POLICY ESSAY

THE ILLS OF GERRYMANDERING AND INDEPENDENT REDISTRICTING COMMISSIONS AS THE SOLUTION

THE HONORABLE ALAN S. LOWENTHAL, PH.D.*

I. INTRODUCTION

There is no greater threat to a democracy than when the voters lack confidence in their political system. This is exactly what we are experiencing today. American voters have an ever-increasing feeling that political institutions do not have voters’ interests at heart.¹ Multiple issues contribute to this, including the increasing influence of dark money in our elections² and the emergence of the twenty-four-hour news cycle.³ However, perhaps no other factor has contributed more to negative public perception about voting and elections than gerrymandering and its side effects. In an age of voter disillusionment with their elected officials, the term “gerrymandering” has come to represent more than the malapportionment of political districts for parti-

¹ Member, United States House of Representatives (D-Cal.). B.A., Hobart and William Smith Colleges, 1962; Ph.D., The Ohio State University, 1967. Congressman Lowenthal is serving his fourth term in the U.S. House of Representatives representing California’s 47th District, which encompasses the cities of Long Beach, Signal Hill, Avalon, and parts of Lakewood in Los Angeles County and the cities of Los Alamitos, Rossmoor, Cypress, Westminster, Garden Grove, Buena Park, Anaheim, Midway City, and Stanton in Orange County. He serves on the House Transportation and Infrastructure Committee and on the House Natural Resources Committee, serving as Ranking Member of the Subcommittee on Energy and Mineral Resources. Before being elected to Congress, Congressman Lowenthal served two terms on the Long Beach City Council, three terms as a California Assemblyman, and two terms as a California State Senator. He is widely regarded as a leader on environmental, transportation, human rights, education, and redistricting issues. He moved to Long Beach, California in 1969 where he taught Community Psychology at California State University, Long Beach until 1998. The author wishes to thank Shane Trimmer, Senior Legislative Assistant on his staff, for his assistance in the research and drafting of this article.


Gerrymandering is the practice of manipulating electoral districts to partisan advantage, securing safe seats for a party, group, or individual at the expense of others. The practice inherently disenfranchises voters. While these maps are often associated with federal congressional districts, this practice is widespread throughout all political maps: state legislatures, city councils, even school boards. While the practice has always been unpopular and widely condemned, politicians’ ability to protect their own positions of power and expand their influence is too tempting for them to give up. History has shown, as I highlight in the coming pages, that state legislatures that have this power will not give it up of their own accord. Therefore, the power to redistrict must be taken from them and given to the voters. When the decisions are made to disenfranchise voters by creating safe seats for incumbents and protecting the party in power, it tears at the fabric of our democratic institutions. Democracy works best when the voters themselves lead it. However, if voters do not believe our democracy is working fairly, they will not participate.

The practice of gerrymandering is not unique to one party. President Ronald Reagan addressed the Republican Governors Club and highlighted the successful Democratic gerrymander of California’s district maps, when he called for “an end to the antidemocratic and un-American practice of gerrymandering congressional districts.” He continued, “The fact is, gerrymandering has become a national scandal.” More than thirty years later, after the 2010 decennial census, the United States saw an unprecedented effort organized by the Republican Party, through their Redistricting Majority Project (“REDMAP”) operation, to draw maps throughout the country to favor Republicans and secure their power for the next decade. This led Pres-

---


8 Id.

ident Barack Obama, in his last State of the Union address, to say, “[W]e’ve got to end the practice of drawing our congressional districts so that politicians can pick their voters, and not the other way around.” He concluded, “Let a bipartisan group do it.”

Historically, partisan gerrymandering has taken two forms. The first creates districts of unequal population sizes. Since numerous Supreme Court rulings in the early 1960s establishing the one person, one vote principle, states have been required to draw congressional districts with equal population and redraw districts following the decennial census to account for any population changes. Prior to these Supreme Court rulings, gerrymanders often created districts with significant population differences. Further, parties redrew some district maps multiple times within a single ten-year period to suit their own needs. Likewise, some districts remained unchanged for decades.

The second form of partisan gerrymandering, more common today, is the manipulation of district lines to ensure electoral success and to maintain the party in power. Map drawers employ two techniques, “cracking” and “packing.” Cracking a district spreads the opposition across many districts so that the majority of districts do not have a large enough population of opposition supporters to enable the opposition to viably compete in elections. Packing a district, alternatively, concentrates opposition supporters in a disproportionately low number of districts, creating safe seats for a single opponent while diluting opposition strength in the other districts. Because gerrymandering reduces competition by creating safe seats, politicians become more polarized, moving to their respective bases. This increased entrenchment of incumbents and parties disenfranchises voters, who as a result become disillusioned and disengaged from the political process.

Gerrymandering is not unique to the United States. Liberal democracies around the world have faced similar problems around partisan district map
drawing. Further, since its inception in the United States it has been unpopular. However, what is unique today is the level of sophistication that new technology has made possible, allowing map drawers to be more accurate and successful in their self-interested pursuits. Gerrymandering, at its core, is fundamentally unfair and unconstitutional, and its evils have been espoused since its first use. It is long past time for the United States to fix this injustice and bring fairer maps and more accurate representation to the voters. The best alternative, in my opinion and experience, has proven to be independent redistricting commissions.

II. HOW I GOT HERE: MY EXPERIENCE WITH REDISTRICTING AND GERRYMANDERING

I was first elected to the California State Assembly in 1998. Accordingly, my first experience in redistricting was in 2001—after the 2000 decennial census. I found this process to be very upsetting. More reminiscent of the Tammany Hall political machine than twenty-first-century governing, incumbent elected officials held private meetings behind closed doors and literally selected their own voters. Those officials prioritized political expediency and incumbency protection over any criteria that considered natural boundaries, compactness, or cohesive communities of interests.

Shortly after, I joined the Assembly Bipartisan Caucus as one of its original members. Within this caucus, I led a task force in reviewing the redistricting process to find a better alternative. I studied redistricting, reviewed proposals, and familiarized myself with existing best practices. We met with representatives from Arizona who had established an independent redistricting commission just prior to their own 2001 redistricting process. To address the problems with California’s redistricting process, I tried to modify and improve upon the Arizona model and drafted legislation in the California Legislature to create an independent commission in California to implement a new, more equitable process.

I was a Democrat in a state controlled by Democrats. My own party largely opposed the concept of independent redistricting because they expected that it would cost Democrats seats and were unwilling to give up their power over the process. Republicans also initially opposed independent redistricting because they felt that by losing their seat at the table, the party leadership would give up its leverage in protecting their incumbents. Due to

15 See Nicholas O. Stephanopoulos, Our Electoral Exceptionalism, 80 U. Chi. L. Rev. 769, 780–86 (2013) [hereinafter Stephanopoulos, Exceptionalism] (discussing problems confronted by other liberal democracies prior to removing redistricting from partisan control).
16 See Sloan — Waldman, supra note 5.
18 See Sloan — Waldman, supra note 5.
the opposition from both sides, I was unable to move my proposal forward through the State Assembly. Only later, while serving in the California State Senate, having reintroduced the proposal in subsequent terms to bring an independent commission to California, was I able to get the proposal passed in that body. However, the Assembly simply refused to take up the legislation. Fortunately, California has a ballot initiative process, which allows voters to bring measures directly to the ballot. It was through that process that I joined with a coalition of groups including Common Cause and the League of Women Voters, both of which used my legislation as a blueprint for their proposal for an independent redistricting commission in California. Under the leadership of Governor Arnold Schwarzenegger, signatures were collected in order to bring the proposal to the ballot. In 2010, Proposition 20 passed with support from 61 percent of voters. Since enacting the Citizens Redistricting Commission in California, there have been no successful court challenges to the maps. Further, California has seen more competitive elections, compared to both previous years in the state as well as compared to the rate of competitive races throughout the country.

I was first elected to the United States House of Representatives in 2012. The first piece of legislation I introduced in Congress was the Let the People Draw the Lines Act, which would have required all states to establish their own independent redistricting commissions, in order to develop their respective congressional redistricting plans. I have also co-sponsored with my colleague Brian Fitzpatrick (R-Pa.) a House resolution expressing the sense of the House of Representatives that congressional redistricting should be reformed to remove political gerrymandering. To date, neither of these efforts have received a hearing or a vote in the House.

In 2014, states’ ability to create independent redistricting commissions was challenged when the Arizona State Legislature brought a case against

---

20 See S. Res. 3, 2005 Leg., Reg. Sess. (Cal.) (resolution to amend the State Constitution relating to redistricting); S. Res. 10, 2007 Leg., Reg. Sess. (Cal.) (resolution to amend the State Constitution relating to redistricting).
21 CAL. CONST. art. 2, § 8.
the Arizona Independent Redistricting Commission, claiming the United States Constitution provided authority only for a state legislature to redistrict. This case made it to the United States Supreme Court, where I led an amicus brief filed by bipartisan Members of Congress in support of a state’s right to form an independent commission to redraw political lines. In a 5-4 decision, the Supreme Court upheld Arizona’s use of an independent commission to adopt congressional districts. This not only allowed other states to keep their independent commissions, it also set the precedent that these bodies are constitutional.

More recently, the Supreme Court had another opportunity to address the issue of partisan gerrymandering by taking up the case of *Gill v. Whitford*. Prior to October Term 2017, Justice Ginsburg predicted that *Gill* may be the most important case that the Court would decide during that term. Previous Supreme Court decisions had broached the constitutionality of gerrymandering but failed to determine a standard as to whether a district was constitutionally drawn. In the end, the Supreme Court did not rule on the standard litigated in *Gill*. Instead, the Court issued a narrow ruling finding that the parties challenging the maps did not have standing. *Gill*, and similar cases, are already making their way back through the federal courts and could be reheard in the Supreme Court as soon as the next term.

I also led the bipartisan amicus brief of current and former Members of Congress in *Gill*. Justice Kagan cited that brief in her concurring opinion, joined by Justices Ginsburg, Breyer, and Sotomayor, stating, “the evils of gerrymandering seep into the legislative process itself. Among the amicus briefs in this case are two from bipartisan groups of congressional members and state legislators. They know that both parties gerrymander. And they know the consequences.” She continued:

---

33 See *Gill*, 138 S. Ct. at 1926–34.
The congressional brief describes a “cascade of negative results” from excessive partisan gerrymandering: indifference to swing voters and their views; extreme political positioning designed to placate the party’s base and fend off primary challenges; the devaluing of negotiation and compromise; and the impossibility of reaching pragmatic, bipartisan solutions to the nation’s problems.37

III. HOW WE GOT HERE: A BRIEF HISTORY OF GERRYMANDERING

Partisan gerrymandering has been consistently denounced as undemocratic and unconstitutional. The vision of the Framers was for the House of Representatives to stand as the legislative body closest to the pulse of the people. To ensure this, they constructed Congress in such a way as to make the House more directly accountable to the voters through regularly-held competitive elections. Partisan gerrymandering acts as a direct threat to this foundational principle, putting a partisan barrier between elected officials and the voters. The Framers were aware of the conflict that could arise if federal elected officials were to draw their own district boundaries. For this reason, they gave the responsibility to the state legislatures. As the American colonies began to fight for their independence and debate the initial ideas surrounding self-governance, the role and scope of a government took shape. In Common Sense, Thomas Paine wrote:

[A]s the colony increases, the public concerns will increase likewise, and the distance at which the members may be separated, will render it too inconvenient for all of them to meet . . . . This will point out the convenience of their consenting to leave the legislative part to be managed by a select number chosen from the whole body, who are supposed to have the same concerns at stake which those have who appointed them, and who will act in the same manner as the whole body would act were they present.38

John Adams, that same year, echoed Paine’s comments:

In a large society, inhabiting an extensive country, it is impossible that the whole should assemble to make laws: The first necessary step, then, is to depute power from the many to a few . . . . [Such a representative body] should be in miniature an exact portrait of the people at large. It should think, feel, reason and act like them. . . . Great care should be taken to . . . prevent unfair, partial, and corrupt elections.39

37 Id.
39 JOHN ADAMS, THOUGHTS ON GOVERNMENT, APPLICABLE TO THE PRESENT STATE OF THE AMERICAN COLONIES 8–10 (1776).
It is clear that the desire was for a representative body close to the pulse of their voters and constituents.

A decade later, after the need for a more robust federal government became clear due to the Continental Congress’s lack of power and authority conferred by the Articles of Confederation, the role of the legislature and its relationship to the people once again figured prominently in the discussion. Fears of congressional district manipulation and malapportionment arose as the role of the proposed House of Representatives was debated at the Constitutional Convention and sent to the states for ratification. In Massachusetts, Anti-Federalists objected to ratification because of the possibility of “making an unequal and partial division of the states into districts for the election of representatives.” During the same debate, Anti-Federalists also raised concerns that without federal oversight and protection, representatives “will not be chosen by the people, but will be the representatives of a faction of that [s]tate.” In Federalist No. 52, James Madison stated, “As it is essential to liberty that the government in general should have a common interest with the people, so it is particularly essential that the branch of it under consideration should have an immediate dependence on, and an intimate sympathy with, the people.” In Federalist No. 57 Madison wrote, “The electors are to be the great body of the people of the United States.” Further, he frequently repeated the point that the primary check on members of the House of Representatives was their “dependence on the people.” Even noted Anti-Federalist George Mason agreed. Arguing during the debates over the drafting of a new Constitution, Mason said, “The requisites in actual representation are, that the [representatives] should sympathize with their constituents; [should] think as they think, [and] feel as they feel.”

It was the bipartisan concern for manipulation of elections that led the founders to grant Congress the authority, under Article I of the Constitution, to regulate the times, places, and manner of congressional elections. The House of Representatives was to be the most democratic federal body, closest to the people. Two-year terms ensured that voters would regularly hold their representatives accountable. With the election of the president by the Electoral College and election of senators by state legislatures until the Seventeenth Amendment was ratified in 1913, the House was the only federal

40 See Articles of Confederation of 1781, art. IX.
41 Debates and Proceedings in the Convention of the Commonwealth of Massachusetts, held in the year 1788, and which finally ratified the Constitution of the United States 125 (Bradford K. Peirce — Charles Hale eds., William White 1856).
42 Id. at 50.
45 See, e.g., id. at 220.
47 U.S. Const. art. I, § 2.
48 U.S. Const. amend. XVII.
Policy Essay

body directly elected by the people. The Framers, by constructing the House to be first and foremost accountable to its constituents, also saw it as a bulwark against the harmful effects of factional partisan control.\(^49\) So, when politicians and political parties draw their own districts, manipulating them to select those who will vote for them, it goes counter to the intent of the Framers to have voters select their representatives in the most direct method possible.

Immediately after the idea of our democratic institution went from theory and into practice, the United States saw its first gerrymander. Historians generally identify the first gerrymander in the United States to be in 1788, during the nation’s first congressional election. Patrick Henry and the Anti-Federalists controlling the Virginia House of Delegates deliberately included James Madison’s home county in a congressional district that also included nearby Anti-Federalist-leaning counties. It was an attempt by Henry and his faction to cut Madison out of the newly formed House of Representatives.\(^50\)

The most iconic gerrymander, however, and the one that lends its name to the practice itself, took place in 1812, when Elbridge Gerry was serving as the Democratic-Republican Governor of Massachusetts. He signed the state legislature’s redistricting proposal that included a state senate district which, to many Federalists—as well as the Boston Gazette newspaper and its cartoonist—resembled a salamander-like monster. The Gazette dubbed it the “Gerry-mander.”\(^51\) The Federalists took control of the state legislature a decade later and, with the now notorious “Gerry-mander” map still fresh in the minds, passed new political districts manipulated to strongly favor the Federalists.\(^52\)

Regardless of the state, or which political factions were in charge, gerrymandering continued unabated throughout the eighteenth, nineteenth, and twentieth centuries. This occurred despite steps calling for equality among voters, such as the passage of the Fourteenth Amendment and its Equal Protection Clause. Formal calls continued to be made to address the issue. Future president James Garfield won his 1870 congressional race with help from a beneficial gerrymander, though he would decry the practice and call for its elimination. In a speech on the House floor he said, “It is the weak point in the theory of representative government, as now organized and administered, that a large portion of the voting people are permanently disen-

\(^49\) See The Federalist No. 10 (James Madison).


\(^51\) The Gerry-mander, a New Species of Monster, which Monster, which Appeared in Essex South District in Jan. 1812, Bos. Gazette (Mar. 16, 1812).

franchised.”53 In his annual address to Congress, President Benjamin Harrison called gerrymandering “political robbery” and called on Congress to take action, stating, “The power of Congress is ample to deal with this threatening and intolerable abuse.”54

Lengthy delays in redistricting led to stark disparities in population between districts. For seventy years between 1841 and 1912, congressional district boundaries in Connecticut stayed the same.55 Conversely, the Ohio legislature redistricted congressional districts seven times between 1878 and 1892, and after six consecutive elections at one point during this period.56 “[A]t least one state redrew its congressional districts each year” from 1872 to 1896.57 By 1960, Georgia’s Fifth Congressional District had 823,489 people, while the state’s Ninth District had 272,154 people, a difference of more than 550,000 people.58 It was not until the early 1960s and a series of Supreme Court cases that voters saw their first respite from the practice of gerrymandering. In 1962, the Supreme Court heard Baker v. Carr,59 challenging Tennessee’s failure to redistrict since 1901.60 Through challenges based on the Equal Protection Clause, one person, one vote was set as precedent and the justiciability61 of such gerrymanders was established.62 Though the differences in population and redistricting were inconsistent and undemocratic, it was not until the Supreme Court ruled the practices unconstitutional that the practices stopped. Politicians saw population disparity between districts as a tool to maintain factional control and hurt their oppo-

53 Cong. Globe, 41st Cong. 2d Sess. 4737 (1870) (statement of Rep. James Garfield (R-Ohio)).
61 “Justiciability refers to the types of matters that the federal courts can adjudicate. If a case is ‘nonjusticiable[,]’ a federal court cannot hear it. To be justiciable, the court must not be offering an advisory opinion, the plaintiff must have standing, and the issues must be ripe but neither moot nor violative of the political question doctrine.” Justiciability, LEGAL INFO. INST., CORNELL L. SCH., https://www.law.cornell.edu/wex/justiciability [http://perma.cc/P7V4-5VZ6] (last visited Oct. 23, 2018).
nents. That practice seems archaic and inequitable—like gerrymandering today. But there is no reason to believe that politicians would have proportionately apportioned districts without judicial intervention.

Seeking to eliminate discriminatory election rules, Congress passed the Voting Rights Act in 1965. The Act ensured that states with a statistically significant number of minority voters drew districts with enough minority populations within them to allow for a minority candidate to be elected, allowing racial and ethnic minority groups to select representatives from their own communities.

Fractured opinions in recent Supreme Court cases such as *Davis v. Bandemer* and *Vieth v. Jubelirer* have generated confusion about whether partisan gerrymandering cases are justiciable and whether courts can devise manageable standards to review such claims. The lack of any meaningful resolution from Congress or the Supreme Court has resulted in continued control of the process by state political parties. This has led to attempts by voters to remove politicians from the redistricting process. Arizona voters approved the Arizona Independent Redistricting Commission in 2000 and California voters approved the California Citizens Redistricting Commission later that decade. The year 2010 also saw the unprecedented coordination nationally by Republicans through REDMAP. Organized through the Republican State Leadership Committee, REDMAP sought to elect Republican governors and control state legislatures with the distinct goal of controlling the relevant offices in time for the next redistricting process. Moreover, while Democrats did not have a national coordinated campaign, they engaged in their share of gerrymandering in states where they controlled the government.

While partisan gerrymandering has occurred throughout our history, it runs counter to the intentions of the Framers and has always been regarded by the public at large as disenfranchising, unconstitutional, and undemo-

---

68 See Angelo Ancheta, Redistricting Reform and the California Citizens Redistricting Commission, 8 HARV. L. — POL’Y REV. 109, 113 (2014).
Gerrymandering acts as a barrier between the people and their representatives, making politicians answerable first and foremost to their party, not their voters.

IV. How We Are Impacted: The Issues with Gerrymandering

Designing political districts that are in keeping with traditional redistricting principles would create districts with shared concerns and interests. Having a coherent district built on the principle that the representative is elected to serve the constituency emboldens and empowers the community to participate in the electoral process. Protecting political parties and incumbents motivates district mapmakers in many states. There is a positive correlation between the statewide partisan advantage, as measured by the efficiency gap, and a single party’s control of a state government. Gerrymandered districts lead to a lack of competitive elections, more polarized elected officials, and a disengaged constituency. By putting a barrier between representatives and their constituents, gerrymandering makes representatives more responsive to political factions and special interests than to their constituents and voters. As shown in the previous Section, such results run counter to the system of government designed and envisioned by the founders, in which the power rested in the people. Gerrymandering has flipped this system around, allowing politicians and political parties to select the voters who will vote in their elections.

By drawing districts to remove the likelihood of a competitive election, gerrymandering creates safe seats. Done across a state through “cracking” and “packing,” the party drawing the maps tries to secure their majority, ensuring party leadership in a statehouse and congressional delegation. Maps carved out to maximize the number of registered and likely voters for a specific party minimize the chances for turnover and secure the dominance of a single party or candidate. As a result, over the last thirty years the United States has seen the number of competitive congressional elections decrease as the number of safe seats has increased.

A party that has manipulated political districts to create enough safe seats to protect its majority removes its accountability to the people.
ing politicians more accountable to their party than the voters shifts representatives to the ideological poles—diminishing any chance for bipartisan cooperation and compromise and making the party whip more important than the community advocate. Politicians who do not have to face a competitive general election see only two challenges to keeping their positions: alienating party leadership or a primary challenger from within the party. Either of these can pull a representative toward his or her party’s base. State partisan gerrymanders have been shown to “dramatically influence the representational distortion of House delegations.” Further, individual politicians in gerrymandered seats “tend to fall further from the ideological center than do politicians who have to reach out to voters from both parties to get elected.” While redistricting is up to state legislatures, the process is directly influenced by the national political parties and their leadership. There are also examples of political parties using redistricting as a way to keep members in step with party leadership. Recent examples include Republicans threatening to draw members such as Representatives Jim Jordan (R-Ohio) and Justin Amash (R-Mich.) out of their seats. Both representatives are members of the Freedom Caucus, who have been known to counter Republican party leadership. As another example, in 2000 current Rep. Hakeem Jeffries (D-N.Y.) challenged, and was narrowly defeated by, an incumbent Democratic state assemblyman. In response, the New York State Democratic Party redrew the assembly district where Jeffries lived to carve out the block containing his home, thus preventing Jeffries from challenging the incumbent again.

The reduction of competition that results from gerrymandering is severe and results in fewer opportunities for voters to exercise their right to act as a check on their representatives. Safe seats draw fewer qualified opposition candidates, if they attract any challenger at all. The nonpartisan voting-

77 See Stephanopoulos, supra note 73, at 2147.
78 Id. at 2144.
83 See id.
rights organization FairVote reports that the 2016 House of Representatives elections saw landslide\textsuperscript{85} victories in roughly 74 percent of races nationwide, with approximately 15 percent of races being uncontested altogether.\textsuperscript{86} Further, only eight incumbents lost in the 2016 general election, out of 387 incumbents challenged,\textsuperscript{87} despite Congress as a whole having an approval rating of 19 percent at the time.\textsuperscript{88}

Without a meaningful choice in electoral competition, the voters’ voices are diminished. Between 2009 and 2016 Gallup found a steep decline in the percentage of Americans who have confidence in the honesty of our nation’s elections.\textsuperscript{89} In just these seven years, voter trust fell from 59% to 30%.\textsuperscript{90} Further, Gallup found that those lacking faith in our electoral system rose from 40% to 69%.\textsuperscript{91} When constituents are gerrymandered into a district where their voices are not heard by their representatives and their votes do not count at the ballot box, they will stop participating in the process. And democracy fails when citizens fail to turn out.

All of the negative ramifications attributed to partisan gerrymandering previously discussed are exacerbated by the increasingly sophisticated tools at the political mapmakers’ disposal today. Advances in technology and the emergence of big data have allowed gerrymanders to be more accurate and aggressive. The process of gerrymandering has evolved from an art to a science, allowing for a smaller margin of error. Politicians have ceased enlisting the help of cartographers when redistricting and now employ analytic consultants.\textsuperscript{92} While state parties historically had to balance incumbent protection with seat maximization, technological advances have allowed mapmakers to “engineer so much advantage that it is possible to satisfy both of these goals.”\textsuperscript{93} Computer technology “has enabled the practice of this dark art with increasing precision and success.”\textsuperscript{94} A study by the Brennan Center for Justice showed that historically the effectiveness of a gerryma-

\textsuperscript{86} See FairVote, supra note 84.
\textsuperscript{87} Id.
\textsuperscript{89} Id.
\textsuperscript{90} Id.
\textsuperscript{92} Anthony J. McGann et al., Gerrymandering, the Supreme Court and the Constitutional Revolution of 2004 87 (2016).
\textsuperscript{93} Pildes, supra note 17 (discussing role of computer technology in district mapping).
Policy Essay

2019]
der would diminish over time, due to population shifts and incumbents retiring.95 However, today, “mapmakers now know a lot more about voters.”96 Gerrymanders are “becoming much more effective and pernicious”97 as technology has allowed more sophisticated and durable maps. If a party can draw lines with such accuracy that it can confidently control a majority of districts for the next decade, the party can ensure it is in power the next time that lines are drawn. That extended control secures its power not only for this decade but also the next.

V. HOW WE FIX IT: THE SOLUTION TO GERRYMANDERING

As I stated in the Introduction, when serving in the California State Legislature, I closely reviewed the best possible solutions to gerrymandering and my thinking quickly coalesced behind the idea of independent commissions. Arizona had paved the way in the United States and had offered tangible success. Today, most liberal democracies in the world have moved past political redistricting to some form of commission.98 Unconstrained political redistricting regularly benefits a single political party at the expense of fair or accurate representation and to the detriment of our democratic institutions. Independent redistricting commissions are currently the best tool that voters have to mitigate political influence and the negative effects of gerrymandering. States that allow their legislatures to control the redistricting process are more likely to have “higher partisan bias, lower electoral responsiveness, and reduced public confidence in the electoral system,” compared to other countries or states with independent commissions.99

The people also recognize these facts. A 2013 Harris Poll found that 64 percent of total respondents believe that redrawing districts is often used to take power away from voters.100 Further, the same poll found that 71 percent thought that those who benefit from redrawing districts should not have a say in the process.101 The preferred alternative, by an overwhelming margin, was an independent commission.102 A 2017 bipartisan poll conducted on behalf of the Campaign Legal Center showed similar overwhelming support nationally for removing partisan bias from redistricting, with 73 percent of
respondents saying they would support limitations, even if it meant their preferred political party would win fewer seats.103

This issue is not unique to the United States system. Most liberal democracies have faced similar issues in the past, with gerrymandering existing in one form or another for many centuries.104 While some individual states in the United States have adopted independent commissions, the U.S. system as a whole remains an outlier among liberal democracies by allowing its political actors to control their own redistricting, which most states in the Union still do.105 Most other countries leave redistricting to commissions.106 Redistricting commissions in Canada, Japan, and the United Kingdom are composed of members appointed by politicians and the commissions’ final plans require legislative approval.107 In contrast, redistricting commissions in Australia and New Zealand are entirely autonomous, with nonpartisan commissioners.108 They do not require approval from a court or legislative body.109 France has the most similar system to the United States: their parliamentary government is also comprised of single-member districts, although their districts are drawn at the federal level by the executive branch and reviewed by a constitutional council.110 Further, research into the French parliamentary redistricting process has concluded the level of partisan bias there to be “minuscule.”111

Overwhelming trends show voter dissatisfaction with Congress and its increasing polarization and an ever-increasing desire to see independent commissions replace self-benefitting politicians in the redistricting process. So why have we not seen more states adopt commissions? As I mentioned earlier, politicians and political parties are self-serving and will not give up power or influence easily. Congress—through the power vested in it by the Elections Clause in Article I of the Constitution—may regulate the times, places, and manner of congressional elections, and could compel states to draw congressional districts using independent commissions.112 Such legisla-

---

105 See Stephanopoulos, Exceptionalism, supra note 15, at 780–86. R
106 See id. at 780.
107 See Stephanopoulos, Exceptionalism, supra note 13, at 783–84. R
108 See id. at 784.
109 See id.
110 See Prokop, supra note 106.
112 See U.S. CONST. art. 1, § 2, cl. 3.
tion requires support from many of the same politicians who directly benefit from gerrymandering themselves or fear what may happen if they no longer had a seat at the table. As stated in the Introduction, when I was in the California State Legislature, initially I could not get a vote in the assembly to create a redistricting commission. The first piece of legislation I introduced when I arrived in Congress, the Let the People Draw the Lines Act, never received a vote in the House of Representatives. Neither has my bipartisan resolution expressing the sense of the House that congressional redistricting should be reformed to prevent political gerrymandering.

Likewise, I previously highlighted that partisan gerrymandering is counter to the Framers’ intent and is unconstitutional. Because politicians abused the process to benefit one faction over another, it was the Supreme Court that had to put limits on malapportioned districts by establishing that new lines must be drawn every decade.113 The Court has also ruled racial gerrymandering unconstitutional unless narrowly tailored to meet a compelling governmental interest.114 They have yet to rule partisan gerrymandering unconstitutional or even put barriers to limit extreme partisan gerrymandering. Justice Kennedy’s retirement after the last term jeopardizes the potential for judicial intervention in the short-term.115

This leaves state-level change as the best hope for movement on this issue. As in Arizona and California, the change must come from the people. Twenty-four states have some form of voter initiative or ballot referendum process, with fourteen states having a direct initiative process.116 Currently, there are grassroots efforts in Arkansas, Michigan, Missouri, Oklahoma, Utah,117 and other states to bring some form of redistricting change to their states through their respective initiative processes. Further, there have been recent successful grassroots efforts by coalitions of local advocates to bring change to the ballot through their local state legislatures.118

In the United States, Arizona and California have been national leaders in successfully removing gerrymandering from their redistricting processes and serve as models for others.\textsuperscript{119} Provisions governing both the Arizona Independent Redistricting Commission and the California Citizens Redistricting Commission are similar: both seek to build representative constituencies that incorporate traditional redistricting principles while minimizing partisan drivers and political influence over the process. I believe that they serve as the current best practice and a model from which other states can build.

Arizona’s commission is made up of five members, with no more than two members of any one political party.\textsuperscript{120} No member may have recently held or been involved in political office. Building on this, California, to ensure that there was no partisan influence, increased its number of commissioners to fourteen. Five commissioners come from the largest political party, five from the second largest, and four from neither of those two parties. The chair must be appointed from the last group. In addition, California commissioners are prohibited from participating in overt political activity.\textsuperscript{121} Both states’ commissions require that all of their meetings be transparent and open to the public, with ample opportunities for public input.\textsuperscript{122} The proposed maps for each must be publicly displayed and available for public comment before final adoption.\textsuperscript{123} Final adoption of the maps must include a majority from each group of commissioners.\textsuperscript{124}

Both commissions must comply with federal requirements, such as equally apportioned districts and the Voting Rights Act; additionally, both must consider specified criteria in their map drawing including communities of interest, geographic contiguity, existing local government boundaries, and compactness.\textsuperscript{125} District lines that are drawn to incorporate communities of interest, follow natural geographic features, and consider existing county lines “are linked to higher voter participation, more effective representation, and . . . lower legislative polarization” compared to those that are drawn to pursue a partisan advantage.\textsuperscript{126} Redistricting criteria serve not only to deter gerrymandering, but also to build a coherent political and geographic community that can be represented effectively. University of Chicago Law School professor, gerrymandering scholar, and co-creator of the efficiency

colorado-redistricting-resolutions-head-to-the-ballot/73-549670538 [https://perma.cc/NV4S-8HZX].
\textsuperscript{120} See Ariz. Const., art. IV, pt. 2, § 1.
\textsuperscript{121} See Cal. Const., art. XXI, § 2(c)(2).
\textsuperscript{122} See Ariz. Const., art. IV, pt. 2, § 1; Cal. Const., art. XXI, § 2(c)(2).
\textsuperscript{123} See Ariz. Const., art. IV, pt. 2, § 1; Cal. Const., art. XXI, § 2(c)(2).
\textsuperscript{124} See Ariz. Const., art. IV, pt. 2, § 1; Cal. Const., art. XXI, § 2(c)(2).
\textsuperscript{125} See Ariz. Const., art. IV, pt. 2, § 1; Cal. Const., art. XXI, § 2(c)(2).
\textsuperscript{126} Stephanopoulos, Exceptionalism, supra note 15, at 806.
Policy Essay

2019]

gap theory Nicholas O. Stephanopoulos stated, “The rules for how districts are drawn shape constituencies’ internal complexions, which in turn shape the makeup of the legislature as a whole—and thus the very character of representative democracy.”

Lastly, both California and Arizona explicitly remove any consideration of an incumbent’s residence, requiring that the place of residence of any incumbent or political candidate not be considered when drawing the maps. Further, resident party registration and voting history data are also excluded from consideration in drawing maps in California. In Arizona, the commission must consider district competitiveness. Accordingly, party registration information is initially withheld from the commission and then incorporated before final approval. In a partisan redistricting process controlled by the state legislature, securing party power and incumbent protection are the two primary goals. In California, however, they are not to be considered at all. In Arizona, incumbent residence information is not to be considered, while voter registration information is only to be considered when seeking greater competition. Additionally, Idaho, Iowa, and Montana also ban consideration of an incumbent’s home and limit the use of political data.

VI. Conclusion

If the practice of gerrymandering is allowed to continue unfettered in many states, as we see today, the next round of redistricting after the 2020 decennial census will likely be the worst ever experienced. Both political parties—Democrats and Republicans—will be emboldened by the indecisiveness and inaction of the Supreme Court and Congress. Looking ahead to 2020 and beyond, Republicans have re-launched their REDMAP operation. They continue to seek victories for Republicans in governorships and in state legislatures, in part in order to maximize their ability to control redistricting efforts in 2020. Likewise, in response to the Republicans’ previous success with REDMAP, Democrats—not to be caught off guard again—have put forward their own effort with the National Democratic Redistricting Committee (“NDRC”), led by former Obama administration Attorney General

127 Id. at 814.
128 See Ariz. Const., art. IV, pt. 2, § 1; Cal. Const., art. XXI, § 2(c)(2).
129 See Cal. Const., art. XXI, § 2(c)(2).
131 See id.
133 Iowa Code Ann. § 42.4(5) (West 2018).
135 See Ruger, supra note 80.
Eric Holder. While built to help Democratic statewide and state-legislative efforts, particularly in the lead up to 2020, the NDRC states that their mission is to “ensure the next round of redistricting is fair and that maps reflect the will of the voters.” Concerns have been raised that this is simply a Democratic BLUEMAP operation.

Further, there are concerns with ongoing efforts by the Trump administration to undermine the upcoming decennial census. Earlier this year the administration announced its decision to include a citizenship-status question in the 2020 census questionnaire. While many states have sued the administration for this policy shift, challenging it as unconstitutional, there is a chance that the next census will see lower participation as a result of the administration’s efforts. The census governs the allocation of representatives, electoral votes, and government funding. Lower participation and a resulting undercount in the census will result in malapportionment in congressional seats for states, as well as risking federal funding in communities. Under the Constitution, congressional districts are based on total population and not voters or voting-eligible persons. Undercounting immigrants or minority communities could have disastrous consequences on state congressional allocation figures.

Partisan gerrymandering historically has been unpopular and criticized as an undemocratic, unconstitutional practice. It is long past time for the practice to end. As shown above, there are multiple ways to put checks and limits on political map drawers or simply remove politicians from the process altogether and implement independent commissions. Voters and grassroots advocates working at the state level have had the most recent success in response to inaction from Congress and the Supreme Court. By working together, collecting signatures, and getting measures on local and state ballots, advocates and everyday voters have successfully brought independent redistricting commissions to states like Arizona and California, where politi-

---

136 About the NDRC, NAT’L DEMOCRATIC REDISTRICTING COMM., https://democraticredistricting.com/about/ [https://perma.cc/V5QX-JQP4].
137 See Daley, supra note 70.
cians no longer draw their own districts to prioritize incumbent protection and party power at the expense of communities of interest and disenfranchising voters. We must make our democracy work again. We need fair maps. We need independent commissions.
ARTICLE

LOBBYING AGAINST THE ODDS

KIRSTEN MATOY CARLSON*

I. INTRODUCTION ........................................ 24

II. EMPIRICALLY INVESTIGATING THE PREVAILING NARRATIVE:

A. American Indians and Legislative Advocacy ........ 29

B. The Study ........................................... 35

1. Data Collection .................................. 35

2. Methodology .................................... 36

3. Limits to the Study .............................. 37

III. CHANGING STRATEGIES: THE RISE OF AMERICAN INDIAN LEGISLATIVE ADVOCACY .................. 38

IV. INTEREST GROUP AND SOCIOLEGAL APPROACHES TO UNDERSTANDING ADVOCACY ................ 46

A. Interest Group Approach ......................... 46

B. Sociolegal Approach ............................... 50

V. A GENERALIZABLE APPROACH FOR UNDERSTANDING ADVOCACY CHOICES ............................ 52

VI. HOW AND WHY AMERICAN INDIANS CHOOSE LEGISLATIVE ADVOCACY ............................. 55

A. Political Context: The Tribal Self-Determination Policy Sets the Stage for American Indian Legislative Advocacy ........................................ 56

B. Political Capacity: The Senate Committee on Indian Affairs Creates Access for American Indians ........ 61

C. Shifting Institutional Dynamics: Legislative Advocacy and the Courts ..................................... 63

D. Resources: Legislative Advocacy and the Rise of Indian Gaming ............................................ 69

E. Opposition: The Backlash to Gaming and Legislative Advocacy ........................................... 71

* Associate Professor of Law and Adjunct Associate Professor of Political Science, Wayne State University. Ph.D. 2007 (political science), The University of Michigan; J.D. 2003, The University of Michigan Law School; M.A. 1999 Victoria University of Wellington, New Zealand (Fulbright scholar); B.A. 1997, The Johns Hopkins University. I thank Sarah Abramowicz, Bethany Berger, Scott Cummings, Peter Hammer, Chris Lund, Steve Melti, William Ortman, Kevin Washburn, Jon Weinberg, Steve Winter, and participants at the 2018 Yale/Stanford/Harvard Junior Faculty Forum, the 2016 Law and Society Annual Meeting, and the 2016 Midwest Law and Society Retreat for their comments on earlier versions of the article. The National Science Foundation Law and Social Science Program funded this research (#SES 1353255).
Conventional narratives maintain that groups lacking political power litigate because they cannot attain their goals politically. They predict that American Indians—who remain the most impoverished group in the United States and are hardly electorally significant with two percent of the population spread across 35 states—will pursue litigation strategies. In contrast to the prevailing narrative, lawyers, tribal advocates, and political scientists have suggested that American Indians have increasingly engaged in political advocacy since the mid-twentieth century.

In this Article, I extend my earlier research investigating the relationships among groups, courts, and political processes by starting to evaluate the widely-accepted proposition that groups that lack political power litigate through a case study of American Indian advocacy at the end of the twentieth century. American Indian advocacy is a particularly rich setting for investigating how and why groups craft advocacy strategies over time. American Indians, and especially Indian nations, have a long and rich history of engaging with the United States government. Since its formation, the United States has established legal relationships with American Indians, treating them as separate political communities or tribes. During the first century of its existence, the United States government and Indian nations entered into over 400 treaties. These treaties acknowledge the tribes’ preexisting and ongoing rights and governmental authority. Treaties, federal legis-

2 Kirsten Matoy Carlson, Congress and Indians, 86 U. Colo. L. Rev. 78, 82–83 (2015) (questioning whether advocates should assume that legislatures treat Indians more favorably than courts without systematic, empirical data on the amount and kinds of Indian-related legislation enacted by Congress and producing the first comprehensive study of Indian-related legislation).
3 See generally Fredrick E. Hoxie, This Indian Country: American Indian Activists and the Place They Made (2013); Edward Lazarus, Black Hills/White Justice: The Sioux Nation versus the United States, 1775 to the Present (1991) (documenting over two decades of Sioux advocacy for a congressional act authorizing the bringing of the Black Hills claim in federal court).
islation, and Supreme Court decisions form the basic legal framework governing Indians in the United States today. The key elements of this framework include: federal recognition of inherent governmental authority possessed by Indian tribes, which usually supplants state powers; a federal trust obligation toward and special federal powers over Indian tribes and their citizens; and federally protected lands for designated Indian tribes.6

Indian nations have interacted with the federal government in various ways over the past five centuries. Some Indian nations have petitioned and sent delegates to Washington, D.C. to meet with members of the Executive Branch and Congress.7 Others have litigated, protested, occupied federal lands, or exercised treaty rights. Over time, different tribes have used various combinations of these strategies.8 Despite this variation, few scholars have studied, and to my knowledge none have systematically documented, the different strategies used by American Indians.9

This Article starts to fill the gap in existing knowledge about American Indian advocacy by presenting the first comprehensive study of reported lobbying by American Indian tribes and organizations. Part II describes the study and situates it within the history of American Indian advocacy more generally. Part III reveals a dramatic 600 percent increase in reported lobbying by American Indian tribes and organizations from 1978 to 2012.10 This increase exceeds the rise in legislative advocacy that has occurred more generally in the U.S. population over the past five decades.11 My findings, however, mirror those of other scholars who have documented other politically

6 See generally Goldberg-Ambrose, supra note 4.
9 See infra Part II.A.
10 The author generated this percentage from original quantitative data she collected and analyzed. For a full discussion of the author’s data collection methods, see infra Part II.
powerless groups, including individuals with disabilities and marriage equality activists engaged in legislative strategies. Moreover, my research suggests considerable variation among Indian tribes in their lobbying efforts and qualitative changes in lobbying as some Indian tribes have expanded their lobbying efforts beyond traditional Indian law topics. The extraordinary increase in the use of legislative strategies by American Indians, especially when considered as related to the larger trend towards increased lobbying in the United States, calls into question the prevailing narrative about when and why groups advocate in particular institutions.

Contrary to the dominant narrative, my data reveal that some groups may choose to lobby even if the odds seem stacked against them. Lobbying gained momentum as a viable strategy for American Indians in the late 1970s even though they had little political influence, the courts were still largely receptive to their claims, and gaming had yet to infuse some tribes with new financial resources to support their political efforts. American Indians did not turn to Congress because the courts were unavailable, and they did not abandon their litigation strategies in appealing to Congress. Rather, they appear to have been using legislative advocacy as a parallel or complement to other strategies. My findings, thus, suggest that groups may turn to legislative advocacy when they appear to lack political power even if another institution is receptive to their claims. Moreover, they indicate that groups may utilize legislative advocacy in a broader array of circumstances than usually thought. These circumstances include but are not limited to using legislative advocacy as a parallel, an alternative, a complement, or a precursor to litigation. My data, thus, indicates that the common wisdom is either just plain wrong or substantially oversimplifies how, when, and why groups chose advocacy strategies.

See Neta Ziv, Cause Lawyers, Clients, and the State: Congress as a Forum for Cause Lawyering During the Enactment of the Americans with Disabilities Act, in Cause Lawyering and the State in the Global Era 212 (Austin Sarat & Stuart Scheingold, eds.) (2001); see also Cummings & NeJaime, supra note 11, at 1248–62.


Skepticism about the prevailing narrative is not new. Several different literatures question aspects of this popular narrative and suggest its limited utility as a general proposition. Robert Dahl investigated the assertion that the Supreme Court’s primary role is to protect the rights of the minority against the tyranny of the majority and found that “policy views dominant on the Court are never for long out of line with the policy views dominant among the lawmaking majorities of the United States.” Robert A. Dahl, Decisionmaking in a Democracy: The Supreme Court as a National Policy-Maker, 6 J. PUBL. L. 279, 285 (1957). While political scientists have contested Dahl’s findings, see, e.g., Jonathan D. Casper, The Supreme Court and National Policy Making, 70 AM. POL. SCI. REV. 50, 50–51 (1976), some scholars have grown increasingly skeptical about the usefulness of litigation as a tool for successful law reform or policy change. See, e.g., Michael J. Klarmen, From Jim Crow to Civil Rights
Part IV answers an even more important question raised by problematizing the prevailing narrative: how and why do groups craft advocacy strategies? This question is not new. But in answering it, few scholars have questioned a key assumption underlying the dominant narrative, namely its treatment of advocacy strategies as a dichotomous choice between litigation and legislative strategies. By and large, political scientists and sociolegal scholars have continued to rely on this assumption in formulating frameworks for understanding how groups craft advocacy strategies. Political scientists have focused on explaining why groups lobby while sociolegal scholars have investigated why groups litigate. In contrast, my data demonstrate that American Indians did not face a binary choice when they turned to lobbying strategies in the 1970s. Rather they actively engaged in both litigation and legislation. Like previous empirical studies, my findings suggest that viewing advocacy strategies as an either/or choice oversimplifies how advocates actually craft advocacy strategies. I argue that abandoning this assumption and combining insights from the political science and sociolegal literatures allows for the development of a more generalizable theory about how and why groups craft advocacy strategies in multiple institutions over time. The value of a more generalizable theory cannot be overstated in an era when advocates increasingly pursue advocacy strategies in non-judicial contexts and in multiple institutions simultaneously.

Part V then proposes a more generalizable and relational approach for understanding how and why groups craft advocacy strategies across institutions over time. I combine insights from the existing sociolegal and political science approaches to construct a generally applicable framework for understanding the evolution of advocacy strategies. Consistent with the sociolegal approach, I conceptualize the development of advocacy strategies as an ongoing and interactive process. Treating the development of advocacy strategies as an interactive process allows for in-depth exploration of the dynamic interplay among the factors influencing how strategies develop over time. From the interest group literature, I borrow the factors found to influence lobbying and integrate them into the sociolegal approach to emphasize the ongoing, dynamic nature of crafting advocacy strategies.

My approach improves upon existing frameworks in three main ways. First, it goes beyond existing sociolegal and interest group approaches to consider a fuller range of factors influencing advocacy strategies and how those factors interact in complex ways over time to shape and reshape strate-
gies. It highlights how changes in one factor may facilitate changes in others, either spurring or deterring the adoption of a particular strategy. For example, a change in the political context, such as the loss of the support of political elites due to electoral losses, may encourage a group to reconsider a legislative strategy. Moreover, it emphasizes how feedback loops may develop among factors, reinforcing existing conditions supportive of a particular advocacy strategy. Second, my approach allows for in-depth exploration of the relational aspects of making advocacy decisions. Advocates choose strategies while considering the alternatives. The receptivity of one institution may encourage or discourage advocacy in another institution, but that receptivity may shift over time, causing advocates to reconsider or alter their strategies. For example, marriage equality advocates considered litigating, but court decisions unsupportive of gay rights encouraged them to rethink that strategy and consider legislative ones instead. Examining institutional interactions reveals the relational aspects of advocacy choices, including how a group’s strategy may depend on how it perceives its chances of success in another institution and how such perceptions change over time. Third, my approach models a way for scholars to devise more complex understandings of advocacy strategies and how they develop in and across different institutions over time.

In Part VI, I demonstrate the value of my approach by using it to explain the explosion in legislative advocacy by American Indians from 1978 to 2012. American Indians increased lobbying after a dramatic shift in federal Indian policy in 1975. The creation of a new federal Indian policy signaled a change in the receptivity of members of Congress toward Indian claims and opened the door for American Indians to lobby. Around the same time, the Senate created the Senate Committee on Indian Affairs, which provided access to Congress and new opportunities for American Indians to engage in legislative strategies. Lacking in financial resources and electoral clout, American Indians nonetheless used these opportunities to build relationships with members of Congress and their staffers, which further reinforced their access to the legislative process. American Indians continued to use legislative strategies as the Supreme Court grew increasingly hostile towards their claims in the 1980s. In the 1990s, the emergence of Indian gaming further encouraged and reinforced American Indian legislative advocacy. The American Indian case study illustrates how strategies develop in response to a confluence of factors that evolve over time, including changes in the political context, the development of a group’s own political capacity, a

---


dramatic shift in institutional receptivity, and the infusion of new financial resources.

Part VII concludes by considering the implications of the approach for studies of advocacy strategies, interest groups, and federal Indian law. My research demonstrates that groups may turn to legislative advocacy more frequently and in different ways than previously thought, which suggests that scholars need to think more carefully about how interactions among different factors affect advocacy strategies over time. It also raises important questions about the role of American Indians in the political process, including the impact of their lobbying on substantive policy and the success of their legislative advocacy.


This Part describes how I investigated the prevailing narrative that politically powerless groups choose litigation strategies by empirically documenting American Indian legislative advocacy at the end of the twentieth century. Part II.A situates the study in the history of American Indian advocacy generally. Part II.B explains the data collection, methodology, and limits of the study.

A. American Indian Advocacy in Historical Perspective

American Indian advocacy predates the formation of the United States. Indian nations started petitioning colonial governments almost as soon as Europeans landed on American soil.21 They have continually advocated for the recognition and protection of their tribal sovereignty and land rights since then. During the nineteenth century, Indian nations used the treaty-making process to retain their existing governmental and property rights.22 Indian nations also petitioned and sent delegates to Washington, D.C. to meet with members of the Executive Branch and Congress.23 They continued to petition and send delegations to Washington, D.C. after Congress unilaterally terminated treaty-making in 1871.24

21 See Herman Viola, Diplomats in Buckskins: A History of Indian Delegations in Washington City 13–21 (1995); Carpenter, supra note 7, at 349.
22 See Cohen’s Handbook, supra note 5, at § 1.03[1]; Indian Law Res. Cent., supra note 5, at 123. For example, Anishinaabek nations in present day Wisconsin and Minnesota negotiated the retention of their hunting, gathering, and fishing rights even when they ceded lands to the United States government. 1837 Treaty with the Chippewa, Chippewa-U.S., July 29, 1837, 7 Stat. 536; see also Minnesota v. Mille Lacs Band of Chippewa Indians, 526 U.S. 172, 175 (1999).
23 See Hoxie, supra note 3, at 70–71 (documenting Choctaw delegations to Washington, D.C. in the 1820s); see also W. Dale Mason, supra note 7, at 40; Carpenter, supra note 7, at 349.
24 See Viola, supra note 21, at 190–99.
Indian nations have used different advocacy strategies over time. As early as the 1830s, Indian tribes started utilizing United States courts to protect their land and sovereignty rights.25 Others have gone to war, protested, occupied federal lands, exercised treaty rights, or used various combinations of these strategies to protect their interests and resist encroachments on their lands and rights.26

American Indian advocacy continued into the twentieth century.27 The twentieth century, however, brought tremendous changes to American Indian identities and ways of life, which in turn influenced their advocacy efforts. American Indians left the reservations at an unprecedented rate during World War II to join the armed forces or seek jobs supporting the war effort.28 As a result, Indians met Indians from other tribes.29 They slowly came together to form pan-Indian organizations and encourage advocacy across tribal lines.30 These pan-tribal efforts, while not always successful, reflected an emerging pan-tribal Indian identity.31 They culminated in the formation of the National Congress of American Indians (“NCAI”), the oldest existing national pan-Indian organization in 1944.32 In the 1950s, the NCAI mobilized resistance against the termination policy, which sought to end the federal government-to-government relationship with Indian tribes and dissolve the trust status of Indian reservations.33

The relocation of American Indians to urban centers in the 1950s further facilitated a resurgence in Indian identity and advocacy.34 The relocation program failed to secure steady jobs for Indians but led to the creation of urban Indian organizations, which further brought Indians from different tribes together.35 Federal resources flowed to these new urban centers, and

26 See, e.g., Wilkins & Stark, supra note 8, at 189–206.
27 For a more in-depth discussion of American Indian political action in the twentieth century, see Stephen Cornell, The Return of the Native 187–201 (1988).
29 See id. at 119–20.
30 The first pan-Indian organizations emerged in the early twentieth century with the creation of the Society of the American Indian and the National Council of American Indians. See Dexter Fisher, Zitkala-Sa: The Evolution of a Writer, 5 AM. INDIAN Q. 229, 235 (1979). These organizations advocated for Indian citizenship, Indian employment in the Bureau of Indian Affairs, equitable settlement of tribal land claims, and legal reforms relating to Indians. See id.; see also Wilkins & Stark, supra note 8, at 195 (describing the Society of the American Indian). These organizations no longer exist.
34 See Nagel, supra note 31, at 119–21 (explaining how relocation increased contact among American Indians, and thus, identification of common issues and goals).
35 See id.
they generated urban American Indian identities and communities.\textsuperscript{36} Urban Indian centers cultivated Indian activism in the early 1960s, as disillusioned urban Indians founded the American Indian Movement to end police mistreatment.\textsuperscript{37} The red power movement, the American Indian corollary of the civil rights movement, would emerge out of the early efforts of these urban Indian centers.

