago,* 747 F. 2d 384, 398 (7th Cir. 1984).

define those same elements, it is clear that the RICO statute, though broad in scope, must not be considered unconstitutionally vague or ambiguous in the context of urban street gang prosecutions.

Reframing the Bitcoin Problem

Jacob Romeo†

Though the use of bitcoin was concentrated in small networks for the first few years following its inception in 2008, it has recently gained favor outside of the young and tech-savvy community in which it was incubated. Today, it is estimated that between eighty and one hundred thousand bitcoin transactions take place daily.[1] These transactions are anonymous and are thus untraceable as far as governmental agencies are concerned, offering nearly free range for illegal activities. It is largely for this reason that bitcoin has faced such a high degree of scrutiny from regulators—and because it has the power to render many traditional services provided by banks, payment services, and even lawyers obsolete.[2] Some ardent constitutionalists argue that virtual currencies obstruct state power. Still, decentralized virtual currencies like bitcoin have great potential to positively change the way banking and commerce has been conducted since the 15th century. The purpose of this article is twofold: first, to briefly discuss the nature of virtual currency and its role in the global economy; next, to highlight how, given the multiplicity of categories into which bitcoin may be placed, and the regulatory confusion that results, as tempting as it may be to try to fit the metaphorical square peg into a round hole, it is necessary to establish a new statutory framework specific to bitcoin's oversight.

---

†: Jacob Romeo is a third-year in the College, majoring in Philosophy.

1: Mark Wetjen, *Bringing Commodities Regulation to Bitcoin*, Wall St. J. (Nov. 3, 2014).

2: Michael Casey, *Why Bitcoin's Erratic Price Doesn't Matter,* Wall St. J. (Dec. 21, 2014).

I. Introduction to Bitcoin

The Gold Standard of the U.S. dollar fell to the wayside in the latter half of the twentieth century as President Richard Nixon pushed for a free-floating exchange rate. With the overhaul of the Bretton Woods System, paper money cut its ties with precious metals, "freeing the greenback to find its own level in currency markets."[3] About forty years later, the monopoly of paper currency is being threatened by the introduction of virtual currency into to the global economy. Thanks to the swift rise in the prominence of e-commerce, virtual currency has quickly gained strong footing with tech-savvy consumers and is steadily diffusing to the public. Bitcoin, the virtual currency that has received the most attention to-date, was introduced in 2008 in response to the financial crisis. The crisis, which involved the collapse of several systemically important financial institutions due to risky loan portfolios and inadequate capital reserves, shook the confidence of many in the ability of the banking system to serve and protect the interests of its users. Bitcoin is a "purely peer-to-peer version of electronic cash [that allows] online payments to be sent directly from one party to another without going through a financial institution."[4] Bitcoin was initially created to solve a few of the problems of traditional fiat currency: the need for financial intermediaries, inflationary risk, service fees, and more.[5]

Perhaps the greatest advantage of virtual currency is its universality. Bitcoin promotes global commerce by allowing transactions to occur without exchange fees or the headache of conversion across currencies. However, its universality has been taken advantage of. Aided by the ability to hide behind the veil of the Internet, bitcoin has made it easier for users to skirt the law by laundering money, dealing in the underground economy, and evading taxes by bypassing financial intermediaries. Virtual currency does not have legal tender status in any particular jurisdiction,[6] meaning it is neither subject to taxation at source nor at destination. Also, even though the entire history of bitcoin transactions is available on a public ledger known as the "blockchain,"

3: Bejamin Cohen, "Bretton Woods System," in R.J. Barry Jones, ed., *Routledge Encyclopedia of International Political Economy* 100 (2001).

4: Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System,* online at https://bitcoin.org/bitcoin.pdf.

5: Jon Matonis, *ECB: 'Roots of Bitcoin Can Be Found in the Austrian School of Economics*,' Forbes (Nov. 3, 2012), online at http://onforb.es/1rcryT8.

6: Financial Crimes Enforcement Network, *Application of FinCEN's Regulations to Persons Administering, Exchanging, or Using Virtual Currencies,* Dep't. Treasury (2013), online at http://fincen.gov/statutes_regs/guidance/pdf/FIN-2013-G001.pdf.

accounts are entirely anonymous. Rather than being held in an account issued by a FDIC-insured bank or credit union, bitcoin is held in virtual "wallets" that do not require any personal information. These two qualities—less taxation and more anonymity—are likely the two most appealing features to those who look to protect assets from taxes and purchase illegal goods.[7]

While steps have been taken to mitigate the prevalence of assets held abroad, virtual currency may prove to be too elusive. As it stands, there is no central authority that can fundamentally change the entire the bitcoin economy's behavior.[8] In fact, bitcoin is not even minted by any one agency. The United States Federal Reserve controls the flow of cash and value into the U.S. economy through monetary policy and open market operations. No such organization exists to control the flow of bitcoins.[9] Instead, bitcoins are "mined" through a process in which computers solve complex algorithms and mathematical problems[10] at a predictable rate such that the eventual total number will be 21 million.[11] Supply and demand forces determine the price of each individual bitcoin, which has settled around $250 following its high of $1145 in late 2013. Upon generation, the miner is provided with cryptographic keys as both public and private proof of ownership and only the user with the claim to the key for a particular bitcoin is able to spend it.[12] In many ways, this makes virtual currency much more secure than paper money. This is particularly enticing in light of recent hacker attacks on large institutions such as Target, Home Depot and Chase Bank, in which confidential credit card and personal information was stolen, leaving consumers vulnerable to large financial losses.[13] Bitcoin does not require such sensitive information, and even if it did, as a result of the decentralized nature of the system, there are no major concentrated sources of information from

7: Omri Marian, *Are Cryptocurrencies Super Tax Havens?,* 112 Mich. L. Rev. First Impressions 38 (2014).

8: Fergal Reid & Martin Harrigan, *An Analysis of Anyonymity in the Bitcoin System*, online at http://arxiv.org/pdf/1107.4524v2.pdf.

9: Dan Stroh, *Secure Currency or Security? The SEC and Bitcoin Regulation*, Univ. Cinc. L. Rev. Blog (Nov. 18, 2014), online at http://uclawreview.org/2014/11/18/secure-currency-or-security-the-sec-and-bitcoin-regulation/.

10: Nakamoto, *Bitcoin* at 8 (cited in note 4).

11: Reid & Harrigan, *An Analysis of Anyonymity in the Bitcoin System* at 6 (cited in note 8).