With the advent of the civil rights movement, Indian political activism continued to grow in the 1960s and 1970s.\textsuperscript{38} Indian tribes persevered in resisting restrictions on their hunting, fishing, and resource rights.\textsuperscript{39} American Indians also borrowed organizational forms, practices, and rhetoric from the civil rights movement and tailored them to their own grievances against the federal government.\textsuperscript{40} American Indian activists reclaimed land by occupying Alcatraz Island,\textsuperscript{41} marched on Washington, D.C. to reiterate the importance of the United States honoring its treaties,\textsuperscript{42} occupied the Bureau of Indian Affairs (“BIA”) building in Washington, D.C. to express dissatisfaction with federal regulation of Indian affairs,\textsuperscript{43} and protested during the standoff at Wounded Knee.\textsuperscript{44}

American Indians also turned to legislative activities in the 1960s, but these focused primarily on a highly visible campaign for the U.S. government to abandon its policy of terminating tribal governments and assimilating American Indians.\textsuperscript{45} Congress remained committed to the termination policy throughout the 1960s.\textsuperscript{46} The Indian lobby, however, remained weak as Indian nations had limited financial resources and little experience in federal policymaking.\textsuperscript{47} Historically, tribes had approached Congress to resolve tribe-specific issues, but significant changes were on the horizon. In the

\textsuperscript{36} See Nagel, supra note 31, at 126–27 (describing how an influx of funding from federal programs “sparked a dramatic growth in urban Indian social, economic, and political programs”).

\textsuperscript{37} See Cornell, supra note 27, at 189–90.


\textsuperscript{39} Id.; Cornell, supra note 27, at 189.

\textsuperscript{40} See Nagel, supra note 31, at 130.

\textsuperscript{41} See id. at 131–35.

\textsuperscript{42} See id. at 135–37.

\textsuperscript{43} See id. at 131–37. For a fuller discussion of the emergence of the red power movement, see id. at 158–86.

\textsuperscript{44} See Cornell, supra note 27, at 113–15.

\textsuperscript{45} Some tribes engaged in aggressive lobbying both historically, see Carpenter, supra note 7, and at this time to achieve specific tribal goals. Two examples of successful lobbying efforts in the late 1960s and early 1970s include the almost seventy-year struggle of the Pueblo of Taos to restore sacred lands around Blue Lake, Castile, supra note 38, at 93, 100–03, and the Menominee campaign to reverse their termination, id. at 148–52. Castile suggests that these tribal lobbying efforts encouraged other tribes to pursue their claims against the federal government legislatively. Id. at 100–03.

\textsuperscript{46} See Castile, supra note 38, at 21.

\textsuperscript{47} The BIA had denied tribes the ability to use their own trust funds to finance lobbying visits to Washington well into the late 1940s. See Castile, supra note 38, at xxv.
1950s and 1960s, with the resurgence in Indian identity and the development of pan-tribal organizations, tribes started to combine their advocacy efforts and shift their focus to national level issues. As a result, pan-tribal organizations, led by the NCAI, started taking American Indian concerns directly to Congress in the 1950s and 1960s.

Litigation served as another advocacy strategy used by American Indians throughout the twentieth century. Even though they were not always successful, Indian nations had long resorted to courts as a check on detrimental U.S. policies. In the 1950s, Indian nations litigated and won several influential cases, which acknowledged their land claims along the East Coast, recognized their civil jurisdiction, and validated their fishing and hunting rights. These wins encouraged American Indians to bring more cases to protect their self-determination, lands, and water rights. As a result, American Indians largely perceived the courts to be friendly to their interests in the 1960s, 1970s, and 1980s. And they were right. The Supreme Court handed down a few worrisome decisions in the late 1970s and early 1980s, but found in favor of American Indian litigants almost 60 percent of the time from the late 1950s to the late 1980s.

While American Indians have used various advocacy strategies over time, few studies have tried to document American Indian advocacy systematically. Most political scientists routinely omit American Indians from...
studies of American politics entirely. At least one scholar tried to document changes in collective action by American Indians over time but did not systematically consider legislative strategies largely because Indian nations did not yet have the resources to pursue them.

The few other studies on American Indians do not represent the full range of advocacy strategies utilized by American Indians. These studies have focused on more limited subjects including voting rights, Indian protest movements after World War II, campaign contributions made by Indian nations involved in gaming enterprises, and Indian engagement in state and local politics.

Of these, a few studies have examined legislative advocacy by American Indians. These studies often frame their research within the theoretical frameworks used by interest group scholars. They expect, and often find, that American Indians act like organized interests. Most of these studies
have documented the increased use of interest group strategies, especially lobbying and campaign contributions, by Indian tribes at the federal, state, and local levels.67

Professors Witmer and Boehmke conducted the most comprehensive study to date on legislative advocacy by Indians at the federal level. They reported an increase in federal lobbying expenditures made by American Indians from 1997 to 2000.68 Witmer and Boehmke argued that resource constraints limited American Indians’ potential for legislative advocacy historically and that gaming has altered that constraint since the late 1980s.69 They concluded that gaming has provided opportunities in terms of resources for tribes to increase their lobbying activities.70 The insights from the study, however, are limited because of its narrow focus on the influence of one factor (gaming), and its short time frame (only three years). Due to the short time frame of their analysis, they do not really compare American Indian legislative advocacy before and after the rise of gaming and thus cannot assess fully how gaming may have altered lobbying.71

To my knowledge, only one other study has considered American Indian lobbying activities on the federal level.72 A recent study investigated lobbying strategies employed by Indian groups seeking federal recognition from 1977 to 2012.73 It demonstrated that the political context helped to shape Indian groups’ advocacy strategies over time.74 While it analyzed a longer time period than Witmer and Boehmke, the study was more limited in scope in that it only analyzed lobbying by non-federally recognized Indian groups. As a result, the findings may be limited only to that subset of Indian groups.

Similarly, scholars have yet to systematically study the use of litigation strategies by American Indians. Several scholars have tried to determine how successfully tribes have litigated before the U.S. Supreme Court over time,75 but I have yet to identify a study that looks at how frequently tribes litigate in federal, tribal, or state courts.

68 Witmer & Boehmke, supra note 65, at 134–35.
69 Id. at 130–31.
70 Id. at 140.
71 Id. at 130–31. They include a short description of American Indian legislative advocacy prior to the Indian Gaming Regulatory Act, but they do not include any quantitative data on reported lobbying or contributions by American Indians before 1988. Id.
72 See generally Carlson, supra note 67 (investigating the advocacy strategies of 124 Indian groups seeking federal recognition from 1977 to 2012).
73 See generally Bethany Berger, Hope for Indian Tribes in the U.S. Supreme Court? Memoriee, Nebraska v. Parker, Bryant, Dollar General, . . . and Beyond, 2017 ILL. L. REV. 1901 (2017); Fletcher, supra note 57; Getches, supra note 56; Alexander Tallchief Skibine, The
Lobbying Against the Odds

The dearth of studies on Indian advocacy indicates an acute need for more research on Indian advocacy strategies and how they have developed over time. It motivated the study on reported lobbying by American Indians described in the next section.

B. The Study

This study is a first step towards developing more comprehensive knowledge about American Indian advocacy strategies. It seeks to describe reported lobbying by American Indians from 1978 to 2012 and to start to fill some of the gaps in the existing literature on American Indian advocacy.

1. Data Collection

I accessed primary data on lobbying by American Indians from several sources, including serial publications and lobbying disclosure reports, publicly available and archival materials on legislation introduced in Congress from 1975 to 2012, and interviews with lobbyists for American Indian nations and organizations. I collected data on reported lobbying by American Indians from serial publications and lobbying disclosure reports. I wanted to collect data starting in 1975 so that it would coincide with the adoption of the Tribal Self-Determination Policy by Congress and with a related database on Indian-related federal legislation. I chose to use 1978 as the baseline year instead, after reviewing several serial publications and determining that it was the first year for which I could find consistent and reliable data.

My use of serial publications is consistent with other interest group studies. See generally Jeffrey M. Berry, Lobbying for the People: The Political Behavior of Public Interest Groups (1977); Kay Lehman Schlozman & John T. Tierney, Organized Interests and American Democracy (1986). I identified tribes and Indian organizations with lobbyists by searching the subject matter index. For the years 1978-1992, I used the subject matter category “minorities” to identify Indian tribes and organizations (none of the other subject matter indices seemed relevant). For 1993-1996, I used both the “minorities” and “Native American” subject matter categories. Most of the tribes and Indian organizations were identifiable by name, but occasionally, I looked up an entity online. If I could not tell if an organization was made up of American Indians and advocated for their benefit, I did not include it on the list. I excluded charitable organizations that appeared to be non-Indian friends of the Indians, such as the Bureau of Catholic Indian Missions.
tions on which they lobbied.78 For the years 1997 to 2012, data was collected from the Open Secrets website run by the Center for Responsive Politics.79

To facilitate comparative analysis over time, data was collected on the name of the organization or tribe reporting lobbying; the years and congressional sessions in which each reported lobbying; the issue areas in which they lobbied; and the lobbyists, if any hired by the organization or tribe; and the years it used a specific lobbyist.80 Organizations had to start reporting lobbying expenditures in 1997, and data was collected on the amount of money each organization reported spending on lobbying by year and congressional session beginning in that year. Additional data was collected from publically available sources on the geographical location of each Indian tribe that reported lobbying during the time period studied and whether the tribe engaged in gaming during the years they reported lobbying.81

2. Methodology

This study investigates the use of lobbying strategies by American Indian nations and organizations from 1978 to 2012. It is the first to use empirical methods to look systematically at reported lobbying and lobbying expenditures by American Indians over a thirty-five year time period. This time period allows for the identification and comparison of trends in reported lobbying and lobbying expenditures over time.

I use quantitative methods to describe and compare reported lobbying and reported lobbying expenditures by American Indian organizations and tribes from 1978 to 2012. Drawing on the few existing studies of Indian legislative activity and the general trend towards increased lobbying by groups over time, I expected to find an increase in reported lobbying and reported lobbying expenditures by Indians over time. I investigated this expectation by looking at the frequencies of reported lobbying and reported lobbying expenditures by American Indians nation and organizations and comparing them over time. I collected and analyzed additional data on gaming, the issues Indians lobbied on, and geography to provide context and a broader understanding of trends in Indian lobbying behavior. I supplemented this quantitative analysis with archival research and interviews with lobby-
ists both to substantiate my findings and to determine qualitative changes to lobbying behaviors.

3. Limits to the Study

Several limits exist to this study. First, the data only represents American Indian organizations and tribes that reported lobbying from 1978 to 2012 and reported lobbying expenditures from 1997 to 2012. They may underrepresent lobbying and lobbying expenditures by American Indian organizations and tribes for several reasons. First, the Lobbying Disclosure Act only requires the reporting of direct lobbying not indirect or grassroots lobbying. Second, some organizations and tribes may not have reported lobbying prior to 1995.82 Third, after 1995, some organizations and tribes may not have conducted enough lobbying to report their activities under the Lobbying Disclosure Act of 1995 or may have chosen not to do so.83 Fourth, after 1997, some organizations and tribes may not have spent enough money on lobbying expenditures to report their activities. The underreporting of lobbying and lobbying expenditures raises concerns that the data may underrepresent actual lobbying by American Indian nations and organizations.

The benefit of the data presented here (even if imperfect) is that it presents a much broader and more systematic view of reported lobbying and lobbying expenditures by American Indians over time. The study spans thirty-five years—a much longer time period than previous studies of lobbying by American Indians at the federal level—and thus, allows for the identification of trends and changes in lobbying behaviors over time. Moreover, it includes the collection and analysis of data on reported lobbying and lobbying expenditures by American Indians both before and after the rise of Indian gaming. This allows for important comparative analysis over time, missing from earlier studies.84

A second limit to this study is that it does not attempt to evaluate the effects of American Indian lobbying. While that question is important and merits close investigation, this study has a more limited purpose. It seeks to investigate and explain changes in lobbying trends over time.

82 Several scholars have noted possible inaccuracies in data reported by lobbyists. See, e.g., WILLIAM N. ESKRIDGE, ET AL., LEGISLATION AND STATUTORY INTERPRETATION 198–200 (2d ed. 2006); Timothy M. LaPira, Lobbying in the Shadows: How Private Interests Hide from Public Scrutiny and Why That Matters, in INTEREST GROUP POLITICS 224, 238–40 (9th ed. Allan J. Cigler, et al. eds., 2016). Congress first started requiring lobbyists to register in the Federal Regulation of Lobbying Act of 1946, but many interest groups may have felt that they did not have to register after the Supreme Court seriously limited the Act’s application in 1954 in United States v. Harriss, 347 U.S. 612, 619 (1954). ESKRIDGE, ET AL., supra, at 201. Lobbying registrations increased after enactment of the Lobbying Disclosure Act of 1995. Id. But they have recently dropped. LaPira, supra, at 238–40. For a discussion of the limits of data collected under the Lobbying Registration Act, see LaPira, supra, at 233–39.

83 LaPira, supra note 82, at 224–26, 238–39 (noting some of the reasons why lobbyists do not report their lobbying activities).

84 Witmer & Boehmke, supra note 65, at 132.
Third, this article evaluates strategic decisionmaking on a macro rather than a microlevel. It describes trends in lobbying over time but has not attempted to disaggregate or analyze the data on an individual, tribal level. Thus, it cannot explain any particular tribe or Indian organization’s decision to lobby on a specific issue at a given moment in time. Similarly, the analysis cannot directly compare strategy choices or why an Indian tribe or organization decided to lobby rather than, or in addition to, litigating any particular issue.\textsuperscript{85} The article attempts to situate the increase in lobbying by American Indians generally with trends in litigation by American Indians but acknowledges the limits of doing so given the lack of data on how frequently American Indian tribes and organizations engage in litigation.\textsuperscript{86}

Another limit to the study is that it only measures changes in legislative advocacy from 1978 onward. Due to the difficulty of collecting data, I can only compare lobbying starting in 1978 and lobbying expenditures starting in 1997. This date is somewhat arbitrary because historical studies and accounts document Indian legislative advocacy from contact. These historical studies suggest that my data are part of a larger, ongoing narrative of Indian advocacy.\textsuperscript{87} Unfortunately, I have yet to find a way to measure systematically such advocacy efforts.

III. CHANGING STRATEGIES: THE RISE OF AMERICAN INDIAN

American Indian advocacy started to change dramatically in the 1970s as American Indians increasingly engaged in legislative strategies. This shift continued over the next thirty-five years with American Indian legislative advocacy growing significantly from 1978 to 2012.

Six times as many American Indian organized interests reported lobbying in 2012 as in 1978.\textsuperscript{88} Figure 1 depicts the frequency of reported lobbying

\textsuperscript{85} For examples of this kind of microlevel analysis of group decisionmaking, see, e.g., Carlson, supra note 72; Cummings & NeJaime, supra note 11.
\textsuperscript{86} See supra Part II.
\textsuperscript{87} See generally Hoxie, supra note 3; Hermann J. Viola, Diplomats in Buckskins: A History of Indian Delegation in Washington City (1995); The New Indian Politics, supra note 28.
\textsuperscript{88} Thirty-six American Indian organizations, including tribes, reported lobbying in 1978 and 242 in 2012, an increase of over 600 percent (36/242). The numbers are even more startling for tribes: 24 tribes reported lobbying in 1978 and 177 in 2012. This is an over 700 percent (24/177) increase in the number of tribes lobbying (and 2012 does not even represent the highpoint for tribes reporting lobbying over this time period). While lobbying rates have generally increased over the past five decades, it is hard to compare the increase in American Indian lobbying with other groups due to differences in variables and measurements. Leech et al. included Indian affairs as an issue area, but they only reported the average number of interest groups per issue area and the average number of lobbyists per issue area over a four-year period (1996-2000), and they did not break down the numbers by year. Leech, et al., supra note 11, at 24–25. According to their data, very few interest groups lobbied on Indian affairs in comparison with other issue areas. \textit{id.} at 24. They did not, however, explain whether they included Indian tribes as interest groups in their data set. Moreover, they reported that the
over time by American Indian interest organizations, including Indian nations, tribal consortiums, American Indian non-profit organizations, Alaska Native for- and non-profits under the Alaska Native Claims Settlement Act, Alaska Native villages, and Native Hawaiian groups. Political scientists have documented a similar rise in lobbying, but these numbers exceed the increases in reported lobbying found by political scientists studying the general population. For example, in 2004, Professors Loomis and Schiller noted that “registered lobbyists rose in numbers from . . . almost 6000 in 1981 to an astonishing 20,000 in 2002”—an increase of over 300 percent. 89 Recent studies, however, have documented a decline in lobbying after 2007. 90

![Graph showing lobbying over time from 1978 to 2012.](image)

Figure 1. American Indian organizations, including Indian nations, reporting lobbying over time, 1978-2012.

Of the American Indian interest organizations reporting lobbying during this time, Indian nations filed the vast majority, 71 percent, of the reports. 91 Figure 2 displays reported lobbying by Indian nations over time. Over half—a total of 325 Indian nations or 57.4 percent 92 of the 566 feder-

---

89 LOOMIS & SCHILLER, supra note 11, at 38. Loomis and Schiller reported an even greater 400 percent increase in the number of Washington-based interest representatives (from 4,000 in 1977 to about 17,000 in 2002). Id. They also noted that the rise in interest group advocacy has not been spread evenly across groups. Id.

90 See, e.g., LaPira, supra note 82, at 240 fig.1.

91 I generated this number by dividing the number of tribes filing reports by the total number of reports filed by American Indian organizations from 1978 to 2012. American Indian organizations filed a total of 5,111 annual reports from 1978 to 2012, but some of these organizations filed more than one report. Indian nations filed 3,646 of these reports.

92 I generated this percentage by dividing 325 by 566. It does not accurately represent the actual percentage because the number of tribes increased rather than remained constant over the time period studied. As a result, the actual percentage may be lower.
ally recognized tribes in the United States at the time—reported lobbying at least once during this time period. While the overall trend for Indian nations indicates a 700 percent increase in the number of tribes lobbying, it has varied some over time.\textsuperscript{93} Reported lobbying by tribes rose dramatically in the late 1970s, leveled off in the early 1980s, and then grew significantly from 1984 to 1986 before dipping slightly in the late 1980s. From 1990 to 1995, the number of tribes reporting lobbying steadily increased again. It declined in the late 1990s, surged in the early 2000s, decreased with the economic recession of 2008, and appears to be rebounding in the past few years.\textsuperscript{94}

![Graph showing Indian nations reporting lobbying over time, 1978-2012.](image)

Not all Indian nations reported lobbying during this time period. In fact, 43.7 percent of federally recognized Indian nations did not report lobbying at all during the time period studied.

The tribes that did report lobbying varied in the frequency of their lobbying. Some tribes reported lobbying almost every year, such as the Makah Indian Tribe and Quinault Indian Nation, while others reported lobbying only once or twice during the time period studied.\textsuperscript{95} On average, tribes re-

\textsuperscript{93} I generated this number by dividing the number of tribes reporting lobbying in 1978 by the number reporting lobbying in 2012. Twenty-four tribes reported lobbying in 1978 and 177 in 2012.

\textsuperscript{94} I have not tried to explain these variations in the lobbying pattern. Several factors could influence fluctuations in individual tribal decisions to lobby, including the growth of gaming in the early 1990s and changes in party alignments in Congress and the Executive Branch. In contrast, reported lobbying in general has been decreasing since 2007. LaPira, supra note 82, at 240 fig.1.

\textsuperscript{95} Other Indian nations that report lobbying frequently include, but are not limited to, the Standing Rock Sioux Tribe, the Tulalip Tribes of Washington, the Three Affiliated Tribes of the Fort Berthold Reservation, the Navajo Nation, the Miccosukee Tribe of Indians of Florida, the Menominee Tribe of Wisconsin, the Lummi Indian Nation, and the Hoopa Valley Tribe. It is
ported lobbying during 11 of the 34 years in the data set. A quarter of tribes, however, reported lobbying during five years or less, and another quarter reported lobbying for 18 years or more.96

As Figure 3 shows, very few of the tribes that reported lobbying were engaged in gaming prior to 1991. The number of Indian nations engaged in gaming and reporting lobbying increased dramatically over the time period studied as more tribes opened gaming operations in the mid-1990s. Indian nations engaged in gaming exceeded non-gaming Indian nations in reporting lobbying for the first time in 1995—the same year that most of the tribes in the dataset started gaming operations. Gaming tribes have outnumbered non-gaming tribes in reported lobbying every year since then. Over the entire time period studied, the majority—60 percent—of the Indian nations that reported lobbying operated gaming establishments at the time they reported lobbying.97

![Figure 3. Reported lobbying by gaming and non-gaming tribes over time, 1978-2012.](image)

Both gaming and non-gaming Indian nations reported more lobbying over time. As Figure 3 shows, reported lobbying by gaming and non-gaming

---

96 In terms of congressional sessions, on average, tribes reported lobbying six congressional sessions. A quarter of tribes reported lobbying three or fewer congressional sessions, and a quarter reported lobbying nine or more congressional sessions.

97 The median year that tribes in the dataset started gaming was 1995. A tribe was coded as operating a gaming establishment if it engaged in gaming the year(s) in which it reported lobbying. To code tribes as gaming or non-gaming, data was collected from public sources (e.g., Tiller’s Guide, tribal websites, etc.). Tribes were coded as gaming for years in which they operated a gaming establishment and as non-gaming for years in which they did not.
tribes has followed the same trend line—lobbying by both dipped in the late 1990s and then rebounded. The rate in reported lobbying for gaming tribes, however, has increased more steadily and by a larger amount over time than the rate increase for non-gaming tribes.98

The amount of money American Indian organizations reported spending on lobbying has fluctuated more than reported lobbying but overall supports a trend toward increased legislative advocacy during the past thirty-five years. Figure 4 reports the amount of money spent on lobbying by American Indian organizations over time, starting in 1997 (the first year in which the amount spent is available).99 In nine of the fifteen years for which data is available, reported lobbying expenditures by American Indian organizations increased.

Unlike reported lobbying, the money spent by American Indian organizations did not consistently increase over time. In this respect, reported lobbying expenditures by American Indians do not mirror trends in reported lobbying expenditures in the U.S. population generally. Recent studies show that reported lobbying expenditures increased from 1998 to 2010.100 American Indian organizations reported spending a particularly high amount in 1999, but then spending leveled off at a much lower level until 2004 when it decreased dramatically.101 The most significant decrease, however, occurred in 2005. Since 2010, reported spending has increased inconsistently, but appears to be on the rise generally. Most likely the decrease in reported lobbying expenditures from 2005 to 2010 responded to the Abramoff lobbying scandal that broke in 2006 and revealed that lobbyists had swindled Indian tribes out of millions of dollars.102 It may also have reflected the economic downturn. The inconsistencies in reported lobbying expenditures by American Indians could also indicate that money is a less consistent resource for tribes, that effective lobbying costs less, or that tribes rely more on lobbying by tribal members or in-house counsel.

98 Reported lobbying by non-gaming tribes has always been more variable than for gaming tribes. This may reflect the fewer resources that non-gaming tribes have to spend on lobbying and choices that they make in prioritizing what issues to lobby on and when.

99 Like the lobbying data, these numbers only include reported spending. They may underrepresent the amount spent on lobbying by American Indian organizations as some may choose to lobby on their own behalf or not conduct enough lobbying to report such activities under the Lobbying Disclosure Act of 1995.

100 LaPira, supra note 82, at 240 fig.1 (noting that reported lobbying expenditures have increased in general from 1998 to 2010 and then decreased from 2010 to 2013).

101 I have not tried to investigate the reasons for this dip in lobbying after 1999. This dip is not reflected in data on reported lobbying among the general population. Reported lobbying continued to increase during the late 1990s. LaPira, supra note 82, at 240 fig.1.

102 See Wilkins & Stark, supra note 8, at 168–69.
American Indian organizations reported spending on average $94,411 a year with a median of $40,000.\textsuperscript{103} A quarter of all organizations did not report spending any money on lobbying.\textsuperscript{104} Another quarter reported spending over $112,000 with the top spender (an Indian tribe) reporting spending $3,205,000 in a single year.

Indian nations reported spending significant amounts of money on lobbying and more than American Indian organizations in general. On average, Indian nations spent $109,408 on lobbying per year, but this number conceals significant variation among tribes. A quarter of all tribes did not report spending any money on lobbying. Half of all tribes reported spending between $40,000 and $120,000. Another quarter reported spending $120,000 or more. The Mississippi Band of Choctaw Indians topped the list of high spenders and reported spending over $3 million in 1999. While a few tribes annually reported spending over $200,000 a year on lobbying,\textsuperscript{105} most did

\textsuperscript{103} While the mean is $94,411, there is a large standard deviation (201.644). Here, American Indian organizations include Indian nations.

\textsuperscript{104} Tribal consortiums reported the lowest levels of spending of the American Indian organizations.

\textsuperscript{105} Tribes consistently reporting spending over $200,000 annually on lobbying include the Agua Caliente Band of Cahuilla Indians, the Jicarilla Apache Nation, the Gila River Indian Community, the Mashantucket Pequot Tribal Nation, the Mississippi Band of Choctaw Indians, the Oneida Indian Nation, the Pechanga Band of Luiseno Mission Indians, the Seminole Tribe of Florida, the Tunica-Biloxi Tribe of Louisiana, and the Viejas Band of Kumeyaay Indians.
not consistently report spending high amounts of money on lobbying but tended to increase spending in a particular year. For example, the Osage Nation reported spending over $2 million dollars in 2011 but returned to its much lower regular spending level (under $100,000) in 2012. Such variations in spending most likely reflect changing tribal priorities.

Indian nations engaged in gaming reported spending more on lobbying than non-gaming tribes. On average, gaming tribes reported spending twice as much—almost $135,000—on lobbying per year as Indian nations not engaged in gaming, which reported spending only $51,500 on lobbying per year.

Consistent with earlier studies on lobbying by American Indians on the state and federal levels, the evidence on reported lobbying and lobbying expenditures shows that American Indians increased their lobbying in the 1970s even though they had not overcome the constraints the prevailing narrative suggests should have prevented them from lobbying. Anecdotal evidence from secondary sources, archival research, and interviews with Indian law advocates confirm these findings. In addition to reporting lobbying and expenditures, Indian tribes started opening offices in Washington, D.C. to support their lobbying efforts in the 1980s. Tribal organizations, including the National Tribal Chairman’s Association and the NCAI, launched a major initiative to defeat anti-Indian legislation in the late 1970s and have continued to closely monitor legislation ever since. Indian tribes have also testified extensively before Congress since the early 1970s.

American Indians had neither gained in electoral significance nor become less impoverished when they increased their lobbying efforts in the

106 ANOVA tests indicate that this difference in spending between gaming and non-gaming tribes is significant at the 0.00 level.
107 The median amount of money reported as spent on lobbying was significantly less for both gaming and non-gaming tribes. The median amount reported as spent by gaming tribes was $60,000 per year and $20,000 per year for non-gaming tribes.
108 See supra Part II.
109 See supra fig.2.
110 I conducted research on Indian-related bills with files retained by the Senate Committee on Indian Affairs and the House Committee on Interior and Insular Affairs at the Center for Legislative Archives. I also reviewed files on or related to legislation in the National Congress of American Indians archives (1933-1990) at the Smithsonian National Museum of the American Indian Archives Center.
1970s. In fact, American Indians remain largely electorally insignificant to this day.\footnote{113} While gaming eventually introduced new financial resources that tribes could use on lobbying, the data demonstrate that most tribes did not start gaming operations until 1995—almost two decades after American Indians initially increased their lobbying efforts. Contrary to the view expressed in popular culture, the media, and some scholarship,\footnote{114} Indian gaming could not have caused the dramatic increase in the lobbying of federal officials by American Indians over the past thirty-five years. Rather, the data show that American Indians have been gaining momentum as lobbyists over time. Gaming appears to have encouraged and reinforced the use of legislative strategies by American Indians at the end of the twentieth century, but it was not the primary or only influence leading to it and lobbying expenditures have not consistently increased over time with the advent of gaming. Moreover, American Indians chose to engage in legislative advocacy even as their claims continued to be recognized and validated by the Supreme Court.\footnote{115} American Indians turned to lobbying even when the odds

\footnote{113} American Indians rarely affect electoral outcomes because they do not comprise a majority population in any state, TINA NORRIS, PAULA L. VINES, & ELIZABETH M. HOEFFEL, U.S. CENSUS BUREAU, THE AMERICAN INDIAN AND ALASKA NATIVE POPULATION: 2010 6–7 (Jan. 2012) (noting that California has the highest percentage of Indians living in any state at 14 percent), and constitute a majority in only one or two congressional districts. See McCOOL ET AL., supra note 61, at 176–91 (noting the few instances in which Native voters may have made a difference in elections in Western swing states); TOVA WANG, DEMOS, ENSURING ACCESS TO THE BALLOT FOR AMERICAN INDIANS & NATIVES: NEW SOLUTIONS TO STRENGTHEN AMERICAN DEMOCRACY 3 (2012) http://www.demos.org/sites/default/files/publications/IHS%20Report-Demos.pdf [http://perma.cc/SB2N-V2CJ] (reporting low voter turnout rates among American Indians). Indian issues rarely, if ever, decide congressional elections. See McCOOL ET AL., supra note 61, at 176–91. The issues that American Indians care about are not highly salient to the general public and are not among the issues upon which most non-Indian constituents base their voting decisions. See, e.g., Owen G. Abbe et al., Agenda Setting in Congressional Elections: The Impact of Issues and Campaigns on Voting Behavior, 56 Pol. Res. Q. 419, 422 (2003) (identifying education, social security, health care, and the economy as the most important issues in the 1998 House Elections). Public opinion polls also routinely find that Americans do not rank Indian issues as important. See, e.g., Most Important Problem, GALLUP, http://www.gallup.com/poll/1675/most-important-problem.aspx [http://perma.cc/JFMG-MUKB] (last visited Nov. 4, 2018). To the extent that non-Indians care about Indian issues, it may jeopardize a member of Congress's reelection due to an increased backlash movement against Indians in recent decades. See WILKINS & STARK, supra note 8, at xxx–xxxi, 169. Thus, as a collection of election-minded politicians (rather than conscientious lawmakers), members of Congress have few incentives to pay attention to Indian issues. See CHARLES C. TURNER, THE POLITICS OF MINOR CONCERNS: AMERICAN INDIAN POLICY AND CONGRESSIONAL DYNAMICS 130–33 (2005). More often than not, politicians may gain political support from non-Indians by disfavoring Indian interests. See, e.g., CASTLE, supra note 38, at 166–68 (describing how Senator Henry Jackson (D-Wash.) and Congressman Lloyd Meeds (D-Wash.) waived their support for tribal self-determination due to fishing rights controversies in Washington state). Financially, most Indian nations and individuals have historically been impoverished and unable to contribute to electoral campaigns. See Boehmke & Witmer, supra note 63, at 179, 181 (stating that prior to gaming, tribes previously did not have financial resources to participate politically).

\footnote{114} See, e.g., House of Cards, Chapter 21 (Netflix 2014). For a discussion of the scholarly literature on gaming and lobbying see infra Part IV.A.

\footnote{115} See supra Part II.A.
seemed to be against them. This reality suggests that the prevailing narrative does not fully explain the rise in American Indian advocacy. The next Part reviews the interest group and sociolegal literatures in search of a more accurate explanation of how groups craft advocacy strategies.

IV. INTEREST GROUP AND SOCIOLEGAL APPROACHES TO UNDERSTANDING ADVOCACY

The dramatic rise in American Indian legislative advocacy raises an important question: What compels groups to engage in some advocacy strategies but not others? In this Part, I evaluate the interest group and sociolegal approaches to understanding why groups pursue different advocacy strategies.

A. Interest Group Approach

Interest group scholars study lobbying.116 Their primary focus has been on why groups lobby legislatures.117 They have paid significantly less attention to the lobbying of other institutions and have only recently begun to consider how venue choice affects lobbying decisions.118 Groups lobby legislatures to attain a particular policy goal,119 but the tactics and strategies available to them are constrained by external conditions, including their political

116 See e.g., ANTHONY J. NOWNES, TOTAL LOBBYING: WHAT LOBBYISTS WANT (AND HOW THEY TRY TO GET IT) 5 (2006).
2019] Lobbying Against the Odds

capacity and resources, the political context, and active opposition by other advocates or policymakers.120

Interest group scholars have defined different aspects of these external conditions. Political capacity refers to the characteristics of the group, including its access to policymakers,121 its electoral influence,122 and the kind of claims it makes.123 Resources include the number of members in a group, the degree to which the public supports the initiative, the group’s financial resources, the number of other organizations allying with the group, and the number of staff (including in-house lobbyists).124 The political context includes issue salience, party control of government, and support of political elites but rarely institutional alternatives to lobbying a legislature.125 Opposition means active opposition by others, such as blocking legislative action or mobilizing a countermovement. It may come from other organized interests, administration officials, members of important committees in Congress, other members of Congress, or unorganized individuals.126

Interest group studies have sought to determine the relationships among these variables and lobbying activities. A few studies have determined that some of these factors influence lobbyists’ institutional choices.127 Groups with political capacity, including electoral influence and access to policy-

---

120 See BAUMGARTNER, ET AL., supra note 117, at 110–11.
121 Holyoake, supra note 118, at 333–34.
123 See ESKRIDGE ET AL., supra note 82, at 62. Political scientists have found that groups seeking to challenge the status quo are more likely to mobilize than groups seeking to defend the status quo. BAUMGARTNER ET AL., supra note 117, at 164 (“In general, supporters of the status quo need to do much less than do status quo challengers.”). Status quo defenders may adopt a watchful approach and choose not to mobilize unless they feel they have to. See id. at 57–58, 152–53. They may only mobilize if the issue gains attention and merits a response. In contrast, status quo challengers may have to use legislative advocacy to increase attention to the issue to get policymakers to act on it. See id. at 147, 164.
124 BAUMGARTNER ET AL., supra note 117, at 194.
125 McQuide, supra note 118, at 12–13 (measuring the political environment by looking at issue salience, party in control of government, presidential party, issue conflict, stage of the President’s term, and presidential agenda priority); Baumgarter, et al., supra note 11, at 13 (measuring government activities in terms of congressional hearings, federal spending, and presidential attention to an issue).
126 Scott L. Cummings, Litigation at Work: Defending Day Labor in Los Angeles, 28 UCLA L. REV. 1617, 1624 (2011). Some political scientists have defined opposition more broadly to consider other obstacles in the policy making process, including a lack of interest, effort, support, concern or attention from policy makers. See BAUMGARTNER ET AL., supra note 117, at 78. They have considered the influence of active opposition on policymaking and find that it is not a predictor of policy success or failure. See id. at 76. Some scholars have considered active opposition in their studies of American Indian politics. In their study of contemporary challenges to indigenous nationhood, Corntassel and Witmer identify opposition or backlash, and particularly backlash based on misperceptions of American Indian nations and their people as one of, if not the, greatest threat to Indian nations today. CORNTASSEL & WITMER, supra note 64, at 3–6, 24–28. Their study focused on state-tribal interactions and indicated that these misperceptions have motivated state officials to oppose tribal initiatives and that Indian nations have responded by mobilizing politically on the state level. Id. at 6.
127 See generally Holyoake, supra note 118, at 325; McKay, supra note 118, at 123; McQuide, supra note 118, at 5–6.
makers, are more likely to engage in legislative advocacy. Similarly, political scientists have found that higher levels of resources, including personnel, allies, and money, increase the likelihood that a group will lobby a specific institution or across multiple institutions.

Studies have shown that a positive external political context fosters lobbying. Lobbying may increase during divided partisan government as it provides more opportunities to lobby against proposed policy changes. The support of political elites also encourages lobbying activities. Several forms of political support motivate groups to lobby, including congressional hearings, federal spending, and presidential attention to an issue. In response to increased levels of federal activities, affected interests lobby to fight off new federal incursions, encourage the activity, or attempt to modify the proposals before they are complete.

Scholars have found opposition to both encourage and discourage lobbying. While some suggest that opposition may deter political mobilization, at least one study argues that “[c]onflict seems to propel groups into political activity.” Other scholars have noted that by challenging the status quo, groups encourage mobilization by opponents. This oppositional organization, however, may help draw attention to the issue, and thus, benefit the

---

128 See Holyoake, supra note 118, at 333–34. Political scientists have consistently found that groups lobby their allies or legislators that are already friendly to their causes. See, e.g., Ken Kollman, Inviting Friends to Lobby: Interest Groups, Ideological Bias, and Congressional Committees, 41 Am. Pol. Sci. Rev. 519, 519 (1997). Access may be linked to financial resources as studies have shown that money buys access to decision makers but not policy outcomes. See Baumgartner, et al., supra note 117, at 193–94 (summarizing the literature on the relationship between money and policymaking). Groups, however, may be able to overcome the financial obstacles to access by building expertise and relationships with policymakers over time. See Evans, supra note 64, at 74, 88–97. Thus, they gain access because policymakers identify them as informational resources and turn to them for advice.

129 See, e.g., Holyoake, supra note 118, at 333–34 (finding that membership in a coalition increased the likelihood of lobbying in a venue); McKay, supra note 118, at 135 (explaining that “the more money an organization spends on overall lobbying, the more likely the lobbyist is to direct his or her efforts toward Congress or both branches over agencies.”); McQuide, supra note 118, at 14 (reporting that level of resources affected lobbying behavior with groups having their own in-house lobbyist more likely to lobby beyond Congress).

130 See McQuide, supra note 118, at 16 (finding that groups take partisanship into account in determining which institutions to lobby).

131 Legal mobilization scholars have also noted the influence of elite support in advocacy decision making. See Cummings & NeJaime, supra note 11, at 1257-67 (explaining how the support of Gov. Davis influenced marriage equality strategies). In her assessment of the Tribal Self-Determination Policy, Professor Gross notes that Indian advocacy was successful when tribal leaders gained the support of political elites. See Emma R. Gross, Contemporary Federal Policy Toward American Indians, 80, 85–86 (1989).

132 See Baumgartner, et al., supra note 11, at 13. Further, they argue, “Clearly, federal government activities send strong cues to interested constituencies.” Id.

133 Id.


135 Baumgartner et al., supra note 117, at 80.
group challenging the status quo. Another scholar found that the presence of opposition, either alone or in a coalition, increased the likelihood of an organization lobbying in a particular institution. These findings align with the predictions of other scholars, who have suggested that interest organizations mobilize politically when their interests are threatened. At the same time, however, some groups may avoid conflict with opposing groups by choosing to lobby in a different institution.

Interest group scholars have identified many factors affecting groups’ decisions to lobby, but they have historically treated decisions to lobby in isolation instead of considering how institutional dynamics could affect lobbying decisions. Only recently have these scholars started to consider how, when, and why lobbyists lobby particular institutions. Thus, interest group scholars have not explored thoroughly how institutional alternatives may affect a group’s decision to lobby.

Interest group scholars have also more generally overlooked interactions among variables and how those interactions could affect advocates’ decisions to engage in legislative advocacy. Take resources as an example. Groups vary in terms of financial resources, group membership, public support of their cause, allies, and personnel. A group with few financial resources but a large, active membership and a network of other allied groups may have incentives to engage in legislative advocacy. The combination of variables involved as well as how they interact may determine the incentives and disincentives that groups face in deciding whether to engage in a particular strategy. For example, although some causes may not be salient to the public, a group may choose to engage in legislative advocacy because they have the support of political elites. Moreover, changes in one variable may affect others and encourage a group to reformulate its advocacy strategy over time. For instance, Witmer and Boehmke have argued that the increase in resources provided by Indian gaming altered the opportunities tribes had to lobby and contributed to their increased legislative advocacy. While interest group scholars acknowledge the complexity of strategic decision-making and suggest that lobbyists make choices in response to constraints,
they have yet to explore adequately when and how opportunities and constraints interact to influence strategic decisionmaking over time.142

B. Sociolegal Approach

Sociolegal scholars investigate how and why groups use the law as a political and strategic resource to challenge the status quo, change institutional rules, and redistribute power.143 Advocates use the rhetorical resources of the law—legal arguments, frameworks, and practices—to argue for social change in multiple institutions, including but not limited to courts, agencies, and legislatures.144 Most sociolegal studies, however, have focused on courts and litigation activities.

Litigation strategies develop out of an ongoing interplay among advocates and the broader social and political environment. Advocates make choices in “respon[se] to distinct political opportunities and constraints.”145 Constraints may include limited organizational and material resources;146 existing institutional and political barriers;147 the advocate’s own experience, skills, and understandings; countermobilization or opposition; perceptions of efficiency, effectiveness, and the moral acceptability of various institutions;148 and views of the law.149 Sociolegal scholars have demonstrated

142 See Austin Sarat & Stuart A. Scheingold, State Transformation, Globalization, and the Possibilities of Cause Lawyering: An Introduction, in CAUSE LAWYERING, supra note 12, at 13–14 (noting that merely listing factors is not enough to explain strategic decisionmaking).

143 See generally Scheingold, supra note 15; McCann, supra note 18; Joriz, P. Handler, SOCIAL MOVEMENTS AND THE LEGAL SYSTEM: A THEORY OF LAW REFORM AND SOCIAL CHANGE (1978); Cummings & Rhode, supra note 14, at 605.

144 See generally Handler, supra note 143; McCann, supra note 18; Gerald Rosenberg, supra note 15; Gordon Silverstein, LAW’S ALLURE: HOW LAW SHAPES, CONSTRAINS, SAVES, AND KILLS POLITICS (2009); Helena Silverstein, UNLEASHING, RIGHTS: LAW, MEANING, AND THE ANIMAL RIGHTS MOVEMENT (1996).


148 See Silverstein, LAW’S ALLURE, supra note 144, at 131; Wasby, supra note 147, at 66–67.

149 See Silverstein, UNLEASHING RIGHTS supra note 144, at 21. This list is exemplary not exhaustive. Other variables can also shape decisions to litigate, and scholars have produced varying lists of them. See, e.g., Sarat & Scheingold, in CAUSE LAWYERING, supra note 12, at 12–13 (listing the goals of the cause or the movement, the resources that it can make available or that lawyers can mobilize, the possibilities at the practice site, the lawyer’s own experience, skills, and understandings, the lawyer’s social capital and networks, the nature of existing social, political, and legal arrangements); see also Chen & Cummings, supra note 145, at
through case studies how oftentimes these variables interact to facilitate the use of a litigation strategy. Moreover, studies have shown that advocacy decisions are fluid rather than static; advocates frequently adapt their strategies to fit changing social, political, and institutional conditions. Strategic decisionmaking, thus, is an ongoing, reiterative process that evolves over time.

In studying advocacy over time, sociolegal scholars have increasingly examined the interplay between litigation and legislative strategies as well as the linkages between courts and legislatures. But the focus has been on explaining why groups litigate and understanding the effects of political mobilization on the development of litigation strategies. As a result, sociolegal scholars have yet to question and test the conventional wisdom about when groups choose to lobby. Rather, their studies have continued to suggest that groups litigate unless they are foreclosed from doing so and focused on the effects of political mobilization on the success of litigation strategies. They have yet to fully consider how legal mobilization could positively or negatively affect advocates’ use of a legislative strategy. In short, sociolegal scholars have yet to unpack fully the complex relationships between litigation and legislative strategies.

Sociolegal scholars have recently suggested the need for further investigations into the utility of applying sociolegal theories in non-judicial contexts identifying political support, political background, social context, institutional settings, and background as influencing advocacy decisions); Scott L. Cummings, Litigation at Work: Defending Day Labor in Los Angeles, 28 UCLA L. Rev. 1617, 1623 (2011) (listing legal capacity, receptivity of the judiciary, mechanism of legal enforcement, and extent of rights saturation); Cummings & Rhode, supra note 14, at 615 (explaining that “the effective use of litigation requires a strategic analysis of the forces that shape its outcome, including organizational capacity, the likelihood of success on the merits, the challenges of enforcement, and the possible political responses.”).

150 See Cummings, supra note 149, at 1623–24; see generally Handler, supra note 143.
152 See, e.g., McCammon, supra note 151, at 12; see generally Silverstein, Law’s Allure, supra note 144.

153 Some scholars have developed case studies of why groups choose particular strategies, but they have not tried to extrapolate a broader understanding of the factors influencing institutional choices from these cases. See, e.g., Cummings & NeJaime, supra note 11, at 1247–56 (arguing that marriage equality lawyers chose legislative strategies to avoid litigating).

154 See, e.g., Silverstein, Law’s Allure, supra note 144, at 15–41; Cummings, supra note 149, at 1623. Sociolegal scholars have also emphasized the importance of political mobilization to enforce court decisions.

155 Some sociolegal scholars have repeated rather than questioned the conventional wisdom. For example, one sociolegal scholar recently hypothesized that a group will only choose a legislative strategy if it has a high degree of political mobilization and does not have the option of using a litigation strategy. Cummings, supra note 149, at 1625.

156 Carlson, supra note 67, at 935–36; Cummings, supra note 149, at 1623. A few scholars have started to consider the effects of legal mobilization on legislative advocacy. See, e.g., Silverstein, Law’s Allure, supra note 144, at 35–41.
texts and explored some initial ways of doing this.\textsuperscript{157} These studies suggest that the interactive approach used by sociolegal studies may enhance existing understandings of groups’ advocacy choices. But they have yet to extend the sociolegal approach to address fully when and why groups engage in legislative and other non-judicial forms of advocacy and to explore the relational aspects of institutional choices.

V. A GENERALIZABLE APPROACH FOR UNDERSTANDING ADVOCACY CHOICES

This Part combines insights from the interest group and sociolegal literatures to construct a generally applicable framework for understanding how groups choose advocacy strategies and how their strategies evolve across institutions over time. It treats institutional choices as relational and evolving rather than binary and emphasizes that advocates craft strategies by considering their alternatives and options.

In devising this new approach, I start from the premise, shared by many political scientists and sociolegal scholars, that advocates choose strategies in response to social, political, and institutional constraints. Then I borrow the factors found by political scientists to influence lobbying and integrate them into the interactive approach for understanding advocacy choices used by sociolegal scholars. These factors include, but are not limited to, resources, opposition, political capacity, legal capacity, institutional receptivity, and the political context. By including factors relevant to both legislative and judicial processes, my approach allows for a fuller discussion of the factors influencing advocacy choices and for a more nuanced investigation of how advocates consider institutional alternatives in crafting strategies. Once these factors are identified, I, like other sociolegal scholars, consider the interactions among them and how those interactions affect strategic choices. Finally, I highlight the relational aspects of advocacy decisions by considering the interactions among the different strategies that groups face.\textsuperscript{158} Thus, my approach emphasizes the ongoing, dynamic nature of crafting advocacy strategies over time.

My approach improves on the existing frameworks in several ways. First, by integrating the two literatures, it provides for a fuller discussion of the conditions influencing the development of advocacy strategies. Recent political science and sociolegal studies suggest that political capacity, resources, the political context, and opposition influence the development of lobbying strategies.\textsuperscript{159} They have predicted how each category of variables affects the likelihood that advocates will choose a legislative strategy but

\textsuperscript{157} Carlson, supra note 67, at 962.

\textsuperscript{158} See generally Silverstein, Law’s Allure, supra note 144, at 35–41 (considering some of the negative effects of legal mobilization on legislative advocacy); Cummings, supra note 149, at 1623 (noting the potential relationships among political and legal mobilizations).

\textsuperscript{159} See supra Part II.
have yet to consider the dynamics among these different factors. Sociolegal theory expands understandings of strategic decisionmaking by viewing the development of advocacy strategies as ongoing, interactive processes. Adopting this view suggests that groups continually respond to opportunities and constraints and may respond by reshaping their advocacy strategies. But its singular application to questions of why groups litigate overlooks how institutional differences may influence advocates’ choices. Advocates cannot make informed choices about which strategy to use without understanding the different factors that influence the various strategies that they may pursue. Integrating the factors identified by interest group scholars into the sociolegal framework expands it to include information about legislative advocacy strategies and thus, provides for a more comprehensive consideration of the conditions influencing advocacy strategies as they develop across institutions.

Second, my approach allows for deeper investigation into how various conditions interact with one another to influence how and why groups use different strategies over time. The interest group literature has identified many of the factors influencing decisions to lobby but has yet to analyze fully the dynamics among them. The sociolegal approach suggests how to uncover the interactions among variables that shape and reshape advocacy decisions. As sociolegal scholars have demonstrated, contextualized investigations are needed to understand how variables interact in any particular case. Variables may exist in various combinations and have different effects on advocacy choices. Some variables may correlate and have a multiplicative impact. For example, a favorable political context and resources may produce a stronger political capacity and facilitate the use of a legislative strategy. Other combinations of variables could have crosscutting influences on an advocacy strategy. For instance, a favorable political context in terms of a legislative champion may encourage the use of a legislative strategy but a group may be limited in its ability to mobilize if it lacks resources, and thus the group may choose not to lobby. The dynamics of other factors may produce feedback effects, reiterating or undermining legislative advocacy. For example, changes in the political context may alter a group’s

---

160 See supra Part II.
161 See supra Part II.
162 Chen and Cummings note that:

As this work illustrates, careful case studies of law and social change campaigns can provide rich descriptive accounts of the strategic and tactical decisions that lawyers and other movement leaders have to make and whether those decisions prove effective, what barriers arise, which institutional settings are the battlegrounds for the campaign and why, how groups collaborate and work against opponents to reach their desired goals, and what happens in the aftermath of even a successful campaign.

CHEN & CUMMINGS, supra note 145, at 500.

163 For an exploration of how variables act in crosscutting ways in affecting the success of social reform movements, see generally HANDLER, supra note 143.
political capacity, making legislative advocacy an option and encouraging a
group to lobby. Over time, the group’s decision to lobby may create a feed-
back loop in which the group’s increased political capacity enables it to fur-
ther improve the political context by generating allies within the
legislature.\textsuperscript{164} Contextualized investigations will allow scholars to examine
how advocacy decisions evolve through various pathways and will lead to
the identification of those pathways.

Third, my approach provides a more comprehensive view of the
choices advocates make and the constraints they face in crafting strategies
because it includes consideration of institutional differences. Some of the
variables affecting decisions to lobby either differ from those influencing
decisions to litigate or operate differently depending on the institutional con-
text.\textsuperscript{165} While some factors transcend institutional targets and play a role in
advocates’ strategic decisionmaking across institutions,\textsuperscript{166} other factors influence
one institution more than another. For example, electoral concerns may
matter considerably less, if at all, to a court but can play a significant role in
the legislative process.\textsuperscript{167} By considering differences in factors based on in-
stitutional venue, my framework enhances current understandings of why
groups craft specific advocacy strategies and change them over time.

Fourth, this approach emphasizes the relational nature of advocacy de-
cisions. It builds on existing insights by considering how various institu-
tional options relate to one another and how those relationships affect
institutional choices. Advocacy strategies do not develop in isolation but in
the shadow of the other available options.\textsuperscript{168} The most well-known example of
this is the conventional wisdom that groups litigate because they have no
other avenue for redress due to their exclusion from the political process.\textsuperscript{169}
Some scholars, however, have suggested that the inverse may also be true:
Advocates may choose legislative strategies because of judicial risk or hos-

\textsuperscript{164} This situation may produce more of a feedback spiral than a feedback loop because the
advocacy choices were made at different points in time and the advocacy at time one affects
the advocacy decision made at time two (in this particular example, the decision at time one
increases the likelihood of the use of the same strategy at time two).

\textsuperscript{165} For a review of the factors sociolegal scholars have found to influence advocates’ deci-
sions to litigate, see supra note 149.

\textsuperscript{166} For instance, almost all advocates consider their financial resources in crafting a strat-
egy regardless of the institutional target.

\textsuperscript{167} \textit{BAUMGARTNER ET AL., supra} note 117, at 325–26 n.10. Other examples of factors that
play out differently in courts than legislative processes include: access, which may affect legis-
latures more because courts have to accept cases filed in them as long as the jurisdictional and
pleading requirements are met; electoral concerns, which may only affect courts with elected
judges and may impact them less than legislators; and support of political elites, which may
influence courts less due to judicial independence. A few factors play special roles in one
institution but not the other. For example, courts are bound by precedent and the rules of stare
decisis in ways that legislatures are not.\textit{SILVERSTEIN, LAW’S ALLURE, supra} note 144, at
68–70. Thus, legislatures may have more freedom to depart from precedent than a court does.

\textsuperscript{168} Sociolegal scholars have long noted that institutions do not operate in isolation. \textit{See, e.g.,}

\textsuperscript{169} \textit{See, e.g., HANDLER, supra} note 143.
Lobbying Against the Odds

While these accounts suggest that an institution’s receptivity (or lack thereof) to a groups’ claims may affect venue choice, institutional relationships may influence advocates’ choices in other ways as well, especially as advocates turn to multi-institutional strategies. Advocates may choose a strategy that seeks to maximize coordination among institutions. A common instance of this kind of strategy is advocates’ choosing to litigate to raise awareness of an issue and get it on a legislative agenda. Alternatively, advocates may craft strategies that play institutions off one another. For example, advocates may develop legislative strategies to highlight problems in judicial or administrative processes and use the legislature either to pressure these institutions or to reverse their decisions. By considering various factors and how they interact across institutions over time, contextualized investigations illuminate the social dynamics in which groups shape society through the claims they make but, at the same time, are shaped by the institutions and social contexts in which they make those claims.

Finally, the framework presented here expands theoretical understandings by suggesting an approach to studying advocacy that applies across institutions and does not just explain why groups litigate or lobby. This approach acknowledges differences in interest communities rather than attempts to develop a one-size-fits-all theory. It offers a straightforward explanation, namely that advocates choose strategies in response to opportunities and constraints. The approach, however, also allows for complexity by integrating factors specific to particular institutions and adopting the contextualized investigations of specific cases used by sociolegal scholars. This approach, thus, models a way for scholars to devise more complex understandings of advocacy strategies and how they develop across institutions over time.

VI. HOW AND WHY AMERICAN INDIANS CHOOSE LEGISLATIVE ADVOCACY

This Part uses the approach outlined in Part V to provide a more nuanced and complete explanation of the rise in legislative advocacy by American Indian tribes and organizations over the past thirty-five years. It

170 See, e.g., Silverstein, Law’s Allure, supra note 144, at 30–33; Chen & Cummings, supra note 145, at 524.


172 Steinman describes this model as: “legal mobilization → political pressure → gains.” Steinman, supra note 171, at 763.

173 See, e.g., Carlson, supra note 67, at 948–50.
identifies some of the factors influencing American Indians to choose legislative strategies and the mechanisms by which these factors interacted to encourage American Indian legislative advocacy. The overarching theory here is that American Indians have chosen legislative strategies in response to a combination of factors, including changes in the political context, the availability of resources, the development of their political capacity, and the receptivity of federal institutions towards Indian claims.

Changes in congressional policy and process set the stage for the increase in legislative advocacy by American Indians at the end of the twentieth century. In the 1970s, Congress enacted the Tribal Self-Determination Policy, which acknowledged the importance of tribal input and participation in Indian affairs policymaking. American Indians further gained access to the political system with the institutionalization of Indian affairs within the congressional committee system. These changes in the political context in turn altered the political capacity of American Indians. As Indian tribes engaged more in the policymaking process, they gained expertise and access to policymakers, enabling them to secure a more favorable political climate. This feedback loop encouraged American Indians to choose to advocate legislatively. Changes in the receptivity of the Supreme Court to rights claims further bolstered American Indian legislative advocacy. An increasingly hostile Supreme Court pushed American Indians to consider alternatives to litigation and to seek congressional fixes to detrimental Supreme Court decisions. Finally, the introduction of new resources into Indian country, especially the rise in economic development due to gaming, further contributed to American Indian legislative advocacy. It both enabled and encouraged some American Indians to expand their lobbying and fueled new opposition to American Indian legislative advocacy.

A. Political Context: The Tribal Self-Determination Policy Sets the Stage for American Indian Legislative Advocacy

A tremendous shift in federal Indian-affairs policy occurred in the 1970s and set the stage for the explosion in American Indian legislative advocacy.174 Congress, historically the primary actor in the creation of Indian policy, had relegated Indian affairs a backseat in the legislative process by 1951.175 The Legislative Reorganization Act of 1946 reduced the status of the Indian Affairs Committees in both houses.176 This change in committee status along with a resurgence of assimilationist voices in Congress negatively affected Congress’s ability to legislate appropriate solutions to

174 For a full discussion of the role of the Executive Branch in this shift, see generally CASTILE, supra note 38.
175 WILKINS & STARK, supra note 8, at 89.
176 Id.
problems affecting Indian people. Congress continued to promote termination and relocation policies aimed at undermining tribal self-determination and encouraging Indian assimilation throughout the 1960s.

Forces outside the legislative branch precipitated the shift in Indian affairs policy. The inclusion of Indians in general legislation and programs had the unintended consequence of encouraging both Indians and Executive Branch officials to rethink Indian affairs. Despite a lack of presidential interest and a weak Indian lobby, the Kennedy and Johnson administrations wrote Indians into various bills meant to improve the lives of the poor and underserved. Central to the War on Poverty, the Economic Opportunity Act created the Office of Economic Opportunity (“OEO”) and its Community Action Program to empower poor people on a local level to reform institutions to end poverty. The Johnson Administration wanted to include Indians in these programs. In 1964, American Indian leaders lobbied for OEO funding to go directly to tribes (rather than the states). As a result, the Economic Opportunity Act of 1964 made tribal governments eligible for OEO grants. Bypassing the BIA, the OEO programs channeled significant amounts of federal money directly to Indian tribes. Indian tribes could spend this money and administer programs on their own. As a result of this transfer of authority for local decisionmaking, tribal governments gained competence in planning and running their own programs. These experiences empowered Indians, encouraging them to consider policy proposals similar to the OEO programs that could replace the termination policy. Moreover, the success of these programs in Indian country prompted federal officials to reconsider their approach to Indian affairs. The policy climate

177 Id.; see also Castile, supra note 38, at xxi–xxvii (discussing the forces facilitating the creation of the termination policy).
178 Castile, supra note 38, at 58–60.
179 Id. at 4 (explaining that the Kennedy Administration showed very little interest in Indian affairs).
180 To be fair, the NCAI launched a campaign against termination in the early 1950s and scored some important victories. Cowger, supra note 33, at 116–19 (noting that the NCAI managed to modify the first termination bill and prevent termination of certain tribes). But the Indian lobby more generally remained weak. Castile, supra note 38, at 4.
181 Castile, supra note 38, at 4, 24–25.
182 Id. at 29.
183 Charles F. Wilkinson, Blood Struggle: The Rise of Modern Indian Nations 127–28 (2005) (identifying this as the first time in American history that “Indian people had conceived of a provision to be inserted in national legislation and then lobbied it through Congress into law.”).
185 Castile, supra note 38, at 31–33.
186 Id. at 33 (“Prior to the entry of the Community Action programs, virtually all funds on reservations were directly administered by the federal agencies that allocated them.”).
187 Wilkinson, supra note 183, at 191–94.
188 Castile, supra note 38, at 48.
189 Id. at 68–69.
was shifting from one of termination to self-determination by the end of the Johnson Administration.\footnote{Id. at 73–74.}

President Nixon embraced the self-determination approach in his 1968 presidential campaign and his support stimulated the adoption of this approach as federal Indian policy.\footnote{GROSS, supra note 131, at 34–38. According to Gross, Nixon attributed his positive stance towards American Indians to the influence of his college football coach. Id. at 70–71.} By 1970, President Nixon had publicly repudiated the negative Indian policies of the 1950s and replaced them with the Tribal Self-Determination Policy.\footnote{CASTILE, supra note 38, at 92–98, 155–56. Unable to secure passage of their legislative proposals, the Nixon administration sought to implement the self-determination policy by making changes to and within the BIA. Id. at 87–91.} During his presidency, Nixon sent several legislative proposals to Congress, which renounced termination in favor of a policy of tribal self-determination.\footnote{CASTILE, supra note 38, at 105–06. Senator Jackson’s pro-Indian stance waivered in the 1970s, but lasted long enough to secure passage of self-determination legislation. Id. at 166–68.}

While resistance from a Democratic Congress initially undermined many of President Nixon’s initiatives, his enduring commitment to tribal self-determination may have helped soften such opposition in this area. Aiming to run for the presidency, Senator Henry Jackson (D-Wash.) reversed his position on termination and sponsored a self-determination bill that the Senate passed in 1972.\footnote{CASTILE, supra note 38, at 105–06. Senator Jackson’s pro-Indian stance waivered in the 1970s, but lasted long enough to secure passage of self-determination legislation. Id. at 166–68.} Senator Jackson continued to sponsor self-determination legislation with Representative Lloyd Meeds (D-Wash.) and Senator James Abourezk (D-S.D.) supporting his efforts in 1974.\footnote{CASTILE, supra note 38, at 166–68.}

Strong executive branch support along with the failure of the termination policy and supportive leadership in Congress encouraged Congress to embrace a tribal self-determination policy.\footnote{GROSS, supra note 131, at 75–86; CASTILE, supra note 38, at 165–68. Lobbying may have also played a role in Congress’s adoption of the Self-Determination Policy, but scholars disagree about the extent to which Congress consulted tribes in formulating the Indian Self-Determination and Education Assistance Act (“ISDEAA”). Wilkinson admits that Indian people did not play a direct role in formulating the Self-Determination Policy, WILKINSON, supra note 183, at 189, but notes that the arrival of American Indians as staffers to the Senate Interior Committee and the House Interior Committee in the early 1970s contributed to the change in policy and explosion of Indian legislation, id. at 195. Strommer and Osborne similarly report a lack of tribal input on the ISDEAA. Geoffrey D. Strommer & Stephen D. Osborne, The History, Status, and Future of Tribal Self-Governance Under the Indian Self-Determination and Education Assistance Act, 39 AM. INDIAN L. REV. 1, 20 (2015). In contrast, Delaney suggests that tribal advocates initially suggested that the Kennedy administration “use the government contracting process as a mechanism for transferring control over federal funds and programs from agencies to tribes.” Danielle Delaney, The Master’s Tools: Tribal Sovereignty and Tribal Self-Governance Contracting/Compacting, 5 AM. INDIAN L.J. 308, 328 (2017). It is unclear whether Delaney is referring to tribal lobbying for Office of Economic Opportunity (OEO) funding to go directly to tribes (rather than the states) in 1964 or something else. Tribes and tribal coalitions testified at committee hearings held prior to the enactment of the ISDEAA, but it is not clear whether they played any other role in the policy’s development. See, e.g., Indian Self-Determination and Education Assistance Act Amendments of 1987: Hearing on S. 1703

\footnote{Id. at 70–71.}

\footnote{CASTILE, supra note 38, at 92–98, 155–56.}
Determination Policy with the enactment of the Indian Self-Determination and Education Assistance Act ("ISDEAA") in 1975. The Act declared, "[T]he United States is committed to supporting and assisting Indian tribes in the development of strong and stable tribal governments, capable of administering quality programs and developing the economies of their respective communities." It enabled tribes to build their institutional capacities and economies by transferring control over federal programs to the tribes. The ISDEAA required the Secretaries of the Interior and Health and Human Services, upon the request of any Indian tribe, to contract with tribal organizations to operate federal programs for Indians.

Before the S. Select Comm. on Indian Affairs, 100th Cong. 37 (1987) (statement of Joe DeLaCruz, President, Affiliated Tribes of Northwest Indians) (explaining that he was "directly involved in the policy discussion leading to the original Self-Determination Act"); Indian Self-Determination and Education Program Hearing on S. 107 and Related Bills Before S. Comm. on Interior and Insular Affairs, 93d Cong. 106–12, 133–37 (1973) (statements of Valentino Cordova, Chairman, All Indian Pueblo Council, Joe Upicksoun, Chairman, Education Committee of the National Tribal Chairman’s Association); Indian Self-Determination Act of 1972: Hearing on S. 3157, S. 1573, S. 1574, and S. 2238 Before the S. Subcomm. on Indian Affairs of the S. Comm. on Interior and Insular Affairs, 92d Cong. 69–76, 78–79 (1972) (statements of Franklin Ducheneaux, Legislative Consultant, NCAI, Buffalo Tiger, Chairman, Miccosukee Business Committee, and S. Bobo Dean, Council for Association on American Indian Affairs). Their testimony may have influenced some members of Congress. Even if the testimony did not sway any legislators to vote for the ISDEAA, it demonstrated that Indians could participate in the policymaking process and it probably impacted future lobbying efforts, especially as the political context shifted. A reinforcing cycle emerges: Indians testify before Congress, which impacts substantive policy decisions and opens the door for more political participation leading to more Indian lobbying. This feedback loop contributed to the rise in American Indian lobbying at the end of the twentieth century.

The adoption of the Tribal Self-Determination Policy signaled a dramatic shift in the federal government’s position on American Indian policy.\textsuperscript{200} For the first time since the 1930s, Congress promulgated a policy that supported Indian nations as separate governments and invited their participation in federal policymaking.\textsuperscript{201} Under the ISDEAA, money flowed directly to tribal governments, bypassing the BIA and enabling Indian nations to make important decisions regarding their welfare. Tribes, which had begun to administer programs under the OEO, embraced the ISDEAA and complained about the BIA’s reluctance in implementing it.\textsuperscript{202}

Despite some initial resistance, the new perspective on Indian affairs embodied in the Tribal Self-Determination Policy “eventually came to pervade [Congress’s] entire approach to Indian affairs.”\textsuperscript{203} Throughout the rest of the 1970s and early 1980s, Congress enacted several key bills expanding and enhancing the Tribal Self-Determination Policy.\textsuperscript{204} This legislation institutionalized the Tribal Self-Determination Policy and reinforced the growing capacity of Indian nations as governments.\textsuperscript{205}

This change in the political context facilitated the development of Indian political capacity and lobbying. By encouraging Indian political participation, the Tribal Self-Determination Policy increased opportunities for tribes to lobby. As tribes engaged in more lobbying, they gained experience and developed more sophisticated lobbying strategies, which further reinforced their ability to lobby. After decades of enacting policies destructive to American Indians, Congress appeared much friendlier to American Indian interests.

\textsuperscript{200}\textsuperscript{GROSS, supra note 131, at xvi.}

\textsuperscript{201}\textsuperscript{CORNELL, supra note 27, at 204–05.}


\textsuperscript{203}\textsuperscript{GROSS, supra note 131, at 78.}


\textsuperscript{205} The Tribal Self-Determination Policy may have also stimulated American Indian legislative advocacy by helping some Indian nations build their internal capacities to deliver programs and services to their citizens. The stakes Indian nations had in the implementation of these programs and services increased as they gained direct managerial control over them. Tribes, thus, had more incentives both in protecting these programs and in advocating for improvements in them. Moreover, as their internal capacities grew, Indian tribes gained expertise in these areas, making them more knowledgeable about the programs that work in their communities and more able to advocate effectively to policymakers. \textsuperscript{GROSS, supra note 131, at 108.}
Political Capacity: The Senate Committee on Indian Affairs Creates Access for American Indians

While the Tribal Self-Determination Policy signaled a new receptivity to American Indian participation in policymaking, the formation of the Senate Select Committee on Indian Affairs helped create the basic conditions for American Indians to engage in legislative advocacy. The Senate Select Committee on Indian Affairs emerged out of the recommendations of the American Indian Policy Review Commission (“AIPRC”), a bipartisan committee charged with investigating the historical and legal relationship between Indians and the government and proposing ways to revise Indian policy and programs to benefit Indians. The AIPRC specifically recommended the creation of permanent Indian Affairs committees in both houses of Congress. The AIPRC expressed grave concerns about the results of the termination of such committees in 1946. It justified its recommendation on congressional plenary power over the administration of Indian affairs. In response to the AIPRC recommendations, the Senate extended the duration of a temporary Select Committee on Indian Affairs in 1977 and then made the committee permanent in 1984.

The existence of a longstanding Senate Committee on Indian Affairs has created unprecedented opportunities for American Indian legislative advocacy. Structurally, it gives “Indian constituencies a unique opportunity to make their policy preferences known.” American Indians have seized this opportunity by directly lobbying members of the committee and hiring Washington lobbyists to represent them on important matters.

A spillover effect of these increased opportunities for American Indian legislative advocacy has been that congressional perceptions about who speaks for American Indians have changed dramatically. Historically, members of Congress had ignored Indian constituencies in formulating Indian policy. If they paid attention to Indian interests at all, members of Congress turned to the BIA or non-Indian friends of the Indians for advice. The influx of American Indians lobbying members of Congress, however, has altered this dynamic and created a feedback loop supportive of Indian legislative advocacy. As American Indians gained access to members of Congress and their staffs, they built relationships with them.

206 Id. at xx.
207 WILKINS & STARK, supra note 8, at 90.
208 Id.
209 Id.
210 Id. at 91.
211 GROSS, supra note 131, at 77, 103.
212 Id. at 99–105.
213 Id. at 79 (“In fact, before 1970, Indian constituencies were so insignificant in policy development that they were virtually invisible.”).
214 Id. at 77.
215 EVANS, supra note 64, at 7, 74.
these relationships, members of Congress and their staffs have increasingly learned to trust the information and advice they obtain from American Indians. As a result, they now identify American Indian organizations and tribal governments as experts and, thus, seek their advice on policymaking. Thus, relationship building has reinforced the ability of American Indians to influence policy by providing them with increased and continued access to legislators and their staffs.

The institutionalization of the Senate Committee on Indian Affairs over the past forty years has demonstrated political support for Indian affairs and further encouraged American Indians to choose legislative strategies. The Senate Committee on Indian Affairs has broad authority to develop and oversee polices related to American Indians. In setting its own agenda, it has been exceptionally active and committed to implementation of the Tribal Self-Determination Policy. The committee has “examined a plethora of issues, including land-related topics, health concerns, housing, education, economic development, water claims, land claims, trust funds, gaming, recognition, natural resources and environmental concerns, religious freedom, the committee administrative tasks, Alaska Natives, Indian child welfare, and tribal-state relations.” The hearings and other congressional attention to these myriad issues has stimulated legislative advocacy by signaling to American Indians that the committee, or at least some of its members, support their issues and want to hear their concerns.

Moreover, this government support led to results, and thus, further spurred American Indians to use legislative strategies. Members of Congress

---

216 Id. at 74–97. In effect, American Indians have become repeat players in legislative policymaking. They gain many of the benefits that repeat players obtain in other policymaking venues, such as the courts. For a discussion of repeat players, see generally Marc Galanter, supra note 146.

217 Gross, supra note 131, at 99 (“Indian tribes played a central role in developing the political agenda and in bringing about the policy successes of the seventies.”).

218 Wilkins & Stark explain:

The Committee on Indian Affairs is the authorizing committee for programs of the BIA in the Department of the Interior, the Indian Health Service and the Administration for Native Americans in the Department of Health and Human Services, and the Office of Indian Education in the Department of Education. Furthermore, the Committee has oversight responsibility for programs affecting Indians in all other federal agencies, including the Indian Housing program of the Department of Housing and Urban Development. These responsibilities dovetail with those specified in Senate Resolution 4, which include matters relating to tribal and individual lands; the federal government’s trust responsibilities; and Indian education, health, Indian land claims, and natural resources. In effect, this committee (like the subcommittee in the House) is charged with an enormous task: the oversight of Congress’s continuing historical, constitutional, and legislative responsibilities to 565 distinctive indigenous entities.

WILKINS & STARK, supra note 8, at 92–93.

219 Id. at 93.

220 Id.

221 See Baumgartner et al., supra note 11, at 13 (showing that governmental support of an issue increases lobbying on that issue).
have proposed, and Congress has enacted a tremendous amount of legislation related to American Indians.\textsuperscript{222} Moreover, a significant proportion of these bills supported or expanded the Tribal Self-Determination Policy.\textsuperscript{223} In 1994, Congress dramatically broadened the self-determination program, expanding the tribal role beyond the federal responsibilities of the Department of Interior to include health, housing, and environmental protection.\textsuperscript{224} Congress also entertained multiple legislative proposals aimed at improving the health, education, and housing of American Indians.\textsuperscript{225} The receptivity of Congress to pro-Indian bills both reflected and encouraged American Indian legislative advocacy. It appeared that American Indians had little to lose and much to gain by going to Congress.

C. Shifting Institutional Dynamics: Legislative Advocacy and the Courts

Indian nations had several options for pursuing their claims when they initially turned to Congress. Indian tribes had successfully gone to the courts for recognition of their rights, and the courts remained a viable option. As Congress opened its doors to American Indian advocacy, however, the courts increasingly closed theirs. The rise of American Indian legislative advocacy occurred as federal courts, and especially the Supreme Court, became less receptive to American Indian litigants.\textsuperscript{226} Federal courts were never consistent protectors of American Indian rights, but their decisions started trending against American Indian litigants in the late 1970s.\textsuperscript{227} The Court handed down some particularly devastating cases in the late 1970s and early 1980s.\textsuperscript{228} These early losses, however, did not dissuade American Indi-

\textsuperscript{222} See Carlson, supra note 2, at 119 (showing that Congress enacted close to 20 percent of all Indian-related bills from 1979 to 1991 and close to 15 percent from 1991 to 1997 as compared to less than 10 percent of all bills generally during this time period).
\textsuperscript{225} See Cornell & Kalt, supra note 223, at 22 (“Over the same period [1973-2010], there have been 2,405 sponsors of 305 legislative measures aimed at improving conditions for American Indians, typically through increased spending on health, education, housing and the like.”).
\textsuperscript{226} Fletcher, supra note 57, at 935 (showing that the success rate of tribal litigants in the Supreme Court has not improved since 2001); Getches, supra note 56, at 280–81 (2001) (finding that tribes lost 82 percent of the cases decided by the Supreme Court from 1991–2000).
\textsuperscript{228} Oliphant, 435 U.S. at 195; Moe, 425 U.S. at 478–83; Montana v. United States, 450 U.S. 544, 547–50 (1981); Washington v. Confederated Tribes of Colville Indian Reservation, 447 U.S. 134, 152–58 (1980). In hindsight, these cases appeared to forewarn that a shift was coming in the Supreme Court, and Indian advocates were displeased with them at the time. See, e.g., Sarah Krakoff, Mark the Plumber v. Tribal Empire, or Non-Indian Anxiety v. Tribal
ans from litigating, and American Indians continued to win 60 percent of the cases heard by the Supreme Court until 1987. It would take almost another decade of decisions adverse to Indian country for American Indians to question their faith in the Supreme Court.

A growing number of losses in the Supreme Court, however, eventually pushed American Indians to consider alternative strategies and engage in legislative advocacy. By the mid-1980s, American Indians were accumulating losses in the Supreme Court. The Supreme Court stripped tribes of their criminal jurisdiction over non-Indians and non-member Indians, limited tribal civil adjudicatory jurisdiction over non-Indians, permitted state taxation on Indian lands, and refused to recognize some tribal water rights.

The extent of the losses further encouraged Indian tribes to go to Congress. The Supreme Court started to "veer[] away from the foundations of Indian law." It ignored precedents favorable to tribal rights and abandoned its earlier approach of relying on Congress "to decide clearly the bounds of Indian sovereignty." As a result, the Court started remaking federal Indian law on terms much less favorable to American Indians and placing the

Seventy? The Story of Oliphant v. Suquamish Indian Tribe, in Indian Law Stories 283–85 (Carole Goldberg, et al. eds., 2011) (describing reactions to Oliphant). But in the same year that the Court decided Oliphant, the Court also handed down two cases that upheld tribal sovereignty. See United States v. Wheeler, 435 U.S. 313, 326–32 (1978) (holding that tribes retain the inherent power to try their own members); Santa Clara Pueblo v. Martinez, 436 U.S. 49, 52, 55–56 (1978) (holding that the Indian Civil Rights Act did not allow suits against tribes in federal courts except for petitions for habeas corpus).

229 Fletcher, supra note 57, at 942. R
230 Getches, supra note 56, at 273–74. R
231 Oliphant, 435 U.S. at 195. R
236 Id. at 206, 208. Prior to the Oliphant decision, only Congress had plenary power over Indian affairs and could determine the sovereign rights of Indian nations. The Court in Oliphant seemed to be allocating plenary power to itself.

237 Getches, supra note 56, at 276. R
238 Id. ("[T]he present Supreme Court is shunning its own legal traditions and creating new rules that conform to its own perceptions of current realities, instead of staying its hand and forcing the political branches to deliberate the difficult choices.").
burden on American Indians to convince Congress to clarify its position on Indian affairs in legislation.\textsuperscript{239}

These shifts in the Court’s Indian law jurisprudence deterred some American Indians from litigating\textsuperscript{240} and provided incentives for tribal lawyers and advocates to turn to Congress to protect tribal interests, especially tribal self-determination and jurisdiction.\textsuperscript{241} Adverse decisions ceased being outliers, and scholars started reporting the abysmally low win rates of American Indians.\textsuperscript{242} One influential study revealed that American Indians lost in the Supreme Court over seventy-five percent of the time—more frequently than convicted felons—from 1986 to 2000.\textsuperscript{243} A later study confirmed the continued hostility of courts towards American Indians, reporting that American Indians were still losing seventy-five percent of their cases in the Supreme Court in 2009.\textsuperscript{244} The risks of losing seemed to outweigh the benefits of litigating cases.

These changing institutional dynamics encouraged American Indians to avoid the courts and engage more in legislative processes. By the 1990s, American Indians had responded to the devastating losses they were facing in the courts by actively coordinating and launching legislative campaigns to overturn unfavorable Supreme Court decisions.\textsuperscript{245} For example, less than a month after the Supreme Court held that Indian nations did not have criminal jurisdiction over non-member Indians in \textit{Duro v. Reina},\textsuperscript{246} the NCAI “had convened a meeting of tribal, Bureau of Indian Affairs, and congressional representatives to discuss the implications of the case and discuss possible legislative responses.”\textsuperscript{247} Meetings with tribal leaders ensued over the summer to develop a legislative proposal and mobilize Indian country behind it.\textsuperscript{248} Congress acted swiftly to reverse the Court. Six months after the Court handed down the \textit{Duro} decision, Congress enacted temporary legislation restoring inherent criminal jurisdiction over non-member Indians as part of a defense appropriations bill.\textsuperscript{249}

\textsuperscript{239} Philip P. Frickey, \textit{(Native) American Exceptionalism in Federal Public Law}, 119 \textit{Harv. L. Rev.} 431, 483 (2005) (noting how, in the past, tribal success in the courts placed the legislative burden on tribal opponents, so that tribes were in the easier position of trying to kill reactive legislation rather than seeking legislation on their own behalf).

\textsuperscript{240} Burton, \textit{supra} note 55, at 39.

\textsuperscript{241} Berger, \textit{supra} note 58, at 12 (detailing advocacy in the \textit{Duro} fix legislation); Getches, \textit{supra} note 56, at 276 (suggesting that the legislative process has advantages over adjudication).

\textsuperscript{242} Getches, \textit{supra} note 56, at 280–81.

\textsuperscript{243} Id. (“Convicted criminals achieved reversals in 36 percent of all cases that reached the Supreme Court in the same period, compared to the tribes’ 23 percent success rate.”).

\textsuperscript{244} Fletcher, \textit{supra} note 57, at 935.


\textsuperscript{246} 495 U.S. 676 (1990).

\textsuperscript{247} Berger, \textit{supra} note 58, at 12; Newton, \textit{supra} note 245, at 110.

\textsuperscript{248} Berger, \textit{supra} note 58, at 12.

\textsuperscript{249} Berger, \textit{supra} note 58, at 12; Newton, \textit{supra} note 245, at 111; \textit{see also} Skibine, \textit{supra} note 75, at 767–68.
This temporary Duro fix expired at the end of a year so Indian advocates and tribal leaders mobilized to amend the law and remove the expiration date.250 Congressman Bill Richardson (D-N.M.) introduced a bill, which quickly passed in the House.251 Two similar bills, introduced by Senator Inouye (D-Haw.) and supported by Senators John McCain (R-Ariz.), Pete Domenici (R-N.M.), Paul Simon (D-Ill.), and Paul Wellstone (D-Minn.), encountered opposition from Senator Slade Gorton (R-Wash.), who argued that because of its constitutional nature, Congress could not alter the Duro decision.252 As a result, the Senate amended one of the bills to include another temporary two-year extension, and the amended bill passed the Senate.253 The bill went to conference committee, but the committee could not reconcile the differences in the House and Senate versions of the bill and it died.254 This initial failure led Indian advocates and tribal leaders to redouble their efforts. The temporary Duro fix expired during their efforts to secure a permanent one, making the issue more pressing.255 Indian advocates stepped up their lobbying inside the halls of Congress and mounted an outside lobbying effort by mobilizing “an avalanche of telegrams from tribes expressing outrage and concern about the passing of the deadline and the need for a permanent solution.”256 The Senate Committee on Indian Affairs marked up the second Duro fix bill, and it passed the Senate.257 With two bills having passed the Senate and one the House, the conference committee reconvened—this time to agree on a bill that would make the Duro fix permanent.258 President George H.W. Bush signed the Duro fix into law on October 28, 1991.259

The success of the Duro fix affected American Indian legislative advocacy in several significant ways. First, it established a precedent of using legislative advocacy to reverse negative Supreme Court decisions. The Duro fix demonstrated that American Indians could use the law to reframe “the formulation of congressional acts regarding tribal power,” and thus, reconceptualize the power relations between Congress and Indian nations.260 In the past, Congress had delegated power to Indian nations rather than acknowl-

---

250 Newton, supra note 245, at 114; see also Berger, supra note 58, at 13 (“Opponents of the bill were overwhelmed by witness after witness from Indian organizations and tribes.”).
251 Id. at 114.
252 Id. at 115.
253 Id. at 115–16 (providing details on the compromise).
254 Id. at 116. Newton attributed this impasse to neither side wanting to compromise. The House members insisted on making the Duro fix permanent while the Senate members of the conference committee felt bound to honor the compromise they made with Senator Gorton. Id.
255 Id. at 116.
256 Id. at 117.
257 Id.
258 Id.
259 Id.
260 Berger, supra note 58, at 13.
edged the inherent sovereignty possessed by them. Delegation had the potential to limit tribal power by subjecting it to constitutional restrictions not applicable to Indian nations otherwise. Savvy Indian advocates saw the danger of a Duro fix that delegated authority to Indian nations. As an alternative, they proposed new language that would reiterate the status of Indian nations as separate governments with their own inherent powers. Consequently, instead of delegating power to Indian nations and suggesting that they occupy a lesser position as sovereigns, Congress “recognized and reaffirmed” the “inherent power of Indian tribes . . . to exercise criminal jurisdiction over all Indians.” This language ensured that the Supreme Court “could only hold that Congress intended that tribes exercising jurisdiction over non-members under the Duro fix were exercising inherent tribal, not federal, power, and the Constitution did not prevent this result.” In effect, by adopting this language, Congress reasserted its importance, and historical primacy, in the area of Indian affairs and signaled to American Indians and the courts that it could and would act to restore tribal sovereignty. Moreover, American Indians learned that they could effectively use the law to leverage institutional dynamics. Knowing the power of the Court to undermine a legislative victory, they had lobbied for—and Congress adopted—language that limited the Court’s ability to interpret the statute against their interests. This lesson would shape future legislative efforts by American Indians.

Second, the success of the Duro fix encouraged Indian nations and tribal organizations to embark on additional campaigns to reverse negative Supreme Court decisions. Since the Duro fix, advocates have proposed legislation to overturn several other Supreme Court decisions. After the Supreme Court restricted tribal court jurisdiction in Nevada v. Hicks, tribal advocates, including the NCAI, sought congressional reaffirmations of tribal criminal and civil jurisdiction. Similarly, Indian nations have sought to overturn Salazar v. Carcieri and allow all Indian nations to take land into trust under 25 U.S.C. § 465 since 2011. Most significantly, however,

261 Id.; see also Rice v. Rehner, 463 U.S. 713, 715 (1983) (delegating federal power to Indian tribes).
262 Berger, supra note 58, at 13.
263 Berger, supra note 58, at 13; Newton, supra note 245, at 112.
264 Berger, supra note 58, at 13.
266 Berger, supra note 58, at 13.
268 533 U.S. at 363–70 (holding that tribe could not hear civil rights claims against state police arising out of search on-reservation for evidence of off-reservation crimes).
269 See Tribes Seek to Overturn Supreme Court, INDIANZ.COM (Feb. 27, 2002), http://www.indianz.com/News/printme.asp?ID=law02/02272002-1 (describing the legislative proposals drafted to restore tribal sovereignty after Hicks).
American Indian advocates commenced a long-term campaign to partially reverse the Supreme Court’s decision in *Oliphant v. Suquamish*,271 prohibiting tribal criminal jurisdiction over non-Indians, as part of the reauthorization of the Violence Against Women Act.272 Advocates adopted language similar to that used in the *Duro* fix and argued for a section restoring the inherent power of tribal governments to exercise special criminal jurisdiction over all persons committing specific intimate-partner-related crimes in Indian country.273 These efforts came to fruition with the Violence Against Women Reauthorization Act of 2013.274 Once again, astute Indian advocates used a legislative strategy to reconfigure the power dynamics between Indian nations and the U.S. government and to protect Indian country from further encroachment on its sovereignty by hostile courts.

Moreover, the success of the *Duro* fix stimulated American Indians to use legislative advocacy more broadly. American Indians interpreted the legislative success of the *Duro* fix as tribal advocacy paying off in a Congress more supportive of Indian issues than the Court.275 Scholars, tribal leaders, and advocates started suggesting that Congress may be more responsive than the courts to Indian interests.276 Some even argued that Congress is the most appropriate institution within the U.S. government to make federal Indian law and policy and that the courts should defer heavily to Congress.277

This perception of Congress as a more favorable institution encouraged Indian nations and organizations to engage in legislative strategies across a wide range of issues. For example, Indian nations have turned to Congress to argue for the congressional restoration of their government-to-government status,278 to acquire homelands,279 to protect and repatriate their cultural ob-
Lobbying Against the Odds

jects and remains, to settle natural resource claims, and to gain federal recognition as Indian tribes. The institutional dynamics around American Indian policymaking shifted as the courts’ receptivity towards American Indian claims declined in the early 1980s. American Indians had increased their lobbying efforts in the 1970s, so they had gained considerable experience and skill in lobbying by the time the Supreme Court grew hostile to their claims. With little hope of vindication of their rights in court, American Indians had no choice but to continue to lobby Congress for help. Indian advocates built on their earlier experiences and expanded their lobbying efforts. They used legislative advocacy to enact statutes that redefined their relationship with the U.S. government and protected their rights from hostile courts.

D. Resources: Legislative Advocacy and the Rise of Indian Gaming

For most of the twentieth century, almost all American Indians lacked the financial resources to engage in lobbying. The rise of gaming by Indian nations in the late twentieth century has provided some but not all tribes with financial resources to invest in legislative strategies.

American Indian legislative advocacy predates gaming, but has increased steadily and consistently with the growth of gaming. In 1987, the Supreme Court held that Indian nations could operate gaming establishments free of state regulation in California v. Cabazon Band of Mission Indians. Congress enacted the Indian Gaming Regulatory Act (“IGRA”) a year later. Gaming in Indian country has grown tremendously ever since, with

---

279 See, e.g., Miller, supra note 58, at 434–38 (describing the successful efforts of the Timbisha Shoshone to obtain legislation providing them with homeland from 1994 to 2000).
280 See, e.g., Dumont, supra note 58, at 10 (discussing the efforts of Indian nations, pan-tribal organizations, and Native Hawaiians to secure passage of the Native American Graves and Repatriation Act).
282 For example, the number of non-federally recognized Indian groups seeking recognition legislatively spiked after the Duro fix in 1991. See Kirsten Matoy Carlson, Congress, Tribal Recognition, and Legislative-Administrative Multiplicity, 91 Ind. L.J. 955, 972 (2016). For a discussion of why non-federally recognized Indian groups chose legislative strategies, see Carlson, supra note 67, at 941–58.
283 See Cornell, supra note 28, at 129–30 (describing lack of Indian financial resources to engage in politics until 1970s).
284 Many scholars have linked rises in political mobilization, both at the state and federal levels, to gaming. See, e.g., Cornassel & Witmer, supra note 59, at 522; Witmer & Bochenke, supra note 65, at 139; Steven Andrew Light & Kathryn R.L. Rand, Indian Gaming and Tribal Sovereignty: The Casino Compromise 65–69 (2005).
285 See infra Part III fig.3.
286 See infra Part III fig.3.
just under half of all tribes engaging in gaming operations today.\textsuperscript{289} The effects of gaming remain uneven—with only a few tribes experiencing spectacular success—but it has provided some tribes with new resources to invest in lobbying.\textsuperscript{290}

By introducing new financial resources into Indian country, gaming has enabled Indian nations with profitable gambling operations to engage in legislative strategies.\textsuperscript{291} Gaming revenues provide these tribes with resources “to employ skilled lobbyists and savvy public relations firms . . . ‘to win influence, make friends, and crush opponents’ in a manner heretofore unknown.”\textsuperscript{292} Thus, as gaming revenues have skyrocketed from a few million dollars in 1985 to 28 billion dollars in 2015, expenditures spent on lobbying have also risen, albeit inconsistently.\textsuperscript{293} Indian nations engaged in gaming have exceeded non-gaming Indian nations in reporting lobbying since 1995.\textsuperscript{294} Gaming tribes lobby more consistently over time than non-gaming tribes\textsuperscript{295} and report spending more on lobbying expenditures than non-gaming tribes. On average, Indian nations engaged in gaming report spending twice as much annually on lobbying as Indian nations not engaged in gaming.\textsuperscript{296} The Indian nations that report spending the most money on lobbying—over $200,000 annually—run some of the most profitable gaming operations. For example, the Agua Caliente Band of Cahuilla Indians, the Jicarilla Apache Nation, the Gila River Indian Community, the Mashantucket Pequot Tribal Nation, the Mississippi Band of Choctaw Indians, the Oneida Indian Nation, the Pechanga Band of Luiseño Mission Indians, the Seminole Tribe of Florida, the Tunica-Biloxi Tribe of Louisiana, and the


\textsuperscript{290} See id. (reporting that forty percent of all gaming revenue is concentrated in the hands of a few small tribes located near major urban areas and that other tribes with gaming operations “are only marginally profitable”).

\textsuperscript{291} LIGHT & RAND, supra note 282, at 65–66; Witmer & Boehmke, supra note 65, at 132 (explaining resource mobilization theory and how it applies to gaming tribes).

\textsuperscript{292} WILKINS & STARK, supra note 8, at 165–66.

\textsuperscript{293} See infra Part III fig.4; see also Randall K. Q. Akee et al., The Indian Gaming Regulatory Act and Its Effects on American Indian Economic Development, 29 J. ECO. PERSPS. 185, 194 (2015). Figures depicting lobbying expenditures and annual gaming revenues closely mirror one another with consistent growth until about 2006. While lobbying expenditures decreased from 2007 to 2010, see infra Part III fig.5, gaming revenues leveled off around the same time, see Akee et al., supra, at 194.

\textsuperscript{294} See supra Part III fig.4; see also Witmer & Boehmke, supra note 65, at 139 (reporting that “far more gaming tribes than non-gaming tribes actively engage in lobbying members of Congress.”). The increase in legislative mobilization is not limited to Congress. Several recent studies document the influence of gaming revenues on state politics. See MASON, supra note 7, at 70–145; CORNTASSEL & WITMER, supra note 64, at 125–33; see generally HANSEN & SKOPEK, supra note 64.

\textsuperscript{295} See Witmer & Boehmke, supra note 65, at 139.

\textsuperscript{296} See supra Part III.
Viejas Band of Kumeyaay Indians—all tribes with lucrative gaming operations—report spending the most on lobbying annually.\(^{297}\)

The increase in legislative advocacy stimulated by gaming, however, is not limited to lobbying related to gaming or Indian affairs. Gaming tribes have strong incentives to lobby to protect their businesses and governmental interests, but they lobby on other issues as well. They have “seized the opportunity to express their positions on a variety of other issues of importance to them.”\(^{298}\) Thus, gaming has enabled some tribes to weigh in on issues that seem unrelated to gaming or tribal affairs and, as a result, expand their policy influence.\(^{299}\) Tribes have reported lobbying on federal appropriations, taxation, transportation, and natural resources.\(^{300}\) As a result, the effects of American Indian lobbying reach beyond the individual tribes engaging in it.

Gaming, or rather the introduction of new financial resources into Indian country as a result of it, has facilitated the growth in American Indian legislative advocacy. It has provided some Indian nations with the financial resources necessary to advocate legislatively and to expand their lobbying efforts.

E. **Opposition: The Backlash to Gaming and Legislative Advocacy**

The new resources provided to Indian country by gaming, however, have had crosscutting impacts on legislative advocacy. Gaming has not only stimulated the use of legislative strategies by providing the money to hire lobbyists, but also it has encouraged a growing backlash movement against Indian-owned casinos and Indian country more generally.\(^{301}\) Opposition to American Indian interests is not new, but the rise of Indian gaming and the political power thought to accompany it has reinvigorated anti-Indian sentiments.\(^{302}\) These sentiments, however, are not simply lodged against the most successful gaming tribes.\(^{303}\) Rather, they pervade Indian country and often contribute to stereotypes and misperceptions of American Indians.\(^{304}\) Contrary to these misguided notions of Indians as rich,\(^{305}\) American Indians “re-

\(^{297}\) See supra Part III.

\(^{298}\) Frederick J. Boehmke & Richard Witmer, *Tribal Political Expenditures in California and Washington, D.C.,* in *Hansen & Skopek,* supra note 64, at 25, 34; *see also* Carlson, supra note 112, at 7–9.

\(^{299}\) See *Welkins & Stark,* supra note 8, at 165.

\(^{300}\) Boehmke & Witmer, *in Hansen & Skopek,* supra note 298, 34.

\(^{301}\) See *Corntassel & Witmer,* supra note 64, at 6.

\(^{302}\) Id. at 28–46.

\(^{303}\) Some extremely successful gaming tribes have experienced relentless backlash attacks. In particular, the Mashantucket Pequot Tribe has faced extensive, racialized challenges both to their federal recognition and their gaming enterprises. *See, e.g.,* Renee Ann Cramer, *Cash, Color, and Colonialism: The Politics of Tribal Acknowledgment* 137 (2005).

\(^{304}\) See id. at 105; *Corntassel & Witmer,* supra note 64, at 45.

\(^{305}\) Rich Indian racism is a post-IGRA phenomenon, “where false images related to indigenous gaming are created and propagated by governmental and media entities.” *Corntassel & Witmer,* supra note 64, at 4.
main Americas poorest people" and now may need to engage in lobbying to protect the little they have.\footnote{306 Native Am. RTS. FUND, supra note 289. Moreover:

[T]he needs of reservation Indians are so great that even if the total annual Indian gaming revenue in the country could be divided equally among all the Indians in the country, the amount distributed per person would still not be enough to raise Indian per capita income (currently $11,259) to anywhere near the average of $21,587.\cite{Id.}}

Countermobilization to American Indians, largely catalyzed by Indian gaming, has both undermined and facilitated the growth in American Indian legislative advocacy. Some American Indians have responded to this backlash by investing more in legislative strategies.\footnote{307 As Cornassel and Witmer argue, tribes may have to use multiple strategies to challenge rich Indian racism. CORNTASSEL & WITMER, supra note 64, at 134–49. For example, the Tigua Nation (Ysleta del Sur Pueblo) responded to the closure of their casino by hiring a lobbyist. CORNTASSEL & WITMER, supra note 64, at 7. Unfortunately, the lobbyist they hired, Jack Abramoff, scammed them out of millions of dollars. Id.}

\footnote{308 For example, the Tigua Nation (Ysleta del Sur Pueblo) responded to the closure of their casino by hiring a lobbyist. CORNTASSEL & WITMER, supra note 64, at 7. Unfortunately, the lobbyist they hired, Jack Abramoff, scammed them out of millions of dollars. Id.}

For example, after facing extensive opposition from members of the local non-Indian community, the Gun Lake Band of Pottawatomie lobbied to reaffirm their ability to take land into trust.\footnote{309 Gale Courey Toensing, Gun Lake Trust Land Reaffirmed by Congress, INDIAN COUNTRY TODAY MEDIA NETWORK (Nov. 7, 2014), https://web.archive.org/web/20141107091053/http://indiancountrytodaymedianetwork.com/2014/11/07/gun-lake-trust-land-reaffirmed-congress-157741 [https://perma.cc/B48P-SZYL].}

\footnote{310 Mark Edwin Miller, Forgotten Tribes: Unrecognized Indians and the Federal Acknowledgement Process 253 (2004); see also Carlson, supra note 282, at 976 n.92 (noting that many of the non-federally recognized tribes that face tremendous state or local opposition to their recognition have not sought congressional recognition); Carlson, supra note 67, at 952.}

In other cases, strong opposition seems to have prevented Indian groups from engaging in legislative strategies. For example, Ramapough Chief Ronald Red Bone Van Dunk has suggested that his New Jersey-based tribe, which was denied federal recognition by the BIA, has not pursued legislative recognition because of the opposition it would face from Atlantic City.\footnote{311 The number of Indian groups gaining federal recognition slowed tremendously in the mid-1990s, about the same time that reported lobbying dramatically increased. See Carlson, supra note 282, at 974 (reporting the number of Indian groups gaining federal recognition R}
rates of compliance have changed over time, but the data suggest that the trend toward increased lobbying occurred prior to the changes in the lobbying disclosure regime which took effect in 1996. Moreover, the downturn in reported lobbying by American Indians from 1996 to 1998 seems contrary to the assertion that the new regime accounts for the increase in American Indian lobbying. It is also unlikely that the regime change accounts for the dramatic increase in reported lobbying by American Indians in the early 2000s.

Public choice theory also provides an alternative explanation for the increase in American Indian lobbying. In terms of legislative advocacy, either administratively or legislative from 1975 to 2013). Moreover, the federal recognition of Indian groups slowed even more after 2000, see id., yet reported lobbying by American Indians continued to increase, see infra fig. 5.

Further undermining the regime change explanation is the fact that reported lobbying by the general population increased in the late 1990s. ESKRIDGE ET AL., supra note 82, at 203 (noting that reported lobbying nearly doubled within six months after the Lobbying Disclosure Act of 1995 went into effect). The fact that reported lobbying by American Indians decreased suggests that the regime change was not driving their lobbying strategies.

Most scholars suspect that lobbyists underreport lobbying activities. LAPIRA, supra note 82, at 225 (discussing the existence of shadow lobbyists).

Public choice theory suggests that interest groups are more likely to influence Congress to enact client policies, or statutes that concentrate benefits on special interests while distributing their costs to the general public. JAMES Q. WILSON, AMERICAN GOVERNMENT: INSTITUTIONS AND PROCESSES 432–33 (5th ed. 1992). Conversely, interest groups are less likely to influence Congress to enact general or majoritarian policies that distribute benefits and costs broadly across large numbers of people or entrepreneurial policies that benefit large numbers of people at the expense of a small, identifiable segment of society. See, e.g., Wil-
public choice theory suggests that groups will have incentives to engage in legislative strategies when their interests are not salient to the larger public and are perceived as narrow, technical, or nonpartisan.\textsuperscript{316} To the extent that American Indians lobby on narrow issues specific to them, public choice theory seems to provide a viable explanation for the increase in lobbying. The evidence presented here, however, suggests that American Indian lobbying has increased generally across a range of issues affecting tribes both individually and nationally. Some of these issues are highly contested, of interest to the general public, and do not only benefit Indians. This suggests that public choice theories do not adequately explain the general increase in lobbying by American Indians over time.

VII. IMPLICATIONS FOR UNDERSTANDING STRATEGIC ADVOCACY AND FEDERAL INDIAN LAW

This Part explores some of the implications of this research for studies of advocacy strategies, interest groups, and federal Indian law. Part VII.A considers how the framework builds on and improves current understandings of how groups choose advocacy strategies. Part VII.B highlights the new and important questions that the Article raises for the study of federal Indian law.

A. Strategic Advocacy

The implications of this research for sociolegal and interest group studies are significant. The dramatic rise in American Indian legislative advocacy suggests that the conventional wisdom that groups litigate because they are foreclosed from the political process simply does not explain a world in which advocates increasingly rely on multi-institutional and non-judicial advocacy strategies.\textsuperscript{317} More complicated narratives exist for how groups respond to multiple, interactive, and often reinforcing influences in crafting advocacy strategies. As a result, sociolegal and interest group scholars need to develop more complex approaches to understanding how and why groups develop different advocacy strategies over time. This Article contributes to this larger project by presenting a more generalizable and nuanced approach to studying how groups make advocacy decisions across institutions over time.

This Article improves upon the sociolegal approach for understanding why groups litigate by adding insights from the interest group literature to provide more complete explanations of advocates’ strategic choices. This approach starts from the premise emerging in the literature that advocates choose strategies in response to social, institutional, and political opportuni-
ties and constraints. Borrowing from the interest group literature, it identifies the potentially relevant factors influencing the development of legislative advocacy strategies and integrates them into an interactive, sociolegal perspective. Viewing advocacy strategy development through this interactive lens allows for examination of how the dynamics and interactions among factors influence advocates’ strategic choices. This approach highlights the evolving nature of advocacy decisions, including the ways in which groups shape society through the claims they make and how those claims are in turn shaped by the institutions and social contexts in which they are raised. As a result, it provides a more comprehensive view of the choices advocates make and the opportunities and constraints they face in crafting strategies because it considers how institutional differences and relationships influence the choices advocates make. Revealing the multitude of considerations underlying advocates’ choices and the relationships among them, thus, allows for identification of the multiple pathways leading advocates to pursue different advocacy strategies.

My analysis of American Indian legislative advocacy serves as an illustrative example of the utility and richness of this approach. It demonstrates how groups respond to multiple interactive, and often reinforcing, influences in crafting advocacy strategies. Figure 6 portrays these interactions and how they contributed to the escalation in American Indian lobbying at the end of the twentieth century.

Contrary to conventional wisdom, American Indians increased their lobbying when the odds still appeared to be against them. Lobbying emerged as a viable strategy for American Indians in the late 1970s when the courts were still largely receptive to their claims and before gaming infused some tribes with new financial resources to support their efforts. A perceived change in the receptivity of Congress toward Indian claims reflected in the Tribal Self-Determination Policy opened the door for American Indians to lobby. The creation of the Senate Committee on Indian Affairs then provided access to Congress and new opportunities for American Indians to engage in legislative strategies. Lacking in financial resources and electoral clout, American Indians nonetheless used these opportunities to build relationships with members of Congress and their staffers, which further ensured their access to the legislative process. Thus, American Indians were committed to and skilled at using lobbying strategies by the time the Supreme Court became more hostile toward their claims in the late 1980s. The Supreme Court’s increasing reluctance to uphold tribal rights simply further pushed American Indians toward the legislative arena. Similarly, the emergence of gaming and the money it introduced into Indian country provided some Indian nations with the financial resources to lobby Congress more actively and reinforced the existing trend toward legislative advocacy by American Indians.

318 See supra Part III.
A few final observations suggest the broader implications of this approach for sociolegal and interest group studies. First, my approach shows how scholars can develop richer accounts of groups’ strategic decisionmaking by detailing how various factors interact to shape and reshape institutional choices over time. The role of the political context in facilitating American Indian legislative mobilization in the 1970s illuminates how considering the dynamics among variables produces a more comprehensive account. Congress’s support of Indian affairs through the creation of the Senate Committee on Indian Affairs signaled to American Indians that Congress might be receptive to their lobbying. As previous studies have shown, this government support encouraged legislative advocacy by creating a political context conducive to it. But the committee’s formation did more than signal support. It also changed the opportunity structure by providing access for American Indian advocates. These two factors—access and government support—occurred simultaneously and reinforced each other’s influence. The interaction of these two variables developed into a self-reinforcing pattern (or feedback loop) as American Indians used their increased access to garner more governmental support. Once they had access, they built relationships

---

319 See Baumgartner et al., supra note 11, at 3.
with members of Congress and their staff, which enabled them both to shape a political context favorable to their claims and to build their political capacity. Accounts that do not consider the dynamics among the variables often obscure or miss entirely these interactions and reinforcing mechanisms.

A second important implication of my account is that it indicates that scholars and advocates need to think more carefully and critically about the multitude of possible interactions among different institutions and how those interactions affect advocates’ strategic choices. Sociolegal scholars have long recognized the importance of institutional dynamics in the context of legal mobilization.\textsuperscript{320} For example, scholars have recently noted that judicial risk—or the recognition that the courts may not be willing to recognize rights as broad as a group wants—may discourage advocates from litigating.\textsuperscript{321} My findings, however, suggest that institutional dynamics may play more complicated roles in group advocacy decisions. American Indians did not engage in legislative strategies in a vacuum—they made calculated decisions about how best to pursue their legal claims in an ever-changing legal and political world. Like previous interest group studies, my account stresses the importance government support plays in groups’ decisions to lobby.\textsuperscript{322} The rise in American Indian legislative advocacy depended not only upon American Indians’ decisions to act but on a pro-Indian congressional policy and increased political access to members of Congress and the Senate Committee on Indian Affairs. But it diverges from most sociolegal accounts, which describe legislative advocacy as a response to a hostile court or an effort to enforce an important court decision.\textsuperscript{323} Instead, American Indians seized an opportunity to make their claims to Congress even though their chances of success seemed slim and they had alternative institutions open to them. These results contradict the classic argument that minority groups litigate because they are excluded from the political process and suggest that groups may turn to legislative advocacy more frequently and in different ways than previously thought. The traditional sociolegal emphasis on how groups use political strategies to supplement or implement litigation strategies may be too narrow. Advocates may pursue legislative advocacy in a variety of circumstances, including, but not limited to, as a parallel, an alternative, a complement, or a precursor to litigation.

Moreover, my findings highlight how advocates pay close attention to the various institutional options and the dynamics among them in crafting advocacy strategies. In contrast, most scholars have studied advocacy decisions based on advocates targeting one primary institution and paid less attention to how advocates consider the alternatives.\textsuperscript{324} Such singular approaches may be too simplistic and overlook important influences on

\footnotesize{\textsuperscript{320} CHEN & CUMMINGS, supra note 145, at 524; Steinman, supra note 171, at 763.} R
\footnotesize{\textsuperscript{321} CHEN & CUMMINGS, supra note 145, at 524.} R
\footnotesize{\textsuperscript{322} See, e.g., Leech et al., supra note 11, at 21.} R
\footnotesize{\textsuperscript{323} See, e.g., Cummings & Rhode, supra note 14, at 616–17.} R
\footnotesize{\textsuperscript{324} See supra Part II.A.}}
group decisionmaking, American Indians did not initially turn to Congress because they were losing in the Supreme Court, but as the Court and their perceptions of its receptivity towards their claims changed, they had more incentives to continue to use legislative strategies. Shifting institutional dynamics, thus, reinforced American Indians’ decisions to lobby and expanded their legislative strategies to include efforts to overturn unfavorable Supreme Court decisions. But advocates may have other reasons for pursuing multi-institutional venue strategies. They may want to maximize coordination among institutions, highlight problems in another institutional process, or play institutions off one another. My findings suggest that groups may choose strategies in relation to their assessments about other options and that their strategies may shift over time as their assessments change. As a result, scholars need to continually check their assumptions and pay more attention to institutional dynamics—when and why the political process may be more appealing to specific groups at certain times than the courts or other institutions.

My approach devises a way for scholars to study different venue options and the relationships among them in a more integrated fashion by incorporating consideration of the various institutional options and how they relate to one another into investigations of venue choice. This emphasis allows for examination of how institutional relationships affect how and why advocates choose particular strategies.