12: Nakamoto, *Bitcoin* at 2-3 (cited in note 4).

13: Jessica Silver-Greenberg, Matthew Goldstein, & Nicole Perloth, *JPMorgan Chase Hacking Affects 76 Million Households,* DealBook (Oct. 2, 2014), online at http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/.

which data may be taken.

The most fundamental concern regarding the governance of virtual currency is its classification—whether as currency, commodity, security, etc. These classes are not necessarily mutually exclusive, but they are far from synonymous. This ultimately dictates the set(s) of laws to which a financial instrument is subject. Unfortunately, governing agencies have been unsuccessful in designating bitcoin's rightful place—wherever that may be. In 2013, a federal district judge ruled that "for the purposes of U.S. securities regulation, bitcoin is indeed 'money,'" but other uses remain for which the rules are not as clear.[14] As it stands, states can monitor virtual currency service providers through various means, including implementing new laws and/or regulations written explicitly for virtual currency activities, or by interpreting or amending existing ones.[15] Thus far, many agencies have elected the latter route. This method can only be effective insofar as governing agencies are collectively able to agree upon a classification for bitcoin and other such virtual currencies. The remainder of this essay will detail some of the intricacies of a few of the existing acts that are being employed to mitigate the need for a more intrusive systematic overhaul. In the end, I will attempt to provide an objective assessment of each method, ultimately to make the claim that it is not presently possible to govern virtual currencies exactly in the way that many would like. As a result, lawmakers must accept the fact that new technology warrants new legislation. And without a more systematic reconstruction, bitcoin and its virtual peers will continue to slip through the fingers of regulatory agencies.

II. Bitcoin As Alternative Currency

In March of 2013, The Financial Crimes Enforcement Network ("FinCEN") issued guidance defining currency as coin or paper money that circulates, is designated as legal tender, and is customarily used and accepted as a medium of exchange in the issuing country. Virtual currency fulfills only two of these requirements. Specifically, "virtual currency does not have legal tender status in any jurisdiction."[16] Later, in August 2013, judge Amos Mazzant of the

14: Marian, 112 Mich. L. Rev. First Impressions at 39 (citing *Securities & Exchange Commission v. Shavers,* 2013 U.S. Dist. LEXIS 110018 (2013)) (cited in note 7).

15: Conference of State Bank Supervisors, *CSBS Policy on State Virtual Currency Regulation* (2014), online at http://www.csbs.org/regulatory/ep/Documents/CSBS%20Policy%20on%20State%20Virtual%20Currency%20Regulation%20--%20Dec.%2016%202014.pdf.

16: Danton Bryans, *Bitcoin and Money Laundering: Mining for an Effective Solution*,

Eastern District of Texas ruled in *Securities and Exchange Commission v. Trendon Shavers and Bitcoin Savings and Trust* that virtual currency is indeed "a currency or form of money."[17] This ruling is critical, as it extends governance to the use of bitcoin in relation to securities and securities law, since the 1946 case *SEC v. W. J. Howey*[18] ruled that an investment contract requires four elements, the first of which is an "investment of *money*."[19] The other three criteria are that the money is invested into a common enterprise; profits are expected from the investment; and the expected profits are generated solely from the efforts of the promoter or a third party. Trendon Shavers, the defendant in the 2013 case, operated Bitcoin Savings and Trust ("BTCST") as a Ponzi scheme, promising investors "up to 7 percent weekly interest based on BTCST's Bitcoin market arbitrage activity."[20] The SEC charged Shavers with "offering and selling investments in violation of the anti-fraud and registration provisions of the securities laws, specifically Sections 5(a), 5(c) and 17(a) of the Securities Act of 1933, Section 10(b) of the Securities Exchange Act of 1934 and Exchange Act Rule 10b-5."[21] In accordance with the rulings of *Howey* and *Shavers*, the Securities and Exchange Commission convicted Shavers and has been able to bring enforcement actions against a number of alleged bitcoin-related frauds thereafter.

III. Bitcoin As Commodity

A commodity, as defined in section 1a(9) of the Commodity Exchange Act, includes agricultural commodities and physical commodities such as an agricultural product or a natural resource *as opposed to a financial instrument such as a currency or interest rate*.[22] Despite this differentiation, U.S. Commodity Futures Trading Commission ("CFTC") Commissioner Mark Wetjen suggested that the definition of "commodity" under the CFTC's authorizing statute could be read to include bitcoin. If this were to be the case, the agency would have the authority to bring charges against individuals who attempt to manipulate the virtual currency, just as they do with those

89 Ind. L.J. 441, 442 n. 8 (2014).

17: 2013 U.S. Dist. LEXIS 110018 (2013).

18: 328 U.S. 293 (1946)

19: *Id.* at 301 (emphasis added).

20: *SEC Charges Texas Man With Running Bitcoin-Denominated Ponzi Scheme,* Securities & Exchange Commission (2013), online at http://www.sec.gov/News/PressRelease/Detail/PressRelease/1370539730583.

21: *Id.*

22: 7 U.S. Code § 1(a)(9).

who corner commodities in their respective markets.[23] The problem is that the definition clearly precludes financial instruments and currency—exactly what the *Shavers* ruling declared bitcoin to be.[24]

Even though the CFTC may not be able to govern the currency itself, under reforms included in the Dodd-Frank Act of 2010, the CFTC expanded its scope to regulating trading in derivatives called swaps. In September 2014, TeraExchange L.L.C. received approval from the CFTC to begin listing an over-the-counter swap, a derivative product, based on the price of bitcoin. CFTC's approval marks the first time a U.S. regulatory agency approved a bitcoin financial product.[25.] Regulated derivatives trading might ease some of the extreme volatility for which bitcoin is notorious, and thus mitigate one of the primary barriers to its widespread adoption as a currency and payments tool.[26] The price of one bitcoin fluctuated between $13 and $1,200 USD in 2013,[27] and almost $11.3 billion in value has been lost since that peak.[28] Bitcoin has proven to be extremely sensitive to large sudden price swings, as evidenced by the flash crashes triggered by the dissolution of the Silk Road—an online black market—and false reports of a bitcoin ban in China. Without these swap products, there is no way to hedge risk without reverting to the underlying "commodity." Large merchants like Microsoft, Dell, Expedia, and Overstock have made the use of these swaps common practice. The swap contracts provided by TerraExchange are "written, quoted, and settled in dollars" so that neither side of the trade has to deal with the underlying bitcoins directly.[29] Directly exchanging bitcoins for dollars requires compliance with anti-money laundering regulation, which typically includes the need to obtain specific licenses to be a designated money transmitter.[30] This issue will be more extensively explored in the discussion of the money-laundering implications of virtual currency a few sections later.