My results also contribute to a growing literature on how groups use the law in non-judicial settings by highlighting how American Indians have used legal arguments and frameworks in their legislative advocacy to craft substantive federal Indian law and policy beneficial to them. My findings build on these earlier studies by demonstrating how groups consider the relationships among courts and legislatures in their lobbying efforts. American Indians have sometimes paid particularly close attention to how the Supreme Court would interpret legislation during the drafting process. For example, in the case of the Duro fix, they drafted statutory language and developed a legislative history that limited the Court’s ability to interpret the statute against their interests. This indicates that advocates consider not only the options among institutions but also how the institutions may interact with one another later in time in their advocacy strategies. While this insight is not new, my findings illustrate how advocates consider future litigation in their legislative strategies as well as how litigation may affect politics or

325 See supra Part VI.C.
326 Steinman, supra note 171, at 763 (discussing how advocates could use conscious coordination of legal and political strategies to increase their chances of success); Paris, supra note 19, at 631 (same).
328 See, e.g., id. at 958–59.
Lobbying Against the Odds

require later political mobilization for enforcement. Accordingly, future studies should pay more attention to how advocates consider various interactions among institutions over time in crafting and coordinating more long-term advocacy strategies.

Finally, by questioning existing assumptions about the power of groups to use the political process, my results contribute to contemporary debates over how to understand and measure power and powerlessness for doctrinal purposes. They suggest that the dichotomy between powerful and powerless used to identify suspect classes in existing equal protection doctrine may oversimplify reality. Groups may exercise power in some contexts but not others. Thus, my research challenges scholars and judges to think more carefully about what power is, how to conceptualize and measure it, and what it means for groups to exercise it. Moreover, it suggests that future studies should investigate how, when, and why other marginalized groups exercise power in the political process and identify the different conditions leading to power and powerlessness.

B. Federal Indian Law

This Article also has important implications for the study of federal Indian law and American Indian advocacy. First, it debunks the prevailing myth that gaming has led to the increase in American Indian legislative lobbying by presenting original quantitative data on American Indians’ reported lobbying over a thirty-five year period. The collection and analysis of data on lobbying and lobbying expenditures by American Indians both before and after the rise of Indian gaming allows for important comparative analysis over time, missing from earlier studies. The analysis confirms earlier studies’ conclusions that gaming has created opportunities and incentives for Indian tribes to engage in lobbying. It also reveals that the increase in reported Indian lobbying predates the rise of gaming and that reported lobbying expenditures have not consistently increased since the advent of gaming. My findings, thus, indicate that gaming plays a more complicated role in American Indian lobbying on the federal level than the prevailing narrative suggests. They suggest the need for scholars to question their assumptions about the role of gaming in influencing American Indian lobbying and de-

329 See, e.g., Silverstein, Law’s Allure, supra note 144, at 1–3 (considering how litigation may affect politics).
331 See Witmer & Boehmke, supra note 65, at 132 (analyzing lobbying and gaming compacts after signing of IGRA in 1989).
332 See id.
velop more nuanced explanations of when and why American Indians craft advocacy strategies over time.

Second, the data reveal the staggering growth in American Indian legislative advocacy over the past thirty-five years. This incredible growth raises important questions about the efficacy of this advocacy. How successful are American Indians at enacting policies beneficial to them? How successful are they at preventing the enactment of policies detrimental to them? Are some tribes more successful than others? Does success depend on the issue being advocated on? These questions merit investigation. Even if they are not entirely successful in their advocacy efforts, American Indians are clearly playing a larger role in the creation of federal Indian law and policy than they did three decades ago when they lobbied less frequently. The tactics American Indians are using in lobbying deserve closer attention to determine if and how they are affecting the drafting, introduction, and progression of bills through the legislative process.

Third, the infusion of American Indian voices into the legislative process may contribute to substantive changes in the content of legislation governing American Indians, especially if they are playing an increased role in drafting legislation. As my account demonstrates, American Indians largely shaped the language of the legislation restoring tribal criminal jurisdiction over non-member Indians. They intentionally crafted the statutory text to limit the Supreme Court’s ability to overturn the statute. Their efforts extended beyond that statute as they used the language as a template for drafting the section of the Violence Against Women Act of 2013 that restored the inherent power of tribal governments to exercise criminal jurisdiction over all persons committing specific intimate-partner-related crimes in Indian country. American Indians may have an impact on the substantive context of legislation in other areas as well. The evidence shows that they lobby on a broad range of issues so the potential for such influence may extend beyond Indian-related policies. Future studies should more closely examine how exactly American Indians are engaging in legislative advocacy and what impact it could have on the substantive content of federal Indian law.

Finally, like recent studies on Indian nations’ use of interest group strategies on the state and local level, my findings show that Indian nations are increasingly turning to institutions other than the federal courts to influence

333 Public choice theories suggest that the issue lobbied on could affect American Indian legislative success. See supra Part VII.A.
334 For an initial exploration of how tribes have used legislative strategies to argue for policies beneficial to them, see Carlson, supra note 202.
335 See supra Part VI.C.
336 See, e.g., Congress May Restore Tribal Jurisdiction, supra note 272, at 1, 5.
337 For an investigation into how tribal lobbying efforts influence substantive provisions in federal statutes, see Carlson, supra note 202.
338 See supra Part VI.D.
the creation of federal Indian law. These findings run contrary to the bulk of federal Indian law scholarship, which has traditionally focused on court decisions. They suggest the need for more research into how federal Indian law is made in non-judicial settings.

VIII. CONCLUSION

Popular narratives maintain that groups lacking political power, electoral influence, and resources litigate because they cannot use legislative strategies to achieve their goals. Contrary to this conventional wisdom, the new empirical evidence presented in this Article documents a 600 percent increase in legislative advocacy by American Indians in the past three decades. This discrepancy suggests the need for more accurate explanations of how and why groups choose advocacy strategies. This Article presents a new approach for understanding how and why groups engage in advocacy strategies across institutions over time. It integrates the factors identified by interest group scholars as influencing advocates’ decisions to lobby into the interactive approach to strategic decisionmaking formulated by sociolegal scholars. This approach highlights the evolving nature of advocacy decisions, including the ways in which groups shape society through the claims they make and how those claims are in turn shaped by the institutions and social contexts in which they are raised. As a result, it provides a more comprehensive view of the choices advocates make and the opportunities and constraints they face in crafting strategies because it considers how institutional differences and relationships influence those choices.

The Article demonstrates the utility of this approach through a case study of American Indian advocacy from 1978 to 2012. American Indians started lobbying more frequently in the 1970s even though they lacked political power and the odds appeared stacked against them. My account produces a richer narrative about how and why this happened. It reveals that multiple factors influenced American Indians to pursue legislative strategies, and that these factors often interacted with one another to encourage or undercut the development of American Indian advocacy strategies over time. In particular, this account of American Indian legislative advocacy highlights how institutional dynamics may influence lobbying strategies. As a result, it encourages scholars and advocates to question the conventional narrative about when and why groups lobby and to think more carefully and critically about the multitude of possible interactions among different institutions and how those dynamics affect advocates’ strategic choices.

See Hansen & Skopek, supra note 64, at 212.
ARTICLE

THE STAFFER’S ERROR DOCTRINE

JESSE M. CROSS*

ABSTRACT

Over the last forty years, a new type of legislator has arisen in Congress: one who, rather than drafting statutes, instead manages a staff bureaucracy that produces these statutes. By becoming a manager of bills, not a drafter of them, this new legislator has altered a key relationship in our democracy: that between members of Congress and the laws they enact. Yet no study has documented how this modern relationship works—i.e., has chronicled how today’s federal legislators learn the contents of bills—and thereby shown the modern relationship between legislator and law. Nor has any study reflected on whether courts, in light of this altered relationship, need to change the way they interpret federal statutes.

This Article takes up these tasks. First, it reports the findings of an original empirical study—one that, through staffer interviews, chronicles the strategies that members of Congress now use to learn a bill’s contents. Its findings reveal that, in Congress, legislators’ understanding of a bill typically is based on the surprisingly brief memoranda and oral briefings they receive from staff. These sources educate members of Congress on legislative purpose, but they do not address the smaller details of statutory text, which generally are left to staffers.

In light of these empirical findings, the Article then argues that courts should adopt a new “staffer’s error doctrine.” Under this doctrine, before a court applies the plain meaning of a statute, it first confirms that statutory text does not undermine statutory purpose. In the era of managerial legislators, this check provides a useful proxy: it protects legislator decisions from staffer errors. In this way, the doctrine takes a neglected principle from the old scrivener’s error doctrine—that courts interpreting statutes should review the work of unelected legislative staffers for mistakes—and updates it to address modern congressional realities.

This new doctrine has several merits that the Article highlights. First, it is compatible with a wide range of interpretive theories—not only intentionalism, but also at least four prominent varieties of textualism. Second, the doctrine aligns with—and makes sense of—the Court’s recent direction in statutory interpretation. This merit is reflected in King v. Burwell, the landmark case interpreting the Affordable Care Act. In King, the Court essentially (if unwittingly) performed the exact check required by this new doctrine. In so doing, King showed that the staffer’s error doctrine is a workable doctrine—and the doctrine, in turn, shows that the interpretive approach in King enjoys previously unnoticed claims to methodological legitimacy.

* Assistant Professor, University of South Carolina School of Law. Yale Law School, J.D. 2011; University of California, Irvine, M.A. 2006. I particularly wish to thank William Eskridge, Jr., Abbe Gluck, Bruce Ackerman, John Witt, Nicholas Parrillo, Jack Balkin, Claire Priest, and all the participants of the 2016 Legislation Roundtable for their helpful and generous feedback. I also wish to thank my former coworkers at the Office of the Legislative Counsel for the U.S. House of Representatives, as well as the many congressional staffers who generously contributed interviews to this Article.
I. INTRODUCTION .................................................... 84

II. THE STAFFER-DELEGATION MODEL OF LAWMAKING ...... 89
   A. Congress’s Models of Delegation ............................ 89
      1. Preceding Delegation Models ............................. 89
      2. The Staffer-Delegation Model .............................. 91
   B. The Necessity of Delegation .................................. 110

III. TEMPLATE FOR A NEW DOCTRINE: KING v. BURWELL ...... 117
   A. The Court’s Reasoning in King v. Burwell ................. 117
   B. The Court’s Reasoning and the Staffer-Delegation Model ..................................................... 120

IV. THE STAFFER’S ERROR DOCTRINE ............................ 121
   A. Precedent: The Scrivener’s Error Doctrine ................. 121
   B. Administration of the Staffer’s Error Doctrine .......... 124
   C. The Legitimacy of the Staffer’s Error Doctrine .......... 127
      1. Intentionalism and the Staffer’s Error Doctrine ....... 127
      2. Textualism and the Staffer’s Error Doctrine .......... 131

V. CONCLUSION: A NEW (YET OLD) THEORY OF CONGRESS..... 138

I. INTRODUCTION

Over the last forty years, Congress has undergone a fundamental transformation. Prior to this period, Congress could have been regarded—quite accurately—as a quaint institution populated only by legislators and a small cadre of assistants. That no longer is true, however. Today, thousands of legislative staffers populate the halls of Congress—staffers who perform key legislative tasks. As a result, Congress has been transformed into something that it did not resemble prior to the 1970s: a large, modern bureaucratic institution.

The rise of this bureaucracy has led, among other things, to the emergence of a new type of legislator in Congress. Occupying a position atop an expansive legislative bureaucracy, this new legislator now spends much of her time performing a CEO-like task: namely, the task of managing subordinate staffers. This shift by members into a managerial role has fundamentally altered the way in which Congress drafts its statutes—yet courts...
continue to use interpretive rules and methods that assume that Congress drafts its statutes the same way it did a century ago.

This Article aims to redress this state of affairs. To accomplish this, it pursues two specific goals. The first is descriptive: namely, to document the relationship that exists between members of Congress and modern statutes. It is a relationship that differs significantly from that which is assumed to exist in much of the legislation scholarship. Here, the Article observes that, as a result of the shift by members of Congress into a managerial role, these members no longer participate in the drafting of statutory text, and they rarely participate in the pre-enactment review of this text. Instead, the Article reveals, their pre-enactment knowledge about the content of bills typically comes from summaries that outline the purpose of statutory provisions—summaries provided either via memoranda or in-person staffer briefings. As such, the staffer-delegation system that prevails in the modern Congress has produced a situation in which members of Congress, as a routine matter, provide direct review only of the purpose of statutory provisions—a purpose that they expect congressional staffers to flesh out through policy development and statutory drafting.

This descriptive portion of the Article draws on the work experience of the author (who worked for six years as a drafter of congressional statutes), and it is supported by an original empirical study conducted for this Article. This study, which consisted of interviews with current and former congressional staffers, documents this modern division of congressional labor—and it sheds light on its inner workings. In so doing, it reveals certain practices that might be surprising—practices such as the “one-page rule,” which provides that bill summaries for some members should not exceed a page in length, and the “vote rec” strategy, which sometimes restricts bill summaries to a vote recommendation that is written on a single notecard. Through an examination of these and other practices (and, it should be noted, the “one-page rule” and the “vote rec” strategy do not operate in a vacuum), this Article reveals and focuses attention upon a central fact about the modern Congress: namely, that many modern drafting and briefing practices are tailored to provide members, in the preponderance of cases, with knowledge only about the outlines of statutory purpose, thereby entrusting to staffers the fine details of statutory text.

As the following pages will explain, this internal division of labor is not merely the result of members of Congress being busy, distracted, or lazy. Rather, it is a result—at least in part—of the fact that members of Congress reside at the intersection of two fundamental and conflicting forces. On the one hand, members are subject to a constitutional mandate that they operate as generalists. They must represent their constituents with respect to a wide variety of topics and issues—and the Constitution provides no requirement

Shepsle, supra note 2, at 559 (“[A]s a consequence of staff expansion each member of Congress has come to operate as the head of an enterprise . . . .”).
that they bring anything other than a generalist’s knowledge to this position. On the other hand, they are expected to produce legislation that reasonably responds to, and intervenes in, a world of private and governmental institutions that has grown enormously complex. It is a world in which the crafting of legislation, if it is to be done responsibly, requires domain-specific expertise. In response to this conflicting set of imperatives, members of Congress have established a system that makes good sense. Under this system, members typically approve statements of statutory purpose, and they delegate to staffers the task of carrying out this purpose via statutory text.

The first goal of this Article, therefore, is to describe this modern drafting reality. The second goal, meanwhile, is to advance an argument about how courts ought to respond to this new congressional practice. The fundamental change in congressional drafting practice that is documented in this Article gives rise, after all, to an important set of questions in statutory interpretation—questions that the field of legislation largely has failed to confront. Indeed, in the relatively rare instances in which legislation scholars have directed their attention to the increase in delegations to congressional staffers, they have done so simply in order to consider its implications for the debate about whether and how courts should use legislative history in statutory interpretation. Yet this new division of labor has implications that extend well beyond the proper use of legislative history—implications for our basic understanding of how courts should approach statutory text itself. Consequently, in addition to documenting this division of labor, this Article also traces its implications for the ways that courts can, and should, interpret statutes.

In particular, this Article argues that courts ought to adopt a new doctrine—referred to here as the “staffer’s error doctrine”—that acknowledges and addresses this shift in drafting approaches. The Article finds a template for this doctrine in the Court’s opinion in King v. Burwell, the 2015 case interpreting a provision of the Patient Protection and Affordable Care Act. In this case, the Court essentially checked to ensure that statutory plain meaning did not undermine statutory purpose. This particular type of review, this Article argues, can be viewed as legitimate—and perhaps even necessary—in light of the way in which Congress now divides its labor when drafting statutes.

In what sense, it might be wondered, does the division of labor in Congress legitimate this interpretive approach? It does so by revealing that the two relevant elements of the statute—statutory purpose and statutory text—

---

4 See infra note 137 and accompanying text.
are managed and reviewed by different actors within Congress. In a world in which statutory purpose is approved by members of Congress and in which statutory text is drafted by staffers, the review found in King v. Burwell takes on a new dimension: it operates to ensure that a congressional staffer has not inserted an error into statutory text that undermines a decision made by members of Congress.

Once this interpretive method is viewed as a review of the work product of unelected staffers, it acquires new claims to legitimacy. In statutory interpretation, there is a longstanding tradition which holds that courts can review the drafting work of unelected staffers—a review designed to ensure that this work has not undermined the decisions made by elected representatives. In fact, there exists a doctrine of statutory interpretation that, although often misunderstood, is designed to enable courts to perform this exact review: the scrivener’s error doctrine.

As the following pages explain, the scrivener’s error doctrine ostensibly was designed to police the work of Congress’s internal agents—a policing function designed to ensure that these agents do not exceed or undermine the delegation that members of Congress have entrusted to them. As Congress has delegated an increasing number of lawmaking tasks to staffers in recent decades, however, the scrivener’s error doctrine has retained the antiquated assumption that Congress delegates a narrow set of clerical drafting tasks—and only these tasks—to staffers (or “scriveners”). When courts review statutory text to ensure that it accords with legislative purpose, by contrast, they actually do check to ensure that the work of today’s staffers accords with the legislative instructions they were given by elected representatives, thereby performing this policing function. In this regard, they effectively create an updated version of this doctrine—a version that this Article refers to as a “staffer’s error doctrine.” Under this updated doctrine, the principles that animated the original scrivener’s error doctrine are combined with modern congressional realities and, through this combination, are made relevant once again.

By finding this doctrine embedded within the Court’s opinion in King v. Burwell, this Article reveals that King and the staffer’s error doctrine each contribute something to the other. On the one hand, King illustrates that the “staffer’s error doctrine,” as endorsed by this Article, is a workable doctrine. On the other hand, the doctrine provides new theoretical underpinnings that can justify the new approach to statutory interpretation that the Court inaugurated in King.

In further defense of the doctrine, this Article argues that, regardless of whether statutory interpreters consider themselves to be intentionalists or textualists, they should be persuaded that the doctrine is consistent with the underpinnings of their preferred interpretive methodology. As part of this defense, this Article observes that the doctrine is compatible with at least four different variants of textualism.
Finally, this Article concludes by suggesting that, in light of this Article’s findings, it might be useful for the field of legislation to undergo a shift away from the visions of Congress espoused by textualists and intentionalists—a shift instead toward the competing vision of Congress found in the field of administrative law. In administrative law, it has been widely accepted that delegation is a fundamental quality of modern lawmaking. In response to this fact, administrative law has embraced a core principle: namely, that courts ought to review the work of Executive Branch agents in order to ensure that those agents have not exceeded the scope of their delegations from Congress. Due to fundamental changes that Congress underwent in the 1970s and 1980s, this Article explains, delegation to unelected agents has become a similarly defining feature of lawmaking within Congress. Consequently, federal lawmakers—whether outsourced to agencies or conducted within Congress—is now marked by delegation. Building upon this insight, this Article asserts that courts should be reviewing these two types of lawmaking similarly; in each instance, courts should be ensuring that agents of Congress did not exceed the scope of their delegations. It is an interpretive conclusion that is intelligible only when the delegation-based dimension of the modern Congress is made visible—a dimension that is obscured by the theories of Congress offered by textualists and intentionalists.

This argument is made in five parts. Part II documents the historical rise—and the present necessity—of a congressional lawmaking process that this Article describes as the “staffer-delegation model of lawmaking.” Moreover, it reports the findings of this Article’s empirical study, thereby chronicling the processes and documents that, in this staffer-delegation model, are used to educate members of Congress about the contents of bills. In so doing, Part II draws attention to an important consequence of this model of lawmaking: the fact that members of Congress no longer provide, as a routine matter, direct review of statutory text. Part III reviews the Court’s opinion in King v. Burwell, the case that will provide a template for the “staffer’s error doctrine,” and it explains how the Court’s ruling takes on a new dimension when viewed in light of the staffer-delegation model of lawmaking. Part IV shows that the review performed by the Court in King revives, albeit accidentally, a type of review that traditionally has been embedded in the scrivener’s error doctrine—a review that courts have failed to perform effectively, however, as the doctrine has stagnated in the staffer-delegation era. Part IV additionally shows that, once the Court’s review in King is seen as a revival of the principles embedded in the scrivener’s error doctrine, the opinion gains new claims to legitimacy—claims that should appeal to both intentionalists and textualists. Finally, Part V concludes by suggesting that, going forward, the field of legislation may benefit from a shift away from the visions of Congress espoused by textualists and intentionalists—and toward the vision of Congress found in the field of administrative law.
The Staffer’s Error Doctrine  

II. THE STAFFER-DELEGATION MODEL OF LAWMAKING

For the past several decades, Congress has produced statutes through a specific process. Under this process, members of Congress delegate significant lawmaking tasks to staffers within Congress—and these members, rather than being involved in the myriad details of policy development and statutory drafting, instead adopt a managerial role. This role allows members to provide broad policy guidance and oversight throughout the drafting process yet to delegate to staffers (both partisan and nonpartisan) the subordinate tasks of developing policy details and statutory language.

This Part documents the rise, the inner workings, and the necessity of this model of lawmaking—referred to here as the “staffer-delegation model” of lawmaking. In so doing, it highlights the most important consequence of the staffer-delegation model: namely, that most members of Congress are no longer connected, in any direct sense, to statutory text, and instead are engaging with statutes at the level of legislative purpose.

A. Congress’s Models of Delegation

1. Preceding Delegation Models

Congress has not always relied upon the staffer-delegation model of lawmaking. As Section 2 explains, this model did not emerge until the post-New Deal era, and it was not the default method of lawmaking in Congress until at least the 1970s. This does not mean, however, that members of Congress never delegated lawmaking tasks to subordinate entities prior to the 1970s. Rather, there were two models of delegation that preceded the staffer-delegation model.

The first of these models was the committee-delegation model. Under this model, Congress would assign a relatively small number of its members to a committee, and this group would be responsible for performing a variety of lawmaking tasks. In the Founding era, for example, the drafting of statutory text often was assigned to ad hoc drafting committees that were comprised of members of Congress. By the second decade of the 1800s, the House of Representatives had established standing committees that were expected, in addition to drafting statutory text, to perform initial policymaking work on bills. The Senate would mimic this turn to standing committees

---

8 BARBARA SINCLAIR, UNORTHODOX LAWMAKING: NEW LEGISLATIVE PROCESSES IN THE U.S. CONGRESS 5 (2011) (citing Joseph Cooper & Cheryl D. Young, Bill Introduction in the Nineteenth Century: A Study of Institutional Change, 1 LEGIS. STUD. Q. 67 (1989)). See generally Norman K. Risjord, Congress in the Federalist-Republican Era, in 1 ENCYCLOPEDIA OF THE AMERICAN LEGISLATIVE SYSTEM (Joel Silbey ed., 1994). This followed the practice used at the Constitutional Convention, where the full membership of the Convention debated and approved constitutional provisions, while a “Committee of Detail” comprised of Convention delegates drafted and assembled these provisions.

9 SINCLAIR, supra note 8, at 5.
soon thereafter. This reliance upon committees persisted throughout the nineteenth century and, by 1885, it had increased to the point where Woodrow Wilson could make his famous statement that “Congress in its committee-rooms is Congress at work.” This reliance would only further increase during the first half of the twentieth century, to the point where committees became vital policymaking engines within Congress.

As the volume and complexity of congressional legislation increased in the New Deal era, meanwhile, a second model of delegation also emerged: the agency-delegation model. Since each executive agency employed a large collection of individuals who were focused upon a single area of legislative activity, the delegation of lawmaking tasks to these agencies allowed Congress to have its legislation shaped, as James Landis famously phrased it, by “men bred to the facts.” It was an option that Congress increasingly utilized as the demands for complex legislation increased up to, and during, the New Deal era.

Throughout the New Deal, Congress used at least two different strategies to accomplish this delegation to agencies. First, Congress enacted a variety of statutes that provided agencies with subordinate rulemaking authority. By the end of the New Deal, this strategy had become fully entrenched, and early debates about its constitutionality had largely been settled. Second, Congress would delegate the initial drafting of statutes—and, in some cases, the pre-enactment review of statutes—to executive agencies. As Nicholas Parrillo has documented, this method of producing statutory text predominated during the period of 1933 to 1945, and it continued to be utilized to some extent in subsequent decades.

10 Id.
12 See Stephen S. Smith et al., The American Congress 6 (9th ed. 2015) (detailing the ways in which “[i]n the mid-twentieth century . . . many standing committees and their chairmen were the dominant players in designing legislation”). Acknowledging that this model divorced non-committee members from statutory text, Learned Hand observed:

It is of course true that members who vote upon a bill do not all know, probably very few of them know, what has taken place in committee. . . . [Courts] recognize that while members deliberately express their personal position upon the general purposes of the legislation, as to the details of its articulation they accept the work of the committees; so much they delegate because legislation could not go on in any other way.

Sec. & Exch. Comm'n v. Collier, 76 F.2d 939, 941 (2d Cir. 1935).
14 Since the Court’s “switch in time” in 1936, it has upheld every statute that has come before it under the nondelegation doctrine. For a discussion of this post-New Deal history of the doctrine, see, for example, William N. Eskridge Jr., Abbe R. Gluck & Victoria Nourse, Statutes, Regulation, and Interpretation 88 (2014) (“No federal statute has been invalidated by the Supreme Court on nondelegation grounds since the 1930s.”).
The Staffer’s Error Doctrine

Today, however, it can no longer be said that committees or agencies are the primary drafters of statutory text. Instead, the agency-delegation model has given way to yet another model of statutory production: the staffer-delegation model.

2. The Staffer-Delegation Model

i. The Historical Rise of the Model. For much of our nation’s history, lawmaking was conducted without any significant delegation of lawmaking duties to staffers. Congressional staffers did not even exist during the first six decades after the ratification of the Constitution.16 Some staff were hired during the period from 1840 to the 1920s, but their numbers were few and their role was marginal.17 By 1924, the committees of Congress collectively employed a grand total of only 261 committee aides, for example.18

The number of congressional staffers would increase only minimally in the following two decades.19 Moreover, these pre-1940 congressional staffers performed only clerical functions within Congress.20 Such staffers typically did not possess the legislative expertise needed to assist Congress in its lawmaking duties.21 It was estimated that, out of seventy-six committees in existence at this time, only four had appointed any experts.22 In 1941, there were “not more than two hundred persons [employed by Congress] who could be considered legislative professionals.”23 As such, staffers did not perform duties prior to the mid-1940s that could realistically be characterized as lawmaking functions. As one scholar put it, “[u]ntil 1947 the standing committees did not regularly employ professional staffs. When such staff

16 Fox & Hammond, supra note 1, at 15.
17 See id. (describing committees first receiving authorization to hire part-time clerks in the 1840s, two committees receiving authorization to hire full-time clerks in 1856, and senators receiving authorization to hire staffers in 1885); id. at 12; see also Printing Reform Act of 1919, Pub. L. No. 65-314, 40 Stat. 1213 (1919) (regularizing employment of member-based staffers).
18 Fox & Hammond, supra note 1, at 15 (“By 1924, 120 committee aides were employed in the House and 141 in the Senate.”).
19 Id. at 20.
20 Id. at 15; see also id. at 21 (quoting Rep. Ramspeck (D-Ga.) stating that, prior to the 1946 Act, the staff on the Agriculture Committee of the House “had[d] no special training for what they [were] doing, they had[d] no special knowledge of the problems of agriculture, they [were] merely a clerical staff”).
21 The exception to this was the House and Senate Appropriations Committees and the Joint Committee on Taxation, which had nonpartisan professional staffs as early as the 1920s. See Cong. Quarterly Press, How Congress Works 190 (5th ed. 2013); see also Fox & Hammond, supra note 1, at 15 (“At the time of the Legislative Reorganization Act in 1946, committees such as Appropriations had a history of expert, nonpartisan staff—a tradition which did not extend to most other committees.”).
22 Fox & Hammond, supra note 1, at 20 (citing WOL Radio Address by Representative Mike Maroney, April 1943, in Special Joint Committee on the Organization of Congress, 79th Congress, The Organization of Congress: Symposium on Congress 148 (Comm. Print 1945)).
assistance was needed, it was borrowed on detail from an executive agency . . . .”

Moreover, staffers who worked for nonpartisan offices of Congress similarly were confined to clerical duties during this period.

The result was a Congress that, through the mid-1940s, was still largely dependent upon the agency-delegation model of lawmaking. By the 1940s, however, this dependence began to operate as a source of frustration for some members of Congress. As Fox and Hammond put it, “[t]here was a general feeling by 1946 that the executive branch had increased tremendously, and that the imbalance needed to be corrected.” This need was dramatically articulated by Representative Thomas Lane (D-Mass.), for example, who would lament that:

As regularly as the clock strikes the hour, and enforced by that striking, every Member is reminded every day of two facts, as closely allied as the hands of the clock. The first is that the demands on his time are incessant and even oppressive, and that his sources of information and assistance inadequate. . . . [The second is that we] must rely on the very representatives of the Federal agencies for information when we are trying to exercise our supervision of their carrying out of the policy we have prescribed.

In response to these frustrations, Congress enacted the first of several statutes that would allow for a steady, decades-long increase in the number of (and dependence upon) expert congressional staffers: the Legislative Reorganization Act of 1946. This act provided the first authorization for committees to retain professional staff—an authorization that soon was extended to individual Senators as well. In the wake of the act’s passage, the number of congressional staffers—as well as the expertise of these staffers—increased notably.

24 Kammerer, supra note 1, at 3.
25 See Jarrod Shobe, Intertemporal Statutory Interpretation and the Evolution of Legislative Drafting, 114 Colum. L. Rev. 807, 822 (2014) (on Legislative Counsel being limited to formatting duties); id. at 834 (on Congressional Research Service’s narrow original responsibilities).
26 Fox & Hammond, supra note 1, at 20–21; see also id. at 21 (“The committee staffing discussion [in Congress] focused on the need for technical competence [and the need for] an information analysis capability independent of the executive branch . . . .”.
27 See id. at 22 (quoting 92 Cong. Rec. H10054 (daily ed. July 25, 1946)).
29 Fox & Hammond, supra note 1, at 14 (describing this change and noting that it marked “a watershed in the area of staffing”); see also Kammerer, supra note 1, at 4 (similarly describing this change). Professional aides to representatives would not be recognized until 1970. Fox & Hammond, supra note 1, at 12.
30 Fox & Hammond, supra note 1, at 12.
31 Id. at 22 (citing Kammerer, supra note 1, at 34).
Nonetheless, the 1946 Act was not sufficient, by itself, to cause an immediate shift away from the agency-delegation model of lawmaking.\textsuperscript{32} Instead, Congress would need to pass an additional act: the Legislative Reorganization Act of 1970.\textsuperscript{33}

As Jarrod Shobe has put it, the Legislative Reorganization Act of 1970 “paved the way for the modernization of legislative drafting.”\textsuperscript{34} The act accomplished this in a variety of ways. It provided representatives with statutory recognition of their ability, like their counterparts in the Senate, to retain professional staff.\textsuperscript{35} It created the modern Congressional Research Service, granting the research-focused office autonomy and allowing it to increase significantly in size and clout.\textsuperscript{36} It expanded the authority of the House Office of the Legislative Counsel, paving the way for notable increases in the size and capacity of this drafting office (and setting the stage for a similar expansion of the office’s Senate counterpart).\textsuperscript{37} And it enabled changes to the committee system that empowered Congress to hire, retain, and rely upon committee staffers in the performance of its lawmaking duties.\textsuperscript{38}

Throughout the 1970s, Congress would pass a series of measures that would further expand the number of professional staffers working for Congress. The Committee Reform Amendments of 1974 tripled the staffs of most standing committees of the House of Representatives, including with respect to professional staff.\textsuperscript{39} Several other measures were passed in the 1970s that allowed for similar staff increases.\textsuperscript{40} Meanwhile, the Budget and Impoundment Control Act of 1974 created the Congressional Budget Office (“CBO”), an office that would produce important economic analyses for members—and that eventually would play a central role in the legislative process.\textsuperscript{41}

As with the 1946 Act, these new measures were part of a coherent, self-conscious effort in the 1970s to build an internal bureaucracy that would allow Congress to leave behind the agency-delegation model of statutory drafting. The scandals of the Nixon administration had given rise to a new

\textsuperscript{32} See Parrillo, supra note 15, at 338 (citing KAMMERER, supra note 1, at 41–42); Harry W. Jones, Drafting of Proposed Legislation by Federal Executive Agencies, 35 A.B.A. J. 136, 137 (1949).


\textsuperscript{34} Shobe, supra note 25, at 816.

\textsuperscript{35} Pub. L. No. 91-510, §§ 301–305, 84 Stat. 1140, 1175–81 (1970; see also Fox & Ham mond, supra note 1, at 24).

\textsuperscript{36} See Pub. L. No. 91-510, § 21; see also Shobe, supra note 25, at 321–22.

\textsuperscript{37} Pub. L. No. 91-510, §§ 501–531; see also Shobe, supra note 25, at 818–23.

\textsuperscript{38} Pub. L. No. 91-510, §§ 501–306; see also Cong. Quarterly Press, supra note 21, at 192.

\textsuperscript{39} See H.R. Res. 988, 93d Cong. (1974); see also Cong. Quarterly Press, supra note 21, at 188.

\textsuperscript{40} For a detailed accounting of these statutes, see Cong. Quarterly Press, supra note 21, at 190.

wariness of Executive Branch dependence. In response, Congress undertook these efforts to increase congressional staffing—efforts that, as one longtime Beltway journalist has put it, were Congress’s way of “building a bureaucracy to fight a bureaucracy.”

Due to these and other enactments, the number of staffers employed by Congress increased steadily and significantly in the 1970s and 1980s, thereby creating the robust congressional bureaucracy that persists today. The number of subcommittee staffers employed in the House of Representatives increased by 650 percent during the 1970s, for example. Between 1975 and 2011, the number of attorneys employed by the Offices of the Legislative Counsel tripled; by 2011, these offices together employed more than eighty attorneys. The Congressional Research Service would expand into a formidable office that employs more than 400 experts and specialists. The number of committee staffers would more than double between 1965 and 2009, expanding to include over 2200 employees. And the number of personal staffers would nearly double between 1965 and 2009, leaping to over 5800 employees.

The growth of congressional staffs, it should be noted, has not been uniform during these periods; staff numbers for many congressional offices peaked in the mid-1980s, and then subsequently retreated modestly. Nevertheless, as the comparisons to pre-1980s staffing numbers make clear, these

42 See Hedrick Smith, The Power Game: How Washington Works 281–82 (quoting Sen. Moynihan (D-N.Y.) on this motivation); see also id. at 290 (describing CBO as “representing the most important institutional shift of power on domestic issues between the executive branch and Congress in several decades” and quoting a domestic policy chief in the Carter White House on this point).

43 See Fox & Hammond, supra note 1, at 13 (“Change has occurred most rapidly in recent years.”); Salisbury & Shepsle, supra note 2, at 559 (“[I]n the years between 1967 and 1979 the number of congressional committee staff and personal staff employees grew from 7,014 to 13,276; the total number of employees (staffers and others) [by 1981 was] well over thirty thousand.”).

44 Cong. Quarterly Press, supra note 21, at 193. The number of Senate subcommittee staffers increased by only 50 percent during the same period. Id. at 191. S. Res. 60 provided senators with other ways to access increased professional staffing, however. See S. Res. 60, 93d Cong. (1973).


47 Vital Statistics on Congress, Brookings Inst., ch.5, at 10 tbl5-5 (2017), https://www.brookings.edu/wp-content/uploads/2017/01/vitalstats_ch5_full.pdf [https://perma.cc/7CSZ-RGK5]. On the increase in personal legislative staff of members, albeit during a slightly earlier period, see Fox & Hammond, supra note 1, at 25 (describing the drastic increase in the number of legislative aides between 1960 and 1975); see also id. at 12 (“Sizable staffs, capable of sophisticated services, now exist [in 1977] within Member offices and committees.”).


recent staff reductions still have left intact a formidable congressional bureaucracy—one that carries the staffer-delegation model of lawmaking into the present day.

ii. The Consequences of the Staffer-Delegation Model. As these staffer offices grew in size after 1970, their employees were able to specialize to a far greater extent than ever before. In the Congressional Research Service, for example, the increase in size allowed for the Service to be divided and subdivided into areas of intensive expertise, with “each attorney working in just a few legislative areas throughout his or her career,” as one scholar has put it.50 The Offices of the Legislative Counsel underwent a similar specialization,51 as did committee staffs.52

With this new specialization, not surprisingly, came a change in the balance of functions performed by members of Congress, as opposed to those performed by congressional staffers. No longer were these staffers simply performing minor clerical tasks. Instead, staffers began to perform three important lawmaking tasks: policymaking, statutory drafting, and pre-enactment review of statutory text.

Each of these shifts in responsibility is worth considering further. First, staffers had begun, by the late 1970s, to assume significant policymaking responsibilities. This shift was noted in several contemporaneous studies of Congress. A survey of the literature on congressional staffs from 1981, for example, observed that: “[I]n many specific situations, [member policy instructions to staffers] constitute no more than a general guide to action.”53 In such instances, the authors observed, members relied upon “the often important role of staff in negotiating legislative substance.”54

These trends only continued throughout the 1980s. Speaking in 1986, Senator William Cohen (R-Me.) could observe that:

[N]o single person can keep up with the sprawling substance of policy. . . . [A member] will dart into one hearing, get a quick fill-in from his staffer, inject his ten minutes’ worth and rush on to the next event, often told by an aide how to vote as he rushes onto the floor. Only the staff specialist has any continuity with substance. The member is constantly hopscotching.55

---

50 Shobe, supra note 25, at 837.
51 Id. at 822.
52 Id. at 845; see also Fox & Hammond, supra note 1, at 144 (“Today [i.e., in 1977] staff are more expert and are becoming increasingly specialized.”).
53 Salisbury & Shepsle, supra note 2, at 567; see also Fox & Hammond, supra note 1, at 143 (describing Senators as “giving direction to policy and giving staff the responsibility for the details”).
54 Salisbury & Shepsle, supra note 2, at 569.
55 Smith, supra note 42, at 292.
The result of this situation, Cohen added, was that: “You [as a senator] skate along on the surface of things. More and more you are dependent on your staff.” Consequently, Cohen noted, staffers were routinely making significant policymaking decisions by the mid-1980s.

In their efforts to describe this new division of labor within Congress, scholars in the 1970s and 1980s repeatedly turned to the same analogy: members of Congress, they observed, now more closely resembled a corporate CEO than a traditional legislator. Rather than make each minor policy decision for themselves, members would now oversee a set of subordinates who were empowered to make these decisions. These members shifted into a supervisory role—a role that allowed them to retain responsibility for reviewing larger scale decisions that subordinates believed warranted the attention of their principals.

As part of this larger retreat into a managerial role, members of Congress also increasingly shifted responsibility for statutory drafting to staffers. Stuart Eizenstat, former domestic policy chief in the Carter White House, would remark in 1986 that members were routinely delegating statutory drafting of tax bills to the nonpartisan staffers on the Joint Tax Committee. As Eizenstat said:

The actual drafting of [tax] legislation, the fine points, the statement of managers, the explanations, the things that can make millions and millions of dollars of difference have to be done by Joint Tax because of the complexity and technicality of the tax code. . . . They’re the only ones who can deal with it. And every tax conference of necessity leaves significant areas for the Joint Committee to fill in. Significant areas, involving tens of millions of dollars in decision.

Notwithstanding Eizenstat’s observation about the role of the Joint Tax Committee, however, the offices that would assume nearly sole responsibility for statutory drafting during these decades were the Offices of the Legislative Counsel. As these offices became larger and more specialized, they assumed an increasingly large role as the primary drafters of statutory text. Today, Legislative Counsel drafts nearly every bill that is introduced in Congress. In their recent survey of contemporary congressional drafting practices, one staffer told Abbe Gluck and Lisa Schultz Bressman that “99% [of

56 Id. at 289–90.
57 See id. (quoting Cohen as saying: “[S]ometimes you’ll call a senator and ask, ‘Why are you opposing me on this?’ and he’ll say, ‘I didn’t know I was.’ And you’ll say, ‘Well check with your staff and see.’”).
58 See, e.g., Fox & Hammond, supra note 1, at 143 (“Senators, indeed, are functioning more and more like the president or chief operating officer of a corporation, giving direction to policy and giving staff the responsibility for the details.”); Salisbury & Shepsle, supra note 2, at 559 (“[A]s a consequence of staff expansion each member of Congress has come to operate as the head of an enterprise.”); Breyer, supra note 3, at 858–89.
59 Smith, supra note 42, at 279.
statutory text] is drafted by Legislative Counsel,” and they were told by another that “[n]o staffer drafts legislative language. Legislative Counsel drafts everything.”60 Indeed, the Rules Committee of the House frequently will not even consider amendments to bills unless the members of Congress have used the services of Legislative Counsel to draft the text of the amendment.61

As Gluck and Bressman point out, the assumption of drafting responsibilities by Legislative Counsel has a noteworthy implication: it reveals “an important disconnect . . . between members and statutory text.”62 A 2002 study coauthored by Victoria Nourse—a former counsel to the Senate Judiciary Committee—and Jane Schacter uncovered the same disconnect, asserting that: “Most staffers indicated [in a survey] that, as a general rule, senators themselves did not write the text of legislation.”63 Ganesh Sitaraman—a former Senior Counsel for Senator Warren64—similarly has noted in a law review article that: “[S]taff drafts virtually everything; members almost never write or edit legislative text.”65 Sitaraman adds:

The overall picture that emerges from understanding the workings of a congressional office is that [members of Congress] are not drafters but rather decisionmakers. They are managers of a mini-bureaucracy who set the direction for policy and sometimes wade into the details of policy, but who rarely get into the technical work of legislative drafting.66

Ted Kennedy similarly attested to this disconnect between members and statutory text in his 2009 memoir. Kennedy stated that: “Ninety-five percent of the nitty-gritty work of drafting and even negotiating is now done by staff. That alone marks an enormous shift of responsibility over the past forty or fifty years.”67 Reflecting on Kennedy’s observation in light of his journalistic study of the Dodd-Frank bill, Robert Kaiser would add:

In the case of Dodd-Frank, 95 percent might understake staff members’ share of the work . . . [Staffers] were responsible for every aspect of producing the final legislation: writing provisions (most

62 Bressman & Gluck, supra note 60.
66 Id. at 90–91.
Harvard Journal on Legislation [Vol. 56

based on Obama administration drafts), vetting the contents with interest groups of all kinds, looking for glitches or omissions, and hearing out the recommendations and complaints of hundreds of experts, lobbyists and affected parties.68

As the work by the aforementioned scholars makes clear, the delegation that occurred with respect to the Dodd-Frank bill was no exception. Rather, the delegation of statutory drafting to staffers has, today, become a fully entrenched practice within Congress.

Additionally, the shift to the staffer-delegation model also has meant that members of Congress typically will not perform a pre-enactment review of statutory text. As one lobbyist put it when asked if members read most of the bills: “Most of the bills? [They read] almost none of them! Any member that was honest will tell you that.”69

Representative Jim Moran (D-Va.) did indeed acknowledge this fact in an interview about the 2014 National Defense Authorization bill. Moran partook in the following exchange:

[Interviewer]: “Have you read all 1,648 pages of the final text of the National Defense—”
Moran: “Of course not. Are you kidding?”

[Interviewer]: “Are you planning on reading it before you vote?”
Moran: “No.”

[Interviewer]: “How will you know what’s in it?”
Moran: “Because we were, many of these issues we worked on for years and I trust the leadership . . . .”

[Interviewer]: “Do you think Boehner and Reid have read it?”
Moran: “I know their staff has.”70

These anecdotal statements are supported by a qualitative study conducted in 2013 that found that, in the modern Congress, members of Congress typically do not conduct pre-enactment review of statutory text.71

69 LAWRENCE LESSIG, REPUBLIC, LOST: HOW MONEY CORRUPTS CONGRESS—AND A PLAN TO STOP IT 126 (2011).
When asked if members of Congress read bills, one representative journalist responded: “Oh . . . absolutely not, no. Not in the United States Congress.”

Instead, the study found, responsibility for pre-enactment review of statutory text is conventionally entrusted to staffers. One chief of staff said that reading legislation was “mostly a staff thing,” while a Legislative Director said it was “predominantly staff’s job” to read all the legislation. The study found journalists familiar with Congress corroborating this finding.

Summarizing these findings, the study stated that: “[M]embers of Congress employ specialized staff whose job it is to read and write legislation, and report to their bosses on particular points of interest.” In their survey of congressional staffers, Gluck and Bressman similarly observed: “Our findings cast doubt on whether members or high-level staff read, much less are able to decipher, all of the textual details, and illustrate how the different players in the legislative process contribute to the final product in different ways, not all of them text focused.”

This is not to say that members of Congress categorically refuse to read statutory text. Indeed, members may well supplement bill summaries by perusing isolated provisions of actual statutory text. However, such reviews—which are scattershot and de-contextualized—can hardly be considered thoroughgoing “readings” of the statutory text. Instead, the nature of the pre-enactment review process—and the sum of materials used by members therein—suggests that, in essence, most members now provide only an indirect review of the contents of statutes.

In short, a host of scholarly and journalistic studies have documented the fact that, in the staffer-delegation era, members of Congress have shifted responsibility for three important tasks to staffers: the tasks of interstitial policymaking, statutory drafting, and pre-enactment review of statutory text. Due to this shift in responsibility, these studies reveal, members of Congress are performing only indirect reviews of the statutes upon which they vote. This gives rise to questions. What exactly is the nature of this indirect review? What information do members of Congress actually review before voting on a bill?

iii. The Inner Workings of the Staffer-Delegation Model. In the effort to understand the nature and sources of information that members of Congress draw upon to understand bills, interviews were conducted for this Article with twenty-five congressional staffers. Together, these staffers have worked

72 Id. at 14.
73 Id.
74 Id.
75 Id.
76 Id. at 8.
77 Bressman & Gluck, supra note 60, at 742.
in seventy-five different positions within Congress. Interviewees were drawn from both parties, both chambers of Congress, and a wide range of committee and member offices. The interviews were semi-structured in nature; planned questions were posed to interviewees that, while open-ended, nonetheless would permit for coding by response; and spontaneous follow-up questions were posed in order to invite interviewees to provide detail and elaboration on their initial answers. When interviewees answered a question multiple times to account for the different members for whom they had worked, their responses were coded as separate answers.

An interview-based study of this sort has several limitations, of course, a few of which are worth noting at the outset. Such studies always are subject to possible manipulation by interviewees, who might feel professional pressure to provide self-serving answers to sensitive questions about the conduct of their superiors. Moreover, a preference for depth and detail led to the use of open-ended questions. Due to the time-consuming nature of such questioning, the number of interviews that feasibly could be conducted was limited—a tradeoff that prevents the resulting study from operating as a comprehensive or exhaustive survey of congressional offices.

With these caveats in mind, we can turn to the findings of the study. First, interviewees were asked: how do members of Congress learn the contents of bills? Interviewee responses coded as follows:

![Figure 1. How members learn bill contents.](image)

Several findings captured in Figure 1 warrant discussion. First, no interviewee mentioned direct member involvement in bill drafting. In this way, the interviews underscored the fact that members no longer participate per-
sonally in the drafting of statutory text and that, consequently, none learn the contents of bills through this activity. This highlights the level at which members are engaging with bills. As one interviewee put it:

[Members are] not sitting there with a pen and writing the legislation themselves. But they oftentimes think about the concept, at a very general level. . . . [T]he staffer will get the idea—again, very broad-based from [the member]—and it’s your job to take that, run with it, figure out the details.79

Second, the majority of interviewees (nineteen respondents) did not cite direct review of statutes as a method used by members; a smaller number (seven respondents) mentioned it as a method used only in exceptional instances; and very few (three respondents) cited it as a consistently used method. These findings offer several lessons. On the one hand, they counsel against any categorical assertion that members never read bill text. A particular member of Congress was mentioned by several interviewees as reading nearly every bill voted upon, and one committee staffer expressed his belief that most members on his committee did spend some time with statutory text. On the other hand, these heavier readers of statutory text appear to be outliers. One interviewee who worked for a member who read some statutory text, for example, added: “My impression was that was not normal.”80 Another such interviewee estimated that only thirty percent of members look at statutory text.81 This outlier status is reflected in the aggregate interview data, where only three out of twenty-five interviewees described review of bill text as a method consistently used by members.

More commonly, as mentioned above, interviewees either did not list this method at all (nineteen respondents), or they described it as an exceptional practice reserved for abnormal instances (seven respondents). In a typical response flagging this method as exceptional, for example, one interviewee said: “Largely it’s a conceptual conversation with occasional references to sections of legislation.”82 There was no uniform description of these abnormal instances. Various interviewees mentioned the following as abnormal instances that increase the likelihood of some statutory text getting consulted: when a provision addresses a hot-button topic;83 when a provision addresses an issue on which the member has staked a reputation for high knowledge or passion;84 when a larger bill is of particular political or social

79 Interview with No. 23, Cong. Staffer, in Wash., D.C.
80 Interview with No. 22, Cong. Staffer, in Wash., D.C.
81 Interview with No. 18, Cong. Staffer, in Wash., D.C.
82 Interview with No. 3, Cong. Staffer, in Wash., D.C.
83 Interview with No. 4, Cong. Staffer, in Wash., D.C.; Interview with No. 11, Cong. Staffer, in Wash., D.C.
84 Interview with No. 22, Cong. Staffer, in Wash., D.C.
consequence; and when a particular member’s vote is undecided and is essential to passage.

Two interviewees also mentioned a practice that may effectively prevent “off-committee” members from learning bill contents by reading the bill. Sometimes, they asserted, a committee will not share the actual bill text with member offices until shortly before the bill text is posted publicly. This practice exists, interviewees said, due to concern that member offices might, intentionally or unintentionally, leak the bill text before the committee hopes to release it publicly. While bill text must be posted forty-eight hours in advance of action on the bill, it often has been lamented that this provides insufficient time for outside actors to read a bill comprehensively. Based on these interviewees’ remarks, some member offices may labor under similar constraints (though they may not mind this, as explained below).

In short, while it is an overstatement to say that members categorically refuse to read statutory text, interviews made clear that it is not prevailing practice for members to learn bill contents by reading the bill.

Instead, Figure 1 outlines a world in which members of Congress now rely overwhelmingly on two sources for their information about bills: (1) memoranda and summary documents; and (2) briefings and conversations. Twenty-five interviewees reported that memoranda (and other summary documents) were used exclusively or routinely by members for this purpose, and twenty-five interviewees similarly reported that briefings (and other in-person conversations) were used exclusively or routinely by members for this purpose. When it comes to how members learn bill contents, as one interviewee put it: “Conversations and memos are usually [the strategy].”

To shed light on the nature of these information sources, interviewees were asked a number of questions about them. With respect to memoranda and summary documents, interviewees first were asked about the type of summary documents used. In response, they mentioned a number of different documents that might be provided to the member, but the most commonly reported document—by a wide margin—was a staffer-generated memorandum. In the words of a typical respondent: “Almost every personal office that I have worked in or worked with, the [legislative] staff do vote memos for members.” Another remarked: “Whenever we do a bill introduction, we do a memo to [our member] outlining the parameters of [it, which would say]: ‘This is what the bill would do, this is our idea, this is the general concept.’”

---

85 Interview with No. 20, Cong. Staffer, in Wash., D.C.
86 Interview with No. 8, Cong. Staffer, in Wash., D.C.
87 Id.; interview with No. 8, Cong. Staffer, in Wash., D.C.; Interview with No. 12, Cong. Staffer, in Wash., D.C.
88 Interview with No. 10, Cong. Staffer, in Wash., D.C.
89 Interview with No. 21, Cong. Staffer, in Wash., D.C.
90 Interview with No. 23, Cong. Staffer, in Wash., D.C.
The Staffer’s Error Doctrine

As asked to compare the level of detail in these documents to that found in the section-by-section of a committee report, interviewees responded as follows:

![Graph showing level of summary document detail]

**Figure 2. Summary documents, compared to section-by-section.**

Section-by-section breakdowns in committee reports, it must be recalled, are themselves relatively concise. Yet, as the above chart shows, interviewees were unambiguous: memoranda and other summary documents are more brief and purpose-oriented than even a section-by-section. As interviewees put it:

- “[Summary documents are] much more of an overview than a section-by-section. Pithiness is valued very much on the Hill.”[^91]
- “You’re distilling a 60-page bill into 500 words, or 300 words—very top-level.”[^92]
- “We send information upwards—at my level, it’s X amount of information, and then what I give to the staff director is X minus two, because you need to simplify it, because they have less time. And then they take that, and it’s like X minus six for the staff director’s staff director, and then it’s even less when it gets to the member. It then ends up being talking points.”[^93]
- “Less is generally more.”[^94]

[^91]: Interview with No. 9, Cong. Staffer, in Wash., D.C.
[^92]: Interview with No. 16, Cong. Staffer, in Wash., D.C.
[^93]: Interview with No. 18, Cong. Staffer, in Wash., D.C.
[^94]: Interview with No. 15, Cong. Staffer, in Wash., D.C.
“Most members, everything they get is a very, very concentrated look at what the issue is.”

“We had a top-level document that kind of said, ‘Okay, we fixed [the issue]!’ Which is really what [members] cared about. ‘And here are the offsets [i.e., spending reduction policies that offset the bill’s costs].’ That’s what they want to know.”

“In terms of the detail, it will not necessarily be a section-by-section—I think that’s really information that, in the agent-principal relationship, your [member] is deferring to you to know the details in the weeds. It’s usually bigger picture. . . . So a memo would generally set out, very briefly, ‘Here is the contents of what’s in the bill, here are the [members] that support the bill, here are the groups that support the bill and oppose the bill, here is the pathway to passage . . . and here is your staff recommendation and why.’

“If you go to congress.gov, the summary page on each of the bills—that’s how much information [about 70% of the members] will take.”

The brevity of these memoranda was underscored by a noteworthy congressional practice that was mentioned by seven different interviewees: an informal expectation that memoranda should not exceed one page in length. One interviewee referred to this as the “one-page rule.” In the words of interviewees:

“Some members would like to see all the information and read everything, [but] most members want no more than one page on any given issue, on any given bill.”

“If I can’t fit an entire ‘rec’ [i.e., vote recommendation] on one page, then I’m not doing my job well enough.”

“I would say that, for the most part, they don’t know what’s in the contents of the bill. They certainly know what’s in a one-pager, and a summary. I don’t know if I would go as far as to say that they know what’s in a section-by-section of the bill. . . . That’s the vast majority of members.”

“Any time you go past one page, then you need a reason for it. Your typical vote memo for a non-controversial bill is probably a page. For a bill you think your boss is interested in, [it] might be two. I would consider a four-page memo to be a super-long

---

95 Interview with No. 18, Cong. Staffer, in Wash., D.C.
96 Interview with No. 19, Cong. Staffer, in Wash., D.C.
97 Interview with No. 23, Cong. Staffer, in Wash., D.C.
98 Interview with No. 18, Cong. Staffer, in Wash., D.C.
99 Interview with No. 9, Cong. Staffer, in Wash., D.C.
100 Interview with No. 11, Cong. Staffer, in Wash., D.C.
101 Interview with No. 17, Cong. Staffer, in Wash., D.C.
The Staffer's Error Doctrine

memo that I would only send to a substantive member that was really interested."

• “The unwritten rule is, you really want to give [your member] one page. . . . If you’re voting on an amendment on the floor, it could be very much shorter than that. It may literally be a paragraph of three to four sentences. ‘Here is your staff’s recommendation, here’s why.’”

Four additional interviewees mentioned that memoranda for members should not exceed two pages. In the words of one such interviewee: “[Typically,] it’s been literally two pages or less. If you’re feeling really special today, maybe you can go into a third.” There was some additional variation in the responses on memorandum length; one interviewee claimed to insist upon writing five-to-eight-page memoranda, and another mentioned that, while some members would not read past the first page, others would read a five-page memorandum. Even these outlier reports, however, describe documents that, when compared to the length of many federal bills, are very brief. And, overall, the interviews supported the contention of one interviewee that “some members would like to see all the information and read everything, [but] most members want no more than one page on any given issue, on any given bill, so that they are able to understand the gravity of their vote.”

A few other summary documents warrant comment. One—already mentioned above—is the “vote rec,” which provides a member with a brief summary of a bill and a staff recommendation on how to vote. A “vote rec” for the floor of Congress often is limited to a notecard or sheet of information. For some members, this notecard presumably serves merely as a reminder of the bills to be voted upon. Yet staffers also describe these notecards as regularly constituting the entirety of the information that members will consume on bills, saying:

• “When members go down to vote every day, they get a card which is two-and-a-half inches by six, or seven. And it has the recommendations for why they should vote yes or no on every bill that’s going to be on the floor. And the blurb is—imagine that small of a notecard, and that’s how much [roughly seventy percent of members] will usually know. Unless they care about

---

102 Interview with No. 21, Cong. Staffer, in Wash., D.C.
103 Interview with No. 23, Cong. Staffer, in Wash., D.C.
104 Interview with No. 5, Cong. Staffer, in Wash., D.C.; Interview with No. 6, Cong. Staffer, in Wash., D.C.; Interview with No. 24, Cong. Staffer, in Columbia, S.C.; Interview with No. 25, Cong. Staffer, in Columbia, S.C.
105 Interview with No. 6, Cong. Staffer, in Wash., D.C.
106 Interview with No. 14, Cong. Staffer, in Wash., D.C.
107 Interview with No. 21, Cong. Staffer, in Wash., D.C.
108 Interview with No. 9, Cong. Staffer, in Wash., D.C.
it, where they’ll pull a staffer and take them with them and walk and talk on the way to the floor.” 109

• “[A member’s staff] will write up a ‘vote rec.’ So, personal office staffers . . . will write up a vote recommendation, like, ‘Lean yes. Yes. Lean no. No.’ And why—a paragraph why. And so most members, you’ll see them walk on the floor with a card, you know, this big [gestures the size of a notecard], that has their vote recs on it. A lot of them don’t know what they’re voting on. You’re on the floor, they’re like, ‘what am I voting on right now?’” 110

Several interviewees also mentioned “dear colleague” letters as an educational tool, which are letters written in the voice of one member explaining the bill’s goals to another. Other interviewees mentioned a “charge-and-response,” which outlines the opposing party’s likely criticisms of a bill and provides retorts to those criticisms. Others mentioned the fact sheet about the bill that the Leader’s office typically generates. All of these documents were described as similarly concise and overview-oriented in nature, with interviewees saying:

• “Generally a ‘dear colleague’ is broad . . . you might have a bill that’s pretty long—100 pages plus—and you still have a single-page ‘dear colleague.’” 111

• “[In a charge and response,] if you’re defending against a response for three or four pages, methinks you do protest too much.” 112

• “The leadership [fact sheet] is usually much more message focused [i.e., focused on political rhetoric and ‘spin’] than ours is, and it usually ends up being longer. . . . I think their press staff do it. . . . So it’s more of a communications document.” 113

The brevity of these documents can seem disconcerting, but the documents do not exist in a vacuum. Only four interviewees described summary documents as the exclusive material used by members. Rather, most interviewees described a process whereby memoranda and staff-generated summary documents worked in tandem with: (1) additional summary documents (though of a similar overview nature) such as analyses from stakeholders, newspaper articles, Bloomberg and Congressional Quarterly summaries, and whip notices; and (2) in-person briefings of, and conversations with, members on the contents of bills. These briefings and conversations were described as taking various forms, such as:

109 Interview with No. 18, Cong. Staffer, in Wash., D.C.
110 Interview with No. 12, Cong. Staffer, in Wash., D.C.
111 Interview with No. 21, Cong. Staffer, in Wash., D.C.
112 Interview with No. 13, Cong. Staffer, in Wash., D.C.
113 Interview with No. 21, Cong. Staffer, in Wash., D.C.
The Staffer’s Error Doctrine

- Member-to-member conversations.
- Member legislative staff briefing the member.
- Committee staff briefing individual members.
- Committee staff briefing caucuses.
- Committee staff briefing member staffers (who then brief members).
- Whip sessions and other whip briefings to staff and members.

What, then, is the nature of these briefings and conversations? Asked whether these were conceptual or text-focused, interviewees responded as follows:

![Figure 3. Briefing and conversations: conceptual or text focused?](image)

As Figure 3 reflects, interviewees were unambiguous that briefings and conversations are more conceptual in nature. Elaborating, interviewees remarked:

- “I think the majority of members—I would say eighty-five percent of members—it’s going to be a conceptual conversation. Like, you’re going to say, ‘Here’s what the bill does.’ You would never go through the statute with them or say, ‘Here, in [the text of this specific section of existing law].’ . . . [W]hat they do is say, ‘Take me through section by section. What does this do.’ And you walk through. But you’re not going to take out your statute and say, ‘Here’s where it goes.’” 114

---

114 Interview with No. 6, Cong. Staffer, in Wash., D.C.
“Largely, it’s a conceptual conversation with occasional references to sections of legislation.”

“Oh, [it is] conceptual for sure.”

“It’s not text-based. It’s much more conceptual than that. It’s a high-level discussion. There are individual folks who may want to get more engaged and may want to get more into the weeds, but generally speaking, at these staff director or LD [i.e., legislative director] briefings, or member-level briefings, they tend to be ten-thousand foot. . . . All overview stuff.”

“[You do it] conversationally. ‘This is the problem. This is what we’re trying to do to address it. This is generally how we’re addressing it.’ But you make it conversational.”

“Our conversations tend to be very akin to a law clerk-judge relationship, where you’re talking very broadly about the concept, and then you’re delegated the responsibility to implement that—to fill it in, fill in the weeds and the details. If specific issues come up that could be problematic in the weeds, you may revisit and talk to [the member] about it. But for the most part, once you’re given the marching orders, you pretty much execute.”

“Usually it’s at a high level, so usually it’s more of a policy question [that is discussed], and the policy strengths and weaknesses. . . . Who will be impacted, what will be the limitations, how broad will the scope be, how does this reconcile with what I have supported or voted on in the past.”

Just as the brevity of congressional memoranda underscored their general, purpose-focused nature, the time constraints for member briefings emphasize the same quality. Interviewees remarked:

“You’re not going to take out your statute and say, ‘Here’s where it goes.’ Because you don’t get that much time with them. You get, if you’re really lucky, 30 minutes, on a pretty deep topic. It’s really rare that you would get more than that with them. And that’s going to be interrupted. . . . You’ve got to speak fast and speak concisely.”

“You get two minutes up front to make your pitch, and then they ask questions.”

---

115 Interview with No. 3, Cong. Staffer, in Wash., D.C.
116 Interview with No. 6, Cong. Staffer, in Wash., D.C.
117 Interview with No. 7, Cong. Staffer, in Wash., D.C.
118 Interview with No. 13, Cong. Staffer, in Wash., D.C.
119 Interview with No. 23, Cong. Staffer, in Wash., D.C.
120 Id.
121 Interview with No. 6, Cong. Staffer, in Wash., D.C.
122 Interview with No. 21, Cong. Staffer, in Wash., D.C.
• “[The member] would look at [the summary documents and text] overnight, send back comments and say, ‘Let me meet with them for fifteen minutes today. The vote’s tomorrow.’”123
• “You may have a bill that you’ve worked on for three months, and you may have anywhere from five to fifteen minutes to talk to the [member] about everything they need to know. So . . . you don’t have to communicate the whole fact pattern—but it’s knowing, what are the major issues, what is it that the [member] needs to know, and making sure they know that.”124
• “[Conversations are] anywhere from as short as thirty seconds to as long as maybe forty-five minutes to an hour, depending on what’s at stake, depending on how weighty the issue is, how technical it could be.”125

The discovery of these time constraints reinforces the central finding about briefings: namely, that they provide members with a big-picture, purpose-level overview, not with all the fine details embedded in statutory text. In aggregate, interviewees therefore depicted a staffer-delegation model that relies largely on memoranda and briefings to educate members on the contents of bills—materials that were depicted as conceptual and concise. As a result, they depicted a system in which members of Congress typically are engaging with bills at the level of statutory purpose. As interviewees put it:

• “For the most part, my observation is [that members] don’t really care about the details of an individual bill. They would consider themselves more big-picture thinker[s], not really concerned with the details. The staff will handle that.”126
• “Most members do not want to understand the intricacies of [an in-the-weeds reform]. They want to know how it’s going to affect their [district]. So, you can say, ‘Here’s what we’ve done, generally speaking, and here’s how it’s going to affect X, Y, and Z.’ [That] is what they’re really concerned about, not the technical [issues, such as], ‘And then, we inserted a subclause (I)!’”127
• “I think it’s safe to say that senators and representatives are generalists and they’re conceptualists. They’re big-picture people, and they hire staff to fill in the weeds and to know the details—and to be able to communicate with them what those details are, and to tell them what’s important.”128

123 Interview with No. 22, Cong. Staffer, in Wash., D.C.
124 Interview with No. 23, Cong. Staffer, in Wash., D.C.
125 Id.
126 Interview with No. 17, Cong. Staffer, in Wash., D.C.
127 Interview with No. 19, Cong. Staffer, in Wash., D.C.
128 Interview with No. 23, Cong. Staffer, in Wash., D.C.
This information about members of Congress comes with a caveat mentioned by nearly every interviewee: that the internal operations of congressional offices can vary significantly, based on the member. Office practices—including with respect to legislative activities—are designed to accommodate the preferences, interests, and learning styles of the elected representative. Consequently, there is no uniform rule about how members learn the contents of bills. Yet this diversity only makes the trends discovered in this empirical study all the more interesting—and they paint an unambiguous picture of members relying primarily on briefings and memoranda to engage with bills at the level of statutory purpose.

B. The Necessity of Delegation

The rise within Congress of the staffer-delegation model of lawmaking begs an important question: Why have members of Congress, for several decades, consistently delegated certain lawmaking tasks to others? Why have they not simply performed these tasks themselves? Commentators have offered one possible answer to this question. Most often, these commentators suggest, delegation within Congress is due to the fact that members of Congress are busy individuals. In his concurrence to *Bank One Chi. v. Midwest Bank & Trust*, for example, Justice Stevens stated that delegation to committees was necessary because: “Legislators, like other busy people, often depend on the judgment of trusted colleagues when discharging their official responsibilities.” More recently, Judge Katzmann has made the same point, stating “the expanding, competing demands on legislators’ time reduce opportunities for reflection and deliberation. In that circumstance . . . members operate in a system in which they rely on the work of colleagues [to discover] what the proposed legislation means.” As John Manning has observed, these arguments by Justice Stevens and Judge Katzmann are typical. They assume that congressional delegation both is occurring and is permissible because, as Manning put it: “A busy Congress must have the ability to paint with a broad brush and leave the detail work to the committees and sponsors who take the lead in framing solutions to a perceived social problem.”

This assumption—namely, that members of Congress delegate lawmaking tasks simply because they are busy people—has a noteworthy consequence. It suggests to courts that, while this delegation may indeed be occurring, it is not a fact that courts should take into account when interpreting statutes. The assumption has this effect for two reasons. First, the “busy

---

130 Id.
131 KATZMANN, supra note 68, at 17–18.
132 John F. Manning, *Textualism as a Nondelegation Doctrine*, 97 COLUM. L. REV. 673, 693 (2001); see also Jones, supra note 71, at 9 (suggesting that members do not read statutory text because “[l]awmakers are extremely busy individuals”).
The Staffer’s Error Doctrine

Congress” hypothesis suggests that Congress’s use of the staffer-delegation model is the result of workflow patterns in Congress that are variable, uncertain, and difficult to verify. After all, how busy are members of Congress, really? And do they really need to be this busy? It is difficult to measure this—and even more difficult to predict it for the future. As such, delegation to staffers does not appear to be a solid, unwavering, documentable fact about congressional practice—i.e., the sort of fact that courts want to rely upon when making assumptions about the lawmaking process.

Indeed, in his efforts to discourage legislative history usage by courts, Justice Scalia gave voice to this precise concern. As Scalia put it:

The only plausible justification for giving effect to legislative history is that the legislature is far too busy to consider the minute details of the bills that it considers—that it expects, it wishes, them to be resolved by the members and committees that draft the legislation and bring it to the floor. We have no idea whether this assessment of legislative expectations and desires is correct; there are forceful assertions of congressional sentiment to the contrary.133

When delegation is understood simply to be the result of a busy Congress, then this delegation frequently will appear to judges as it appears to Justice Scalia: as too variable and uncertain to warrant judicial recognition.

Moreover, there is a second reason why the “busy Congress” hypothesis might make courts hesitant to acknowledge the omnipresence of staffer delegations: the hypothesis makes these delegations appear to be the product of choice rather than of necessity. If someone is “too busy” to do something, then that fact seemingly reflects the person’s priorities as much as the person’s workload. When delegation is viewed as the result of members being too busy to craft statutes themselves, therefore, it paints a troubling picture of Congress. The drafting, reviewing, and passing of statutes are, after all, the core professional responsibilities of members of Congress. If members of Congress believe that they are too busy to engage in these lawmaking activities, the thinking goes, then perhaps they should simply clear more time on their schedules for this activity.134 And the last thing that courts should do is enable this irresponsible congressional behavior—yet this is precisely what courts would be doing, it seems, by acknowledging these delegations.135 In this sense, a theory that grounds congressional delegation in the “busy Congress” hypothesis dissuades courts from acknowledging these delegations

134 For a scholar making these claims, see Hanah Metchis Volokh, A Read-the-Bill Rule for Congress, 76 Mo. L. Rev. 135, 140–41 (2011).
135 See id. at 152 (criticizing instances where a “court decides to facilitate that improper behavior” of legislators not reading bills).
for a second reason: it seems to incentivize bad behavior by members of Congress.

How accurate is the “busy Congress” hypothesis, however? The answer to this question is somewhat complex. On the one hand, there undoubtedly is some truth to the hypothesis, as members of Congress now have an incredible number of demands on their time. At the same time, however, this hypothesis is notably incomplete. The routine delegation of lawmaking tasks to staffers is not simply the result of members of Congress being busy. Rather, this delegation also is a result, at least in part, of the fact that members of Congress now unavoidably exist at the intersection of two conflicting realities: a Constitution that envisions generalist legislators, on the one hand, and a complex modern world that requires specialized, expert governance, on the other hand.  

Consider each of these conflicting elements in greater detail. On the one hand, the Constitution mandates that members of Congress be generalists. The Constitution makes no effort to divide the legislative power and assign it to individuals who have in-depth knowledge or expertise in subject-matter areas. In fact, the Constitution actively prevents such a division; it does not permit the citizens of a state to elect a senator of military affairs, a senator of health care, a senator of agriculture, and so on. Instead, it mandates that electors choose individuals to represent them—and to make legislative decisions—with respect to the full range of legislative issues that will arise in the federal government.  

On the other hand, members of Congress find themselves in the position of regulating modern institutions, actors, and forces that are incredibly sophisticated and complex. Today, this complexity takes the form of at least three distinct challenges. First, members face the challenge posed by real-world complexity. The complexity encountered on this front has increased  

---

136 In some senses, the concern about this tension is as old as the Founders. See, e.g., The Federalist No. 62, at 379 (James Madison) (Clinton Rossiter, ed., 1961). (“It is not possible that an assembly of men called for the most part from pursuits of a private nature continued in appointment for a short time and led by no permanent motive to devote the intervals of public occupation to a study of the laws, the affairs, and the comprehensive interests of their country, should, if left wholly to themselves, escape a variety of important errors in the exercise of their legislative trust.”) This was a concern, Publius elsewhere noted, because: “No man can be a competent legislator who does not [possess] . . . a certain degree of knowledge of the subjects on which he is to legislate.” The Federalist No. 53, at 332 (James Madison) (Clinton Rossiter, ed., 1961).

Nonetheless, Publius was not envisioning a world in which legislators would need to possess a level of expertise anywhere near the level required today. This fact is highlighted by Publius’s proposed solution to this problem: multi-year terms in the House and Senate. See id. (explaining that two-year terms in the House afforded the “actual experience” necessary to become, and to function as, a “competent legislator” on federal matters); The Federalist No. 62, at 374 (James Madison) (Clinton Rossiter ed., 1961) (“Another defect to be supplied [that is, remedied] by a senate lies in a want of due acquaintance with the objects and principles of legislation.”).

137 U.S. Const. art. I, § 1, cl. 1 (vesting “All legislative Powers” in members of Congress, rather than dividing these powers among different individuals based upon subject matter).
2019] The Staffer’s Error Doctrine

significantly since the dawn of the twentieth century. Specialization during this period meant that many regulated industries and institutions became internally more complex—a fact which makes it more challenging to intervene in these industries in a precise, effective, and minimally disruptive fashion. Moreover, these specialized institutions have provided access to increasingly sophisticated knowledge of the surrounding world—knowledge that, once available, legislators were expected to draw upon in the production of legislation.

Environmental legislation provides a good example. This legislation no longer is motivated by traditional, generalist-accessible concerns about nuisances. Instead, the Clean Air Act outlines a regulatory regime relating to ozone pollutants that creates tiers based on the concentration of such pollution in micrograms per cubic meter;¹³⁸ that weighs treatment of CO compounds as opposed to volatile organic compounds;¹³⁹ and that outlines the balance of benzene, aromatics, and oxygen that may be included in reformulated gasoline.¹⁴⁰ Here, twentieth-century specialization has led to fields of knowledge that, even as they motivated legislation, also rendered that legislation accessible only to a relatively small field of experts.

Generalist legislators have little way of knowing, at the outset, whether such complexities lurk with respect to any given bill or legislative project. A bill that uses seemingly simple terms frequently will belie a thicket of concerns. To take one example: the real-world complexities of the health care industry have led the seemingly straightforward term “hospital” to have different, highly-technical meanings in different areas of federal law (indeed, even within different parts of the Medicare statute).¹⁴¹ When complexities so frequently lurk beneath seemingly straightforward language, responsible legislation essentially requires reliance upon a system of consistent delegation.

Moreover, the modern Congress also must navigate a second challenge that necessitates expertise: the challenge of statutory complexity. Many federal statutory regimes are now both incredibly large and impenetrably dense. First, consider their size: by the early 1980s, to take one example, the federal criminal code was already estimated to address more than 3000 distinct crimes.¹⁴² The Social Security Act, which was thirty-five pages in length when first enacted, now contains an individual title that is itself over 1000 pages in length.¹⁴³ The tax code, meanwhile, is now roughly 2600 pages in

¹³⁹ See, e.g., id. § 187(b)(2).
¹⁴⁰ See id. § 211(k).
length.\textsuperscript{144} While these statistics are not necessarily good proxies for the complexity of the underlying laws, they nonetheless illustrate the vastness of the existing statutory regime.

In addition, these statutes often are dense with formulas and defined terms that ripple throughout the statute in far-flung and hard-to-perceive ways.\textsuperscript{145} Consider again the example of the Medicare statute. Here, the term “star rating system” that appears in Part C of the statute has no formal definition, but it implicitly imports a definition previously developed by the agency to administer Part D of the statute.\textsuperscript{146} Meanwhile, the formula for physician payments in Part B of the statute scaffolds upon prior payment formulas located elsewhere in the statute (but not cross-referenced),\textsuperscript{147} and it makes no mention of a multiplier that appears much later in the statute (which, in some years, can vary the payment amount by as much as eighteen percent).\textsuperscript{148} Describing the resulting statute, the Fourth Circuit once observed: “There can be no doubt but that the statutes and provisions in question, involving the financing of Medicare and Medicaid, are among the most completely impenetrable texts within human experience.”\textsuperscript{149}

Due to statutory complexity of this sort, the Legislative Counsel for the Senate would remark, even decades ago, that it took two or three years of on-the-job training in order for an individual to become a competent drafter even with respect to a single subject-matter area of federal legislation.\textsuperscript{150} In the equivalent drafting office for the House of Representatives, conventional wisdom now holds that it takes closer to six years to reach this level of competence. The notion that a generalist legislator elected for a six-year term (or, in the House, for a two-year term) could navigate this statutory landscape in any competent way for herself, rather than by delegating the task, is farfetched.


\textsuperscript{147} See id. § 1848(d)(1)(B).

\textsuperscript{148} See id. § 1848(b)(1) (outlining certain components of the formula for physician payments); id. § 1848(q)(6)(E) (describing an additional merit-based incentive payment system, which increases or decreases Medicare payments by up to nine percent a year).

\textsuperscript{149} Rehab. Ass’n of Va., Inc. v. Kozlowski, 42 F.3d 1444, 1450 (4th Cir. 1994).

\textsuperscript{150} A Concurrent Resolution Establishing a Joint Committee on the Organization of Congress S. Con. Res. 2 Before the J. Comm. on the Org. of the Cong., 89th Cong. 1182 (1965) (statement of John H. Simms, Senate Legislative Counsel) (“Because of the specialized nature of the work, a new member of the staff is of little value to the office until he has served in it for 2 or 3 years . . . .”)}
Moreover, a third challenge awaits any member of Congress who is drafting legislation: the challenge of agency-implementation complexity. Today, executive agencies are massive, sophisticated institutional operations. The Department of Defense, which technically has over 3.2 million employees, is reported to be the largest employer in the world. These large institutions—and the quantity of their output befits their size. In 2013, agencies collectively published over 75,000 pages of material in the *Federal Register*—a number which includes the publication of anywhere between 2500 and 4500 rules per year. Agencies, in short, are massive bureaucratic undertakings. Legislation that is not attentive to their voluminous products and to their complex internal processes can result in significant and costly disruptions to ongoing implementation operations. A policy that appears wise out of context often appears quite different when viewed in light of agency-level realities, and policies that are uninformed by the minutiae of agency-implementation realities can quickly go astray.

These three challenges—posed by real-world complexity, statutory complexity, and agency-implementation complexity—confront any modern legislator who attempts to draft responsible legislation. The result is a modern drafting landscape that fundamentally differs from the comparatively simple drafting scenarios envisioned in many legislation classrooms. In this complex modern landscape, it is difficult to imagine any option that a legislator has other than to delegate extensive policymaking and drafting decisions to agents. In these situations, direct member involvement leads only to frustrations of the sort articulated by Representative John Conyers (D-Mich.), who asked: “What good is reading the bill if it’s a thousand pages

---


155 Id. at 2.

and you don’t have two days and two lawyers to find out what it means after you read the bill?"\textsuperscript{157}

In these situations, in short, delegation of lawmaking is not merely a byproduct of legislators having busy schedules. It is a necessary adaptation to the complexity of modern lawmaking.

The legal community has long acknowledged that Congress’s use of its prior model of delegation—the agency-delegation model of lawmaking—was a necessary response to the complexity of lawmaking in the modern era.\textsuperscript{158} Indeed, the Court has repeatedly concurred in this assessment.\textsuperscript{159} As the Court put it in \textit{Mistretta v. United States}:\textsuperscript{160}

\begin{quote}
[O]ur jurisprudence [regarding rulemaking delegations to agencies] has been driven by a practical understanding that in our increasingly complex society, replete with ever changing and more technical problems, Congress simply cannot do its job absent an ability to delegate power under broad general directives.\textsuperscript{161}
\end{quote}

In the context of agency delegations, therefore, it has been acknowledged that the complexity of modern lawmaking necessitated the use of a delegation model. So, why has the same acknowledgment not been made with respect to the more recent use of the staffer-delegation model? In the decades since the New Deal, after all, the complexity of Congress’s regulatory tasks has only grown. In their 1981 study of Congress, Salisbury and Shepsle could observe that: “[T]he sheer scope and complexity of the legislative agenda make it impossible for . . . [a] member to master the substance of every issue sufficiently well to reach in isolation a judgment about how to vote.”\textsuperscript{162} More recently, Lawrence Lessig has added that: “[I]t is literally impossible [for members to read all the bills for themselves] today: the complexity of the bills Congress considers is vastly greater than in the past. The Senate version of the health care reform bill, for example, was more


\textsuperscript{158} See, \textit{e.g.}, G. Edward White, \textit{The Constitution and the New Deal} 94–127 (2000) (reviewing repeated assertions since the Interstate Commerce Act of 1887 that agency delegation was a response to the rise in the “complexity” of modern lawmaking).

\textsuperscript{159} See, \textit{e.g.}, Panama Refining Co. v. Ryan, 293 U.S. 388, 421 (1935) (“Undoubtedly legislation must often be adapted to complex conditions involving a host of details with which the national Legislature cannot deal directly.”); United States v. Shreveport Grain & Elevator Co., 287 U.S. 77, 85 (1932) (acknowledging that lawmaking had taken on a “variety and need of detailed statement [that] it was impracticable for Congress to prescribe”); Union Bridge Co. v. United States, 204 U.S. 364, 387 (1907) (writing that to deny Congress the power to delegate “would be ‘to stop the wheels of government’ and bring about confusion, if not paralysis, in the conduct of the public business”).

\textsuperscript{160} 488 U.S. 361 (1989).

\textsuperscript{161} Id. at 372.

\textsuperscript{162} Salisbury & Shepsle, \textit{supra} note 2, at 565.
2019] The Staffer’s Error Doctrine 117

than two thousand pages long when introduced.” 163 Similarly, Judge Katzmann recently observed: “[Members of Congress] cannot read every word of the bills they vote upon, and, indeed, reading every word is often not particularly instructive, to the degree bills contain language amending the United States Code or enacted statutes” because such language often is exceedingly difficult to understand. 164

In short, the delegation-based approach to legislation is not the result of members of Congress simply being busy. Rather, this approach is the method by which Congress has chosen to negotiate two conflicting and stable forces: a set of constitutional imperatives written in a generalist era, on the one hand, and the reality of a hyper-specialized, complex bureaucratic world, on the other hand. Neither of these two forces shows any signs of waning or wavering in the foreseeable future. As such, courts can—and should—be confident that modern statutes are the product of significant delegations to staffers.

III. Template for a New Doctrine: King v. Burwell

The congressional divisions of labor described in Part II—divisions which assign statutory purpose and statutory text to different congressional actors—provide good reason for courts to modify their approach to statutory interpretation. In particular, courts ought to consider adopting a new doctrine of statutory interpretation—one labeled here as the “staffer’s error doctrine.” This Part prefaces the argument that will be made in Part IV for this “staffer’s error doctrine.” It does so by examining a recent case that can provide a useful model for this new doctrine: King v. Burwell.

Section A reviews the Court’s decision in King. Section B then explains how this decision looks different—and useful—once it is viewed in light of modern congressional divisions of labor.

A. The Court’s Reasoning in King v. Burwell

In King, the Court had to interpret a provision in the Patient Protection and Affordable Care Act that awards tax credits to low-income individuals—tax credits that are designed to make it financially feasible for these individuals to purchase health insurance. 165 Under the provision, the amount of the tax credit depends partly on whether the individual has enrolled in an insurance plan through an “Exchange established by the State under section 1311 of the [Act].” 166 If this reference to an “Exchange established by the

163 LESSIG, supra note 69, at 126.
164 KATZMANN, supra note 68, at 18.
165 Affordable Care Act § 1401(a), 26 U.S.C. § 36B (2012 & Supp. V 2017). Individuals must have a household income between 100 percent and 400 percent of the federal poverty line to qualify for the tax credit. See id. § 36B(c)(1)(A).
166 26 U.S.C. § 36B(b)–(c).
State” did not capture exchanges established by the federal government (as well as those established by a state), then low-income individuals residing in states that had federal exchanges would be calculated to receive no tax credit under the provision. However, the Internal Revenue Service (“IRS”) had interpreted this reference to an “Exchange established by the State under section 1311 of the [Act]” to apply more broadly. This reference included federal as well as state exchanges, the IRS argued—an interpretation which assured that all low-income individuals throughout the nation would have access to the tax credit.\footnote{See King, 135 S. Ct. at 2487–88.}

In \textit{King v. Burwell}, the Court was faced with a challenge to this IRS interpretation.

In its opinion, the Court essentially conceded that the plain meaning of the statutory text in the tax credit provision, if considered alone, referred exclusively to state exchanges. As the Court put it: “[T]he meaning of the phrase ‘an Exchange established by the State under [section 1311 of the Act]’ seems plain when viewed in isolation.”\footnote{Id. at 2495.} Consequently, the Court confessed that: “Petitioners’ arguments about the plain meaning of [the tax credit provision] are strong.”\footnote{Id.}

However, the Court was not content to resolve the case based solely upon the plain meaning of the statutory provision. Instead, the Court sought to develop an understanding of the purpose and scheme of the statute—and to determine whether the plain meaning, if applied, would allow that purpose to be realized.\footnote{For the Court’s understanding of the statute’s purpose, see infra notes 173–77 and accompanying text. For the idea of checking whether the purpose would be realized, see \textit{King}, 135 S. Ct. at 2495 (“In this instance, the context and structure of the Act compel us to depart from what would otherwise be the most natural reading of the pertinent statutory phrase.”).}

In conducting this purposive inquiry, the Court discovered that the Affordable Care Act was designed to achieve an overarching goal: to increase the number of individuals who were able to participate in insurance markets.\footnote{See, e.g., \textit{King}, 135 S. Ct. at 2493 (“[T]hese three reforms work together to expand insurance coverage.”).} In order to achieve this goal, the Court found, Congress sought to enact three interlocking reforms.\footnote{See \textit{id.} at 2486–87 (“The Affordable Care Act adopts a version of the three key reforms that made the Massachusetts system successful. . . . These three reforms are closely intertwined.”).} First, Congress sought to implement a requirement that, whenever an insurer sells health insurance, the insurer must make the insurance available to all individuals who want to purchase it (a reform implemented via the “guaranteed issue” and “community rating” requirements).\footnote{As the Court put it: “First, the Act adopts the guaranteed issue and community rating requirements.” \textit{Id.} at 2482. “Guaranteed issue” means that health insurers are required to sell (or “issue”) insurance to individuals; “community rating” means that health insurers must sell that insurance to all individuals (i.e., to the “community”) at a substantially similar price (or “rate”). See \textit{id.; see also} 42 U.S.C. § 300gg (2012).} By itself, however, this reform would allow individuals to
wait until they became ill to purchase health insurance—a practice known as “adverse selection” that could cause health insurance markets to collapse. Consequently, the Court discovered, Congress also enacted a second reform: a mandate that all individuals either maintain health insurance, or else make a payment to the IRS. This “individual mandate” created another problem, however. By itself, it would require low-income individuals to purchase something they could not afford (viz., health insurance). Consequently, the Court found, Congress enacted a third reform. It provided low-income individuals with tax credits that could be used to purchase the health insurance that they were required to buy under the individual mandate.

Through its purposive analysis, therefore, the Court discovered that the tax credit provision was meant to serve a specific purpose: namely, to allow Congress’s first two reforms to take effect—including with respect to low-income individuals—in a manner that would avoid the catastrophic consequences of “adverse selection.” Tax credits would allow the guaranteed issue, community rating, and individual mandate reforms to operate in a way that would not threaten to collapse health insurance markets around the country.

When viewed in light of this intended purpose, the Court explained, it was clear that the phrase “Exchange established by the State under section 1311 of the [Act]” was intended as a reference to both federal and state exchanges. After all, tax credits were Congress’s strategy to prevent its other reforms from collapsing health insurance markets. Every state was subject to these other reforms. The overwhelming weight of the purposive evidence suggested that Congress wanted to achieve the same results in all states. Tax credits were essential to achieving this. As the Court put it: “Those credits are necessary for the Federal Exchanges to function like their State Exchange counterparts, and to avoid the type of calamitous result that Congress plainly meant to avoid.” Consequently, the Court concluded, statutory purpose plainly pointed toward a specific answer to the question at

174 Health insurance companies rely on premiums from healthy individuals to subsidize care for the ill, so an imbalance in healthy-versus-ill individuals can undermine the market. See King, 135 S. Ct. at 2485 (“The guaranteed issue and community rating requirements achieved that goal, but they had an unintended consequence: They encouraged people to wait until they got sick to buy insurance. Why buy insurance coverage when you are healthy, if you can buy the same coverage for the same price when you become ill? This consequence [is] known as ‘adverse selection’ . . . .”).


176 King, 135 S. Ct. at 2487 (“Third, the Act seeks to make insurance more affordable by giving refundable tax credits to individuals with household incomes between 100 percent and 400 percent of the federal poverty line.”).

177 See id. at 2486–87.

178 Id. at 2495.

179 Id. at 2494; see also id. at 2496 (“Congress passed the Affordable Care Act to improve health insurance markets, not to destroy them.”).
issue in the case: the phrase “Exchange established by a State under section 1311 of the [Act]” was meant to capture both federal and state exchanges.180

In King v. Burwell, therefore, the Court confronted a situation in which, by the Court’s own admission, statutory purpose was in tension with statutory text. How would the Court resolve this conflict? “In this instance,” the Court stated, “the context and structure of the Act compel us to depart from what would otherwise be the most natural reading of the pertinent statutory phrase.”181 Statutory purpose prevailed, the Court said—and it prevailed specifically at the expense of the plain meaning of statutory text.182

B. The Court’s Reasoning and the Staffer-Delegation Model

While the Court’s opinion in King v. Burwell took recourse to statutory purpose, its methodology was not one that, under the typical conception, would be considered “purposive.” After all, the Court did not use statutory purpose to fill every gap or to resolve every ambiguity in the statute—situations in which statutory purpose is being extended into marginal applications.183 Instead, the Court in King did something different: namely, it reviewed an isolated provision of statutory text in order to ensure that this text did not wholly collapse the basic purpose of the statute.

When viewed in light of Congress’s current division of labor, this purpose-based review of statutory text takes on an important new light. Within the staffer-delegation model, it will be recalled, members of Congress have delegated to staffers the minutiae involved in producing statutory text. However, members have retained personal responsibility for devising, reviewing, and approving statutory purpose. By reviewing an isolated provision of statutory text in order to ensure that it did not undermine statutory purpose, therefore, the Court effectively reviewed the work product of congressional staffers to ensure that it did not undermine the decisions made by elected members of Congress. Viewed as such, the Court in King v. Burwell developed a method of statutory interpretation that, in the context of the modern

---

180 135 S. Ct. at 2487.
181 Id. at 2495.
183 The label of “purposivism” can be applied to many interpretive theories, of course, and it perhaps is reductive to assert that purposivism is defined as an interpretive model that takes recourse to statutory purpose in order to resolve statutory ambiguities, including those that bear on instances of marginal applications. Since the Court’s renewed emphasis on the plain meaning rule in the Burger era, however, the Court has moved away from the strongest forms of purposivism that might use statutory purpose to overcome plain statutory text and toward narrowed versions that take recourse to purpose in situations of ambiguity. See generally Bressman & Gluck, supra note 60, at 407–09 (recounting the evolution of the Court’s uses of legislative history and statutory purpose since the Burger era).
Congress, serves an interesting function: namely, it provides a means by which courts can protect legislatures from the errors of unelected staff.

IV. The Staffer’s Error Doctrine

A. Precedent: The Scrivener’s Error Doctrine

Several critics have argued that the Court’s interpretive approach in *King v. Burwell* was illegitimate. The most notable of these was Justice Scalia, whose dissenting opinion was quick to criticize the majority for exceeding the scope of its powers within our constitutional system. As Justice Scalia put it: “This Court . . . has no free-floating power ‘to rescue Congress from its drafting errors.’” However, when viewed as rescuing elected members of Congress from drafting errors committed by unelected agents of Congress, there is a time-tested judicial canon which suggests that courts do indeed have this power: the scrivener’s error doctrine.

In statutory interpretation, the scrivener’s error doctrine posits that judges have the authority—and, arguably, the obligation—to identify and correct a narrow set of technical mistakes in statutes (such as punctuation errors). Why have courts believed themselves to be justified in correcting this set—and only this set—of statutory errors? According to Justice Scalia himself (along with coauthor Bryan Garner), this doctrine has historically been justified by a specific theory: one that, in the following pages, will be described as the “division-of-labor theory.”

As described by Scalia and Garner, this division-of-labor theory has two components. First, the theory posits that, under the scrivener’s error doctrine, judges have traditionally been allowed to correct a narrow set of statu-

---


185 *King*, 135 S. Ct. at 2504 (Scalia, J., dissenting) (quoting *Lamie v. United States Trustee*, 540 U.S. 526, 542 (2004)).

186 For differing views on the scope of the doctrine, see, for example, United States v. Locke, 471 U.S. 84, 120–26 (1985) (Stevens, J., dissenting) (applying the doctrine to application dates); Costanzo v. Tillinghast, 287 U.S. 341, 344 (1932) (arguing that the doctrine should apply to syntax as well as punctuation); Barrett v. Van Pelt, 268 U.S. 85, 91 (1925) (saying that the doctrine should only apply to application dates); Amalgamated Transit Union Local 1309 v. Laidlaw Transit Servs., Inc., 448 F.3d 1092, 1097 (9th Cir. 2006) (Bybee, J., dissenting) (arguing that the doctrine should only apply to instances of nonsensical or ungrammatical sentences or erroneous cross-references). For a modern argument that the doctrine should encompass any statutory text that appears on the face of the statute to be an error, regardless of the nature of the error, see SCALIA & GARNER, supra note 133, at 235–36. See generally Ryan D. Doerfler, *The Scrivener’s Error*, 110 NW. U. L. REV. 811 (2016) (discussing and critiquing this iteration of the doctrine). For a discussion of the particular iteration that the Court had been applying leading up to *King*—viz., the doctrine of mistake outlined in *Lamie v. United States Trustee*—see Gluck, supra note , at 103–05.
tory errors because such errors have been “thought to be the work of the engrossing clerk or the printer” rather than of the legislators themselves. In other words, the doctrine has targeted minor typographical features because the insertion of these features into statutory text—a specific drafting task—has been assumed to be a task performed by unelected staffers. Second, the theory holds that the doctrine is further justified by the fact that: “[I]n days of yore . . . because many legislators voted only on the basis of bills that they heard read aloud—without seeing the printed page—they could take no notice of the punctuation marks.” Put differently, the doctrine has been further justified by the fact that staffers, not legislators, traditionally have provided the pre-enactment review of these typographical features. According to Scalia and Garner, therefore, the scrivener’s error doctrine targeted certain typographical errors because the doctrine was designed to allow judges to review—and to correct—those portions of statutes that legislators typically neither draft nor subject to pre-enactment review.

This division-of-labor view of the doctrine is not specific to Scalia and Garner. It is embedded, for example, in the very name that judges and scholars have accepted as the proper label for the doctrine: the scrivener’s error doctrine. This moniker rests upon two assumptions: first, that a division of tasks exists within a legislature (viz., between legislators and “scriveners”), and second, that courts should correct errors made by a specific institutional actor within this internal bureaucracy (viz., the scrivener). These same assumptions are found in the oft-repeated idea that the doctrine allows judges to correct “clerical” errors in statutes—a description which assumes that such errors result from the work of “clerks,” as Scalia and Garner put it,

---

187 SCALIA & GARNER, supra note 133, at 161 (quoting Morrill v. State, 38 Wis. 428, 434 (1875)).


189 Jonathan Siegel has traced the use of this label for the doctrine, saying: ["[The] National Bank case is the first in which the Court invoked the term “scrivener’s error” in exercising the power of statutory correction. Neither the Court nor any Justice used the term before 1985, when Justice Stevens used it in his opinion in United States v. Locke, 471 U.S. 84, 123 (1985) (Stevens, J., dissenting). Older cases from other courts show the term used almost invariably in connection with errors made either by private parties in drafting contracts or similar instruments or by courts or court clerks in connection with judgments. Courts consider themselves empowered to disregard such errors. See, e.g., Christensen v. Felton, 322 F.2d 323, 325 (9th Cir. 1963) (disregarding a scrivener’s error in a contract). But at least some older cases use the term with reference to judicial reform of statutes. See, e.g., In re Deuel, 101 N.Y.S. 1037, 1038–39 (N.Y. App. Div. 1906) (correcting a “scrivener’s error” that resulted in the omission of the term “not” in a statute)."


190 As other scholars have noted, this term is taken from the field of contract law, where it was preceded by the term vitium scriptoris—literally, the “mistake of a scribe.” See David M. Sollors, The War on Error: The Scrivener’s Error Doctrine and Textual Criticism: Confronting Errors in Statutes and Literary Texts, 49 SANTA CLARA L. REV. 459, 461 (2009).
The Staffer’s Error Doctrine

rather than of legislators.\textsuperscript{191} As Scalia and Garner note, these are assumptions that have a long history in statutory interpretation—assumptions that are brought to the fore by these widely used labels.\textsuperscript{192}

According to the division-of-labor theory, therefore, the scrivener’s error doctrine embodies two insights. First, the doctrine acknowledges the reality that a division of tasks regularly occurs within legislatures. Second, the doctrine advances the basic idea that, when legislatures do indeed divide legislative tasks among different actors, courts have a role to play in protecting and preserving the decisions that democratically elected legislators are making within this institution—including by protecting them from the errors of unelected agents who assume some of the tasks of statutory production.\textsuperscript{193}

In order to properly serve this protective function, however, the scope of the review that judges perform under the doctrine must track the actual division of tasks within the legislature. To the extent that it is desirable for the doctrine to serve this function with respect to federal statutes, therefore, it is necessary for the doctrine to be grounded in a realistic, up-to-date understanding of the inner workings of Congress. In particular, the doctrine must be anchored in an accurate understanding of the range of tasks that members of Congress actually delegate to unelected staffers and agents.

In this regard, modern judges have failed in their task of ensuring that the scrivener’s error doctrine remains vital and coherent. While the number of legislative tasks performed by congressional staffers has dramatically expanded in the staffer-delegation era, this expansion has not been accompanied by a corresponding growth in the number of statutory features that judges will review for errors under the scrivener’s error doctrine. Instead, the doctrine has retained a narrow focus upon the drafting issues that were han-

\textsuperscript{191} SCALIA & GARNER, supra note 133, at 161. For descriptions of the doctrine as addressing “clerical” errors, see, for example, Ass’n of Tex. Prof’l Educators v. Kirby, 788 S.W.2d 827, 830 (Tex. 1990) (stating that a refusal to correct such errors would “amount to government by clerical error”); Michael S. Fried, A Theory of Scrivener’s Error, 52 Rutgers L. Rev. 589, 593 (2000) (“’Scrivener’s error’ refers to a typographical mistake or other error of a clerical nature in the drafting of a document.”).

\textsuperscript{192} A cursory review by Scalia and Garner finds these assumptions in a legal textbook from 1948, SCALIA & GARNER, supra note 133, at 161 (citing James DeWitt Andrews, supra note 188), and in a court case from 1875, see id. (citing Morrill, 38 Wis. at 434). Other cases also evince these assumptions. See U.S. Nat’l Bank of Ore. v. Indep. Ins. Agents of Am., Inc., 508 U.S. 439, 462 (1993) (“[W]e are convinced that the placement of the quotation marks in the 1916 Act was a simple scrivener’s error, a mistake made by someone unfamiliar with the law’s object and design.”) (emphasis added); In re Deuel, 101 N.Y.S. at 1038–39.

dled by unelected staffers in the “days of yore,” as Scalia and Garner put it.194

As such, those who wish to see this doctrine continue to perform the function implied by the division-of-labor theory need to do more than simply defend the original, antiquated formulation of this doctrine. They also need to advocate for the doctrine to be expanded in a manner that would permit judges, in the search for errors, to review all aspects of statutes that members routinely assign to staffers in the modern Congress. Under such an expansion, the doctrine presumably would allow judges to review all statutory text (i.e., the work of today’s unelected staffers) for errors.

Interestingly, this is the review that the Court actually performed in King v. Burwell. The Court essentially reviewed statutory text for a specific, particularly egregious type of staffer error. By scanning statutory text for any conflicts with broad statutory purpose, it effectively looked for staffer errors that had undermined a decision that members of Congress had directly and personally made with respect to a statute. By contrast, the Court did not seek out staffer choices that, while perhaps “errors” in the sense that they violated members’ presumed but unexpressed wishes, nonetheless did not violate any concrete decisions made by members—staffer choices that therefore would not present instances of staffers exceeding their delegations. Instead, the Court’s review functioned simply as a check for staffer errors that were so egregious as to undermine core statutory purpose—which is to say, for errors that would undermine the basic goal that members of Congress would have entrusted to them.

Through its use of this review, therefore, the Court in King v. Burwell effectively created a model for a narrow but updated version of the scrivener’s error doctrine—a version that, in order to reflect this updated quality, might be described as a “staffer’s error doctrine.” It is a doctrine that begins with the principles that animated the original scrivener’s error doctrine—and that combines these principles with an awareness that today’s staffers perform a wider range of lawmaking functions than did the “scriveners” of old. By essentially creating such a “staffer’s error doctrine,” the Court (perhaps inadvertently) found a way to make these principles relevant once again—and its interpretive approach should be considered in light of this underappreciated merit.

B. Administration of the Staffer’s Error Doctrine

By revisiting the Court’s method in King v. Burwell, therefore—and re-reading this method in light of both the historical purpose of the scrivener’s error doctrine and the drafting practices of the modern Congress—we can discover the possibility of a new “staffer’s error doctrine.” This doctrine would direct judges to identify instances in which the work product of staff-

---

194 Scalia & Garner, supra note 133, at 161.
The Staffer’s Error Doctrine

The Staffer’s Error Doctrine undercuts rather than advanced decisions made directly by members of Congress (viz., the selection of overarching policy goals). As such, the doctrine gives rise to the question: how, as a practical matter, would courts administer this doctrine?

One potential answer to these questions is found in the template that the Court provided in King v. Burwell. In this case, the Court’s reasoning suggested a three-step review process that a court could conduct after developing its initial interpretation of a statute. In the first step of this review process, courts would identify the overarching policy goals that a statute is designed to achieve. In King v. Burwell, for example, the Court used a number of sources in order to reconstruct the statute’s purpose. It looked at findings that Congress inserted into the statute,195 external sources that documented the historical context of the legislation (including amicus briefs, scholarly articles, and congressional hearings),196 and the operative provisions of the statute.

Having identified the statute’s overarching policy goals, courts then would move to the second step of this three-step review: they would check whether the isolated snippet of statutory text before them, if applied as written, would sabotage those overarching policy goals. This, it should be noted, is a high threshold. The threshold is high for a reason: the process is tailored to catch only those passages of statutory text that represent instances in which staffers have wholly exceeded the scope of their delegation. Given the broad authority that staffers now possess to engage in interstitial policymaking, a staffer who produces a statutory regime filled with exceptions, exemptions, and anomalies can hardly be said to have exceeded the scope of his or her delegation. Consequently, the discovery of a mere policy anomaly or inconsistency would not suffice for purposes of this review. Rather, this step two review would seek to identify instances in which the statutory text at issue, if applied, would so thoroughly undermine the core policy goals of the statute—or carry the statute so dramatically far afield of these goals—that the statute could no longer be characterized, in any accurate sense, as a reasoned implementation of those basic goals.

In such situations, courts could proceed to the third step of the review. When a divide between purpose and text emerges, this final step directs courts to do the difficult work of investigating whether the divide is more

195 King, 135 S. Ct. at 2486–87, 2493 (citing 42 U.S.C. § 18091(2)(I) (2012)).
likely the result of a mistaken view, among the court, of the statute’s purpose—or, instead, of a drafting error in Congress.

Applying this step in King v. Burwell, the Court found significant evidence, on the one hand, that the statutory text at issue resulted from a drafting error. This text was found within a statute that contained many instances of “inartful drafting” and that was subject to unique procedural challenges that limited the opportunities for pre-enactment review. The specific phrases that seemed plain in the relevant provision were used sloppily and inconsistently throughout the bill, and its plain meaning would interact in odd ways with other provisions in the bill. The Court further observed that other provisions of the bill seemed to assume that the provision at issue means something other than its plain meaning. The result, the Court suggested, was a strong suggestion that the statutory text at issue was the result of a drafting error.

On the other hand, the Court found little reason to believe that it had been led astray by the secondary materials relied upon for its purposive analysis. With the exception of the provision at issue, the Court noted, these materials closely and accurately tracked the operative provisions of the bill. This inspired confidence that these materials were designed to elucidate, not to mislead. The more compelling conclusion, the Court therefore found, was that the statutory provision at issue was the result of a staffer’s error.

Through an application of the three-step process of the staffer’s error doctrine, therefore, the Court essentially concluded that the text at issue in King v. Burwell was the result of a staffer’s error. This conclusion, it is worth noting, appears to have been correct; several subsequent investigations have confirmed the text to have resulted from staffer error. As Senator Bingaman (D-N.M.) put it to the New York Times: “As far as I know, [the relevant statutory text] escaped everyone’s attention, or it would have been deleted, because it clearly contradicted the main purpose of the legislation.” With respect to the provision at issue in King v. Burwell, therefore, it seems that the three-step process embedded in the staffer’s error doctrine successfully performed its intended task: it correctly identified an instance of

\[\text{\footnotesize 197 Id. at 2504.}\]
\[\text{\footnotesize 198 Id.}\]
\[\text{\footnotesize 199 Id. at 2498–99.}\]
\[\text{\footnotesize 200 Id. at 2495.}\]
\[\text{\footnotesize 201 Id.}\]
\[\text{\footnotesize 202 Id.}\]
staffer error that undermined core decisions made by members of Congress—and that, consequently, exceeded the scope of staffers’ delegation.

C. The Legitimacy of the Staffer’s Error Doctrine

Through its method of preserving the purpose of the statute, therefore, the Court’s approach in *King v. Burwell* had the effect of restoring values that have long been embedded in statutory interpretation through the scrivener’s error doctrine. Namely, it restored the value of protecting the decisions made by democratically elected members of Congress—including from the errors made by their agents.

Some may be skeptical of the notion that courts should be in the business of advancing this value, however. Such individuals might ask: why should courts attempt to protect Congress from the errors of its agents? What provides courts with the authority to adopt this approach? This Section explores possible answers to these questions. In particular, it examines the reasons why both intentionalists and textualists might be persuaded that courts have the authority to utilize this “staffer’s error doctrine.” To this end, Subsection 1 details the reasons why intentionalists ought to view this doctrine as legitimate—regardless of whether this new doctrine is viewed as a semantic canon or a substantive canon. Subsection 2 then outlines the reasons why, under four prominent versions of textualism, the doctrine similarly appears to be a legitimate tool of statutory interpretation.

1. Intentionalism and the Staffer’s Error Doctrine

One way to conceptualize the staffer’s error doctrine is as an interpretive tool that assists judges in the specific task of uncovering and enforcing legislative intent. According to the division-of-labor theory, the doctrine provides this assistance by directing judges to examine the work product of unelected staffers and, in so doing, to ensure that this work product has advanced, rather than undermined, the intent of legislators. For intentionalists, it is a check premised on the notion that portions of statutes handled exclusively by unelected staffers are uniquely at risk of having veered away from legislative intent. By permitting (and perhaps requiring) judges to conduct this review, the staffer’s error doctrine thereby protects and preserves the intent of legislators, guarding it from incursions by staffers.