23: Wetjen, *Bringing Commodities Regulation to Bitcoin* (cited in note 1).

24: 2013 U.S. Dist. LEXIS at 110018.

25: Douwe Miedema, *Bitcoin Gets Boost as U.S. Watchdog Approves First Swap*, Routers (Sep. 12, 2014), online at http://www.reuters.com/article/2014/09/12/us-usa-bitcoin-cftc-idUSKBN0H71FU20140912?irpc=932.

26: Michael J. Casey, *TeraExchange Unveils First U.S.-Regulated Bitcoin Swaps Exchange*, Wall St. J. (Sep. 12, 2014).

27: *Bitcoin, Price Index Chart,* CoinDesk (2015), online at http://www.coindesk.com/price/.

28: Michael J. Casey, *Bitcoin's Plunge Bites 'Miners,'* Wall St. J. (Jan. 14, 2015).

29: Casey, *TeraExchange Unveils First U.S.-Regulated Bitcoin Swaps Exchange* (cited in note 26).

30: *Id.*

IV. Tax Sheltering

Just as commerce is now more globally integrated than ever, assets held offshore are at an all-time high. The Tax Justice Network estimates that governments lose as much as $100 billion in tax revenue per year due to offshore asset sheltering by multi-national enterprises alone.[31] Traditionally, tax havens have two defining features: small or non-existent tax rates and strong bank privacy laws.[32] As a response to the proliferation of offshore sheltering, the Foreign Account Tax Compliance Act ("FATCA") was implemented in 2010 to reduce the tax benefit provided to those who hold assets, including those holding bitcoins, in accounts overseas without proper documentation and reporting. By building an alliance of Foreign Financial Institutions ("FFIs"), it was expected that they would reduce the prevalence of offshore tax sheltering. However, those countries that are notorious for allowing American offshore accounts to go unreported are the same ones who are expected to comply with the terms proposed by the United States: the proposed terms offer these countries the option of violating their own bank secrecy laws or facing a steep financial penalty for abiding by them. Under FATCA, FFIs are required to identify their U.S. account holders to the IRS. Failing this, the institution will face a gross tax rate of 30% of certain payments. Having this dilemma forced upon them, many FFIs have agreed to help the U.S. Even more have joined in after the U.S. formed intergovernmental covenants geared towards allowing FFIs to comply with FATCA without violating their own bank-secrecy laws.[33] But professor Omri Marion has pointed out that in addition to lesser taxation and anonymity, bitcoin offers an additional advantage: the operation of bitcoin is not necessarily dependent on the existence of financial intermediaries such as banks—so the friends of FATCA are of little use.[34]

Furthemore, it is not clear which country would have the right to tax bitcoin transactions or holdings. Still, the IRS requires a taxpayer receiving virtual currency as payment to include the fair market value of the virtual currency, measured in U.S. dollars, when computing gross income. However, bitcoin-equity swap contracts have made it possible to use tax-exempt buying

---

31: *UNCTAD: Multinational Tax Avoidance Costs Developing Countries $100 Billion+*, Tax Justice Network (2015), online at http://www.taxjustice.net/2015/03/26/unctad-multinational-tax-avoidance-costs-developing-countries-100-billion/.

32: Aaron Sankin, *Bitcoin Is the Offshore Tax Haven of the Future,* Daily Dot (Oct. 10, 2013), online at http://www.dailydot.com/business/bitcoin-offshore-tax-haven/.

33: Marian, 112 Mich. L. Rev. First Impressions at 41 (cited in note 7).

34: *Id.* at 42.

agents to invest in traded securities and commodities.[35] The agent receives commission from the trade, and is thus indifferent to the contract's price, and all the while the investor is fully exposed to the risk and reward just as if he had bought the security itself. This dynamic adds yet another layer of complexity to finding out who is actually liable for the capital gains of bitcoin, as under these contracts the income goes unreported, and, hence, untaxed. Additionally, as long as bitcoin is not transferred or converted through a financial intermediary, users can exchange goods and services without providing any sort of identification or other personal information. So on one hand, if users can purchase goods and services with virtual currency, save virtual currency, and furthermore, invest it just as they would with fiat currency, with a sufficiently large user base, there becomes relatively little incentive to use government-backed currency at all, meaning that a larger portion of exchange would go unreported for tax purposes. On the other hand, the scale of the sheltering might be too small for it to be a main concern of governments: the total possible value of bitcoin held abroad is still low—the total market capitalization for bitcoin is less than one percent of total outstanding Federal Reserve Notes at the time of this writing.

V. Money Laundering

Money laundering, the process of making illegally gained proceeds appear legal, is made easier by virtual currencies.[36] Although with traditional currency transfer, there is often a physical "paper trail" to observe or intercept for proof of illicit activities, no such means are readily available for virtual currencies.[37] The Bank Secrecy Act of 1970 ("BSA") and the Money Laundering Control Act of 1986 pose the greatest risk for Bitcoin developers, exchanges, wallet providers, mining pool operators and businesses that accept bitcoin. These acts require certain kinds of financial businesses, even if they are located abroad, to register with FinCEN.[38] In March 2013, the Department of Homeland Security effectively shut down bitcoin's biggest exchange's operations—Mt. Gox—for not registering itself as a money services business in accordance with FinCEN guidelines.[39]

35: *Id.* at 43.

36: Financial Crimes Enforcement Network, *History of Anti-Money Laundering Laws,* Dep't. Treasury (2015), online at http://www.fincen.gov/news_room/aml_history.html.

37: Bryans, 89 Ind. L.J. at 456 (cited in note 16).

38: Reid & Harrigan, *An Analysis of Anyonymity in the Bitcoin System* at 5 (cited in note 8).

39: Susan A. Berson, *Some Basic Rules for Using 'Bitcoin' as Virtual Currency,* ABA J.