Viewed as such, the doctrine clearly accords with intentionalist notions of the proper judicial role in statutory interpretation. As John Manning puts it: “[I]ntentionalists have] long emphasized that, as faithful agents of Congress, federal courts have a constitutional duty to implement Congress’s ‘in-

---

205 For Justice Breyer suggesting that judges, in situations of statutory ambiguity (as opposed to situations of error), might work to excavate the “general intention” that legislators entrust to drafters in the modern congressional bureaucracy, see Transcript of Oral Argument, *supra* note 193, at 40.
tent.’” 206 The staffer’s error doctrine can be understood as assisting courts in their effort to serve as “faithful agents” of the sort that Manning describes.

By contrast, what is less evident is whether the staffer’s error doctrine—when viewed through this intentionalist lens—should be considered a semantic canon or a substantive canon of statutory construction. 207 The answer to this question depends upon one’s theory of the proper dividing line between these two categories of canons. According to some scholars, the difference between these two types of canons resides in the role that courts perform when using the canon. Semantic canons are designed to assist the courts in their role as faithful agents of Congress, the theory goes. 208 By contrast, substantive canons allow judges to inject extrinsic values into the legal process, thereby allowing judges to operate as cooperative partners that partake, alongside Congress, in the shared activity of lawmaking. 209 Viewed as such, the staffer’s error doctrine is best characterized as a semantic canon, since it assists the court in its role as a faithful agent of Congress.

According to other scholars, however, the difference between the two types of canons resides in the canons’ differing relationships to statutory text. Semantic canons are rules of language usage that reveal the meaning embedded within statutory text, these scholars argue, 210 whereas substantive canons impose values that are external to the text. 211 According to this view of the canons, the staffer’s error doctrine is a substantive canon since, in most iterations, it explicitly goes beyond the text of the statute. 212 Understood as such, the staffer’s error doctrine serves to advance a particular substantive value: the constitutional value of popular sovereignty. 213

This theory of the staffer’s error doctrine relies upon a particular, intentionalist-friendly vision of popular sovereignty. Under this vision, the governmental architecture created by the Constitution is viewed as a means to an end. It is designed to allow the democratic will to flow outward from the

207 For an argument that it is not a doctrine of statutory interpretation at all but instead is an application of strict scrutiny to Article I, Section 7, see Fried, supra note 191, at 609.
208 See ESKRIDGE, GLUCK & NOURSE, supra note 14, at 458 (noting that grammar canons are employed in part because “the legislature is presumed to know and follow basic conventions of grammar and syntax”).
209 See, e.g., Andrew C. Spiropoulos, A Defense of Substantive Canons of Construction, 2001 Utah L. Rev. 915, 919 (2001) (“[B]y expecting judges to articulate and adhere to substantive canons of construction, we encourage judges to think of themselves as partners in the enterprise of lawmaking, rather than simple agents who are forced to accept bad policy.”).
212 For a version that might not go beyond statutory text, see infra notes 219–23 and accompanying text.
213 On the idea that popular sovereignty is the single, overarching idea that animates the Constitution, see, for example, AKHIL REED AMAR, AMERICA’S CONSTITUTION: A BIOGRAPHY 13 (2006) (explaining popular sovereignty as “the Constitution’s bedrock idea”).
public, through and across a variety of governmental actors and institutions, and back to the public as enforceable laws.

Within this constitutional architecture, the argument goes, the Founders outlined four important transition points. These are instances where the democratic will must be transferred from one constitutional actor to another. First, the democratic will must transfer from the public to members of Congress. This occurs primarily through the mechanism of elections. Second, democratic will must transfer from members of Congress to legal texts. This occurs through the process of enacting laws. Third, democratic will needs to transfer from legal texts to the executive and judicial branches. This is accomplished through the interpretation of laws. Fourth, the democratic will transfers back to the public. This occurs through the enforcement of laws.

Under the popular sovereignty view, the underlying purpose of many constitutional provisions is to ensure that these transfers are successful. Consider the first transition point: the transfer of democratic will from the public to members of Congress. A host of constitutional provisions are designed in the effort to ensure that this transfer is successful. The provisions requiring regular elections are the most notable example. As Publius puts it in his analysis of the House of Representatives in Federalist No. 52:

As it is essential to liberty that the government in general should have a common interest with the people, so it is particularly essential that the [House of Representatives] should have an immediate dependence on, and an intimate sympathy with, the people. Frequent elections are unquestionably the only policy by which this dependence and sympathy can be effectually secured.

The election provisions of Article I tether members of Congress to the public in two different senses. First, they require each member to have been recently selected by the public—something likely to happen only if the member is broadly sympathetic with the will of that public. Second, regular elections warn each member of Congress that another election will be occurring soon. The anticipation of such elections is meant to force members

216 U.S. Const. art. I, § 1; id. art. II, § 1.
217 See The Federalist No. 52, at 327 (James Madison) (Clinton Rossiter ed., 1961) ("The scheme of representation [is] a substitute for a meeting of the citizens in person.").
219 The Federalist No. 52, supra note 217, at 327.
220 See The Federalist No. 60 (Alexander Hamilton) (explaining that the public will elect representatives who mirror their interests); see also Jacob E. Gersen & Matthew C. Stephenson, Over-Accountability, 6 J. Legal Analysis 185, 189 (2014) (describing this “selection effect” as a feature of accountability mechanisms in principal-agent relationships).
of Congress to internalize the interests of their communities, since most members would presumably want to be reelected.\footnote{See The Federalist No. 57 (James Madison) (stating that anticipation of elections will create fidelity to public will).}

Elections were not the sole mechanism that the Founders used in order to transform members of Congress into embodiments of the will of their constituents, however. Publius also outlines a variety of other constitutionally induced pressures that would advance this purpose.\footnote{See The Federalist No. 56, at 293 (James Madison) (Clinton Rossiter ed., 1961) (on requirements that members of Congress be inhabitants of the states they represent); id. at 297 (on members being subject to the laws they make); The Federalist No. 63 (James Madison) (on bicameralism); see also Gersen & Stephenson, supra note 220, at 189 (describing this “incentive effect” as a feature of accountability mechanisms in principal-agent relationships).} The idea, therefore, was that a host of constitutionally induced pressures would uniquely transform the members of Congress into vessels for the democratic will of their constituents, and thereby would act as “cords by which [members would] be bound to fidelity and sympathy with the great mass of the people.”\footnote{The Federalist No. 57, supra note 221, at 353.}

Since this host of pressures would converge only upon members of Congress, it was essential to popular sovereignty that these individuals participate properly in the second transfer point in the constitutional scheme: the production of binding laws. It was vital that this group could successfully enact their decisions into law, for this was the activity that would allow the public will to continue on its path. Consequently, Article I vested the power to produce these laws exclusively in the members of Congress.\footnote{U.S. Const. art. I, § 1 (vesting “All legislative Powers in Congress”). Sections 1 and 2 of Article I made it clear that “Congress” was understood here to mean, specifically, the collection of the members of Congress. Id. §§ 1–2. The presentment requirement also involves a democratically elected president in this process, of course.} Moreover, Article I sought to ensure that no external impediments hindered these members in their efforts to successfully navigate this transition point. Consequently, it provided members with broad latitude to decide for themselves the manner in which to convert the public will into binding legal texts.\footnote{Article I, Section 5 of the Constitution grants Congress great flexibility in determining the methods by which it will draft, review, and approve legislation. See U.S. Const. art. I, § 5. Meanwhile, the legislative power under Article I, Section 1 of the Constitution (along with the Necessary and Proper Clause of Article I, Section 8) gives Congress latitude to determine how best to pursue its legislative ends (including the power to create legislative branch officers via statute that might assist Congress with its lawmaking duties). U.S. Const. art. I, §§ 1, 8; see also Buckley v. Valeo, 424 U.S. 1, 90 (1976).}

The question posed by the problem of staffer’s errors is the following: what should courts do when this second transfer point within the constitutional scheme fails? What should courts do when there is compelling evidence that the will of members of Congress—its proxy for the will of the public—was not successfully encoded within statutory text?
As a substantive canon, the staffer’s error doctrine asserts that courts have an active role to play in this situation. They can legitimately act to reconnect statutory text to a decision made by elected representatives and, in so doing, can thereby restore the flow of the public’s will through the architecture of the federal government. In this way, the staffer’s error doctrine functions to advance the constitutional value of popular sovereignty. As such, it can be treated much like other substantive canons that advance constitutional values, such as the constitutional avoidance canon and the federalism canons. Like those other canons, it operates as “quasi-constitutional law,” as Eskridge and Frickey put it. It uses statutory interpretation as a vehicle to preserve a structural constitutional element—precisely the sort of element that Eskridge and Frickey describe as the primary concern of this subset of substantive canons.

2. Textualism and the Staffer’s Error Doctrine

While the appeal of the staffer’s error doctrine for intentionalists is relatively straightforward, the doctrine’s appeal for textualists is slightly more complex. This is true for two reasons. First, the doctrine admittedly violates a tenet that many textualists seek to uphold: the tenet that the plain meaning of the text must be respected. Second, textualism has been defined and justified in a variety of different ways (as discussed below), thereby meaning that no single analysis of the doctrine is likely to endear it to all textualists. Still, there are several important reasons why the staffer’s error doctrine might be appealing to textualists.

i. Textualism’s Respect for Historical Pedigree. The historical pedigree of the scrivener’s error doctrine might be, for some textualists, a sufficient justification for the continued use of an updated staffer’s error doctrine. After all, textualists are particularly likely to view the historical pedigree of the canons as providing justification for interpretive practices that otherwise might offend textualist principles. As Judge Amy Coney Barrett has put it, “[t]extualists have suggested that, in the modern landscape, those principles that we call substantive canons are a closed set of background assumptions justified by their sheer longevity.” Admittedly, some textualists have begun to revise this view, concluding that historical pedigree alone cannot redeem canons that conflict with textualist premises—and concluding also that the scrivener’s error doctrine is fundamentally incompatible with such premises. For others, however—including both textualists and non-textualists—

---

227 *Id.* at 597.
228 *Id.*
229 Barrett, *supra* note 211, at 111.
230 *See, e.g.*, Amalgamated Transit Union Local 1309 v. Laidlaw Transit Servs., Inc., 448 F.3d 1092, 1097 (9th Cir. 2006) (Bybee, J., dissenting) (Sharply limiting the doctrine based on
the historical pedigree of the scrivener’s error doctrine alone may suggest that the staffer’s error doctrine possesses a claim to legitimacy.

Moreover, Judge Barrett has documented a second reason why the longstanding use of the scrivener’s error doctrine might persuade textualists that this doctrine’s successor, the staffer’s error doctrine, is legitimate. According to Barrett, the canons of construction were widely accepted as legitimate exercises of the judicial function at the time of the Founding.231 In this regard, Barrett notes,232 the canons are immune to John Manning’s claim that non-textualist methods of interpretation exceed the original understanding “the judicial Power” conferred upon federal judges by Article III.233

Indeed, Judge Barrett makes this claim as someone sympathetic to Manning’s general claim about “the judicial Power” (a claim that others have disputed).234 Once the Court’s approach in King v. Burwell is understood as a coherent application of the idea underlying the scrivener’s error doctrine, therefore, this approach may elide the constitutional concerns that trouble textualists such as Manning.

ii. Textualism as an Application of Congressional Intent. Some textualists, including both Caleb Nelson and John Manning, have sometimes accepted the idea that courts should seek—and enforce—Congress’s “intent.” According to these scholars, textualism is a preferred interpretive method simply because it is the best way for courts to discern and apply congressional intent. For textualists of this variety, the staffer’s error doctrine should have significant appeal.235 In order to understand this appeal, it is worth examining the claims made by some of these textualists in detail.

a. Textualism as Applying the Only Available Intent. One version of this argument—i.e., the argument that textualism is the best method of discovering congressional intent—is well represented by John Manning. Here, Manning argues that, due to the murkiness of the inner workings of Congress, the intent of Congress is unknowable.236 As Manning puts it: “The legislative process . . . is too complex, too path-dependent, and too opaque to allow judges to reconstruct whether Congress would have resolved any particular question differently from the way the clear statutory text resolves that
The Staffer’s Error Doctrine

2019

question.”237 In the face of this unknowability, Manning argues, interpreters have no option but to abandon the search for actual congressional intent. Consequently, Manning says, interpreters must turn to a second-best form of intent. An appropriate form of second-best intent is found, he then argues, in Joseph Raz’s notion that: “Even without knowing the speaker’s actual intent or purpose in making a statement, one can charge the speaker with the minimum intention ‘to say what one would ordinarily be understood as saying, given the circumstances in which one said it.’”238 In this way, Manning begins with a claim about the unknowability of Congress’s intent, and he concludes from it that, in order to operate as faithful agents plausibly carrying out congressional intent in some manner, courts should turn to textualist methods.

The use of the staffer’s error doctrine is more consistent with this set of interpretive assumptions than it might appear. After all, the point made by Raz is that, when a speaker makes an utterance, the speaker typically can be assumed to possess an intent to speak the uttered words (and, further, to have those uttered words interpreted according to prevailing interpretive conventions). Based on our intuitive awareness that this assumption is valid, Raz argues, we arrive at an everyday interpretive practice under which “the normal way of finding out what a person intended to say is to establish what he said.”239 Even Raz accepts, however, that the default use of this “normal way” of locating intent is subject to rebuttal; it can be overridden by evidence that a specific error occurred in the conversion of an intention into a speech act. As Raz puts it: “An exception is any explanation of what went wrong [in the speech act] which establishes either that one was trying or had formed an intention to say something and failed, or that one did not mean what one said.”240 The staffer’s error doctrine can be understood simply as a check to see whether such an explanation exists. For his part, Raz believes that underlying legislative intentions exist and can be identified (and, therefore, can be used to check for the existence of such a rebuttal).241 It is Manning who adds the additional idea that legislative intent is unknowable. As such, the staffer’s error doctrine is consistent with Raz’s underlying theory. And it is possible that, while Manning presumably would reject the turn to intent in these instances, he nonetheless would acknowledge that

237 Id.; see also Caleb Nelson, A Response to Professor Manning, 91 Va. L. Rev. 451, 459 (2005) (“According to the textualists, judges who try to reconstruct the legislature’s collective understandings (if any) by engaging in case-by-case investigation of individual legislators’ actual semantic intentions are doomed to failure; they do not know enough to identify reliably whatever collective understandings did in fact exist.”).


239 Raz, supra note 238, at 270.

240 Id.

241 Id. at 263–64.
these instances present a case in which this justification of textualism encounters its limit.

b. Textualism as Applying Intent of a Compromise-Driven Congress. Manning also offers a second argument suggesting that textualism is the best method of discovering congressional intent. This argument is based on a descriptive claim about the inner workings of Congress. The legislative process is marked by unprincipled compromises, Manning claims, not by principle or by logic. Due to the prevalence of legislative bargains, Manning suggests, statutory text that appears philosophically inconsistent, erroneous, or even absurd is—more often than not—an accurate reflection of congressional intent. Consequently, Manning argues, the best strategy for capturing congressional intent is simply to apply statutory text in all cases. Indeed, to interpret statutes otherwise would be to trample over bargains struck during the legislative process, Manning suggests—an approach that would undermine the power accorded to factions through the process of bicameralism and presentment. In this way, Manning offers a second argument on behalf of textualist methods—an argument which asserts that these methods provide the best strategy by which to ascertain the intent of Congress.

An awareness of the staffer-delegation model qualifies Manning’s claim that Congress is an institution driven by unprincipled compromises. Under the staffer-delegation model, lawmaking proceeds along two parallel tracks. Members of Congress make general policy decisions—and these decisions may be the result of bargains and unprincipled compromises. At the same time, members task staffers with producing a statute that will carry out these compromises—and, in so doing, the members frequently will award these staffers the power to strike further interstitial compromises. What is prohibited within the staffer-delegation model, however, is for staffers to make compromises that undermine the purposes agreed upon by members. Consequently, Congress is invariably purposive along a particular axis: statutory text is predictably designed to carry out, not to undermine, whichever purposive instructions members pass along to staffers. Put differently, delegations are inherently purposive; they award an agent with authority to pursue a set of goals (or purposes) on behalf of a principal. Due to the fact that Congress now predictably relies upon delegations in the production of statutory text,
there is a narrow sense in which purposivist interpretation will reliably implement congressional intent—even if Congress otherwise is, as Manning claims, a compromise-driven institution. For this reason, even those subscribing to Manning’s brand of textualism should acknowledge that the staffer’s error doctrine, which requires the application of a narrow but rigid brand of purposivism, constitutes a legitimate exception to that interpretive methodology.

c. Textualism as Applying Intent While Accounting for Judicial Strengths and Weaknesses. Caleb Nelson additionally has argued on behalf of a brand of textualism which accepts the idea that statutory interpreters are engaged in the task of discerning congressional “intent.” According to Nelson, a doctrine designed to correct “drafting errors” is perfectly permissible under this variant of textualism.245 Textualists merely are more reluctant than intentionalists to conclude that a drafting error has occurred, Nelson claims, and so they require more compelling evidence of such an error.246

This reluctance has two distinct sources, Nelson observes. First, textualists are reluctant to correct “drafting errors” because textualists hope that this judicial practice will incentivize members of Congress—or their staffs—to review bills more closely.247 As such, Nelson’s claim is premised upon a theory about why drafting errors might typically find their way into statutes: because members of Congress and staffers might lack sufficient motivation to closely review statutory text. As Part II explained, however, members of Congress typically decline to perform pre-enactment review of statutory text for a different reason: in a world of significant statutory complexity, they lack the capacity to conduct this review. In such a situation, increased incentives cannot prove effective, at least with respect to the members themselves. Consequently, it might make particular sense for judges to seek out and correct drafting errors in the staffer-delegation era, since the countervailing interest in incentivizing members of Congress is absent in such instances. This is precisely what the staffer-delegation doctrine encourages them to do.

According to Nelson, textualists also are reluctant to acknowledge drafting errors because textualists “worry about errors [by judges] in the application of tests that require case-by-case exercises of judgment.”248 This concern is relevant, Nelson explains, because any “drafting errors” doctrine

246 See id. (“It is in the practical application of this concept that textualists and intentionalists really part company.”).
247 Id. at 382 (expressing the hope that “the courts’ reluctance to identify and correct ‘drafting errors’ may encourage members of Congress or their staffs to spend more time proofreading and poring over each individual bill”).
248 Id. Nelson further states:

The more one distrusts the ad hoc judgments that tests for drafting errors necessitate . . . the more conservative one’s favored test will be. In particular, one might adopt
is based on an intuitive calculation about the likelihood of false positives (i.e., of judges finding errors where none exist) versus the likelihood of false negatives (i.e., of judges overlooking genuine drafting errors). The concern about case-by-case exercises of judgment is relevant because it alters this equation for textualists, making textualists particularly hesitant to declare a staffer error in any particular case.

Like other doctrines that seek to identify drafting errors, the staffer’s error doctrine does require judges to perform “case-by-case exercises of judgment.” However, an awareness of the staffer-delegation model of lawmaking—and of members’ dependence upon this model—would presumably alter the intuitive calculation described by Nelson (for both textualists and non-textualists). In a world where members of Congress often are not competent to review the work of the staffers who draft statutory text, the risks of errors—and therefore of false negatives—presumably is heightened. This would make any interpreter somewhat less reticent to conclude that a drafting error has occurred—and therefore more likely to apply the staffer’s error doctrine.

iii. Textualism as a Restriction on Source Materials. According to some scholars, textualism’s central tenet is a prohibition on the use of secondary materials—a prohibition that, for one reason or another, limits the interpreter to the text of the relevant statute. John Manning, for example, has argued for such a restriction on constitutional grounds. According to Manning, the nondelegation doctrine imposes a constitutional mandate that judges not stray beyond the text in statutory interpretation. Scalia defended this restriction on practical grounds, meanwhile, arguing that legislative history “provides a uniquely broad playing field” for judges—one that, once allowed in statutory interpretation, permits too much “variety and specificity of result.”

Due to this strain of textualism, several judges and scholars have suggested that it is acceptable for courts to apply the scrivener’s error doctrine any time the relevant error is evident to an interpreter based solely on the text of the statute. Most notably, Scalia argued that the doctrine can legitimately be applied to any situation in which: “[O]n the very face of the statute it is clear to the reader that a mistake of expression (rather than of

---

the position advocated by modern textualists: judges should have leeway to identify and correct ‘drafting errors’ only in what they consider to be very clear cases.

Id. at 381.

249 Id. at 381 (citing Frederick Schauer, The Practice and Problems of Plain Meaning: A Response to Aleinikoff and Shaw, 45 VAND. L. REV. 715, 730 (1992)).

250 Id. at 382.

251 Manning, supra note 132, at 696–99.

legislative wisdom) has been made.” 253 Manning has contemplated such a version of the scrivener’s error doctrine as well, 254 as has Nelson. 255

This version of textualism raises an interesting question: can a version of the staffer’s error doctrine be created that relies solely upon statutory text? It is difficult to know. Nonetheless, it is worth noting that the Court relied surprisingly heavily on statutory text—and surprisingly little on secondary materials—in its reconstruction of statutory purpose in King v. Burwell. Here, the Court relied upon findings within the statute for statements of statutory purpose, and it viewed these statements in conjunction with the statute’s various provisions in order to develop a bird’s eye view of how these provisions were meant to operate in service of the statute’s larger goals. 256 In this sense, the Court’s efforts in King v. Burwell paint an intriguing picture which suggests that a text-based version of the staffer’s error doctrine—a version that potentially would appease source-focused textualists—might be possible.

iv. Textualism as a Rejection of Intent-Based Interpretation. According to the most natural interpretation of the staffer’s error doctrine, this doctrine is designed to discover and enforce the intentions of Congress. Nonetheless, it also is possible to justify this doctrine without any recourse to notions of congressional intent. Such a justification may be particularly appealing to textualists, since many textualists reject the notion that congressional intent is the proper object of statutory interpretation.

According to this justification, the staffer’s error doctrine does indeed rest upon the division-of-labor theory discussed in Section A of this Part. This doctrine, it confirms, is designed to induce courts to review the work of unelected staffers—in particular, to ensure that these staffers did not insert errors into statutory text. However, the work of unelected staffers does not receive additional review simply because it poses an inherent threat to the proper encoding of congressional intent, it argues. Instead, staffer work receives this added review because it makes sense, for some additional reason, to treat the work of unelected staffers differently from the work of elected representatives.

For example, one can imagine a situation in which courts might want to adopt textualist practices in order to incentivize members of Congress to perform their legislative tasks precisely and responsibly. However, textualists might simultaneously hope to refrain from incentivizing staffers, or from incentivizing Congress to rearrange its internal divisions of labor. This reti-

253 Id. at 20.
254 Manning, supra note 206, at 2459–60 n.265 (“[W]hen an internal textual inconsistency or an obvious error of grammar, punctuation, or English usage is apparent from reading a word or phrase in the context of the text as a whole, there is only the remotest possibility that any such clerical mistake reflected a deliberate legislative compromise.”).
255 Nelson, supra note 245, at 356.
256 See King v. Burwell, 135 S. Ct. 2480, 2483 (2015) (“[T]he words of a statute must be read in their context and with a view to their place in the overall statutory scheme.”).
ence might be grounded, for example, in a concern about courts meddling with internal congressional practices and procedures. As such, this approach might be conceptualized as a peculiar form of respect for Article I, Section 5 of the Constitution, which provides members with autonomy to determine the internal rules, procedures, and practices of Congress.257

A similar theory could be offered that is based on retroactive accountability rather than forward-looking incentivizing. Under this view, elections are designed not only to provide accountability for legislators who have advanced policies that are contrary to the public will; rather, they also are designed to allow for the removal of legislators who are incompetent or ineffective. It is a view which assumes that elections are meant to hold representatives accountable for the process as well as the substance of their legislative work. When judges intervene to save legislators from their own errors, this theory would assert, they are undermining the proper operation of this constitutional mechanism. In effect, they are propping up incompetent legislators—individuals who should be allowed the opportunity to show that they can produce effective legislation, or who otherwise should be subject to judgment on that account at the ballot box.

Even if one accepts the idea that members of Congress should be held accountable for their own errors, however, it is not clear that they should equally be held accountable for the errors of their agents. This is particularly true in a period when the delegation of some lawmakers tasks is unavoidable, as it is today—and especially when many of these delegations are to career staffers whose employment cannot be terminated by an individual member of Congress. In this world—a world in which members of Congress inevitably are dependent upon some experts for the execution of their legislative vision—forcing members to stand judgment for staffer errors no longer seems to serve any accountability function. Instead, it seems to achieve quite the opposite: it takes any member who competently relies upon existing congressional delegation practices, and it renders that member randomly subject to the possibility of being undermined by forces largely outside the member’s control.

In short, there are many reasons why courts might uniquely concern themselves with correcting errors by unelected staffers. While the most obvious reasons relate to the preservation of congressional intent, not all of these reasons are necessarily connected to an intentionalist agenda—and textualists should consider these alternative justifications before summarily dismissing the staffer’s error doctrine.

V. CONCLUSION: A NEW (YET OLD) THEORY OF CONGRESS

By outlining and defending a new “staffer’s error doctrine,” this Article has developed an argument that has implications for the practice of statutory

interpretation. At the same time, this argument also has theoretical implications for the field of legislation—one of which is particularly worth highlighting here. Namely, this argument reveals the inadequacy of the theories of Congress espoused by textualists and purposivists—and it points toward an alternative theory that ought to have a more prominent place in the field of legislation.

First, consider the theory of Congress found in purposivist scholarship. This scholarship typically argues that Congress is, above all, a purpose-driven institution. According to many purposivists, this characteristic is a result of the task that Congress must perform. The very nature of legislation, these scholars argue, is to realize broad social goals or objectives. This view is typified by the famous statement by Hart and Sacks that: “Law is a doing of something, a purposive activity, a continuous striving to solve the basic problems of social living.”258 Based on this theory about the nature of lawmaking, purposivists frequently assume that Congress, as a lawmaking institution, must itself be purpose-driven. Hart and Sacks argued, for example, that courts could accurately assume “that the legislature was made up of reasonable persons pursuing reasonable purposes reasonably.”259

By contrast, many textualists hold that Congress should be viewed instead as a deal-making institution. As one textbook has put it: “[Many textualists adopt] a view of the legislative process that is different—and grittier—than the one that underlies strong purposivism. Rather than seeing the legislative process as coherent and reasonable, the new textualists emphasize the rough-and-tumble of political compromise.”260

In distinction to these two theories, this Article has argued that a third theory of Congress is relevant to the field of statutory interpretation: namely, the theory embedded for many decades in the parallel field of administrative law. In administrative law, a defining feature of the modern Congress is assumed to be its delegation of lawmaking tasks to agents (namely, to executive branch agents). Due to fundamental changes that Congress underwent in the late twentieth century, this Article has argued, delegation to unelected agents has become a similarly defining feature of lawmaking within Congress. Consequently, modern federal lawmaking is invariably marked by delegation, regardless of whether it is outsourced to agencies or conducted within Congress.

Based on this observation, the foregoing pages have argued that courts should be reviewing these two types of lawmaking similarly. In each instance, courts should be ensuring that agents of Congress did not exceed the scope of their delegations. In this way, this article has argued that this alter-

259 Id. at 1378.
260 JOHN F. MANNING & MATTHEW C. STEPHENSON, LEGISLATION AND REGULATION 54 (2d ed. 2013).
native theory of Congress leads to a narrow but rigid brand of purposivism. It is a version whereby courts are not necessarily empowered to fill every gap and ambiguity in a statute with their understandings of statutory purpose, but where courts are empowered to ensure that no statutory text—no matter how unambiguous—is allowed to undermine the core purposes of the statute.

Not all will agree that courts’ review of statutory text should more closely resemble their review of agency rulemaking, of course. The review of staffer drafting is, after all, a review of activity conducted prior to members of Congress voting upon a bill—a quality that distinguishes this activity from agency rulemaking. As such, some commentators will conclude that the proper form of review for errors made within the former process is a pre-enactment review by members of Congress, not a post-enactment review by the courts.

However, the staffer’s error doctrine is based upon the idea that, in an era in which members of Congress are not competent to provide pre-enactment review of statutory text, this distinction between pre-vote and post-vote lawmaking is somewhat hollow. In each instance, members of Congress are vulnerable to the threat of having their lawmaking agents undermine the decisions that they have made during the enactment process. Under the staffer’s error doctrine, these comparable threats are met with comparable protections. As the foregoing pages have explained, they are protections that courts have long understood themselves as empowered to provide—and that make particular sense in an era marked by the staffer-delegation model of lawmaking.

261 This claim resembles Judge Posner’s assertion in Archer-Daniels-Midland Co. v. United States, 37 F.3d 321 (7th Cir. 1994), that: “[I]n the case of statutory language [that is] technical and arcane . . . the slogan that Congress votes on the bill . . . strikes us as pretty empty.” Id. at 324.
ARTICLE

FROM THE DIGITAL TO THE PHYSICAL:
FEDERAL LIMITATIONS ON REGULATING
ONLINE MARKETPLACES

BENJAMIN EDELMAN AND ABBEY STEMLER*

ABSTRACT

Online marketplaces have transformed how we shop, travel, and interact with the world. Yet, their unique innovations also present a panoply of challenges for communities and states. Surprisingly, federal laws are chief among those challenges despite the fact that online marketplaces facilitate transactions traditionally regulated at the local level. In this Article, we survey the federal laws that frame the situation, especially § 230 of the Communications Decency Act (“CDA”), a 1996 law largely meant to protect online platforms from defamation lawsuits. The CDA has been stretched beyond recognition to prevent all manner of prudent regulation. We offer specific suggestions to correct this misinterpretation, so that state and local governments can appropriately respond to the digital activities that impact physical realities.

I. INTRODUCTION .......................................... 142
II. THE MARKETPLACES AT ISSUE ................... 143
   A. Craigslist and General-purpose Listings .......... 144
   B. Uber and Ride-hailing ............................. 146
   C. Airbnb and Short-term Rentals ................... 148
   D. StubHub and Ticket Resale Marketplaces ........ 150
III. ENFORCEMENT EFFORTS .............................. 151
   A. Verification ......................................... 153
   B. Design .............................................. 153
   C. Tax .................................................. 155
   D. Disclosure of User Information ................... 157
IV. ARGUING AGAINST STATE AND LOCAL REGULATION .... 159
   A. The Prima Facie Case Under § 230 ............... 162
   B. Fellow Travelers .................................... 163
      1. 1996 Federal Telecommunications Act .......... 163
      2. Stored Communications Act ..................... 164

* Edelman: Associate Professor, Harvard Business School. Stemler: Assistant Professor, Kelley School of Business, Indiana University. Edelman advises eBay on questions of advertising fraud and in 2017 advised the American Hotel & Lodging Association on questions of online travel agency fees and bias, but neither entity requested, supported, or approved this article. The authors thank Paul Kominers for extraordinary research and contributions to this paper. We are also grateful for the careful assistance of Melissa Latini, feedback on this article from scholars at the 2018 Huber Hurst Research Seminar in Business Law and Ethics, and the editors of the Harvard Journal on Legislation.
I. INTRODUCTION

Imagine you serve on the city council of a mid-size college town. Your constituents are outraged because local landlords are buying small houses near the downtown square just to list them on a short-term rental site. This behavior is raising housing prices and disrupting the peace and character of neighborhoods, while lining the pockets of landlords. As you try to deal with angry phone calls, tweets, and emails, you are surprised to learn that federal law restricts what actions you can take to address this intimate, local issue. What’s the problem? Perhaps the most revered twenty-six words in the United States Code: § 230 of the Communications Decency Act (“CDA”), which states in relevant part: “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”1

Section 230 is a 1996 law originally designed to protect online websites (such as message boards and AOL chatrooms) from defamation lawsuits for

From the Digital to the Physical

user-generated speech. It has been credited as the law that gave us the modern Internet and launched the digital revolution. More recently, however, online marketplaces have tried to use § 230 to protect themselves from everything from product liability to obligations under a myriad of state and local rules. As a result, § 230 is now the first line of defense for online marketplaces seeking to avoid onerous regulations.

This Article challenges existing interpretations of § 230 and highlights how it and similar federal laws interfere with state and local government efforts to regulate online marketplaces—particularly those that dramatically shape our physical realities, such as Uber and Airbnb. This Article also provides a framework for assessing when a marketplace should be held accountable for the activities it facilitates. In line with this theory, it offers specific suggestions to assure that state and local governments can appropriately respond to the challenges presented by online marketplaces.

Section 230 is sacred to many technology companies and tech law scholars, and this Article does not intend to discount the contributions the law has made to the modern Internet. However, as Congress and courts revisit the language of the CDA, it is more important than ever to critically examine its purpose, its benefits, and its harms. Changes are needed to § 230’s interpretation, if not its text, to assure that online marketplaces are accountable for the negative consequences of their actions—and to assure that state and local governments have the tools needed to appropriately govern.

II. The Marketplaces At Issue

Historically, most businesses followed a linear business model, focused primarily on creating goods and services to sell to distributors or customers. However, in the past decade, sophisticated electronic communications platforms have brought a massive shift. Today, many of the
world’s largest companies incorporate or are built on platforms\(^8\) that act as intermediaries between producers and customers or otherwise standardize and shape consumer activity.\(^9\)

A marketplace is a type of platform that facilitates a commercial transaction—be it the exchange of cash for a ride to the airport, or a night on someone’s couch.\(^10\) Marketplaces create the digital space for commerce, reducing transaction costs and allowing people to easily interact and transact.\(^11\) Unlike traditional firms, marketplaces do not create or even acquire the products or services to be sold. Rather, they serve to connect buyers and sellers.\(^12\)

The following subsections briefly describe four online marketplaces—Craigslist, Uber, Airbnb, and Stubhub—which are widely used and well-known, and which motivate the regulatory questions discussed in the remainder of this Article.

A. Craigslist and General-purpose Listings

Online communications can facilitate all manner of transactions. Among the most longstanding and most flexible platforms is Craigslist, arguably the predecessor and inspiration for the specialized marketplaces that followed.

Craigslist was founded in San Francisco in 1995 by contrarian and philanthropist Craig Newmark when he first shared upcoming city events with a dozen friends by email.\(^13\) The emails eventually evolved into an online classified ad service and enjoyed exceptional success. By 2017 it offered comprehensive online listing services in more than 700 cities in seventy-eight

---


\(^10\) See *Moazed & Johnson*, supra note 6, at 29.

\(^11\) Transaction costs include costs of searching, standardizing terms of trade, and distilling vast amounts of information.

\(^12\) *Moazed & Johnson*, supra note 6, at 6.

countries. Craigslist’s categories range from household goods to jobs to housing to friends and personal listings.

Despite its wide scope in both subject matter and geography, Craigslist retains its barebones design—just simple single-colored text on a white background, without graphics or even a logo. The site’s architecture remains similarly simplistic—blank screens where users can describe the products they are selling, apartments they are renting, or workers they wish to hire, with Craigslist serving primarily to store and distribute the information. To find the listings they want, most users rely on Craigslist’s search feature, which searches the full text of posts. Craigslist also collects limited metadata, self-reported by the users providing listings, such as the condition of an item and the geographic location where it was available. Users can search these criteria, if desired.

Most Craigslist users do business in person, which provides an opportunity for in-person inspection of goods, reducing many kinds of disputes. With in-person transactions, Craigslist saw no need for tracking or reporting reputations of buyers or sellers, offering any kind of insurance or guarantees, or facilitating payment on the platform. Foregoing these features, Craigslist ends up less intertwined in transactions that its site facilitates, arguably reducing Craigslist’s responsibilities under the frameworks developed in the following sections.

Consistent with Craigslist’s limited features, the site collects unusually low fees from users. Before 2004, Craigslist was entirely free in all cities except San Francisco. In 2004, Craigslist began charging for certain listings in Los Angeles and New York, ultimately expanding to fees for brokered apartment rentals in New York City and job listings, car sales, and select “gigs” and “services” in particular areas. All other categories are provided without charge to buyer, seller, or anyone else. Craigslist also does not seek additional revenue through advertising. Rather, Craigslist


17 Id.


keeps its costs low by employing few people (less than fifty as of 2017)\textsuperscript{21} and adding only the simplest new features, which yields ample profits\textsuperscript{22} despite providing most of its services without charge.

Craigslist makes do with a skeleton staff in part because most customer functions are automated. Users can add or delete listings, or reset their passwords, without assistance. Craigslist staff notably does not handle day-to-day disputes about improper listings. Instead, links at the top of each listing let other users “flag” posts that are prohibited for violating the Craigslist Terms of Use.\textsuperscript{23} When a post receives several such reports, Craigslist software removes it automatically. Craigslist’s internal software also identifies some posts for automatic removal, and Craigslist staff have the ultimate authority to remove posts as they see fit.\textsuperscript{24}

Despite its flagging system, Craigslist faced a variety of complaints about certain categories of listings. Most controversial were the “Adult” listings where critics said the company was facilitating prostitution, sex trafficking, and abuse. Under pressure, including a congressional inquiry, criticism from attorneys general, and campaigns from advocacy groups, Craigslist shut down this category in 2010.\textsuperscript{25} Craigslist also faced periodic complaints about the content of listings. For example, housing listings expressing a preference for tenants of a particular race were the subject of 2006 litigation discussed in Section V.B.

B. Uber and Ride-hailing

While Craigslist can be used to buy or sell almost anything, Uber and other ride-hailing services specialize in a single type of transaction: transportation.\textsuperscript{26} Similar services include Lyft and Fasten in the United States, Grab in Southeast Asia, and Didi Chuxing in China. Uber historically presented


\textsuperscript{26} Uber has also begun other logistics services which we do not discuss.
itself to users with the motto “everyone’s private driver,” and most passengers think of the company as a substitute for taxis. But in legal documents and proceedings, Uber structures its role and responsibility more narrowly, arguing that it is a technology company that creates a marketplace connecting passengers and drivers. With its role framed in that way, Uber claims it is not responsible for the acts or omissions of drivers. Passengers might have all manner of complaints against drivers, from poor choice of route to vehicle condition to assault, but Uber says that passengers must bring legal claims arising from these complaints against drivers and not the company. At the same time, Uber recognizes that passengers expect assurances about driver safety and reliability. Uber therefore advertises that it offers “a ride you can trust,” including driver background checks, driver ratings, insurance, and around-the-clock support.


29 U.S. Terms of Use, Uber, https://www.uber.com/legal/terms/us/ [https://perma.cc/4ABV-DEMP] (last visited Oct. 14, 2018) (“The Services comprise mobile applications and related services (each, an ‘Application’), which enable users to arrange and schedule transportation, logistics and/or delivery services and/or to purchase certain goods, including with third party providers of such services and goods under agreement with Uber or certain of Uber’s affiliates (‘Third Party Providers’)... YOU ACKNOWLEDGE THAT YOUR ABILITY TO OBTAIN TRANSPORTATION, LOGISTICS AND/OR DELIVERY SERVICES THROUGH THE USE OF THE SERVICES DOES NOT ESTABLISH UBER AS A PROVIDER OF TRANSPORTATION, LOGISTICS OR DELIVERY SERVICES OR AS A TRANSPORTATION CARRIER.”) (capitalization in original).


32 Trip Issues and Refunds, Uber, https://help.uber.com/h/595d429d-21e4-4c75-b422-72affa33c5c8 [https://perma.cc/SJ6L-V282] (last visited Oct. 14, 2018) (providing resources for various issues with an unprofessional driver, including if the driver’s behavior made the rider feel unsafe).

33 U.S. Terms of Use, supra note 29.


36 A Guide to Uber: Rating a Driver, Uber, https://help.uber.com/h/7b64ddaf6-78f5-4575-b7da-3c9e40d2c816 (as it stood through at least August 30, 2018, as preserved by Archive.org http://web.archive.org/web/20170707215004/https://help.uber.com/h/7b64ddaf6-78f5-4575-b7da-3c9e40d2c816 [https://perma.cc/33KD-FNNX]).

Compared to other online marketplaces, Uber exercises more control over transactions. Many marketplaces let sellers post asking prices, allowing independent sellers to set differing prices in light of their costs, impatience, and other factors. For example, each seller on Craigslist independently sets an asking price, as do sellers with fixed prices or buy-it-now prices on eBay. In contrast, Uber both specifies the base price (per minute and per mile rate) in each city, and sets minute-by-minute “surge” price increases based on local supply and demand. Furthermore, many marketplaces let the parties to a transaction choose or approve each other. For example, a buyer on Craigslist or eBay can decide which seller to buy from. Moreover, a Craigslist seller can decline to do business with an unwanted buyer and eBay sellers can decline to do business with buyers with low reputation scores. By contrast, Uber assigns drivers to passengers and vice versa. In theory, a driver dissatisfied with a proposed passenger can decline the ride request or cancel, and a passenger dissatisfied with a driver can do the same, but Uber tracks cancellations by both drivers and passengers and views cancellations unfavorably. Taken collectively, these factors put Uber in a distinct position of control over transactions between passengers and drivers, notably more so than other marketplaces, which could impact the application of § 230 as discussed in Section V.C.

C. AirBNB and Short-term Rentals

Similar to Uber’s narrow scope, short-term rental platforms focus on informal accommodations. They provide an online environment where hosts and guests can find each other, communicate, and transact.
Short-term rental platforms have proven popular. Best known and largest by market share is Airbnb. As of October 2018, Airbnb offered over 5 million listings in 191 countries, ranging from a room shared with others to entire apartments, homes, mansions, and palaces. Competitors include Flipkey, HomeAway, and VRBO. Airbnb has grown sharply; in the summer of 2010, roughly 47,000 guests stayed with Airbnb hosts, but by the summer of 2015 the number had grown more than 300-fold to 17 million.

Consistent with demands from both guests and hosts, and in an effort to streamline the process and avoid or resolve all manner of disputes, short-term rental platforms typically offer a wide range of features. They provide systems for payment, including calculating the amount payable, holding deposits, and sometimes collecting and remitting applicable taxes. They provide mechanisms to receive, process, and report reviews and reputations, including adjudicating which reviews are trustworthy and removing those that fail standards. They assist with communication between guests and hosts, culminating in dispute resolution when requested by either side. Additionally, they provide structured communication systems to help guests search for properties and to help hosts describe and market their properties, sometimes even sending professional photographers to present a property at its best. Additionally, they provide certain insurance and guarantees, protecting guests against certain malfeasance by hosts and vice versa.

Despite the popularity of short-term rentals, most such rentals appear to violate a variety of state and local laws. Critics flag a series of concerns. For one, hosts often offer short-term rentals in properties zoned only for ordinary residential use. In many jurisdictions, zoning offers no exception even for de minimis commercial use. In any event, many Airbnb hosts offer properties

---

52 For example, Airbnb’s Host Guarantee will reimburse eligible hosts for up to $1,000,000 in damages. The $1,000,000 Host Guarantee, AIRBNB, https://www.airbnb.com/guarantee [https://perma.cc/XDC7-YJR3] (last visited Nov. 8, 2017).
continuously, exceeding any notion of *de minimis* exceptions.\(^5^4\) Second, many jurisdictions impose substantial taxes on short-term rentals\(^5^5\) as voters rationally elect to tax outsiders, both because outsiders lack the political mechanisms to oppose such taxes, and because they perceive that visitors will not respond to such taxes by traveling elsewhere.\(^5^6\) Yet short-term rentals largely have not paid these taxes.\(^5^7\) Third, most jurisdictions have higher safety requirements for commercial properties. For example, hotels are often required to install automatic fire suppression systems such as sprinklers\(^5^8\) as well as provide nonflammable bedding.\(^5^9\) These protections respond to a series of incidents in which hotels suffered disasters of exceptional severity,\(^6^0\) but these protections also increase the costs for new entrants seeking to provide accommodations. As applied to hosts offering accommodations in their own homes or in residential units they coordinate, short-term rental marketplaces often argue that these laws are outdated or inapplicable.\(^6^1\)

D. StubHub and Ticket Resale Marketplaces

A variety of ticket resale sites connect buyers and sellers of tickets for athletic, cultural, and entertainment events. Best known is StubHub, founded...
in 2000 and acquired by eBay in 2007. Competitors include Razorgator, SeatGeek, Ticket Liquidator, and Vivid Seats. Each marketplace shows a range of tickets from third-party sellers, lets a buyer choose the desired seats, and charges a commission on each purchase. Marketplaces typically add payment processing, customer service, and insurance to streamline service to buyers and increase buyer confidence.

Marketplaces face tensions when sellers seek to sell tickets above face value ("scalping"). In many states, such resales are broadly unlawful. Further disputes arise when ticket brokers use "bots" to buy tickets en masse, then resell them at a markup through online ticket marketplaces, violating state laws as well as the terms and conditions of the original ticket sale contract.

Critics suggest that ticket marketplaces normalize this unlawful activity and profit from it (since marketplace fees are largely proportional to purchase price). However, marketplaces defend themselves as neutral actors, allowing sellers to list tickets at, above, or below face value as they see fit. Marketplaces typically require ticket sellers to assure that sales and selling prices are legal. For example, Stubhub’s user agreement instructs: "When setting the sale price of your tickets, it is your responsibility to comply with all applicable laws, statutes, and regulations." Nonetheless, disputes arise, particularly when consumers feel they paid too much.

III. Enforcement Efforts

Online marketplaces increasingly facilitate and coordinate activities that impact the physical world—particularly in the spheres traditionally regulated by state and local governments, such as transportation, housing, and tourism. Yet online marketplaces are often able to bypass existing regulations. The application of existing regulations may be unclear under the law. For example, regulations may anticipate only direct relationships between individuals rather than those facilitated by online marketplaces or formal

---

business entities. Thus, enforcement may be impractical, particularly where enforcement agencies lack the information, resources, or experience necessary to pursue a large number of small suppliers. Lastly, as this Article explains, federal law may limit, or may be perceived to limit, responses by state and local governments. The absence of enforcement can introduce a series of market failures. For example, casual service providers may avoid paying a variety of taxes and fees, including income and payroll taxes.68 Indeed, some of the cost advantages of ride-hailing services come from avoiding the fees that taxis pay for the wear on roads and for use of other public spaces. Furthermore, lower prices for ride-hailing services can induce passengers to forego public transportation,69 with resulting externalities including pollution and congestion.70 Meanwhile, casual service providers may find it tempting to discriminate against customers of disfavored races,71 genders,72 sexual orientations,73 or other classifications.

In addition, regulators and enforcement agencies may reasonably be concerned about the apparent asymmetry in regulation of incumbent firms versus online marketplaces and the casual providers they coordinate.74 Even if existing regulations are an imperfect fit for new marketplaces, a complete absence of regulation could cause even larger distortions that push activity to new unregulated platforms and away from existing firms, which are conditioned to better internalize the costs associated with consumer protection. Typically, these market failures do not directly affect marketplace users, who often enjoy and benefit from marketplace transactions. Rather, the transactions at issue create externalities: aggrieved parties tend to be outsiders and not parties to the transactions. The main factor pushing marketplaces to follow applicable law is the prospect of enforcement action—raising the

---

70 See generally Katrina M. Wyman, Taxi Regulation in the Age of Uber, 20 N.Y.U. J. LEGIS. & PUB. POL’Y 1 (2017) (arguing that regulators have not been responsive to taxi apps and new regulations need to encompass both traditional taxis and app-dispatched taxis).
72 Id. at tbl.4.
75 For example, California instituted fines of up to $5,000 for each active and contracted driver who was not subjected to a background check or who failed one. CAL. PUB. UTIL. CODE
question of how state and local governments can regulate these marketplaces. The following subsections survey approaches taken and challenges faced by select state and local governments: asking marketplaces to verify compliance, shaping marketplace design, imposing taxes, and gathering certain user information.

A. Verification

Regulators frequently seek to require marketplaces to verify that users comply with various regulatory requirements. For example, in several states ride-sharing services must ensure that their drivers pass background checks. However, compliance is not guaranteed. During nine months of 2015, the City of San Francisco required short-term rental hosts to register, pay a fee, and comply with additional requirements; but only 1082 of approximately 5378 properties registered as required.

Since short-term rental marketplaces conceal information about hosts (such as names, contact information, and street addresses), it is particularly difficult for regulators to find offending hosts and enforce applicable requirements. And, as described more fully in the next section, short-term rental marketplaces often argue that § 230 shields them from any verification requirements—even in cases where the marketplace’s only obligation is not to accept a fee for illegal transactions.

B. Design

Online marketplaces are built environments, arguably with no “natural” or “necessary” design. Therefore, regulators find it natural to seek specific changes to a marketplace’s composition. The simplest addition to a marketplace is to require the inclusion of certain disclosures. In principle,


the information could be as simple as a static disclosure on a log-in or other type of screen. For example, the State of New York requires each Transportation Network Company ("TNC"), such as Uber and Lyft, to display complaint procedures and the timeframe for the resolution of complaints on the main page of the TNC’s site.79

In other circumstances, a regulator or enforcement agency might ask a marketplace to add more complex information. Massachusetts, for instance, requires TNCs to display fare estimates to riders.80 In California, TNCs must allow disabled passengers to indicate whether they need an accessible vehicle, which requires TNCs to add buttons or similar mechanisms to receive such requests.81

Marketplaces vary in their responses to regulators’ requests for additions. The examples in the preceding paragraphs were largely straightforward, seeking at most increased prominence of features that the marketplaces presumably already provided. But marketplaces oppose requirements they consider too intrusive or otherwise burdensome, and they typically fight them through lobbying and litigation.82 For example, in 2014, Portland, Oregon began to require that short-term rental hosts obtain permits, and that hosting marketplaces display hosts’ permit numbers.83 Homeaway refused to display the numbers and Portland sued. Once again § 230 was central to the litigation, leading the city to abandon the effort.84

Other regulations might reasonably ask marketplaces to withhold certain information.

80 MASS. GEN. LAWS ANN. Ch. 159A 1/2, § 2(d) (2016).
81 CAL. PUB. UTILS. COMM’N, R. 12-12-011, BASIC INFORMATION FOR TRANSPORTATION NETWORK COMPANIES AND APPLICANTS (Rev. July 6, 2015), http://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/Licensing/Transportation_Network_Companies/BasicInformationforTNCs_7615.pdf (“TNCs must allow passengers to indicate whether they require a wheelchair-accessible vehicle”).
84 See City of Portland v. HomeAway, Inc., 240 F. Supp. 3d 1099, 1114 (D. Or. 2015) (stating that the City admitted in supplemental pleadings in the case that § 230 prevents it from holding Homeaway liable for its failure to provide hosts’ registration numbers). Airbnb also lobbied against this requirement before it was enacted; then, after it was enacted, Airbnb sought its retraction. See Steve Law, Airbnb Lobbying Portland to End City Inspection of Short-Term Rentals, PORTLAND TRIB. (Aug. 10, 2017), http://portlandtribune.com/pt9-news/368939-251546-airbnb-lobbying-portland-to-end-city-inspections-of-short-term-rentals [https://perma.cc/CS7F-WZC3].
This approach is most plausible in the context of discrimination. Employers have long been limited in their ability to ask certain questions of job applicants, as the discriminatory impact of such questions is understood to outweigh any proper purpose. Marketplaces similarly collect and distribute a wide range of information, which can facilitate discrimination.

C. Tax

Cities and states seek to tax marketplace transactions for several reasons. Every increase in the tax base allows correspondingly lower taxes across the board. Conversely, failing to tax reduces the tax base, requiring correspondingly larger taxes on other goods and services. Taxing marketplace transactions is necessary for equity with taxed competitors. Consider a potential guest’s choice between a hotel room versus a short-term rental. If a guest thinks a $100 hotel room is comparable in quality to a $110 Airbnb rental but the hotel room is subject to a 20% tax, the guest will choose the latter. Yet, had the purchases been taxed similarly, the hotel would have prevailed.

A concerned state or municipality could attempt to collect taxes directly from marketplace sellers, such as Airbnb hosts. But this approach has obvious challenges. For instance, there are a large number of hosts, requiring correspondingly large enforcement efforts. As discussed in Section II.A, marketplaces widely conceal the names and contact information of their sellers, impeding enforcement efforts premised on communication with those who purportedly owe tax. San Francisco’s 2015 experience attempting to collect transient occupancy taxes from Airbnb hosts, without cooperation from Airbnb, illustrates the difficulty and predictably low compliance.

Despite these challenges, states and municipalities have nonetheless had some success seeking assistance from marketplaces. Airbnb now concedes that taxes are due and assists in collecting them on the behalf of local governments. That said, Airbnb only offers this benefit if a jurisdiction

\[\text{REFERENCES}\]

85 See Adam Samaha & Lior Jacob Strahilevitz, Don’t Ask, Must Tell—And Other Combinations, 103 CAL. L. REV. 919, 946 (2015); Lior Jacob Strahilevitz, Reputation Nation: Law in an Era of Ubiquitous Personal Information, 102 NW. L. REV. 1667, 1711–12 (2008).


87 POLICY ANALYSIS REPORT, supra note 77.

otherwise accedes to Airbnb’s favored regulatory scheme, such as Airbnb’s approach to short-term rentals generally, zoning, enforcement, and more. So while a jurisdiction may be able to tax short-term rentals, it then foregoes other policies contrary to the online marketplace’s preferences.

Similar tax disputes arise in the context of ride hailing. During the period in which Uber and other TNCs operated without regulatory authorization, the company generally failed to collect or remit the various taxes and fees that apply by law, such as airport fees and commercial tolls. In due course, TNCs typically agreed to pay certain airport fees—but only as jurisdictions approved of their overall approach.

Amazon Marketplace also demonstrates the challenges that result from ignoring taxes on marketplace transactions. As of 2017, approximately half of Amazon’s sales came from “Marketplace” sellers. These sellers list their products on Amazon’s site, and many store their goods in Amazon’s warehouses and pay Amazon to pack and send their products. Nonetheless, Amazon presents Marketplace sellers as independent entrepreneurs who are individually responsible for collecting their own taxes. While Amazon in 2017 promised to collect and remit state sales taxes in all applicable states, Marketplace sellers are beyond the scope of that promise, prompting renewed battles between Amazon and tax authorities.

---

89 Airbnb Policy Tool Chest, supra note 88, at 4 (describing “Voluntary Collection Agreements” in which Airbnb collects taxes a jurisdiction requires, a benefit Airbnb offers only if a jurisdiction broadly allows Airbnb hosting).


91 Massachusetts is illustrative: As part of Massachusetts broadly accepting Uber, the company gained permission to operate at Logan International Airport, and began paying fees for each such pickup. Adam Vaccaro, Uber Gets Permission to Operate at Logan, Boston Globe (Jan. 31, 2017), https://www.bostonglobe.com/business/2017/01/31/uber-gets-permission-operate-logan/LGLwinZSrFBD2zWpQLeXFJ/story.html [https://perma.cc/QZ8N-3Z7M].


95 Enforcement of tax payments is ineffective because Amazon has approximately two million Marketplace sellers and sellers may not know which warehouses hold their inventory.
In an effort to avoid the mammoth task of contacting and pursuing Marketplace sellers, several states force or have attempted to force Amazon to collect and remit taxes by defining the platform as a “marketplace facilitator.”\(^6\) Amazon has resisted these efforts. For example, South Carolina argued that Amazon is responsible for collecting taxes on behalf of businesses that used Amazon Marketplace to sell products to South Carolina residents.\(^7\) Among other factors, South Carolina noted that Amazon itself controls to whom and where items are sent; Amazon sends the items itself; Amazon receives and holds payment; and Amazon controls the transaction, such as customer service and returns.\(^8\) Amazon responded with a protest letter, and litigation is ongoing.\(^9\)

**D. Disclosure of User Information**

Regulators sometimes seek to observe user activity in order to monitor behavior on marketplaces and pursue alleged improprieties. Sometimes, regulators may be content to collect data from a marketplace using “scraper” software that examines listings. But these methods sometimes prove ineffective.

A first challenge is the scale of operation. For example, the eBay auction marketplace hosts approximately 1.1 billion listings at any moment, with tens of millions more added each week.\(^10\) While standard site search tools can focus attention on particular terms, an enforcement agency searching for recalled items, counterfeits, or the like would struggle to find all listings of concern.

Constrained by the same standard search tools available to the public, enforcement necessarily lags behind marketplace activity. For example, a prohibited new listing would be present for at least hours, if not days or weeks, before a periodic enforcement search found it, documented its violations, and demanded that the marketplace remove it. When misconduct creates particularly severe harms, such as risks to health and safety, this delay may be seen as unacceptable. In response, Massachusetts in November 2016 began to require that TNCs pre-submit their proposed new drivers for ad-

---


\(^9\) Id.
A third challenge is that marketplaces sometimes conceal the information of greatest importance to regulators. For example, a host on Airbnb can post a verbose description of the property and unlimited photographs—but cannot provide a full legal name, email address, mailing address, or property address. Many hosts would not want to post such information, but Airbnb has clear business reasons to prohibit such postings: If guests could contact hosts directly, they would circumvent Airbnb’s booking service and the associated fees, which would likely require Airbnb to find a new business model. But once Airbnb conceals this information, enforcement agencies cannot find potentially unlawful listings within their jurisdictions. With no other option available, it is natural for regulators to pursue information, if not assistance, from Airbnb in their enforcement efforts.

In response, regulators sometimes seek superior access to marketplace data. For example, in 2013, the New York Attorney General Eric Schneiderman sought data on 15,000 hosts who appeared to be violating applicable zoning and tax requirements. Airbnb resisted, moving to quash the subpoena as an overbroad “fishing expedition.” Airbnb further argued that the applicable laws were unconstitutionally vague, that the requested production was burdensome, and that hosts’ information was private and should not be subject to subpoena. The court rejected each of these contentions, save for concern about overbreadth, finding that the subpoena must confine itself to listings for jurisdictions and lengths that violate applicable zoning laws. In response, Airbnb and Schneiderman reached an agreement, and Airbnb shared some of the requested user information.

---


109 Letter Confirming Agreement Regarding Compliance with Subpoena from Clark Russell, Deputy Bureau Chief of the Internet Bureau, Office of the Att’y Gen. of the State of N.Y.,
For a marketplace facing a regulator’s demand for user information, a key challenge is that such data could extinguish business models grounded in noncompliance with applicable laws. So long as regulators cannot easily find the names and addresses of hosts violating New York regulations, some hosts will break those laws to increase their profits. But with their names and details available to the attorney general, hosts have reason to pause. Indeed, subsequent to the attorney general’s successful demand for host information, Airbnb dropped hundreds of hosts in Manhattan, suggesting that compelled disclosure sharply affected the marketplace’s prospects there.

IV. ARGUING AGAINST STATE AND LOCAL REGULATION

Whatever the benefits of state and local regulation of online marketplaces, marketplaces primarily argue that regulators simply cannot regulate them. Arguments against state and local regulation build on the federal immunity provided by § 230.111

Section 230 was originally enacted, in part, to encourage providers of interactive computer services (“ICS”) to moderate user-provided content without fear of publisher liability,112 but on the whole, courts have offered a notably broader interpretation of this provision. For one, courts have held that § 230 immunizes not only efforts to moderate user content but also decisions not to moderate.113
Moreover, courts have found that the “robust” immunity of § 230 applies even if providers knew of allegedly unlawful material, encouraged it, altered the design of their services to facilitate it, and charged for the assistance they provided. Courts retain the immunity despite these circumstances because a narrower immunity “would have an obvious chilling effect” in discouraging providers from monitoring their services.

Proponents of these broad interpretations also assert that “the entire Web 2.0 revolution through which thousands of smaller innovative platforms have come to offer a range of socially useful services” would not be possible without § 230. On this view, § 230 is a “masterpiece” and “a remarkable success.” And, “[n]o other sentence in the U.S. Code . . . has been responsible for the creation of more value.” Numerous scholars agree. Tech companies are similarly effusive. For example, the Internet Association (representing forty technology companies such as Amazon, Facebook, Google, and Microsoft) said § 230 was crucial to keeping the Internet “free,

---

114 Carafano v. Metrosplash, 339 F.3d 1119, 1123 (9th Cir. 2003).
117 Doe, 817 F.3d at 22 (finding immunity when defendant designed its service to facilitate sex trafficking, including declining to verify phone numbers, declining to verify email addresses, and removing all metadata from photographs used in advertisements in order to impede investigations).
120 Brief for Chris Cox and Netchoice as Amici Curiae Supporting Plaintiffs and Reversal Motion, HomeAway.com, Inc. and Airbnb, Inc. v. City of Santa Monica at 7, No. 18-55367 (9th Cir. Apr. 25, 2018).
From the Digital to the Physical 161

innovative, and collaborative.” The Electronic Frontier Foundation, arguably the best-known digital rights group, called § 230 “[t]he most important law protecting Internet speech,” and Wired said § 230 was “the most important law in tech.”

Online marketplaces have embraced § 230 when seeking to invalidate legislation and regulation they dislike. Best known is Airbnb’s June 2016 litigation against the City of San Francisco, where Airbnb argued that the city’s amended short-term rental ordinance violates § 230 because it “treats online platforms such as Airbnb as the publisher or speaker of third-party content” and thus is completely preempted. In seeking a preliminary injunction against enforcement of the ordinance, Airbnb emphasized that the ordinance “punish[es] platforms for failing to verify and screen third-party listings” which “directly conflicts with the CDA and is barred under settled law.” Distinguished amici (including the Internet Association and the Electronic Frontier Foundation) sought to file briefs with the court in support of Airbnb’s position. Airbnb also made similar arguments in its subsequent

case against the City of Santa Monica, which substantially copied the San Francisco ordinance.131

A. The Prima Facie Case Under § 230

To invalidate a state or local law under § 230’s immunity, a marketplace must establish three elements: first, that the marketplace operates an interactive computer service within the meaning of § 230; second, that third parties provided the information at issue; third, that the challenged state or local law conflicts with § 230 by treating the interactive computer service as a publisher or speaker of that information.132 Quickly and with little reflection, marketplaces attempt to claim that they easily satisfy all three elements.

First, marketplaces call for a broad interpretation of “interactive computer service.”133 In describing eBay, Airbnb, or Uber, a lay person might refer to them as modern replacements for malls, hotels, or taxis. However, the marketplaces claim that their operations are grounded in the provision of interactive computer services whereby buyers and sellers find one another to transact.

Indeed, the services are “interactive” in the sense that they provide users with the ability to submit, filter, and process information, among numerous other interactive features. People access these services via a “computer” broadly understood (notably including smartphones). Courts are similarly inclusive in their interpretation of “interactive computer service,” so this element is usually straightforward.134

Second, marketplaces allege that the disputed behavior at issue arises from third-party actions. When a seller uploads a listing within the eBay marketplace for a product that is itself unlawful (perhaps a counterfeit, a recalled item, or an item not licensed for sale within the United States), it provides all the information associated with the listing, including the title, description, and photos.

Third, marketplaces allege that when regulations hold them liable for illegal transactions, the regulations improperly treat the marketplaces as publishers or speakers of content provided by users.135 Consider a state law disallowing the sale of recalled items or banning ticket sales above face value.

132 Batzel v. Smith, 333 F.3d 1018, 1037 (9th Cir. 2003).
133 See, e.g., Complaint at 13, Airbnb, Inc. v. City of Anaheim, No. 8:16-cv-1398 (C.D. Cal. July 28, 2016); Memorandum of Defendant StubHub, Inc. in Support of Its Motion to Dismiss at 13–17, City of Chicago v. StubHub, Inc., 624 F.3d 363 (7th Cir. 2010) (No. 09-3432).
eBay and StubHub would argue that liability for including such listings in their respective marketplaces is exactly what § 230 prohibits.\textsuperscript{136} Despite the apparent simplicity of prima facie § 230 claims, the additional actions taken by marketplaces to cause, assist, and facilitate the disputed information and activities give rise to many potential disputes. These disputes are the focus of Section V, below.

\textbf{B. Fellow Travelers}

While marketplaces frequently rely on § 230 to challenge state and local regulation, additional federal laws also grant Internet-based companies special privileges and immunities. Since these federal laws are less often used to avoid regulation, this Article explores them only briefly.

1. 1996 Federal Telecommunications Act

Some marketplaces seek shelter in the 1996 Federal Telecommunications Act\textsuperscript{137} (“FTA”), the parent act of § 230. As the first significant overhaul of federal telecommunications law in more than sixty years, the FTA broadly sought to increase competition among telecommunications firms, in part by reducing regulation.\textsuperscript{138}

In an attempt to fight off regulation of its activities in both Maryland and California,\textsuperscript{139} Uber argued that the purposes and objectives of the FTA give rise to conflict preemption.\textsuperscript{140} Specifically, Uber claimed that in order to establish a “pro-competitive, deregulatory national policy framework,”\textsuperscript{141} the FTA distinguished between “telecommunications services,”\textsuperscript{142} which are subject to regulation by the Federal Communications Commission as com-

\textsuperscript{136} Id.
\textsuperscript{140} Conflict preemption exists whenever the “challenged state statute ‘stands as an obstacle to the accomplishment and execution of the full purposes and objectives of Congress.’” Perez v. Campbell, 402 U.S. 637, 649 (1971) (quoting Hines v. Davidowitz, 312 U.S. 52, 67 (1941)).
\textsuperscript{142} “Telecommunications” is “the transmission, between or among points specified by the user, of information of the user’s choosing, without change in the form or content of the information as sent and received.” 47 U.S.C. § 153(50) (2012). “Telecommunications service” is “the offering of telecommunications for a fee directly to the public, or to such classes of users as to be effectively available directly to the public, regardless of the facilities used.” 47 U.S.C. § 153(53).
mon carriers, and “information services,” which are not. This distinction arguably showed Congress’s intent to “occupy the field of regulation of information services,” and thus “regulations that have the effect of regulating information services are in conflict with federal law and must be preempted.”

Under this line of reasoning, Uber and other online marketplaces must demonstrate that they are “information service” providers in order to fall under the FTA’s preemption umbrella. The FTA identifies two requirements to qualify as an “information service”: 1) the provider must offer “a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information via telecommunications,” and 2) the provider must not use “any such capability for the management, control, or operation of a telecommunications system or the management of a telecommunications service.” The second requirement ensures the mutual exclusivity of telecommunications and information services.

Relying on its familiar claim that “Uber Technologies, Inc. is a technology company,” Uber asserted in a dispute with the Maryland Public Service Commission that it satisfied the first requirement because the core of its business makes information available about third-party providers of transportation services. Similarly, it claimed to satisfy the second requirement because it uses various telecommunications services (wireline and wireless Internet providers) to transmit information and does not offer or manage a telecommunications service itself. Uber’s argument went untested because the case settled.

2. Stored Communications Act

Marketplaces also invoke the Stored Communications Act (“SCA”) to avoid state and local regulations that require marketplaces to share user data with regulators. The SCA limits when and how governments may access information about online marketplace users. The requirements vary based on

---

143 Memorandum on Appeal of Uber Technologies, Inc., supra note 127, at 42.
144 Id. at 45 (quoting Vonage Holdings Corp. v. Minn. Pub. Util. Comm’n, 290 F. Supp. 2d 993, 1002 (D. Minn. 2003) (finding that federal law preempted Minnesota’s imposition of fees on Vonage’s Voice over Internet Protocol)).
147 Id.
2019] From the Digital to the Physical

the type of service and the type of information sought.\textsuperscript{150} For the actual contents of communications,\textsuperscript{151} the SCA requires a warrant, a subpoena with notice to the customer, or a court order based on “specific and articulable facts” with notice to the customer.\textsuperscript{152} State and local regulations often contemplate marketplaces sharing user information without such protections,\textsuperscript{153} leading marketplaces to invoke the SCA in challenging those regulations.

Airbnb’s dispute with San Francisco is illustrative. The August 2016 San Francisco ordinance required Airbnb to periodically disclose to the City the names and addresses of certain hosts, without any requirement that the city obtain a court order.\textsuperscript{154} In challenging the ordinance, Airbnb invoked the SCA; however, rather than litigate this aspect of Airbnb’s defense, the city agreed to demand user data only with a subpoena.\textsuperscript{155}

3. Federal Tax Laws

Marketplaces have attempted to enjoin state and local governments from imposing tax obligations on them under a variety of federal principles and laws,\textsuperscript{156} including the Due Process Clause,\textsuperscript{157} Dormant Commerce Clause,\textsuperscript{158} and the Internet Tax Freedom Act (“ITFA”).\textsuperscript{159}

Constitutional arguments are usually unsuccessful, particularly under the Due Process Clause. In South Dakota v. Wayfair,\textsuperscript{160} the Supreme Court found that the Dormant Commerce Clause does not demand a physical nexus in order for a state to impose tax obligations on companies.\textsuperscript{161} Prior to 2018,

\begin{footnotesize}
\begin{enumerate}
\item[150] 18 U.S.C. § 2703 (2012). If a marketplace is considered a provider of an electronic communications service (“ECS”), meaning it provides users with the ability to send or receive wire or electronic communications, then law enforcement must have a search warrant to obtain “content information” stored on its servers for 180 days or less. Id. § 2510. If the government wants content information stored for more than 180 days by an ECS or information retained by a remote computing service regardless of the length of storage, the government can obtain that information based on a warrant, subpoena with notice to the customer, or a court order based on “specific and articulable facts” with notice to the customer. Id. § 2703(b)(1)(B).
\item[155] See, e.g., Complaint, supra note 156, at 3.
\item[156] Id.
\item[158] 138 S. Ct. 2080 (2018).
\item[159] Id. at 2099 (overturning Quill v. North Dakota, 504 U.S. 298 (1992)).
\end{enumerate}
\end{footnotesize}
a jurisdiction could impose tax obligations on a company only if the company had an actual presence in that jurisdiction. The structure of online marketplaces made it particularly easy for marketplaces to argue that they had no such presence. However, after Wayfair, states can tax a marketplace as long as it “avails itself of the substantial privilege of carrying on business in that jurisdiction.” This can be established based on both “economic and virtual contacts”—a relatively easy standard for states to satisfy.

Marketplaces have also attempted to use the ITFA, which forbids “[m]ultiple or discriminatory taxes on electronic commerce,” though, contrary to the suggestion in its title, the ITFA did not make transactions on the Internet tax-free. “Multiple” taxes refer to transactions that are taxed by two states without an appropriate tax credit. “Discriminatory” taxes treat Internet commerce differently than other types of commerce. Marketplaces have tried to invoke these protections. For example, StubHub argued that a Chicago tax on its ticket sales violated the ITFA because the tax was discriminatory, posing a distinctive burden on online marketplaces compared to offline resellers. However, the court found that the ordinance did not turn on “the role of a computer server or the provision of electronic services” because the ordinance considered all resellers the same regardless of the form of their services. Thus, the tax was not discriminatory. Nonetheless, the ITFA may still help protect other marketplaces from specially tailored taxes, as suggested by the pending litigation related to Chicago’s tax of online streaming services.

V. Arguing for State and Local Regulation

Despite the appeal of alternative theories, § 230 remains by far the leading authority for marketplaces seeking to avoid state and local regulation. This Section turns to the core arguments about applying that statute to

---

162 See Quill, 504 U.S. at 311–12.
163 Consider a host providing short-term rentals through Airbnb. Airbnb needs no physical presence in a state to offer this service. Therefore, Airbnb predictably argued that the host acts on his or her own—certainly not as Airbnb’s agent nor otherwise on Airbnb’s behalf—in its in-state activities.
164 Wayfair, 138 S. Ct. at 2099.
165 Id.
168 Internet Tax Freedom Act § 1105(6).
169 Id.
170 City of Chicago v. StubHub Inc., 624 F.3d 363, 366–67 (7th Cir. 2010).
171 Id. at 367.
172 Id.
online marketplaces. In total, there are five mechanisms whereby a court could avert application of § 230.

A. Rely on the Legislative History of § 230

Section 230 was written when the Internet was still in its youth, and the interpretation and importance of the law were not yet apparent.\footnote{When the CDA was enacted in 1996, less than eight percent of Americans had access to the Internet, and those who did went online for just thirty minutes a month. Farhad Manjoo, Jurassic Web, SLATE (Feb. 24, 2009, 5:33 PM), http://www.slate.com/articles/technology/technology/200902/jurassic_web.html [https://perma.cc/XJ8M-2CA2].} Moreover, the title of the statute, “Protection for Private Blocking and Screening of Offensive Material,” indicates that § 230 was not meant to provide unlimited immunity for marketplaces. Rather, it was designed to encourage website operators to remove “smut” from their systems without fear of liability.\footnote{Section 230 was written when the Internet was still in its youth, and the interpretation and importance of the law were not yet apparent. Moreover, the title of the statute, “Protection for Private Blocking and Screening of Offensive Material,” indicates that § 230 was not meant to provide unlimited immunity for marketplaces. Rather, it was designed to encourage website operators to remove “smut” from their systems without fear of liability.} Specifically, the text of § 230\footnote{Also known as the Cox/Wyden Amendment. 141 CONG. REC. H8471-2 (daily ed. Aug. 4, 1995) (statement of Rep. Goodlatte (R-Va.)). The CDA is “by no stretch of the imagination a libertarian enactment.” Danielle Keats Citron & Benjamin Wittes, The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity, 86 FORDHAM L. REV. 401, 404 (2017).} responded to Senator James Exon’s (D-Neb.) campaign to use the CDA to prohibit “obscene or indecent” messages online.\footnote{47 U.S.C. § 223(a)(1)(B)(ii), (d) (2012) (prohibiting the “knowing” transmission to children under the age of 18 “that, in context, depicts or describes, in terms patently offensive as measured by contemporary community standards, sexual or excretory activities or organs.”). These provisions were ultimately declared unconstitutional by the Supreme Court. Reno v. Am. Civil Liberties Union, 521 U.S. 844, 885 (1997). As stated by Senator Exon, “[t]he fundamental purpose of the Communications Decency Act is to provide much needed protection for children.” 141 CONG. REC. S8087–88 (daily ed. June 9, 1995) (statement of Sen. Exon); Robert Cannon, The Legislative History of Senator Exon’s Communications Decency Act: Regulating Barbarians on the Information Superhighway, 49 FED. COMM. L.J. 51, 57 (1996).} This “Exon Amendment” was the most threatening attempt to truly censor the Internet, and, fearing that it went too far, Representatives Ron Wyden (D-Or.) and Christopher Cox (R-Cal.) strove to find a private sector solution to the problem of online pornography.\footnote{Also known as the Cox/Wyden Amendment. 141 CONG. REC. H8471-2 (daily ed. Aug. 4, 1995) (statement of Rep. Goodlatte (R-Va.)). The CDA is “by no stretch of the imagination a libertarian enactment.” Danielle Keats Citron & Benjamin Wittes, The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity, 86 FORDHAM L. REV. 401, 404 (2017). As Representative Cox stated “We want to encourage [companies] like Prodigy, like CompuServe, like America Online, like the new Microsoft network, to do everything possible for us, the customer, to help us control, at the portals of our computer, at the front door of our house, what comes in and what our children see.”} Thus misconstruing § 230 to create a “lawless no-
man’s-land on the Internet”\footnote{Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1164 (9th Cir. 2008).} is a mistake and in direct contradiction with the statute’s history and purpose.\footnote{Brief for Internet, Business, and Local Government Law Professors as Amici Curiae Supporting Appellee and Affirmance of the District Court at 21, HomeAway.com, Inc. and Airbnb, Inc. v. City of Santa Monica 21, No. 18-55367 (9th Cir. Apr. 25, 2018).}

B. \textit{Focus on Substance, Not Form, When Classifying Interactive Computer Service Providers}

In extending far beyond computer service to real-world transactions with real-world implications, marketplaces may exceed the boundaries of § 230’s safe harbor because they are not purely providers of interactive computer services. This is especially true when the majority of a marketplace’s activities are outside the scope of immunity intended by Congress.

By its terms, § 230 limits its benefits to “provider[s] of an interactive computer service.”\footnote{47 U.S.C. § 230(c)(1) (2012).} The statute defines an interactive computer service as “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server.”\footnote{Id. at 213.} The definition of an interactive computer service has been interpreted broadly and includes not only Internet service providers and bulletin boards, but mobile applications and various websites.\footnote{Kosseff, supra note 122, 6–8; see also Zango, Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169, 1175 (9th Cir. 2009) (finding a provider of anti-malware software an ICS); Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1125 (9th Cir. 2003) (dating website); Gentry v. eBay, Inc., 99 Cal. App. 4th 816, 836 (Cal. Ct. App. 2002) (auction website); Milgram v. Orbitz Worldwide, Inc., 16 A.3d 1113, 1126–27 (N.J. Super. Ct. 2010) (travel website).} As a result, courts rarely pause to fully consider the definition’s outer limits.\footnote{Milgram, 16 A.3d at 1121 (“There is no issue that defendants qualify as an ‘interactive computer service’ as defined by the CDA.”); Gibson v. Craigslist, Inc., No. 08 Civ. 7735 (RMB), 2009 WL 1704355, at *3 (S.D.N.Y. June 15, 2009) (“Plaintiff does not appear to dispute that Craigslist is a provider of an interactive computer service.”). However, \textit{Federal Trade Commission v. LeadClick}, 838 F.3d 158 (2d Cir. 2016), is a notable exception. LeadClick managed a group of advertisers through tracking software, which recorded customer clicks and purchases. These advertisers engaged in deceptive trade practices, and the FTC attempted to hold LeadClick responsible. LeadClick argued that § 230 immunized it from liability, but the Second Circuit Court of Appeals found that § 230 did not apply because, among other reasons, LeadClick’s tracking software “was wholly unrelated to its potential liability under § 230.” Id. at 175.}

To the extent that [§ 230] represented an understanding of the complex future economic implications of ISP liability regimes, it demonstrated commendable foresight by its sponsors. It appears more likely, however, that the sponsors of the Cox/Wyden Amendment saw Stratton Oakmont as a political vehicle by which they could undermine the CDA.
From the Digital to the Physical

protection under § 230, a marketplace may struggle to claim to be a provider of an “interactive computer service” protected by § 230. The same would be true for a company that mostly operates offline and uses the Internet only to enhance its service. For example, consider a fitness studio that allows users to log on to its website to swap available spots with other participants in its group fitness classes. Despite the role of online scheduling, no one suggests that that the studio is a provider of interactive computer service or would be able to use § 230 to avoid local health and safety regulations.

Marketplaces’ own statements may undermine their providers’ claims to be interactive computer services, to the exclusion of broader functions. Consider Uber’s assertions about its scope and role. Uber historically presented itself to users with the motto “everyone’s private driver,” and then-CEO Travis Kalanick wrote on Uber’s official blog: “[W]e’re rolling out a transportation system in a city near you.” Elsewhere, Uber claimed that the company “provides the best transportation service in San Francisco.”

Having held itself out as a “transportation system” and “transportation service,” Uber struggles to establish that it is only, primarily, or importantly a provider of a computer service. The tension between § 230’s limited scope of “computer service” and Uber’s operations continues to increase in light of the company’s growing ambitions—including new lines of business in driverless transportation, health care, and food delivery.

When courts have considered marketplaces’ combination of computer service and other roles outside the context of § 230, they have been corre-

---

186 Zango presents the rare situation in which an anti-malware company was considered a provider of an interactive computer service simply because it tangentially used the Internet to update its software. The main purpose of the company’s software was to remove malware, and when the company’s program blocked the appellant’s website, the company used the CDA to protect itself from liability. The Ninth Circuit Court of Appeals was dismissive of the appellant’s argument that the word “interactive” requires that the platform provide people with access to the Internet. Instead, the court asserted that a provider of an ICS only needs to provide access to a computer server and not the Internet itself. However, regardless of the court’s interpretation of “provide” or “computer access” by “multiple users to a computer server,” the plain meaning of the term “provider” could still be interpreted to require a platform to primarily provide access to computer servers. Zango, 568 F.3d at 1179–80.

187 For companies that offer services primarily online, courts may take a different approach.

188 Thomas, supra note 27.


spondingly skeptical that computer service predominates. For example, in
denying Uber’s motion for summary judgment in a case about the classification
driver’s motion for summary judgment in a case about the classification of drivers, one court remarked that “Uber does not simply sell software; of drivers, one court remarked that “Uber does not simply sell software; it sells rides.” 192 The court continued:

Uber is no more a ‘technology company’ than Yellow Cab is a ‘technology company’ because it uses CB radios to dispatch taxi cabs, John Deere is a ‘technology company’ because it uses computers and robots to manufacture lawn mowers, Domino Sugar is a ‘technology company’ because it uses modern irrigation techniques to grow its sugar cane. . . . If . . . the focus is on the substance of what the firm actually does . . . it is clear that Uber is most certainly a transportation company. 193

Regulatory proceedings have taken a similarly dim view marketplace’s claims to provide only computer service. For example, the California Public Utilities Commission (“CPUC”) criticized Uber’s claim to be exempt from CPUC’s jurisdiction because it was an information service. CPUC explained: “We reject Uber’s assertion that TNCs are nothing more than an application on smart phones, rather than part of the transportation industry.”194

Indeed, a marketplace’s efforts may extend beyond the provision of interactive computer services in a variety of directions. A marketplace may seek to guarantee quality through insurance, investigations, make-goods, and credits, bringing the marketplace that much closer to the substance of the transaction.

When a marketplace provides incentives to bring in service providers, the marketplace’s role becomes that much larger and its involvement in the transaction that much deeper. When a marketplace sets detailed rules—what type of car an Uber driver may drive or what safety features an Airbnb host must provide—the marketplace similarly tests the boundaries of “computer service.” When these factors combine, a marketplace may end up looking like the organizing force behind a series of transactions: a far cry from simply providing an interactive computer service.

While courts have never seriously considered limiting the application of § 230 based on the statute’s “provider” language, congressional intent suggests that they should. The findings and policy in § 230(a)–(b)—written into the law itself—indicate Congress’s intent to “promote the continued development of the Internet,”195 and its intent to encourage the exchange of infor-

193 Id.
From the Digital to the Physical

171

196 Therefore, when a marketplace extends into wholly offline behavior (such as staying in a short-term rental), the stated purposes of § 230 call into question whether that statute should be read to immunize the offline behavior.

C. Define Marketplaces as Information Content Providers

Some marketplaces may be considered information content providers (“ICP”), exempting them from § 230 protection. If a person or company is “responsible, in whole or in part” for “the development or creation” of a given piece of information, the plain language of § 230 instructs that the person or company is an ICP rather than a provider of an “interactive computer service,” and hence not protected by the § 230 immunity.197 In their many efforts to facilitate and streamline transactions, marketplaces may cross this line and thereby lose the protections of § 230.

ICP status is the most litigated prong of § 230, and courts have applied different tests to it.198 An initial test asked whether an intermediary develops or creates content in a way that exceeds “traditional editorial functions,”199 such as withdrawing, postponing, altering, or organizing.200 If the intermediary’s role extends beyond editorial functions and includes making “material substantive contribution[s] to the information that is ultimately published,” this test finds the intermediary is an ICP outside the scope of § 230 protection.201

The activities of some online marketplaces cross the line into editorial functions. Consider marketplaces that set prices and provide guarantees, insurance, credits, incentives, and dispute resolution mechanisms for users.202

Beginning in 2008, some courts shifted to a new test for ICP status. Looking beyond editorial functions, Fair Housing Council of San Fernando

196 Id. § 230(a)(3) (“The Internet and other interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity”); id. § 230(b)(3) (“[T]o encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services.”).


199 Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997); see also Carafano v. Metosplash.com, Inc., 339 F.3d 1119, 1124 (9th Cir. 2003) (finding that a matchmaking website did not constitute a ICP because user profiles had no content until users actively created them, even if some content was generated from a user questionnaire); Blumenthal v. Drudge, 992 F. Supp. 44, 50 (D.D.C. 1998) (holding AOL was an ICP because there was “no evidence . . . that AOL had any role in creating or developing any of the information” in defamatory statements).


201 Donato, 865 A.2d at 727.

202 See supra Section V.A.
Valley v. Roommates.com instructed that “a website helps to develop unlawful content, and thus falls within the exception of § 230, if it contributes materially to the alleged illegality of the conduct.” Under this test, a roommate search site was found to materially contribute to illegal conduct because its tools (such as dropdown menus) forced individuals to specify preferences based on unlawful criteria (such as gender and family status).

Subsequent cases have interpreted the Roommates.com test. Some courts offered a narrow interpretation known as the “solicitation standard,” which finds an intermediary to be an ICP, outside the protections of § 230, only if it specifically solicited illegal behavior. The leading case for the solicitation standard is Federal Trade Commission v. Accusearch, Inc., in which a defendant ran a website that collected consumer requests for phone records, then obtained those phone records from third-party providers in violation of the Telecommunications Act. The court held that “a service provider is ‘responsible’ for the development of offensive content only if it in some way specifically encourages development of what is offensive about the content.” Because Accusearch specifically encouraged third-party providers to commit illegal acts, it was deemed an ICP without § 230 protection.

Under Accusearch’s solicitation standard, marketplace actions can enter into the specific encouragement that Accusearch disallows. For example, Uber employees coached drivers on avoiding detection by airport police who sought to ticket unauthorized commercial pickups. A closer question arises when most or substantially all user activity is unlawful, for example when a jurisdiction disallows short-term rentals. An intermediary might reasonably argue that continuing to operate its standard service does not “encourage” the misbehavior and certainly does not “specifically encourage” it. Yet some intermediary actions are reasonably un-

203 Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1168 (9th Cir. 2008) (emphasis added).
204 Id.
206 570 F.3d 1187 (10th Cir. 2009).
207 Id. at 1191–92.
208 Id. at 1199 (emphasis added).
210 Tonytee, Comment to FLL Warning to Drivers, UBERPEOPLE.NET (Nov. 4, 2014), https://uberpeople.net/threads/fll-warning-to-drivers.4509/page-2#post-74768 [https://perma.cc/L8JY-BX3D] (publishing the content of an email sent by Uber Miami employees to drivers).
212 See Accusearch Inc., 570 F.3d at 1199.
derstood as targeting specific hosts and transactions, arguably “specifically encoura[ing]” the corresponding transactions.\footnote{See id.} Consider sign-up bonuses to guests and/or hosts, as well as free photography for hosts—in each case, encouraging a specific transaction or set of transactions. One might be similarly skeptical of the marketplace’s provision of services in a jurisdiction that disallows short-term rentals.\footnote{Contra NPS LLC v. Stubhub, Inc., No. 06-4874-BLS1, 2009 WL 995483 (Mass. Super. Ct. Jan. 26, 2009) (holding that StubHub was an ICP because it used “improper means,” such as providing training materials and incentives, to intentionally induce or encourage ticket sellers to violate anti-scalping laws. This was a broader interpretation of Roommates.com, the inducement standard, which captures general encouragement); cf. Hill v. Stubhub, Inc., 727 S.E.2d 550, 563 (N.C. Ct. App. 2012) (refusing to follow NPS because the court did not find the reasoning persuasive); Milgram v. Orbitz Worldwide, Inc., 16 A.3d 1113, 1127 (N.J. Super. Ct. 2010) (asserting that NPS contradicts the spirit of Donato).} New marketplaces often create the architecture that systematizes the illegality. The details depend on what the statute or regulation requires, especially how it apportions responsibility between marketplace and service provider. But consider a statute that requires every short-term listing to publish its exact street address on every advertisement, versus a marketplace that bans such publication, designs software to block such information, and employs staff to further check for and remove such information. On those facts, surely the marketplace would be “responsible in whole or in part” for the omission and indeed would have “specifically encouraged” the omission that is offensive—having required the omission by both marketplace policy and technical and manual enforcement.

Where a marketplace all but creates a market, there is a fair argument that the marketplace is “responsible in whole or in part” for what follows. Of course, travelers stayed in strangers’ homes before Airbnb, including in the original bed-and-breakfasts, and passengers paid to ride in private cars, including in licensed taxis, in pirate taxis, and with friends and family. Yet Airbnb and Uber dramatically expanded the size of these markets by designing software to make these relationships easy and routine, and by promoting these transactions through advertising, incentives, and pricing. They further formalized these markets through screening, insurance, customer service, and dispute resolution to increase the number of customers and service providers inclined to participate. They also lobbied and litigated to advance and defend the practices they favor.\footnote{See Abbey Stemler, Platform Advocacy and the Threat to Deliberative Democracy, 78 Mo. L. Rev. (forthcoming 2018) (describing the lobbying and userbase mobilization techniques used by marketplaces to create favorable legal environments), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3022409 [https://perma.cc/BRT5-UK32].} Where a marketplace takes such far-reaching actions to make a market, it arguably should be considered responsible in “whole or in part” for the resulting behavior of its users.\footnote{See Olivier Sylvain, Intermediary Design Duties, 50 Conn. L. Rev. 203, 233 (2018).}
D. Limit § 230 Immunity to Defamation-Related Claims

Section 230 incorporates the words “publisher” and “speaker,” both terms of art in defamation claims. This language suggests that § 230 should be limited to defamation claims or at least to claims resulting from an intermediary’s publication. If a marketplace’s illegality comes from some action not fairly traced to conduct as publisher or speaker, § 230 is reasonably understood to offer no protection. Moreover, even considering that many marketplaces could be said to “publish” listings, they often do substantially more, as discussed in Section V.A.

Sometimes a provider’s actions are plainly removed from its actions as publisher or speaker of third-party content. The leading case establishing this principle is Barnes v. Yahoo!, Inc. In Barnes, a Yahoo! employee promised the plaintiff that Yahoo! would remove sexually explicit posts made by her ex-boyfriend. When Yahoo! eventually refused to take down the post, the plaintiff sued for a variety of claims including promissory estoppel. Since promissory estoppel claims seek to hold individuals liable under contract law, the court reasoned that the cause of action did “not seek to hold Yahoo![!] liable as a publisher or speaker of third-party content, but rather as the counterparty to a contract.” Generally, when an ICS’s own conduct is at issue and that conduct “does not turn on holding an Internet service liable for posting or failing to remove content by a third party,” § 230 will not provide protection.

But in other disputes, it can be awkward, if not maddening, to assess whether a law “inherently requires the court to treat” an ICS as “a publisher or speaker” because § 230 was originally designed to respond to defamation or defamation-related torts. In those situations, the illegality is clear—the underlying defamation—and it is apparent that the communication service’s role is primarily, if not solely, to redistribute and “publish” the content. In contrast, for claims such as negligence, support for terrorism, and dis-

---

218 See id. at 211.
219 570 F.3d 1096 (9th Cir. 2009).
220 Id. at 1109.
221 Id. at 1106.
222 Id. at 1107.
223 Airbnb, Inc. v. City and Cty. of San Francisco, 217 F. Supp. 3d 1066, 1074 (N.D. Cal. 2016); see also McDonald v. LG Elecs. USA, Inc., 219 F. Supp. 3d 533, 538 (D. Md. 2016) (declining to give Amazon § 230 immunity for “its own tortious conduct”); Lansing v. Sw. Airlines Co., 980 N.E.2d 630, 638 (Ill. App. 2012) (holding that “section 230 (c) ‘as a whole cannot be understood’ as granting blanket immunity to an ICS user or provider from any civil cause of action that involves content posted on or transmitted over the Internet by a third party”).
224 See Airbnb, Inc., 217 F. Supp. 3d at 1109–02; Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1163 (9th Cir. 2008).
225 Doe v. Internet Brands, Inc., 824 F.3d 846, 847 (9th Cir. 2016).
From the Digital to the Physical

criminal, the roles of the independent user and communication service do not neatly track what § 230 anticipates.

In two cases, courts found a distinction between imputing liability for third-party activity on ICSs and holding them directly liable for their own bad acts. The courts found that § 230 did not prevent San Francisco from imposing civil and criminal liability on Airbnb for its own actions in providing a booking service. In particular, the ordinance at issue regulated the plaintiffs’ “own conduct as Booking Service providers and cares not a whit about what is or is not featured on their websites.”

The Ninth Circuit also distinguished a website’s own acts from its users’ acts in Doe v. Internet Brands, Inc., finding that § 230 did not prevent claims that a website negligently failed to warn a plaintiff of a known scheme to lure women into situations where they could be assaulted. The failure, the court found, was a wrong independent of third-party user behavior and was unrelated to any obligation to moderate or remove dangerous postings.

Some might question whether CDA truly protects marketplaces whose actions include charging a fee. A marketplace’s efforts to seek and accept payment are separate from “treat[ing it] as the publisher or speaker.” When illegality targets payment, rather than publishing, the applicability of § 230 is less clear. Targeting payment has additional virtues: it tracks common law instincts about benefit as an indicator of liability; it exempts non-commercial and zero-fee services (most of Craigslist notably included); and it places liability on the large commercial marketplaces that seem particularly well-positioned to act.

---


228 The awkwardness is apparent in the controversial California Court of Appeal decision, Hassell v. Bird, 247 Cal. App. 4th 1336 (Ct. App. 2016). There, a business had received a negative review on Yelp, and it complained of defamation by the anonymous reviewer. Id. at 1342. After a default judgment, the court ordered Yelp to remove the disputed review. Id. at 1345. Yelp objected, citing § 230 for the proposition that it was not responsible for the underlying defamation and thus could not be required to remove the post. Id. at 1346–47. But the court found that Yelp was not treated as a publisher when a trial court ordered it to remove the post. Id. at 1358. In particular, the court found that since Yelp was not a party to the original case, the removal order did not impose liability on Yelp for its role as a publisher. Id. Yelp complained of possible contempt liability for violating the court’s order, but the court said Yelp offered no authority that § 230 guaranteed Yelp safety from contempt proceedings. Id. at 1365. The court said any contempt sanction would be for Yelp’s own actions, not for its role as publisher or distributor of third-party content. Id.


231 Doe v. Internet Brands, Inc., 824 F.3d 846, 849 (9th Cir. 2016).

232 Id. at 853 (finding that § 230 “does not provide a general immunity against all claims derived from third-party content”).

Proponents of broad § 230 protections question the wisdom of narrowing § 230 to defamation-related claims because they fear that any narrowing of § 230 will damage the vibrancy of the Internet. But where a marketplace’s efforts extend importantly beyond publishing, such as processing payments or otherwise intertwining themselves with transactions, the marketplace may face liability on theories predicated on its own acts, not on its role as publisher or speaker of the material that comes from others.

E. Use Canons of Construction to Support a Narrow Reading of CDA Immunities

The scope of § 230 preemption of state law is arguably ambiguous under the plain language of the statute. But the correct interpretation of ambiguous law is informed by canons of statutory interpretation, under which the courts interpret ambiguous federal statutes narrowly to minimally preempt state law and to reduce any encroachment on traditional areas of state regulation. These canons call for a narrow interpretation of § 230 because a broad interpretation would conflict with state law, at least in certain fields; the canons thereby suggest that § 230’s protection to marketplaces may be correspondingly narrower.

1. The Statutory Language

Section 230(e)(3) provides a preemption clause that declares, “[n]othing in this section shall be construed to prevent any State from enforcing any state law that is consistent with this section. No cause of action may be brought, and no liability may be imposed under any state or local law that is inconsistent with this section.”

This preemption clause at first glance may seem unnecessary. The Constitution’s Supremacy Clause instructs that federal law preempts inconsistent state law. Consequently, there is no need for a federal statute to reiterate the point. It is a “cardinal principle of statutory construction that we must give effect, if possible, to every clause and word of a statute.”

---


236 U.S. Const., art. VI, cl. 2.

ple disfavors an interpretation that renders § 230’s preemption clause meaningless.

We must search, then, for an interpretation that gives genuine meaning to the preemption clause. The most natural one is that Congress sought to emphasize that there are indeed consistent state laws which will and should remain in effect despite § 230. Such emphasis favors a narrower reading of § 230 instead of an expansive one.

2. Avoidance Canons

Rules of statutory interpretation provide an additional basis to question the scope of § 230’s immunity against state law. Absent a clear statement of legislative intent, stare decisis protects the values of federalism, which ensure that federal laws do not “intrude upon the usual balance of state and federal power.” Therefore, if there is doubt about the scope of preemption, principles of federalism would let state law stand. There is also a presumption that constructions of federal statutes that produce “federal encroachment upon a traditional state power” should be avoided. On subjects traditionally left to the states, these presumptions would be particularly deferential to state law.

Broad readings of § 230 run afoul of these values of federalism when applied to certain marketplaces. First, when there are ambiguities in a federal law, as there are with § 230, courts should construe the law so as to not limit states’ rights to impose liability on marketplaces for the offline harms they produce. Second, the presumption against encroachment lends particular force in the context of marketplaces that facilitate transactions that are traditionally regulated under state police power (whether wielded by state or local governments). Areas traditionally regulated by the states and cities in

238 John F. Manning, Federalism and the Generality Problem in Constitutional Interpretation, 122 Harv. L. Rev. 2003, 2026 n.93 (2009) (stating that there is a “presumption against preemption of state law by a federal statute”).


240 See supra Section V.A–C.
clude land use\textsuperscript{241} as well as car services and taxicabs.\textsuperscript{242} Federalism principles therefore suggest that § 230 should not block state and local regulation of the local activities of such marketplaces.

\textbf{F. Turning Away from the CDA Caselaw}

A final line of reasoning turns away from § 230 caselaw in light of broader principles informed by the statute’s plain language and common law traditions. Indeed, many scholars suggest that § 230 is misguided or overbroad.\textsuperscript{243} Some courts have been equally skeptical, favoring common law notions of liability and questioning why Congress (supposedly) instructed a different result online.\textsuperscript{244} To date, the Supreme Court has not taken an opportunity to interpret § 230. It is not implausible that an appellate court would chart a new course, particularly as online information systems come to intermediate more transactions,\textsuperscript{245} as the divergence between § 230 versus common law standards becomes increasingly stark,\textsuperscript{246} as public sentiment shifts

\textsuperscript{241} This can include zoning, real estate, and other aspects of short-term rentals. See SWANCC, 531 U.S. at 174 (recognizing states’ “traditional and primary power over land and water use”); FERC v. Mississippi, 456 U.S. 742, 767 n.30 (1982) (“[R]egulation of land use is perhaps the quintessential state activity.”).


\textsuperscript{244} See, e.g., Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 29 (1st Cir. 2016) (“If the evils that the appellants have identified are deemed to outweigh the First Amendment values that drive the CDA, the remedy is through legislation, not through litigation.”); Fair Hous. Council of San Fernando Valley v. Roommates.com, L.L.C., 521 F.3d 1157, 1189 n.15 (9th Cir. 2008) (“The Internet is no longer a fragile new means of communication that could easily be smothered in the cradle by overzealous enforcement of laws and regulations applicable to brick- and-mortar businesses. Rather, it has become a dominant—perhaps the preeminent—means through which commerce is conducted. And its vast reach into the lives of millions is exactly why we must be careful not to exceed the scope of the immunity provided by Congress and thus give online businesses an unfair advantage over their real-world counterparts, which must comply with laws of general applicability.”).


against tech platforms, and as scholars mount an increasingly vigorous assault against § 230 overbreadth.

A first line of attack narrows the § 230 safe harbor by taking seriously the words “good Samaritan” and “decency” (§ 230(c)’s section heading and the statute title, respectively). Citron and Wittes examine the origin and import of these words and their implications for interpreting the statute. A marketplace is a plausible “good Samaritan” advancing “decency” to the extent that it acts (or attempts to act) to block objectionable material. But those terms are correspondingly inapt where the marketplace seeks to retain, or indeed promote, misconduct. Where a marketplace rejects an easy solution that would increase compliance with applicable law, its status as “good Samaritan” seems particularly far-fetched.

A second line of attack would question the applicability of § 230 when an intermediary knows about a problem or set of problems. Courts have long found that § 230 protects intermediaries from liability for unlawful user content even if they knew of its unlawfulness. However, the stated purposes of § 230 nowhere suggest any intent to protect intermediaries who knowingly and willfully violate state or local law. One might protest that § 230 intentionally grants protection even when an intermediary knows about a problem, in order to encourage and support intermediaries. Yet some intermediaries are on specific notice of violations, such as a marketplace that knows that every transaction in a given jurisdiction is unlawful. It is particu-

---

ett-legal-immunity [https://perma.cc/C7WQ-BD2G].

248 See, e.g., Brief for Legal Momentum et al., supra note 243; Citron & Wittes, supra note 175; Doug Lichtman & Eric Posner, Holding Internet Service Providers Accountable, 14 SUP. CT. ECON. REV. 221 (2006).

249 Citron & Wittes, supra note 175, at 416–17.


251 The stated purposes of § 230(b) are:

(1) to promote the continued development of the Internet and other interactive computer services and other interactive media;

(2) to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation;

(3) to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services;

(4) to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children’s access to objectionable or inappropriate online material; and

(5) to ensure vigorous enforcement of Federal criminal laws to deter and punish trafficking in obscenity, stalking, and harassment by means of computer.

larly difficult to see a proper purpose resulting from allowing those intermediaries to continue with impunity.

The next sections broaden the preceding analyses by considering the fundamental principles that arguably should shape whether, when, and why marketplaces are or should be liable for the transactions they facilitate.

VI. A FRAMEWORK FOR ASSESSING MARKETPLACE LIABILITY

Two normative systems predominate § 230 literature and caselaw: efficiency and fault. Generally, discussions of efficiency ask how assignments of liability maximize goods and minimize harms; discussions of fault ask whether an intermediary is culpable.\(^{252}\) In other words, efficiency asks whether imposing liability is reasonable. Fault asks whether it is appropriate.

In many cases, the two normative systems are in accord. Often it would also be inefficient for the intermediary to prevent the problem, and the intermediary is also essentially blameless. In other cases, the intermediary is relatively well-positioned to make the problem stop and simultaneously bears a significant share of the blame.\(^{253}\)

The next Section, below, considers efficiency and fault rationales that inform when it is reasonable and appropriate for liability to attach to online marketplaces.\(^{254}\) Be mindful of the limitations of this effort: for one, assigning liability requires fact-finding beyond the scope of this Article. Furthermore, the listing of factors is not exhaustive and these factors are not always the most important. Nevertheless, this Section offers these factors to provide a useful and intuitive framework for assessing marketplace liability.


\(^{253}\) The two systems are most obviously aligned when the market itself is considered problematic and when the website on which transactions are taking place is designed to facilitate precisely those transactions. Websites specifically designed to cultivate criminal markets are the most obvious example. See, e.g., Nick Bilton, *Silicon Valley Murder Mystery: How Drugs and Paranoia Doomed Silk Road*, VANITY FAIR (May 2017), https://www.vanityfair.com/news/2017/04/silk-road-ross-ulbricht-drugs-murder [https://perma.cc/SUE7-9LEE]. The fear that the hosts of such markets would evade liability by citing § 230 is presumably why § 230(e)(1) carves out violations of federal criminal law from § 230’s protections. 47 U.S.C. § 230(e)(1).

\(^{254}\) This approach borrows heavily from the language of tort law. This is more than coincidental. Tort law addresses liability to non-contracting third parties, and disputes about the obligations of marketplaces often similarly arise from harms to users other than the marketplaces’ customers.
A. Efficiency

Typically, efficiency analyses ask how to maximize goods and minimize harms—a consequentialist focus grounded in outcomes.\textsuperscript{255} The details are often problematic. Looking at the same limited information, different people can reach divergent conclusions about the goods and harms in play, particularly when the goods or harms are debatable or difficult to measure.

This is also true in § 230 efficiency literature. Looking at § 230, some see the protector of a thriving Internet economy and ecosystem, bringing benefits to innovators and platforms in their vast operations.\textsuperscript{256} In contrast, those who are more concerned about harms often ask which party to an intermediated transaction is in the best position to avert those harms.\textsuperscript{257} Many in the latter group focus on particular harms that intermediaries could work harder to combat and attempt to design regimes that would help limit those harms without unduly burdening intermediaries.\textsuperscript{258}

In an efficiency analysis, the bottom-line question is: Does imposing liability produce a net good? To be tractable, this question must be broken down into smaller pieces. One might begin with importance: how much does it matter whether the objectionable transactions are controlled? Next, cost, effectiveness, and feasibility: what tools does the marketplace have to control these transactions? What do they cost? How well do they work? If the marketplace controls these transactions, what cost would that impose on unobjectionable transactions and on its own operations generally? Finally, consider secondary effects: what other options do participants in the problematic market have? When these transactions flow through the marketplace, does that make law enforcement’s job easier or harder?\textsuperscript{259} When one player in the market is held liable, will other participants in the same marketplace self...

\textsuperscript{255} See, e.g., JEREMY BENTHAM, A FRAGMENT ON GOVERNMENT, at i (1776) (“[I]t is the greatest happiness of the greatest number that is the measure of right and wrong . . . .”); Posner, supra note 252, at 33 (defining “efficient” as “cost-justified”).

\textsuperscript{256} See supra notes 121–23.

\textsuperscript{257} See generally, e.g., Mann & Belzley, The Promise of Internet Intermediary Liability, 47 WM. & MARY L. REV. 239 (2005); Citron & Wittes, supra note 175, at 404.


\textsuperscript{259} Similar questions were debated when intense public pressure led Craigslist to shut down its “erotic services” section, which facilitated prostitution. See, e.g., Claire Cain Miller, Craigslist Says It Has Shut Down Its Section For Sex Ads, N.Y. TIMES (Sept. 16, 2010), http://www.nytimes.com/2010/09/16/business/16craigslist.html [https://perma.cc/PKA7-R4DZ]; William Saletan, Pimp Mobile: Craigslist Shuts Its “Adult” Section, Where Will Sex Ads Go Now?, SLATE (Sept. 7, 2010), http://www.slate.com/articles/news_and_politics/frame_game/2010/09/pimp_mobile.html [https://perma.cc/J2FZ-GAMH]. A working paper found that Craigslist’s erotic services section increased the size of the prostitution market, but also made it safer. Scott Cunningham, Gregory DeAngelo & John Tripp, The Effect of Online Erotic Services Advertising on Prostitution Markets, Pricing, and Murder, (Ctr. for the Econ. Analysis of Risk, Working Paper Series 2018), http://cear.gsu.edu/files/gravity_forms/45-9a8e75f1713c...
regulate out of fear of liability, or will they take advantage of the vacuum left behind in the market?

These questions often pose serious difficulties. The importance, costs, and feasibility of having an intermediary regulate content can be hotly contested, and effects and responses of the market are challenging to predict. Additional challenges arise when harms are difficult to measure, as is often the case for defamation,\textsuperscript{260} psychological distress,\textsuperscript{261} and invasions of privacy.\textsuperscript{262} The harms resulting from traffic in unlawful weapons are, at least in part, more readily measured. In contrast, it is more difficult to quantify the subtler psychological harms of traversing bilious comments on YouTube.\textsuperscript{263}

In the context of marketplaces, at least one efficiency problem is generally simplified. Marketplaces tend to have superior capabilities to monitor and detect user misbehavior, as they have the best information about activities on their systems as well as information from users’ own submissions, customer feedback, and complaints. For example, California requires ride-hailing services, like Uber, to follow a “zero tolerance” policy for drunk driving.\textsuperscript{264} Upon receiving a drunk driving complaint, a service must suspend the driver until further investigation.\textsuperscript{265} Indeed, California imposes fines if a service does not do so, reflecting the judgment that ride-hailing services have the best information about this problem and the most effective tools to remedy it and thus, should be required to act.\textsuperscript{266}


\textsuperscript{262} See, e.g., Dorsey D. Ellis, Jr., \textit{Damages and the Privacy Tort: Sketching a “Legal Profile,”} 64 \textit{IOWA L. REV.} 1111, 1152–53 (1979) (noting the conceptual confusion introduced in privacy torts “result[ing] from the failure to articulate a functional means of evaluating the interests protected by privacy” and logically subsequent ambiguity with respect to appropriate damages).