After several companies requested additional guidance, FinCEN responded, stating that a money services business ("MSB") is "a person that accepts currency, funds, or any value that substitutes for currency, with the intent and/or effect of transmitting currency, funds, or any value that substitutes for currency, to another person or location."[40] These same rules govern other financial service products like credit cards, as well as payment services like PayPal. As MSBs, the entities are subject to the accompanying reporting, recordkeeping and monitoring requirements of the BSA.[41] But bitcoin can be transferred without changing from bitcoin to cash, and as long as bitcoin is held as bitcoin, these transactions are not subject to legislation governing money service businesses. Under this guidance, a user who simply obtained virtual currency and used it to purchase real or virtual goods or services would not be subject to FinCEN regulations.[42] Furthermore, these regulations still allow users to remain unidentified until bitcoin is, if ever, exchanged for a government-backed currency.[43]

In response to FinCen's updated framework, new technologies have been created to render FinCen's progess nearly null and void. As previously mentioned, while bitcoin transactions are anonymous in the sense that one need not provide or verify any personal information, they are public in the way that all exchanges are made available on the "blockchain". Recently developed software titled Dark Wallet—with reference to its use for the Black Market—attempts to make each individual transaction even harder to trace. It does this by integrating a laundering-type mechanism into each and every transaction that occurs using a process called CoinJoin.[44] CoinJoin combines

---

(Jul. 1, 2013), online at http://www.abajournal.com/magazine/article/some_basic_rules_for_using_bitcoin_as_virtual_money/.

40: Financial Crimes Enforcement Network, *Request for Administrative Ruling on the Application of FinCEN's Regulations to a Virtual Currency Trading Platform,* Dep't. Treasury (2014), online at http://www.fincen.gov/news_room/rp/rulings/pdf/FIN-2014-R011.pdf.

41: Financial Crimes Enforcement Network, *Bank Secrecy Act/Anti-Money Laundering Examination Manual for Money Services Businesses,* Dep't. Treasury (2008), online at http://www.fincen.gov/news_room/rp/files/MSB_Exam_Manual.pdf.

42: Jeffrey D Neuberger, *FinCEN Releases Two Rulings Classifying a Bitcoin Payment System and Virtual Currency Trading Platform as Money Services Businesses (MSBs)*, Nat'l. L. Rev. (Oct. 28 2014), online at http://www.natlawreview.com/article/fincen-releases-two-rulings-classifying-bitcoin-payment-system-and-virtual-currency.

43: Berson, *Some Basic Rules for Using 'Bitcoin' as Virtual Currency,* (cited in note 39).

44: Andy Greenberg, '*Dark Wallet' Is About to Make Bitcoin Money Laundering*

the transaction of a user with the transaction of another random user who happens to be selling or purchasing at the same time. This way, purchases and sales are more difficult for regulators to match, and money laundering becomes much easier as several transactions are recorded as if only one took place. Dark Wallet will undoubtedly face extreme scrutiny from regulators as the program matures.

VI. Solutions

While thus far I have been operating under the assumption that a new regulatory framework would be implemented to allow the technology to thrive in ways that have positive social and economic impacts, and monitor its behavior in the areas in which it presently conflicts the law, it is equally possible for governments and lawmakers alike to take a more dire stance against its existence. First, it is possible to make compliance significantly more expensive—in terms of both money and time. For example, bitcoin exchanges on which investors directly exchange bitcoins for dollars have had a tough time gaining regulatory approval in the U.S., in part because they are typically designated as MSBs and must obtain specific licenses from many of the states in which they operate. That process, which involves assuring compliance with rules designed to prevent money laundering, can often be an extensive and arduous process.[45]

Other, stricter actions could be implemented if updates to regulation continue to be ineffective or circumvented. It would be possible for the government to simply devote the computational resources necessary to mine all remaining bitcoin, thereby controlling a large portion of its total market value. In so doing, the government could take billions of dollars in value off the market that may have been used for illegal activities. Additionally, it seems possible that legislators could simply disallow payments in bitcoin altogether. While this would not directly solve the issue of the possibility tax evasion, it would significantly damage bitcoin's popularity, which would thus reduce its liquidity and value, rendering it less effective for tax-evasion purposes. But to employ either of theses strategies would amount not only to removing the problems associated with cryptocurrencies, but all of their benefits as well. A potential solution to this dilemma could be to focus on formalizing the link between the virtual economy and the real economy. Since virtual wallets

---

Easier Than Ever, Wired (Apr. 29, 2014), online at http://www.wired.com/2014/04/dark-wallet.

45: Casey, *TeraExchange Unveils First U.S.-Regulated Bitcoin Swaps Exchange* (cited in note 26).

must to be funded via credit or bank transfer, or via a payment service such as PayPal, there may be a way to mandate the provision of user identification upon government request.[46]

VII. Conclusion

Amidst all this legal uncertainty, the attitude of regulators matters at least as much as the specifics of the law. If regulators view a new technology as basically benign, they will look for ways to interpret the law to allow it to flourish. If they view the technology as a threat, they will look for interpretations of the law that could allow them to shut it down.[47] The most pressing challenge for regulators will be how to monitor the dynamic and complex virtual currency systems without restraining their growth and development.[48] Despite these shortcomings, virtual currencies can offer operational and economic efficiencies over other forms of electronic payment that rely upon the traditional banking system.[49] Bitcoin has the potential to provide tremendous benefits to the under-banked and unbanked parts of the world, particularly in emerging markets where traditional financial services often are not available. While many are calling for its end—currently, Iceland, Bolivia, Ecuador, Kyrgyzstan and Vietnam have already banned its use—others are rushing to get involved.[50] Bitcoin start-ups have attracted some of Wall Street's top bankers, while hedge funds are looking for ways to profit from virtual currencies. Bitcoin's price is now quoted on Bloomberg terminals along with stocks, bonds, currencies, and commodities. Several banks have suggested virtual currency's use for interbank lending.[51] Political parties have accepted virtual currency for campaign funding.[52] Even so, bitcoin might not be the final resting place for virtual currency. Regardless of what regulators elect to do, the underlying technology will continue to grow its influence as

---

46: Matonis, *ECB* (cited in note 5).

47: Timothy B. Lee, *New Money Laundering Guidelines Are A Positive Sign For Bitcoin,* Forbes (Mar. 19, 2013), online at http://onforb.es/1SyJTES.

48: Steven Weitzel, *Bitcoin and Virtual Currency Regulation,* Law (Sep. 4, 2014), online at *http://www.law.com/sites/articles/2014/09/04/bitcoin-and-virtual-currency-regulation/.*

49: *Id.*

50: *Is Bitcoin Legal?,* CoinDesk (Aug. 19, 2014), online at http://www.coindesk.com/information/is-bitcoin-legal/.

51: Casey, *Why Bitcoin's Erratic Price Doesn't Matter* (cited in note 2).

52: Taylor Tyler, *First State Republican Party Accepting Bitcoin*, CryptoCoins News (Sep. 17, 2014), online at https://www.cryptocoinsnews.com/first-state-republican-party-accepting-bitcoin-donations/.

developers create innovative versions of virtual currency.[53]

53: Michael J. Casey & Paul Vigna, *Bitcoin and the Digital-Currency Revolution*, Wall St. J. (Jan. 23, 2015).

The America Invents Act: First-to-File and America's Innovation Climate

Katherine Dannenmaier†

I. Introduction

On September 16, 2011, H.R. 1249, the Leah-Smith America Invents Act, was signed into law, but only on March 19, 2013 did the most controversial section of the bill take effect.[1] Thought to be "the biggest shakeup" of the patent system in fifty years, the America Invents Act (AIA) includes reforms to patent fees, post-grant review proceedings, and prior user rights.[2] One change, however, stands out: the move from "first to invent" to "first to file." As entrepreneur Dr. Ron Katznelson puts it, "now it's the filing date that counts, not the invention date."[3] There are two specific amendments to 35 U.S.C. §§ 102-103 that effect this shift. First, in sections 102(g), 135, and 291, the AIA removes references to "interferences proceedings," the process by which the Patent and Trademark Office determines "priority of invention…[by] respective dates of conception and reduction to practice of the invention."[4] Second, an amendment to section 103 changes the requirements for the novelty of an invention. Previously, an invention had to be non-obvious at "the time the invention was made." Under the new rules it must be non-obvious "before the effective filing date."[5] Thus, prior art is determined from the date of filing, instead of from the date of conception.

For years, the American patent system has been the subject of heated debate. After Canada's switch to first-to-file in 1989, the U.S. became one of only two countries in the world to retain the first-to-invent rule.[6] The primary effect of the change is straightforward: it is an increased incentive

---

†: Katherine Dannenmaier is a third-year at Stanford University, majoring in Economics.

1: Nathan Hurst, *How the America Invents Act Will Change Patenting Forever*, Wired (Mar. 15, 2013), online at http://www.wired.com/design/?p=146445.

2: *Id.*

3: Ron Katznelson, Phone interview with author, May 17, 2013.

4: *Comparison of Selected Sections of Pre-AIA and U.S. Patent Law*, Intellectual Property Owners Association 1 (2011), online at http://www.ipo.org/wp-content/uploads/2013/03/Patent_Reform_Chart_Comparison_of_AIA_and_Pre-AIA_Laws_FINAL.pdf.

5: *Id*.

6: Bernarr R. Pravel, *Why the United States Should Adopt the First-to-File System for Patents*, 22 St. Mary's L. J. 797 (1990).

for inventors to file for patents quickly. This incentive isn't new—there are already many reasons to file early, including fundraising, ease of sale, and ensuring rights in the international market.[7] So, why switch at all? And why the uproar over the switch?

Proponents of the first-to-file system maintained that America's stubborn refusal to switch lowered the efficiency of our system, hurt our ability to trade and export ideas, and delayed information disclosure that could have led to faster innovation. Proponents of the first-to-invent system argued that it improved patent quality and promoted the interests of the small businesses that lead innovation. With the passage of the AIA, it appeared that first-to-file has won the day. But two provisions in the Act suggest that Congress was not completely sure of its decision: subheadings (l) and (m) under Section 3 both call for studies of the effects of the switch to first-to-file, to be completed and reported back to Congress within the next year.[8]

Congress does not need to wait a year. Using current research, this article examines the question of whether the United States should return to its first-to-invent system, or join the rest of the world in using first-to-file. This article begins by laying out the relevant background on the United States patent system, and uses this context to determine which factors are most important in making decisions about the patent system. Next, the paper will analyze the three major differences between the systems: the overall efficiency, the effects on disclosure, and the effect on small businesses. Finally, the paper will conclude with a recommendation.

II. Background

In its report to the House of Representatives, the Committee on the Judiciary wrote, "H.R. 1249 modernizes U.S. patent law to improve the operation of the U.S. Patent and Trademark office, inhibit frivolous patent lawsuits, protect the rights of all inventors, and spur innovation as a means to create American jobs and raise standards of living."[9] The best way to assess the AIA is to assess its ability to achieve these objectives. And the most important objective to consider is the last one, that of "spurring innovation."[10]

The American government has always understood patents, first and

7: Steve Perlman, Phone interview with author, May 21, 2013.
8: H.R. 1249, 112th Cong. (2012).
9: U.S. House Committee on the Judiciary, *America Invents Act: Report Together with Dissenting Views and Additional Views*, 112th Cong., 1st sess., 2011, H. Rept. 112–98, 73.
10: *Id.*

foremost, as a way to stimulate innovation. We see this in the original provision of the Constitution concerning patents, which reads: "The Congress shall have the power…to promote the progress of science and useful arts by securing for limited ties to authors and inventors the exclusive right to their respective writings and discoveries."[11] This wording makes it clear that the founders conceived of the patent system as a way to encourage progress. Professor Mark Lemley of Stanford Law School emphasizes that "the purpose of the patent system is to encourage innovation; it has no moral purpose—there is nothing intrinsically wrong with patent infringement."[12] Thus, when deciding between the first-to-file and first-to-invent systems, our most important consideration should be which system induces more progress.

III. Efficiency

The most common argument from supporters of first-to-file is straightforward: first-to-file is an easier rule to understand and enforce, leading to savings in both the private and public sectors. This is the first argument the House Report on the America Invents Act cites when discussing reasons to switch to first-to-file.[13] Theoretically, the difference between first-to-file and first-to-invent should only come into play when two inventors independently produce the same thing around the same time. Then, there may be a question as to whether the first person to file was also the first person to invent. Professor Lemley explains why the situation suggests a first-to-file rule:

> There are a lot of records of nearly simultaneous inventions. When both inventors are working in parallel, and neither is aware of the other, there's no moral sense in which one is entitled to a patent and the other isn't…We're better off with a simple rule for choosing between them.[14]

The first-to-file system is simple, and its simplicity leads to efficiency. In order to enforce the first-to-invent system, the Patent and Trademark Office allowed for interference proceedings, in which the documentation of the current patent holder and the challenger were exhaustively examined

---

11: *U.S. Constitution.* Art. I, Sec., 8.

12: Mark Lemley, Interview with the author, Stanford, CA., May 13, 2013.

13: U.S. House Committee on the Judiciary, *America Invents Act* at 41 (cited in note 9).

14: Mark Lemley, Interview with the author, Stanford, CA., May 13, 2013.

to determine who invented what first. Studies find that the cost of these proceedings ranges from $100,000 to $500,000, and, on average, take two and a half years.[15] Fewer than two in one thousand patents enter into interferences, but, given that the United States issues over 250,000 patents a year, the total cost of the interference system is almost $277 million.[16] To put this number in perspective, the total budget of the U.S. Patent and Trademark Office was almost $3 billion in 2013.[17] So, the switch should result in substantial savings to the government.

However, the AIA also adds two new types of proceedings: derivation proceedings, to determine if a patent was based in another invention,[18] and prior user defenses, to determine if an entity had created and was using an invention before another entity patented it.[19] These are necessary in a first-to-file system in order to a) ensure that the patent goes to the correct person, and b) ensure that other independent inventors of the same innovation can profit from their ingenuity.[20] The USPTO estimates that the cost of a derivation proceeding will be around $322,000 to each party, similar to the cost of an interference proceeding.[21] Since the prior user rights defense proceeding requires many of the same steps as an interference proceeding—as in an interference, defenders must establish that they conceived and reduced their invention to practice prior to the patent being filed—their cost will likely be similar to that of interference proceedings.[22]

15: Ryan K. Dickey, *The First to Invent Patent Priority System: An Embarrassment to the International Community*, 24 B. U. Int. L.J. 283, 303–304 (2006).

16: *U.S. Patent Statistics Chart, Calendar Years 1963–2012,* U.S. Patent & Trademark Office (2013), online at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm. Also see Dickey, 24 B. U. Int. L.J. at 303 (cited in note 15).

17: Dennis Crouch, *USPTO Budget Cuts*, PatentlyO (Apr. 18, 2013), online at http://www.patentlyo.com/patent/2013/04/uspto-budget-cuts.html.

18: *Derivation Proceedings Under the America Invents Act,* Schwegman, Lundberg, & Woessner, online at http://www.slwip.com/videos/aia-topics/?play=12&?icamp=aia3.

19: David J. Kappos & Teresa Stanek Rea, *Report on the Prior User Rights Defense*, U.S. Patent & Trademark Office 47 (2012), online at http://www.uspto.gov/aia_implementation/20120113-pur_report.pdf.

20: Ron D. Katznelson, *Surviving the America Invents Act's Overhaul of U.S. Patent Law—Startup and Small Business Perspectives* (paper presented at Utah IP Summit, Salt Lake City, Utah, Feb. 2013), online at http://works.bepress.com/rkatznelson/71.

21: *Changes to Implement Derivation Proceedings*, 77 Fed. Reg. 7028-41 (Feb. 10, 2012).

22: Kappos & Rea, *Report on the Prior User Rights Defense* at 56-57 (cited in note

So, what is the final number? Projections from the Congressional Budget Office find that the cost to create these two new procedures and amend the interference proceedings would "increase spending by $1.1 billion over the 2012 to 2021 period."[23]

Additional costs may also appear in terms of wasted productivity. Since firms can't appeal to a first-to-invent system, they have a strong incentive to file every invention they come up with, even if the purpose of the invention is not currently clear.[24] Thus, businesses waste time and money patenting inventions that they will never use. A survey of Japan's first-to-file patent system found that only 27% of patents were actually used by owners.[25] Similarly, a survey of patent applications made under first-to-file systems in Europe found that only 42% of applications were used by their owners.[26] In the United States, 270,000 patents are granted every year, for a conservatively estimated average cost of 15,000 dollars per patent.[27] [28] An increase of a single percent in unused patents—which, the research suggests, is likely to happen—would result in a loss of more than $40 million in wasted effort.

There's another way in which first-to-file improves efficiency in a meaningful way—that of international harmonization. As mentioned earlier, before the American Invention Act passed, the United States was one of only two countries that retained the first-to-invent system; everyone else was on a first-to-file system.[29] Proponents of harmonization point to reduced costs in obtaining and defending patents internationally, and better global enforcement of patent protection.[30] While the United States has taken steps to harmonize its patent system with the rest of the world—including signing

---

19).

23: U.S. House Committee on the Judiciary, *America Invents Act* at 68 (cited in note 9).

24: Steve Perlman, *Letter to Senator Dianne Feinstein* (Mar. 1, 2011).

25: Hiroyuki Odagiri, "Advance of Science-Based Industries and the Changing Innovation System of Japan," in *Asia's Innovation Systems in Transition* 218 (2006).

26: Katznelson, *Surviving the America Invents Act's Overhaul of U.S. Patent Law* (cited in note 20).

27: Gene Quinn, *The Cost of Obtaining a Patent in the U.S* (Jan. 28, 2011), online at http://www.ipwatchdog.com/2011/01/28/the-cost-of-obtaining-patent/id=14668/.

28: *U.S. Patent Statistics Chart, Calendar Years 1963–2012* (cited in note 16).

29: Pravel, 22 St. Mary's L. J. at 800 (cited in note 6).

30: Vito J. DeBari, *International Harmonization of Patent Law: A Proposed Solution to the United States' First-to-File Debate*, 16 Ford. Int'l. L. J. 687, 693 (1992). Also see U.S. House Committee on the Judiciary, *America Invents Act* at 43 (cited in note 9).

several treaties and a series of recent amendments—its first-to-invent rule remained a bone of contention.[31] The switch to first-to-file is not sufficient to fully harmonize the American system—the U.S. still differs in terms of its year-long grace period, rules for publication of applications, and willingness to extend the patent term—but proponents say that it would still be meaningful progress towards that goal.[32]

Some experts, however, question the benefits of harmonization; indeed, they point to the first-to-invent system as a reason for the United States' dominant position in global patenting. The current system preserves flexibility. As long as American companies follow other countries' rules, they get the benefits of international patent protection.[33] However, the U.S. first-to-invent rule preserves a slower system for those who want to wait to file. Steve Perlman, a prolific inventor involved in this issue, summarizes the situation: "Start-ups can take their time—they could lose rights in other countries; that's a risk they take. But their rights will be preserved in the U.S."[34] In the meantime, foreign inventors can file in the United States. Perlman furthers notes, "Most inventors in the world file first in the United States—we are the world's patent office…There was never a pall over the United States patent system from the [first-to-invent system]."[35] The numbers back him up: in 2012, more than half of patents granted by the USPTO were foreign in origin.[36] Perlman attributes the U.S. patent system's popularity to its unique rule. While the AIA was being considered by Congress, he notes, "there were a number of editorials by people from other countries arguing that the U.S. should keep the patent system the way it is because other countries rely on it."[37]

From this section, we see that the goal of efficiency is not enough to motivate this major change. While there will be savings from the lack of interference proceedings, there will also be increased costs from the new types of proceedings and diminished productivity. While this brings us closer to international harmonization, there are also convincing arguments that our unique patent system attracts foreign investment and patenting in our economy. Ultimately, since it is unclear whether the change in costs will increase innovation, this argument about efficiency does not answer our main

---

31: DeBari, 16 Ford. Int'l. L. J. at 690 (cited in note 30).

32: Mark Lemley & Colleen Chien, *Are the US Patent Priority Rules Really Necessary?*, 54 Hast. L. J. 1299, 1305 (2003).

33: Ned L. Conley, *First-to-Invent: A Superior System for the United States*, 22 St. Mary's L. J. 779 (1990).

34: Steve Perlman, Phone interview with author (cited in note 7).

35: *Id.*

36: *U.S. Patent Statistics Chart, Calendar Years 1963–2012* (cited in note 16).

37: Steve Perlman, Phone interview with author (cited in note 7).

question: does the switch to first-to-file spur innovation?

IV. Disclosure

Let us consider the next argument proponents of first-to-file make: increased disclosure.[38] An obvious consequence of faster filing is earlier disclosure. Economic models of disclosure find that it encourages innovation since it allows companies to have access to the technical progress already made by their competitors, helping them to avoid wasting time re-discovering the same idea.[39] Of course, for this very reason, inventors do not want to disclose their inventions, especially if they are in the intermediate stages of a longer product development; they do not benefit from their competitors catching up. In a first-to-invent system, inventors do not have to disclose. Instead, they can wait until they have a finished product, and then patent all the relevant innovations at the same time, preventing their competitors from using their intermediary research.[40] The same rationale does not hold in a first-to-file system: since companies do not know where their competitors are in the innovation process, they need to patent their innovations as they develop them, at whatever stage they are in the process, in order to secure their ownership.[41] Thus, first-to-file encourages a) earlier disclosure and b) more disclosure, since companies will patent innovations at all stages, rather than just the innovations that contribute to their final product. This earlier and greater disclosure, say proponents of first-to-file, will encourage innovation.[42]

The literature, however, does not agree. Researchers suggest that patents do not encourage disclosure because companies only disclose information in patents that is non-concealable, information that would have been disclosed anyway. If the firm had been able to keep it as a trade secret, it would have preferred to do so.[43] Using a game theoretical model, Bessen finds that patents do not encourage the diffusion of information, and may, in some cases, impede it.[44] Furthermore, researchers suggest that disclosures made on patents do not encourage innovation since competitors do not pay attention

38: *U.S. House Committee on the Judiciary, America Invents Act* at 38 (cited in note 9).

39: Suzanne Scotchmer & Jerry Green, *Novelty and Disclosure in Patent Law*, 21 RAND J. Econ. 131, 144 (1990).

40: *Id.* at 136.

41: Steve Perlman, *Letter to Senator Feinstein* (cited in note 24).

42: Scotchmer & Green, 21 RAND J. Econ at 144 (cited in note 39).

43: James Bessen, *Patents and the Diffusion of Technical Information*, 86 Econ. Letters 121 (2005).

44: *Id.*

to them. Surveys of firms have found that they a) rarely read the patents of their competitors; and b) when they do, their purpose is usually not to piggyback on their competitor's technology.[45] Indeed, the patent system incentivizes against competitors reading patents. Heidi Lubin, CEO of the tech start-up HEVT, says, "Reading through patents can set you up for a patent infringement. If we've been asked to collaborate with someone, we look at their patents. Otherwise, it's a risky thing to do."[46] The model does not match with the reality. The incremental increase in patent disclosure does not encourage innovation, so it does not matter whether or not first-to-file encourages disclosure.

V. Small Businesses

Finally, we need to consider the argument made by opponents of the America Invents Act: that it negatively affects small businesses. Sheel Tyle, an associate at the venture capital firm New Enterprise Associates, explains:

> The America Invents Act doesn't help start-ups. They're resource-constrained initially. When the system is first-to-invent, they can prove [their invention] out, raise some money, and then go and patent it. Now that the system is first-to-file – well, patenting costs money, and they're trying to hire people, trying to build a business, and they're not going to put that money into patents.[47]

It's an argument that makes sense, and one that many people have made. In numerous letters to congress, individual inventors, associations of small businesses and startups, and even professional groups have reiterated this point.[48] But is it true, or just a good line?

Before we get deep into the research, we should establish why this question is important. A drop in small business innovation matters because small businesses matter. Small businesses, especially start-ups in new industries, are more innovative on a per dollar basis than other companies, and patents help them get the funding they need to keep innovating.[49] Small companies

45: *Id.* at 128.

46: Heidi Lubin, Phone interview with author, May 20, 2013.

47: Sheel Tyle, Interview with author, Palo Alto, CA, April 30, 2013.

48: Steve Perlman, *Letter to Senator Feinstein* (cited in note 24), and see Ron Katznelson, *Letter on Loss of Grace Period* (2010), online at http://works.bepress.com/rkatznelson/58/.

49: Joshua Lerner, "Small Business, Innovation, and Public Policy," in

backed by venture capital have been found to be especially productive. In their landmark survey of innovation across twenty industries, Samuel Kortum and Josh Lerner found that every dollar of venture capital funding has an effect *three* times that of a dollar of R&D spending: even though venture capital funding accounted for less than 3% of funding, it accounted for 8% of innovation.[50] But venture capitalists only back start-ups who have proven themselves by getting patents. Tyle, speaking with industry experience, says that when considering funding for hardware companies in industries like clean technology, "No matter what, [I look] at their patent portfolios, even if the company is just in the idea stage."[51] Indeed, a study of 370 U.S. semiconductor start-ups found that a "doubling in patent application stock [was] associated with a 24% boost in funding-round valuations beyond what would otherwise be expected."[52] So, the question of whether the switch to first-to-file hurts small businesses is also question of whether first-to-file hurts the American climate for innovation. Does it?

Let's begin by considering interference proceedings, which are unique to a first-to-invent system and thought to be the unique way the system is enforced. Do interference proceedings help small businesses to the point that small businesses will be hurt by their removal? A study of U.S. interference proceedings found that around 18% were initiated by individuals or small businesses, while the majority, 77%, were initiated by large entities.[53] From this evidence, Ryan Dickey concluded that small businesses were not taking advantage of the legal rights that the "first-to-invent" system gave them.[54] A separate study, which examined the results of interference proceedings, found that these interference proceedings harmed small businesses as frequently as it benefited them.[55] Given the high cost of interference proceedings, these results make sense—small entities simply do not have the cash to engage in them.[56] So, in this respect, the first-to-invent system does not help small businesses.

Understanding the Digital Economy: Data, Tools and Research 203 (2002).

50: Samuel Kortum & Josh Lerner, *Assessing the Contribution of Venture Capital to Innovation*, RAND Journal of Economics, 674 (2000).

51: Sheel Tyle, Interview with author (cited in note 47).

52: David H. Hsu & Rosmarie H. Ziedonis, *Patents as Quality Signals for Entrepreneurial Ventures*, Acad. Management Proceedings 1 (2008).

53: Lemley & Chen, 54 Hast. L. J. at 1299 (cited in note 32).

54: Dickey, 24 B. U. Int. L.J. at 303 (cited in note 15).

55: Gerald J. Mossinghoff, *The U.S. First-to-Invent System Has Provided No Advantage to Small Entities*, 84 J. Pat. & Trademark Off. Soc'y 425, 427–428 (2002).

56: *Id.* at 428.

But interference proceedings are not the only things changed by first-to file. Small businesses are more worried about the other amendment: what counts as prior art. While the America Invents Act accommodates a one-year grace period in which an inventor can disclose her own work, inventors say that that does not go far enough.[57] "In a first-to-invent world," says innovator Steve Perlman, "if a piece of prior art is submitted a week before my patent application, I can submit my notebook, and show that I conceived and reduced my invention to practice before the prior art, and I can receive my patent. In a first-to-file world, that's not true."[58] The strategy changes—no longer can an inventor wait to go through all the iterations of an idea before she files the final, best one. In a first-to-file environment, she is pressured to file for patents at all stages of the development of a product, even at stages where it's unclear whether the innovation will contribute to the final profitable product.[59] The problem is not having to file faster—it is having to file more. Small businesses argue that, given the expense of patents—commercial-grade patents can cost up to $30,000—first-to-file makes innovation prohibitively expensive for many companies.[60] Heidi Lubin, the start-up CEO, confirms that "It's more of a race now. Before, if we could prove that we were the first to invent, whether or not we had a patent was less important. Now, patent strategy is part of our product development…Organizations are going to have to be a lot more tactical and forward thinking in patent filing and innovation."[61]

This comes through in the data. A recent paper from the National Bureau of Economic Research compared Canadian patent assignations under a first-to-invent system with assignations after they switched to a first-to-file system. It found that after the Canadian reform, the percentage of patents owned by small businesses declined while the percentage owned by large corporations grew.[62] Controlling for the general business environment by comparing U.S. individual inventor patenting rates with those of Canada, another study corroborated the same conclusion.[63] These studies are extremely compelling

57: *U.S. House Committee on the Judiciary, America Invents Act* at 43 (cited in note 9).

58: Steve Perlman, Phone interview with author (cited in note 7).

59: *Id.*

60: Perlman, *Letter to Senator Feinstein* (cited in note 24).

61: Heidi Lubin, Phone interview with author (cited in note 46).

62: Shih-tse Lo & Dhanoos Sutthiphisal, *Does It Matter Who Has the Right to Patent: First-To-Invent or First-To-File? Lessons from Canada*, National Bureau of Economic Research Working Paper 14926, 26–27 (2009).

63: David S. Abrams & R. Polk Wagner, *Poisoning the Next Apple? How the America Invents Act Harms Inventors*, 65 Stan. L. Rev. 517 (2013).

since they studied the number of patent applications both before and after the passage of a law in the same countries, allowing them to eliminate confounding variables and isolate any changes as a result of the law. The initial results from the switch in the United States suggest the same conclusion. Before March 16th, the day the system switched, around 25 percent of those filing patent applications were small entities, a little above average.[64] After the 16th, small entities only made up 8% of filers.[65]

So, it is clear that the change to first-to-file hurts small businesses by increasing their costs to innovate, and that, given small businesses' unique position as key innovators, this shift hurts innovation. Are there any solutions that both help small businesses and keep the first-to-file system?

The most obvious answer would be for the USPTO to lessen the financial burden on small entities. First, the USPTO offers steep discounts of 75% and 50% respectively on patent application fees for micro- and small- entities.[66] Second, the Act allows for provisional applications, which are less rigorous and arguably less difficult to prepare. Business owners, however, feel that neither of these solutions is adequate. The fee to the USPTO is a negligible part of the costs of filing a patent—it's frequently less than 1% of the total cost—so lowering it has little effect.[67] Furthermore, provisional applications do not lower their costs since, in order for the provisional application to be of use, it must be of the same quality—i.e., cost the same amount of legal work—as a regular patent.[68] Otherwise, it does not serve as sufficient protection. Thus, it appears that a first-to-file system cannot be made consistent with protecting small business innovation.

VI. Conclusion

In this article, we've examined the three main arguments for and against the America Invents Act. We began by setting the terms of the discussion: since the purpose of the patent system is to increase innovation, we should judge the AIA by its effect on innovation. We started with efficiency, the main argument of proponents of the Act, and concluded that a) the marginal increase in economic efficiency from the Act was small, and b) there's no evidence linking efficiency in the patent system to increased innovation, so

64: Ron Katznelson, *Surviving the America Invents Act's Overhaul of U.S. Patent Law* (cited in note 20).

65: *Id.*

66: *Setting and Adjusting Patent Fees*, 78 Fed. Reg. 4212-13 (Jan. 18, 2013).

67: Quinn, *The Cost of Obtaining a Patent in the U.S.* (cited in note 27).

68: Steve Perlman, Phone interview with author (cited in note 7).

it should matter less in our decision. Next, we considered disclosure, the argument which links first-to-file with increased innovation. However, we found no evidence to back this claim. Finally, we focused on the impact to small businesses. We began by establishing that small businesses were key to innovation, and thus could be a deciding issue in the debate. Then, we weighed the empirical evidence and found that a switch to first-to-file has hurt small businesses in other countries. From the analysis of these three issues, a clear conclusion emerges: the America Invents Act ought to be amended to restore the first-to-invent system in order to protect America's unique innovation climate.

UNIVERSITY OF CHICAGO
UNDERGRADUATE LAW REVIEW

UCULR.COM