\textsuperscript{263} Cf. Amelia Tait, \textit{Why Are Youtube Comments the Worst on the Internet?}, NEW STATESMAN (Oct. 26, 2016), https://www.newstatesman.com/science-tech/internet/2016/10/why-are-youtube-comments-worst-internet [https://perma.cc/E4XV-7Y9J] (“For years, YouTube has notoriously been the home of the worst comment section on the internet, and if you Google ‘Why are Youtube comments. . .’ the search engine will helpfully complete your sentence with the options ‘so bad’, ‘so racist’, and ‘so toxic’.”).


\textsuperscript{265} Id.; see also Florida’s zero tolerance policy. \textit{FLA. STAT.} § 627.748 (2017), https://www .flsenate.gov/Session/Bill/2017/221/BillText/er/PDF [https://perma.cc/2BH6-9EYV].

\textsuperscript{266} In April 2017, the CPUC recommended assessing Uber over $1.3 million for 150-plus violations of the zero-tolerance rules, Order Instituting Investigation And Order To Show Cause Why The Commission Should Not Impose Appropriate Fines And Sanctions On Raiser- Ca LLC, Cal. Pub. Utils Comm’n, Order No. I.17-04-009 (Apr. 11, 2017), http://docs.cpuc.ca
On the flip side, several scholars writing about intermediary liability online have argued that service providers will tend to overregulate activities on their platforms when they (a) are subject to liability and (b) have incentives that differ from their users'.

For example, if a blog-hosting platform were liable for libelous posts, it might block or remove all manner of entries—some that are truly libelous, but also many that are at most borderline—in an effort to reduce its liability, with little regard for interests of authors and readers. Based on this concern, those scholars question liability rules that would exacerbate the divergent incentives of providers and users. In this regard, online marketplaces present less of a problem because their incentives tend to align broadly with their sellers': both marketplaces and sellers want to increase the number of transactions.

Certain extreme enforcement efforts have clear efficiency consequences. One might imagine regulations that are so costly, or prevent such marginal harms, as to be unreasonable. It would be enormously costly for eBay to proactively prevent the sale of counterfeit goods, since products ordinarily do not come into eBay’s hands. Requiring eBay to take physical custody of all goods, and inspect them in detail, would surely create costs disproportionate to the problem. Yet that does not mean all anti-counterfeiting efforts are fatally flawed on efficiency grounds. Smaller interventions at eBay, perhaps focused on frequently counterfeited products or high-risk sellers, might pass efficiency scrutiny. Meanwhile, the architecture of Amazon Marketplace, with many goods sent from Amazon’s own warehouses, might make policing comparatively easy because many goods pass through Amazon’s warehouses. Thus, Amazon is in a position to efficiently inspect for counterfeits, recalled serial numbers, and the like, unlike eBay, which does not routinely receive physical products brokered by its marketplace.

---

267 See, e.g., Asaaf Hamdani, Who’s Liable for Cyberwrongs, 87 CORNELL L. REV. 901, 918 (2002) (fearing that ISPs will be overcautious because ISPs “do not capture the full value of the conduct they are entrusted with policing”); Neal Kumar Katyal, Criminal Law in Cyberspace, 149 U. PA. L. REV. 1003, 1007–08 (2001) (“Because an ISP derives little utility from providing access to a risky subscriber, a legal regime that places liability on an ISP for the acts of its subscribers will quickly lead the ISP to purge risky ones from its system.”); Neil Netanel, Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer File Sharing, 17 HARV. J.L. & TECH. 1, 13 n.30 (2003) (“ISPs and their subscribers have asymmetric incentives. ISPs do not fully share the benefits its subscribers derive from placing material, whether infringing or non-infringing, on the network. As a result, imposing liability on ISPs for subscribers’ infringing material induces ISPs to overdeter, purging any material that a copyright holder claims is infringing.”); Felix T. Wu, Collateral Censorship and the Limits of Intermediary Immunity, 87 NOTRE DAME L. REV. 293, 296–97 (2011).


269 eBay has, in fact, implemented some such measures. See Tiffany (NJ) Inc. v. eBay Inc., 600 F.3d 93, 98–99 (2d Cir. 2010).
B. Fault

Analyzing intermediary liability from the perspective of fault entails assessing blameworthiness—asking whether an intermediary deserves to have liability imposed on it, usually based on some disputed aspect of its design or conduct.\textsuperscript{270}

In the context of marketplaces, fault analysis typically looks for indicia of specific blameworthy conduct. One can ask: Did the marketplace do something to cultivate problematic activities?\textsuperscript{271} Did the marketplace fail to intervene where it could have?\textsuperscript{272} Did the marketplace intervene clumsily, making matters worse?\textsuperscript{273}

Here too, the questions often pose serious difficulties. For example, a question about who “did” a given deed is not straightforward when multiple users collaborated—such as when an intermediary provided a tool that one user employed to harm another. There are two main ways to view such a scenario: the intermediary is liable on a root cause theory or not liable on the theory that an independent user’s bad acts are intervening causes. Meanwhile, fault is similarly muddied when a given type of harm has long occurred, with or without intermediaries, or via prior intermediaries.

Nonetheless, fault analysis is usually more tractable than efficiency analysis because it typically requires understanding only one company’s behavior,\textsuperscript{274} whereas thorough efficiency analysis often requires assessing responses by buyers, sellers, competitors, and beyond. Consider a regulation targeting a general-purpose tool that could be used to facilitate discrimination. For example, Craigslist allows sellers to list almost anything, and a landlord could post a listing for tenants that expresses racial preferences. Because the listing tool is general in its purpose, allowing listing or sale of almost anything and with little or no tailoring to a particular use case, it appears unlikely that Craigslist “did” much to support its listings. Blame is arguably further reduced by the ineffectiveness of plausible interventions. Landlords could probably evade most keyword filters via synonyms, euphemisms, misspellings, or the like.\textsuperscript{275} And landlords could move to an alterna-

\textsuperscript{270} See, e.g., FTC v. Accusearch, Inc., 570 F.3d 1187, 1198 (10th Cir. 2009) ("That is, was it responsible for the development of the specific content that was the source of the alleged liability?").

\textsuperscript{271} See, e.g., id. at 1198–1200.

\textsuperscript{272} See, e.g., Zeran v. Am. Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997) ("[U]nder [Stratton Oaknorn]'s holding, computer service providers who regulated the dissemination of offensive material on their services risked subjecting themselves to liability, because such regulation cast the service provider in the role of the publisher").


\textsuperscript{274} See, e.g., Accusearch, 570 F.3d at 1198–1200.

tive platform if Craigslist proved too strict. On a blame theory, then, many analysts may be prepared to forgive Craigslist’s inclusion of some listings expressing a racial preference for tenants.

C. Factors for Assessing Marketplace Liability

Applying the principles of efficiency and fault, there are five factors that can guide assessments of marketplace liability. These are specificity versus generality, scale, sales-enhancing activities, unlawful design, and location.

1. Specificity Versus Generality

Some marketplaces facilitate transactions in exceptionally specific markets. For example, Uber’s local transportation business operates in certain places, in defined classes of vehicles, at specific prices, and on specific terms. In contrast, other marketplaces can be used for almost anything. Consider the exceptional range of both goods and services on Craigslist.276

Constraints along a given dimension are specificity, and their reverse, generality. A high degree of specificity tends to suggest that a given marketplace should be held liable under both efficiency and fault doctrines. Fundamentally, a marketplace’s generality on a given dimension suggests that the marketplace engages in less oversight of the activities of market participants along that dimension. In that case, it may be distinctively burdensome to impose new oversight obligations, raising efficiency concerns. Generality also tends to reduce apparent culpability. All else being equal, a marketplace that has not chosen to facilitate specific problematic activities bears less blame than one that has. On the other hand, if a marketplace chooses to target a specific category of goods or services, it can more reasonably anticipate and police abuses related to that particular market.277

A potential twist in this analysis arises from abuses associated with a specific marketplace that are unanticipated or otherwise not blameworthy.


277 A variation of this point is in the Localism portion of this Article. See infra Section VI.C.5; see also Accusearch, 570 F.3d at 1200 (10th Cir. 2009) (imposing liability in part because “[b]y paying its researchers to acquire telephone records, knowing that the confidentiality of the records was protected by law, [Accusearch] contributed mightily to the unlawful conduct of its researchers . . . . [T]he offensive postings were Accusearch’s raison d’etre and it affirmatively solicited them.”).
Consider people using Airbnb for prostitution.\textsuperscript{278} If short-term rentals are generally attractive to sex workers,\textsuperscript{279} Airbnb’s specificity as to the short-term rental market may mean the platform facilitates prostitution. But unless evidence is adduced that Airbnb attempts to foster this use of its service, or perhaps that Airbnb negligently or recklessly ignored some evidence of the problem, there would be little logic in imposing liability on Airbnb for any facilitation of prostitution that might take place.

In principle, a marketplace might respond by creating potential generality—a technical capability for the marketplace to serve broader markets, even as the marketplace turns a blind eye to how its service is systematically used. Potential generality is probably not enough. When a marketplace is general in theory but specific in practice, an informed analysis should look at the actual situation at hand and assign liability accordingly. If Craigslist was used overwhelmingly for sale of counterfeit goods, recalled goods, illegal services, or other misdeeds, Craigslist’s generality alone should not impede accountability.\textsuperscript{280}

Finally, the specificity-generality spectrum reveals limits to the meaning of a marketplace. Some services might be so specific that they are best understood not as marketplaces but rather as sellers of goods or services. In that case, secondary liability doctrines fall away, and primary liability doctrines take center stage.\textsuperscript{281} Consider first the ordinary case: if a single independent operator offered on Craigslist to drive passengers around town for a fee, we would not think that Craigslist was functionally equivalent to a car service. Rather, we would take that as an inevitable consequence of Craigslist’s generality.

Uber’s specificity, by contrast, gives rise to a different inference. Ordinarily, we would use a term like car service or taxi dispatch to describe an intermediary that dispatches vehicles to transport customers from point to point within a city, taking a portion of each driver’s earnings, providing incidental assistance such as customer service, billing, record-keeping, and perhaps insurance. We would not think that such a company was anything but a business selling transportation services directly to customers. Uber’s functional similarity to a longstanding business model suggests that it is, in fact,
a business selling services rather than a marketplace connecting drivers and riders.\textsuperscript{282}

2. Scale

When objections to a marketplace are best addressed through a solution with large up-front costs but low marginal costs, it may be efficient to impose liability only on especially large marketplaces.

Consider ContentID, the YouTube feature that scans video and audio content to check for reproduction of copyrighted material, automatically paying a rights-holder for use (or, if the rights-holder so instructs, removing an infringing video altogether).\textsuperscript{283} Building ContentID required not only the technical capability to match similar audio but also an inventory of copyrighted works and information about the corresponding rights-holders. Despite the apparent challenge in setting up such a system, once running, it enjoys economies of scale: running ContentID on one hundred thousand YouTube videos is not substantially cheaper than running it on one billion YouTube videos. In this way, YouTube benefits from its exceptional scale. A YouTube without ContentID might rightly be accused of cutting an important corner. Yet the same accusation would ring hollow if directed toward a tiny startup.

Similar concepts apply to other marketplaces. A short-term rental marketplace might consider developing and maintaining a directory of municipal laws that apply to short-term rentals. A large marketplace could spread these costs across its thousands of properties and millions of transactions. In contrast, a small marketplace would probably struggle to do the same work.

3. Unlawful Design

A platform may be designed in a way that simplifies, assists, or encourages unlawful behavior. Roommates.com, discussed earlier, is the prototypical example. The company structured its matching service to make it dramatically easier for users to unlawfully discriminate in their housing choices. In so doing, it enabled and arguably encouraged unlawful discrimination. Compare the facilitation of racial discrimination of Roommates.com with racial discrimination on Airbnb, which results from preferences expressed by guests and hosts and is unprompted by Airbnb.\textsuperscript{284} While Airbnb

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{283} How Content ID Works, YOUTUBE HELP, https://support.google.com/youtube/answer/2797370?hl=en [https://perma.cc/46XU-YZA8].
\item \textsuperscript{284} For evidence on racial discrimination among guests directed at hosts, see generally Edelman & Luca, supra note 86. For evidence on hosts’ discrimination, see generally Edelman, et al., supra note 71.
\end{itemize}
\end{footnotesize}
could make design choices that discourage or prevent discrimination, its design choices do not appear to facilitate or encourage unlawful discrimination in the manner the court critiqued in Roommates.com.

4. Sales-Enhancing Activities

The Section below considers four sub-factors for assessing how a service’s sales-enhancing activities might favor the imposition of liability: control, enhancement, representation, and market development. At the highest level of generality, the idea is that marketplaces may be liable for activities they undertake to raise the quality and volume of the goods and services consumers purchase through them.

i. Control of the transaction. Intermediaries often seek to standardize the experience of market participants. Take Uber, which gives drivers “tips for 5-star trips” including how to drive, how to communicate with passengers, and even how to dress. Even more importantly, Uber sets prices, much like a taxi company or regulator would, eliminating the possibility of one-on-one negotiation. Uber also requires drivers to carry insurance and sets rules relating to car types. Rather than merely facilitating a transaction between rider and driver, Uber organizes and structures the most important elements of the transaction. Other marketplaces similarly provide precise instructions to sellers to standardize buyers’ experience.

These mechanisms of standardization and control may suggest that the putative online marketplace is not an intermediary between sellers and buyers, but the true seller. For example, in setting prices and other key terms, Uber makes itself look more like a true service provider and less like

---

287 See supra notes 40–41. By contrast, Airbnb only provides pricing advice to its users “based on the listing’s ‘features, location, amenities, booking history, availability, and seasonal supply and demand.’” Johanna Interian, Note, Up in the Air: Harmonizing the Sharing Economy Through Airbnb Regulations, 39 B.C. Int’l & Comp. L. Rev. 129, 156 (2016). Since Airbnb’s suggestion is optional, one can argue that Airbnb is only helping the seller, who may not have a refined sense of the market, instead of controlling the transaction.
291 Similar arguments underlie suits arguing that Uber drivers are employees rather than independent contractors. See O’Connor v. Uber Tech., Inc., 82 F. Supp. 3d 1133, 1148–49 (N.D. Cal. 2015) (citation omitted) (“[T]he principal test of an employment relationship is whether the person to whom service is rendered has the right to control the manner and means of accomplishing the result desired.”).
an independent marketplace. As putative intermediaries exercise greater control over the transactions that they facilitate, they step toward being simple sellers and start being subject to primary, rather than secondary, liability. A seller is more blameworthy for problems that arise from its own service than an intermediary is blameworthy for transactions it merely facilitates.

ii. Enhancement. Online marketplaces often do more than facilitate transactions: they may undertake to enhance the offerings on their platform. For example, from the summer of 2010 through the summer of 2017, Airbnb offered hosts free professional photography services to better present their properties.\textsuperscript{292} Improved photographs helped properties attract more guests and command higher rates, to the mutual benefit of the hosts and Airbnb. Similarly, Uber and Lyft offer drivers a variety of tools to improve ride quality. Lyft, for example, has offered its drivers various hardware to help riders identify their drivers,\textsuperscript{293} making it more convenient to hail a ride. These measures encourage larger tips and future trips.

These efforts are sensible business decisions, but they also suggest that Airbnb, Uber, and Lyft are more than mere marketplaces. In particular, these efforts undermine any claim that the resulting service is solely the work of third-party sellers and the marketplace merely a neutral pass-through. Instead, in these cases, the marketplace is more properly understood as collaborating with sellers. In the language of § 230, the intermediary is involved, “in whole or in part, [in] the creation or development of [product] information,”\textsuperscript{294} and the listings are therefore not “provided by another information content provider” so as to immunize the marketplace.\textsuperscript{295} (This is exactly the logic of FTC v. Accusearch.\textsuperscript{296}) Nonetheless, such efforts to enhance consumer experience—voluntary on the marketplace’s part, and optional on the seller’s—are different from the standardization and control discussed earlier. If marketplaces are to be found liable on the basis of their efforts to enhance offerings on their services, the analysis would rest mainly on the logic of fault. Essentially, the marketplace’s involvement in the offering may


\textsuperscript{295} Id. § 230(c)(1) (emphasis added).

\textsuperscript{296} FTC v. Accusearch, Inc., 570 F.3d 1187, 1187 (10th Cir. 2009); see also supra Section V.C.
also create complicity and enable liability. However, efficiency ideas also support imposing liability in these circumstances. For one, marketplaces that become entangled with their offerings may have greater ability to monitor and control those offerings. For example, it is tenuous for Airbnb to argue that it has sufficient resources to identify and train photographers in each of dozens of cities but insufficient resources to learn the zoning and tax requirements in those same cities. The more the marketplace intertwines itself with individual offerings, the better positioned the marketplace is to monitor and oversee them.

iii. Representations. Online marketplaces sometimes make independent representations about the quality of a specific listing or about their general efforts across all listings. Such representations tend to induce interest from marginal customers, to the advantage of both sellers and the marketplace. When customers buy in reliance on such representations, imposing liability may be appropriate on both efficiency and fault rationales.

iv. Market development. Beyond merely presenting listings or making connections between buyers and sellers, marketplaces can develop the markets themselves. Consider the many steps Airbnb took to encourage short-term rentals in private residences. In addition to litigating and lobbying for the right to operate in certain markets, it led public relations campaigns to normalize and defend its business model and to make the short-term rental market more attractive for both hosts and guests. It offered signup bonuses to hosts and guests. How many transactions would have taken place in the short-term rentals Airbnb envisioned, had it not been for these efforts? Having all but created the market, the company’s efforts prompt a natural instinct that it should bear greater responsibility for resulting problems.

On this line of reasoning, a marketplace is more liable for a market that it created—a market that plausibly would not have existed but for its distinctive efforts—than for transactions that would surely have occurred with or without the support of a given marketplace operator. Compare Airbnb’s efforts to bring about short-term rentals in private residences with Craigslist’s provision of a new way for vacation listings to find customers. Certainly, vacation homeowners had found tenants before Craigslist and would do so

---


without Craigslist. Whatever goes wrong in that market, it is much less clear that Craigslist caused or materially increased the problem.

One might object that liability rooted in market development presents an undue impediment to first movers: If market developers face additional liability, then later market entrants could free-ride on leaders’ efforts. That said, theory and experience reveal the advantages of being a successful first-mover in businesses with network effects. There is little suggestion that such free-riding would discourage meritorious market development activities. Moreover, if a first entrant faced substantial liability for its misdeeds, on a market development theory, it is predictable that the applicable regulators, enforcement agencies, and other legal institutions would ultimately become better-positioned to pursue subsequent similar offenders. This would blunt any supposed advantage from having done less to develop the market.

5. **Localism**

Political processes and scholars alike often debate which level of government—federal, state, or local—is appropriate to manage a particular area of policy. These discussions often explore efficiency concerns. For example, consistency across jurisdictional lines can facilitate centralized compliance by companies. Meanwhile, websites by default enjoy international reach, so a local regulation could catch sites unaware and stymie online activity, good or bad. These principles support broad regulatory scope at the national level, but not at the local level. On this view, § 230 enacts a policy judgment that efficiency interests are poorly served by websites having to police where they are being accessed.

But many communities prefer to set local rules to meet idiosyncratic local preferences and circumstances. Similar concerns about local versus national regulations arise in discussions of the proper regulator for communications technologies including the Internet itself.

---


302 Compare, e.g., Am. Fin. Servs. Ass’n v. City of Oakland, 104 P.3d 813, 823 (Cal. 2005) (finding local regulation of financial services preempted in part because “[c]ommercial reality today would confound any effective regulation of mortgage lending based on potentially hundreds of competing and inconsistent measures at the local level”), with id.at 832 (George, C.J., dissenting) (discussing predatory lending as a “community development” issue). See generally John T. Scholz & Feng Heng Wei, Regulatory Enforcement in a Federalist System, 80 AM. POL. SCI. REV. 1249 (1986).
Moreover, many online marketplaces are not mere websites. Their services reach into diverse communities in tangible ways. eBay sellers ship goods; Airbnb hosts put people up in their residences; Craigslist users transact locally in myriad forms.\(^{303}\) This physicality gives online marketplaces a certain concreteness compared to websites, which only distribute information.

Therefore, it may be particularly reasonable that state or local regulation target marketplaces when they distinctively facilitate transactions that are traditional targets of state and local, rather than national, regulation. Transactions related to land use,\(^ {304}\) taxis,\(^ {305}\) guns,\(^ {306}\) alcohol,\(^ {307}\) and cannabis\(^ {308}\) are all regulated in dramatically different ways across jurisdictions.\(^ {309}\) Businesses that facilitate any of the above should not be surprised when they have to navigate more of a jurisdictional thicket than, say, telecommunications companies. There is no particular reason that an online business in these or similar markets should not have to engage with different rules in different jurisdictions, just as physical businesses in these markets have always had to do.

To some extent, this is a special case of the generality and specificity discussed in Section VI.C.1. It seems both intuitive and fair to require attention to state and local rules when an online marketplace targets specific markets that have long been managed at the state or local level.

Notably, online marketplaces may end up facing somewhat different obligations than brick-and-mortar stores. For one thing, the marketplaces facilitate transactions, while the stores are sellers in their own right. Furthermore, brick-and-mortar establishments choose where to operate, getting advance opportunity to review local laws. In contrast, online marketplaces

\(^{301}\) For a more expansive discussion of this point, and of regulation of the sharing economy as a local problem, see generally Davidson & Infranca, supra note 242.

\(^{304}\) See, e.g., Rapanos v. United States, 547 U.S. 715, 738 (2006) (“Regulation of land use . . . is a quintessential state and local power.”); Ostrow, supra note 301, at 1404 (“[T]he dominant descriptive and normative account of land-use law is premised upon local control.”).

\(^{305}\) See Davidson & Infranca, supra note 242, at 217 (including “taxi medallion requirements” in a list of “distinctly local legal issues”).


\(^{308}\) See, e.g., Bob Salsberg, 100 Massachusetts Towns Have Voted for Weed Bans, Moratoriums, or Zoning Restrictions, ASSOCIATED PRESS (Sept. 19, 2017).

\(^{309}\) Note that all of the above may be regulated at either the state level, the local level, or both. Cf. Town of Telluride v. Lot 34 Venture Co., 3 P.3d 30, 32 (Colo. 2000) (en banc) (noting, in case discussing whether state housing law preempted local rent control law, that “[t]he issue of rent control implicates both state and local interests”). This point is not specific to regulations enacted by state or local government, and the authors express no preference between the two.
often by default operate everywhere, and it is more plausible that an online marketplace could truly be unfamiliar with an idiosyncratic local law. When a jurisdiction has unusual laws or is geographically distant from the marketplace’s core operations, online marketplaces should be granted more leeway in compliance, through generous opportunities to cure violations as well as light sanctions.

VII. A Way Forward

This final Section turns to recommendations for courts, Congress, states, and localities on their approach to liability of online marketplaces.

A. What Courts Should Do

As a starting point, courts should be skeptical that § 230 provides complete immunity. They should take seriously every word of § 230, including every requirement, restriction, and contingency. In particular, they should admit that prior courts have sometimes overlooked such gaps, and they should be prepared to impose the requirements fairly written in the statute. Notwithstanding stare decisis, courts should decline to follow cases that were manifestly ungrounded in the statute.\(^{310}\) The analysis in Section V offers a roadmap of how courts can find their way to answers closer to both the text of the statute and sound policy.

B. What Congress Should Do

Congress should begin by recognizing that § 230 has been misinterpreted—far beyond Congress’s actual intent as of 1996, beyond the plain language of the statute, and most of all beyond wise public policy. With this recognition, the natural response is to narrow § 230 through appropriate statutory revisions.

First, Congress could deny § 230 protections when an intermediary is on actual notice of a specific problem or pattern of problems, particularly when such notice comes from a qualified public official (such as an appropriate regulator) or from litigation. Second, Congress could withhold § 230 protections when an intermediary directly profits from an offending listing, for example by charging a transaction fee for such a listing. Third, Congress could require intermediaries to provide some level of diligence in screening material. A natural objection is that the proper level of diligence varies. Yet

\(^{310}\) See Zango, Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169 (9th Cir. 2009) (finding that a distributor of Internet security software qualified as a provider of an ICS because the distributor facilitated access to a computer server); Doe v. Friendfinder Network, Inc., 540 F. Supp. 2d 286, 292 (D.N.H. 2008) (providing a website with § 230 immunity for using a plaintiff’s image and information on an ad teaser despite the website’s assurances to the plaintiff that it would be removed).
a flexible standard could nonetheless prove useful. For example, a “reasonable care” requirement would ask intermediaries to calibrate their efforts to available resources, the nature of the material, and the popularity of a given submission. Fourth, Congress could limit § 230 protections to intermediaries that are truly and substantially facilitating free expression. Where an intermediary is solely facilitating commerce, such as proposing commercial transactions between buyers and sellers, the broad § 230 immunities could be replaced with additional obligations of the sort described above. Similarly, Congress should clarify that § 230 does not reach transactions of the sort that have traditionally been regulated locally. These limitations, individually or in some combination, would rein in judicial broadening contrary to congressional intent.

Any efforts to narrow § 230 will prompt opposition from the firms that benefit from the current broad immunity. In this respect, much can be learned from Congress’s 2017 efforts to prevent online sex trafficking. In response to § 230 defenses to listings of prostitutes who are victims of sex trafficking, Congress in August 2017 introduced legislation that would allow claims under state and federal criminal and civil laws for certain forms of sex trafficking. As notably, the legislation would remove § 230 protections from certain intermediaries that knowingly facilitate such trafficking. As would be expected given § 230’s popularity among intermediaries and their defenders, even this small narrowing of § 230 immunity attracted widespread criticism. CDA proponent Eric Goldman said this legislation would “ruin” § 230. Tech giants Google and Facebook lobbed against it, as did the Internet Association representing them, along with Amazon, Microsoft, Twitter, and others. This legislation would alter only the far edges of § 230, a space in which the tech giants do not participate. Any effort closer to home—altering regulation of mainstream commercial marketplaces—would surely prompt even stronger objections from the companies directly affected.

C. What States and Localities Should Do

State and local regulators cannot directly change § 230 (though state attorneys general have sought to reduce § 230’s immunity), yet they can nonetheless adjust their approach to policy goals in light of § 230.

---


312 See supra note 311.

313 Goldman, supra note 121.


315 Elizabeth Heichler, U.S. states’ attorneys general to take aim at Internet ‘safe harbor’ law, IDG NEWS SERV. (June 18, 2013), https://www.pcworld.com/article/2042351/us-states-
A sensible first step is to be mindful of § 230 limitations from the outset of regulatory efforts. When seeking to regulate an online market facilitated by an online marketplace, a prudent regulatory scheme should anticipate § 230 defenses and try to proceed accordingly. In this regard, the 2016 San Francisco short-term rental ordinance broke new ground, carefully distinguishing between obligations of a marketplace versus those of its users (hosts).

A nuanced regulatory scheme imposes distinct duties on each party, carefully calibrated to avoid violating § 230’s prescriptions. Regulatory schemes that bear in mind § 230 are also more likely to be appropriately calibrated to the actual capabilities and needs of marketplaces and users. Yet even there, San Francisco fell short in some respects. Faced with litigation, San Francisco reworked the challenged ordinance to eliminate requirements or restrictions on the publication of a rental listing, instead only imposing restrictions on collecting fees for providing booking services without registration. Moreover, in litigation, San Francisco abandoned the ordinance’s vision that rental marketplaces provide information without a subpoena.

With the benefit of San Francisco’s experience, Santa Monica followed suit and enacted similar ordinances. These ordinances were once again upheld by the district court, but this time, Airbnb did not compromise with the City, and the case is now pending in the Ninth Circuit. Seeing this dispute, other cities and towns should design their requirements with an eye to § 230 challenges, thereby narrowing the grounds for opposition.

A second approach attempts to overcome structural weaknesses of state and local regulators in disputes with large marketplaces. Marketplaces tend to be both well-funded and profitable, yielding ample resources that often exceed those available to a state or local government. Further, online marketplaces are well-versed on questions of intermediary liability and regulatory scope because they are fundamental to their operations, whereas generalist state and local government attorneys have little reason to be § 230 experts. Finally, marketplaces tend to confront the same issues in myriad jurisdictions, giving their attorneys the advantage of experience that is correspondingly lacking for state and local government attorneys. In response, state and local governments could wisely collaborate—as district attorneys in Los Angeles and San Francisco did in 2014 litigation against Uber. Yet there is little sign of similar joint litigation, or even common drafting or information-sharing, in other state and local proceedings against Uber.

attorneys-general-to-take-aim-at-internet-safe-harbor-law.html [https://perma.cc/3QG8-232X].

316 See S.F. Ordinance No. 104–16, supra note 153.
318 Settlement Agreement, supra note 148, at 1.
many state and local governments concerned about Airbnb would similarly do well to collaborate.

Third, state and local governments may need to be realistic about institutional capabilities. Here again, San Francisco’s experience overseeing short-term rentals is instructive. San Francisco’s ordinance called for a verification system to validate a host’s authorization to provide short-term rentals. Airbnb complained that the system was not functional, yet it could face criminal penalties for failing to use it. Finding these concerns to be legitimate, the court enjoined enforcement until the system was operational.321 It is important to be mindful of the difficulties state and local governments face in designing software, all the more so with the uncertainty of litigation that might change system requirements.

Nonetheless, San Francisco’s approach was imperiled by the unavailability of the required software, and the city eventually had to collaborate with Airbnb to create a registration system.322 If a regulation requires a government to receive or process information, the government must either be able to do so on the timeline contemplated by the regulation or be aware of its limitations.323

Finally, state and local governments usually need to be realistic about popular support for online marketplaces, and popular opposition to restrictions on marketplaces. For example, when Cambridge, Massachusetts proposed to ban Uber in 2014, many residents spoke in opposition.324 Airbnb mobilizes support for “citizen petitions” in response to unfavorable regulation, and Uber uses its own app to alert customers to regulatory threats.325 A skeptic might reject users’ responses as the fruits of corporate astroturf, not users’ independent evaluation.326 Yet the popularity of Airbnb, Uber, and other online marketplaces is undeniable. A government seeking to regulate such marketplaces must convince the public that such regulation is prudent.

325 See Stemler, supra note 215.
In this regard, Austin’s experience regulating TNCs is instructive: in a referendum, citizens supported regulation, by all indications convinced that the proposed requirements (fingerprinting drivers, among other things) were appropriate and that protests by Uber and Lyft were overblown.327 It appears to be crucial that Austin regulators did not overplay their hand. Had the proposed regulations contemplated a complete ban on TNCs, the referendum would have struggled to achieve a majority.

D. Policy in an Era of Large and Growing Marketplaces

The marketplaces bring both modern ecommerce and national vendors to commercial realms that had previously been solely local and offline. The marketplace operators are large and powerful, often favoring operations without substantial regulation. While there may be longstanding laws and regulations on the books, marketplace operators have found comfort in § 230 protection from those obligations. This is likely a mistake. Marketplaces may develop a culture of lawbreaking if they grow accustomed to exemption from regulation.328 Indeed, untouchable intermediaries not only facilitate bad behavior but also are likely to disproportionately hurt those most vulnerable.329 Moreover, these questions are too important to be left to tech elites unchecked by public policy or rule of law. Fortunately, § 230 does not compel that result and, indeed, is better understood as envisioning precisely the opposite.


ARTICLE

DELAYED JUDICIAL REVIEW
OF AGENCY ACTION

LUIS INARAJA VERA*

ABSTRACT

Should Congress be able to completely exempt the orders issued by a government agency from review by the courts? Some scholars argue that limiting courts’ jurisdiction in this fashion would be unconstitutional. What if, instead, Congress merely delays the point in time when a plaintiff may challenge an administrative order?

In recent years, courts have interpreted many statutes to only allow delayed judicial review of orders. This can cause judicial review to be deferred for several months or even years and, as a result, this mechanism can discourage meritorious challenges of orders. Despite the unfairness of delayed judicial review for potential plaintiffs, it is a widely accepted proposition that this form of review is nonetheless necessary in many contexts. This view rests on the premise that the alternative approach, immediate judicial review, would inevitably lead to two types of problems. First, it would cause a dramatic surge in frivolous lawsuits, which, in turn, would flood courts with challenges. Second, these lawsuits would significantly delay the enforcement of many statutes.

This Article challenges this conventional wisdom. The arguments in support of delayed judicial review, while intuitively appealing, are not as persuasive as they seem. First, allowing immediate review would not lead to a significant increase in frivolous suits. Decisions on whether to challenge an order in court are mostly economically driven. Therefore, with limited exceptions, potential plaintiffs will tend to only bring claims that they believe they can win in court. The empirical evidence on statutes that have transitioned from a delayed to an immediate review approach shows that the increase in the number of challenges resulting from that change is minimal. Second, it is unclear that immediate judicial review would substantially delay enforcement. Orders retain their legal effect after being challenged. In fact, while the court is examining the merits of the case, copious daily penalties for violating the order keep accumulating. Therefore, plaintiffs still have a powerful incentive to comply with the order swiftly. In addition to addressing these misconceptions, this Article also fills a gap in the literature by providing an analytical framework to evaluate the constitutionality of delayed judicial review provisions under the doctrine of constitutionally intolerable choices.

* Conservation Law Fellow, Environmental Resilience Institute at Indiana University; Visiting Scholar, Indiana University Maurer School of Law. For helpful conversations and comments, I would like to thank Vicki Been, Jamison Colburn, Joshua Ebersole, Sophie Heaton, David Kamin, David Reiss, Richard Revesz, Alexander Tsesis, Katrina Wyman, and the participants in the New York University Scholarship Clinic, the New York University Lawyering Scholarship Colloquium, the Furman Center Fellows Presentation Series, Vermont Law School’s Colloquium on Environmental Scholarship, Pace University School of Law’s Future Environmental Law Professors Workshop, and Loyola University Chicago School of Law’s Faculty Workshop. I am very grateful to the editors of the Harvard Journal on Legislation for their editorial suggestions.
200

Harvard Journal on Legislation [Vol. 56

INTRODUCTION .................................................. 201

I. ADMINISTRATIVE ORDERS: THE TENSIONS BETWEEN

   ENFORCEMENT AND JUDICIAL REVIEW ....................... 209
      A. Administrative Orders As Enforcement Mechanisms .... 209
      B. The Importance Of Orders' Timing Of Review .......... 211
         2. Futile Challenge: Delayed Review of Cleanup Orders ........................................ 214

II. THE LEGAL DEBATE OVER DELAYED REVIEW OF

   ADMINISTRATIVE ORDERS .................................... 216
      A. Legal Doctrines In Play .................................... 216
         1. Statutory Preclusion ........................................ 216
         2. The Doctrine of Finality ................................... 217
         3. Procedural Due Process .................................... 218
      B. The State Of The Issue Under The Different Statutes ........................................ 220
         1. Clean Water Act ............................................ 220
         2. Clean Air Act ............................................... 222
         3. RCRA ......................................................... 225
         4. CERCLA ...................................................... 226
         5. The OSH Act and Mine Safety Act ....................... 227

III. THE QUESTIONABLE PUBLIC POLICY FOUNDATION OF

   DELAYED JUDICIAL REVIEW ..................................... 229
      A. Arguments in Support of Delayed Judicial Review ...... 229
         1. The Argument That Allowing Immediate Review Would Cause Important Delays in Enforcement .... 229
         2. Allowing Immediate Judicial Review and the Alleged Flood of Challenges ............................. 232
      B. Evaluating The Convenience Of Delayed Review In Specific Cases ........................................ 234
         1. Appropriate Order/Plaintiff Would Challenge ...... 235
         2. Appropriate Order/Plaintiff Would Not Challenge . 237
         3. Deficient Order/Plaintiff Would Not Challenge .... 237
         4. Deficient Order/Plaintiff Would Challenge .......... 238
         5. Summing Up: Overall Appropriateness of Immediate Review of Environmental Compliance Orders ........................................ 238

IV. CONSTITUTIONAL CONCERNS WITH DELAYED JUDICIAL

   REVIEW .................................................................. 239
      A. Is There A Constitutional Right To Judicial Review? 240
      B. The Doctrine Of Constitutionally Intolerable Choices ..................................... 242
         1. Severity of the Punishment .................................. 244
         2. Is There a Reasonable and Predictable Possibility of Not Being Punished if a Challenge is Unsuccessful? ........................................ 245
INTRODUCTION

Judicial review is one of the most important features of the American administrative state. And it is so for good reason. Courts’ ability to control the activity of government agencies is a necessary corollary to the principle of separation of powers. Without it, citizens would not have a meaningful avenue to ensure that the executive branch complies with the law. Not surprisingly, scholars have debated at length the extent to which the Constitution constrains Congress’s ability to limit the jurisdiction of federal courts, especially in cases involving a challenge of some form of agency action. The answer to this question, however, is still uncertain. The Supreme Court has been vague and elusive in its decisions on this matter, and Congress has responded by being prudent when limiting judicial review of agency action.

1 See, e.g., John J. Coughlin, The History of the Judicial Review of Administrative Power and the Future of Regulatory Governance, 38 IDAHO L. REV. 89, 90 (2001) (noting “the broader significance of the judicial review of administrative action in protecting fundamental constitutional freedoms”); Jonathyn T. Molot, The Judicial Perspective in the Administrative State: Reconciling Modern Doctrines of Deference with the Judiciary’s Structural Role, 53 STAN. L. REV. 1, 14 (2000) (“[T]he judiciary in our modern administrative state is positioned to perform an influential function analogous to the function it was positioned to perform under the Founders’ original plan.”).


3 See Henry M. Hart, Jr., The Power of Congress to Limit the Jurisdiction of Federal Courts: An Exercise in Dialectic, 66 HARV. L. REV. 1362, 1371–74 (1953); see also Raoul Berger, Administrative Arbitrariness and Judicial Review, 65 COLUM. L. REV. 55, 57–58 (1965) (noting that the right to right judicial review is mandated by the Constitution); Louis L. Jaffe, The Right to Judicial Review I, 71 HARV. L. REV. 401, 420 (1958) (stating that “in our system of remedies” an individual has the right to secure judicial review).

4 See Harold H. Bruff, Availability of Judicial Review, in A GUIDE TO JUDICIAL AND POLITICAL REVIEW OF FEDERAL AGENCIES 1, 17 (Michael E. Hertz et al. eds., 2015) (noting that “[t]he Court was avoiding [the] lurking constitutional issue”); see also Nicholas Bagley, The Puzzling Presumption of Reviewability, 127 HARV. L. REV. 1285, 1309–10 (2014) (stating that it is “maddeningly hard to say” how much the Constitution limits Congress’s authority to constrain the jurisdiction of federal courts).

5 See Note, Congressional Preclusion of Judicial Review of Federal Benefit Disbursement: Reasserting Separation of Powers, 97 HARV. L. REV. 778, 791 (1984) (pointing out that courts have been dodging the constitutional question); see also id. (indicating that “in these
This uncertainty has also provided a fertile ground for the proliferation of delayed review provisions, that is, statutory mechanisms that do not completely eliminate the availability of judicial review but instead defer the point in time at which plaintiffs may seek it.

This Article adopts a skeptical position with respect to delayed judicial review. As the following examples illustrate, delaying access to courts can often deter potential plaintiffs from seeking judicial review of agency action. In other words, delayed review may result in a recipient of an administrative order obtaining no review at all, which is troubling in a legal system that embraces the principle of separation of powers. Some scholars have claimed that delayed review is necessary in certain areas of the law—especially environmental law—for public policy reasons. These arguments are not only questionable in many instances but also frequently overlook the important constraints that the Constitution places on delayed review provisions.

The principle that delayed review can easily lead to no review at all is best illustrated through examples. Consider one of the most important environmental and public health challenges of the twenty-first century: controlling the risk associated with exposure to hazardous substances. The Environmental Protection Agency (“EPA”) has estimated that there are more than 530,000 contaminated sites in the United States, covering approximately twenty-three million acres. One of the main sources of this type of cases, [courts' apprehensions have been evident”). For an account of the vagueness of some judicial review provisions, see infra Part II.B.  

Immediate review of regulations raises legal and public policy questions that do not perfectly overlap with those relevant to administrative orders. See Daniel F. McNeil, Pre-Enforcement Review of Administrative Agency Action: Developments in the Ripeness Doctrine, 53 NOTRE DAME L. REV. 346, 348–49 (1977) (explaining the special importance, specifically with regulations, of determining if this form of agency action is fit for review, as required by the doctrine of ripeness).

See Steven G. Calabresi, An Agenda for Constitutional Reform, in CONSTITUTIONAL STUDIES, CONSTITUTIONAL TRAUMAS, CONSTITUTIONAL TRAGEDIES 22, 22 (William N. Eskridge, Jr. & Sanford Levinson eds., 1998) (explaining how separation of powers is one of the pillars of American law and one of the most commonly exported features of this legal system); see also Peter L. Strauss, Formal and Functional Approaches to Separation of Powers Questions A Foolish Inconsistency, 72 CORNELL L. REV. 488, 492 (1987) (pointing out that the Constitution requires the existence of Congress, the President, and the Supreme Court as separate entities that undertake the legislative, executive, and judicial functions, respectively).

See infra Part III.A.  


contamination is the improper handling of hazardous materials in the context of manufacturing activities. 11

Imagine a scenario in which the EPA issues two administrative orders to a paper company, which is located on a highly contaminated property in Scranton, Pennsylvania. The first order requires the paper company to change the way it is storing certain chemical substances. Its lawyers, however, believe that their client is not legally required to make these modifications and, as a result, want to challenge this first order in court as soon as possible. The second order mandates the paper company to clean up certain areas of its property. While the paper company does not object to it, there is an environmental group interested in challenging this second order, claiming that the cleanup method approved by the EPA is ineffective and potentially dangerous. These two plaintiffs would like to have immediate access to court to challenge these orders. The EPA would probably prefer that judicial review be delayed. The relevant question, however, is the following: what does delayed judicial review actually entail in each of these cases?

With the first order, under a delayed review approach, the paper company would only be able to access courts in the context of an enforcement action brought by the EPA. 12 Stated differently, for the paper company to be able to argue before a court that the order is improper, it would have to (i) not comply with it and then (ii) wait for the EPA to bring an enforcement action. If and when this happens, the paper company would be able to question the validity of the order in the same court proceeding that the agency initiated. 13 However, the issue is not just one of timing. If these two conditions are not met, the paper company will never have access to a court. Of course, the fulfillment of the second condition is outside the paper company’s control. But, even meeting the first condition is already very problematic because noncompliance with an administrative order issued by the EPA can lead to very severe punishment. Environmental statutes generally authorize the imposition of penalties of tens of thousands of dollars for every

---

13 See 42 U.S.C. § 9613(h)(2) (2012); see also Garry A. Gabison, The Problems with the Private Enforcement of CERCLA: An Empirical Analysis, 7 GEO. WASH. J. ENERGY & ENVTL. L. 189, 191 (2016) (explaining that “[i]f the EPA elects to go forth with the suit, the private individual cannot influence the process until the lawsuit begins by joining the suit as a plaintiff.”).
day of violation of the order.14 This threat operates as a powerful incentive to comply with the order.15 After the recipient of the order complies, the first condition for judicial review is no longer met, and as a result, a challenge in court ceases to be possible.

Because this first order involves a corporation, it is likely that this type of unfairness would not lead to outrage in many fora.16 However, it is important to avoid the simplistic illusion that this problem can be reduced to a choice between the economic gains of corporations and the protection of the environment or human health—a dichotomy that lends itself to resolution based on ideological or moral grounds. To offer a more balanced view of this problem, the next example explores a very similar timing-of-review issue from a different perspective. And it does so by focusing on the timing of review of the second order introduced above, i.e., the order issued to the paper company, which the environmental group wishes to challenge. Under an immediate review approach, the environmental group would be able to challenge the order immediately after the EPA issues it. In the more common delayed review scenario, however, the environmental group would only be able to challenge the order after the cleanup of the property is finalized.17 In this second example, delayed review can easily dissuade the plaintiff from challenging the order because judicial review at such a late stage becomes futile. Environmental groups often become interested in challenging cleanups when there are delays in the remediation process or when they are being conducted in a way that could lead to harm to the environment or human health.18 Thus, a

14 See, e.g., 33 U.S.C. § 1319(d) (2012) (Clean Water Act); id. § 6928(c) (Resource Conservation and Recovery Act (RCRA)); 42 U.S.C. § 7413(b) (2012) (Clean Air Act); id. §§ 9606(b), 9609(c) (CERCLA).

15 One author has claimed that this can lead to orders being used as “an instrument of intimidation.” Andrew I. Davis, Judicial Review of Environmental Compliance Orders, 24 ENVT. L. 189, 190 (1994). Others have highlighted, in the context of CERCLA, the effectiveness of this approach in securing early—pre-litigation—compliance and incentivizing settlements. Healy, supra note 12, at 333 (recognizing the potential unfairness of delayed review under section 113(h) of CERCLA).

16 See, e.g., Quiggle, supra note 12, at 328–29 (defending delayed review as necessary to deal with the actions of “large corporate polluters” or “giants like [General Electric]”).

17 This results from the timing of review provision in CERCLA. See 42 U.S.C. § 9613(h)(4) (2012) (foreclosing citizen suits of cleanup actions that are “to be undertaken at the site”). Courts have explained that this language should be interpreted to require cleanups to be finalized before a challenge may be brought. See, e.g., Clinton Cty. Comm’rs v. EPA, 116 F.3d 1018, 1023 (3d Cir. 1997) (“[A] citizens’ suit challenging a ‘removal’ action may not be brought even after completion of that removal action . . . .”); Schalk v. Reilly, 900 F.2d 1091, 1093 (7th Cir. 1990) (interpreting the language in § 9613(h)(4) to bar citizen suits to cleanups that are not completed); Alabama v. EPA, 871 F.2d 1548, 1557 (11th Cir. 1989) (noting that actions may be brought “only after a remedial action is actually completed”).

18 See Margot J. Pollans, A “Blunt Withdrawal”? Bars on Citizen Suits for Toxic Site Cleanup, 37 HARR. ENVT. L. REV. 441, 483 (2013) (explaining that citizen suits are often driven by the delays in the cleanup process); see also Healy, supra note 12, at 301 (noting the two types of scenarios in which citizen suits are brought to protect human health or the environment).
Delayed Judicial Review

delayed challenge is generally useless because the court would be reviewing past delay or the risk of occurrence of a harm that has already materialized.

Given the complications that delayed review creates, it is not surprising that both recipients of orders and environmental plaintiffs alike have urged courts to adopt an interpretation of many statutes that would allow immediate challenges. As one might expect, agencies such as the EPA and the Occupational Safety and Health Administration (“OSHA”) have generally advocated for a delayed review approach. The outcome, however, has been puzzling.

Courts, focusing their reasoning mostly on the administrative doctrines of finality and implied preclusion, have reached inconsistent conclusions on this subject. In the environmental context, the Supreme Court held in Sackett v. EPA that orders issued under the Clean Water Act are immediately reviewable. A very similar debate over so-called jurisdictional determinations issued under the Clean Water Act was resolved in a similar fashion by the Supreme Court in its 2016 decision in U.S. Army Corps of Engineers v. Hawkes. The status of this issue under the Resource Conservation and Recovery Act (“RCRA”) is significantly different. The Supreme Court has not addressed this question, but lower courts have deemed orders under RCRA to be subject to delayed review only, even though the language of the provisions in RCRA and the Clean Water Act regarding these types of orders is very similar. In the area of mine safety, on the other hand, the Supreme Court has examined this issue and disallowed immediate review of orders under the Mine Safety Act.

In addition to the inconsistencies within and across different areas of the law, the Supreme Court’s position on the delayed-versus-immediate-review debate remains uncertain for two reasons. First, it is unclear whether the Court’s view in Sackett that orders issued under the Clean Water Act are immediately reviewable applies to other environmental statutes. In an earlier case, Alaska Department of Environmental Conservation v. EPA, the

19 See infra Part II.B (referencing the cases in which plaintiffs sought immediate review of orders under the Clean Water Act, the Clean Air Act, RCRA, CERCLA, Occupational Safety and Health Act (OSH Act), and the Mine Safety Act).
20 See infra Part II.B (in these same cases, the agency opposed immediate review).
21 See infra Part II.B.
22 566 U.S. 120, 127–29 (2012) (finding that the order was final and its review not impliedly precluded by the statute).
23 136 S. Ct. 1807, 1816 (2016) (finding that, as with the order in Sackett, jurisdictional determinations were also final agency actions whose reviewability by courts was not precluded by the Clean Water Act).
24 See infra Part II.B.5; see also Thunder Basin Coal Co. v. Reich, 510 U.S. 200, 218 (1994).
Supreme Court treated an order issued under the Clean Air Act to be final and therefore immediately reviewable. However, there is good reason to believe that the Court’s statement in that regard was merely dicta given that the EPA had conceded this end, albeit only for the purposes of that particular case. As for CERCLA, the statutory provision explicitly barring immediate judicial review—which is not present in the Clean Water Act—could justify an outcome different from that reached in *Sackett*. Second, the extent to which delayed review frameworks are constrained by due process or the principle of separation of powers is also uncertain. The Supreme Court has not addressed this question under any of the federal environmental statutes. In the mine safety context, although the Court deemed that it was constitutionally proper to interpret the Mine Safety Act to bar immediate review, the Court’s ruling in that particular case was extremely narrow.

One of the underlying concerns with delayed judicial review is that, by using these limitations on the timing of review of agency action, the legislative branch may be insulating the executive from effective judicial oversight. At a time when the role of courts as a check on the Executive Branch is being amply debated, especially in areas such as immigration and environmental law, it is critical to address the undertheorized aspects of delayed judicial review. Up to this point, most of the literature has sided with the executive on this issue, arguing that delayed review is necessary for public policy reasons. As this Article explains at length, these arguments are misguided. The urgency of examining delayed judicial review in more depth is also exacerbated by the need to reconcile the inconsistent court decisions on the availability of immediate review and resolve the uncertainty with respect to the constitutionality of delayed judicial review provisions.

To address these problems, this Article makes two main contributions to the existing literature. First, it challenges the prevailing view that, from a

27 Id. at 482.
28 See Brief for Respondents at 16, Alaska Dep’t of Envtl. Conservation v. EPA, 540 U.S. 461 (No. 02-658) (accepting that the order was final after having disputed it at the court of appeals level); see also Transcript of Oral Argument at 43–44, Alaska Dep’t of Envtl. Conservation v. EPA, 540 U.S. 461 (No. 02-658) (noting that this conclusion should only apply to this order given the uncommon circumstances of the case).
30 See *Jason D. Nichols, Towards Reviving the Efficacy of Administrative Compliance Orders: Balancing Due Process Concerns and the Need for Enforcement Flexibility in Environmental Law*, 57 ADMIN. L. REV. 193, 195 (2005) (explaining that the Court decided not to resolve the potential constitutional problems associated with the judicial review of administrative compliance orders under the Clean Air Act); see also *Sackett v. EPA*, 556 U.S. 120, 132 (2012) (Alito, J., concurring) (noting that the Court did not comment on whether delayed review of orders issued under the Clean Water Act would raise constitutional issues and expressing his opinion that disallowing immediate review would violate due process).
31 *Thunder Basin Coal Co. v. Reich*, 510 U.S. 200, 218 n.22 (1994) (stating that the petitioner’s claim was not an “abstract challenge to the Mine Act’s statutory review scheme” but was limited to the “present situation”); see infra Part II.B.5.
32 See *infra* Part IV.B.
public policy perspective, delayed review is always the superior option.\textsuperscript{33} This idea is often premised on the erroneous belief that immediate judicial review inevitably leads to substantial delays in the enforcement of statutes and regulations,\textsuperscript{34} and that, as a result, its adoption would cause courts to be flooded with challenges.\textsuperscript{35} While these arguments possess intuitive appeal, an in-depth analysis reveals that they are not as persuasive as they initially seem. Merely allowing a plaintiff to challenge an order does not likely substantially delay enforcement.\textsuperscript{36} Challenging an order does not make it unenforceable\textsuperscript{37} because the plaintiff must seek and successfully obtain a preliminary injunction.\textsuperscript{38} However, this is unlikely to happen given the current test for preliminary injunctions, especially in the important cases, i.e., those in which serious harm would be expected to ensue if the injunction were granted.\textsuperscript{39} In the absence of an injunction, the threat of accumulating daily penalties is a very powerful incentive to comply swiftly. Moreover, courts would not be flooded with these types of challenges because, as the economic models predict and the available empirical evidence confirms, the litigation costs and increasing daily fines for non-compliance would deter frivolous claims.\textsuperscript{40}

This Article makes a second contribution to the existing literature by examining the constitutional limitations on the legislature’s authority to enact statutes that bar immediate review of administrative orders.\textsuperscript{41} The main constraint on the design of delayed review provisions is the doctrine of constitutionally intolerable choices, which emanated from the Supreme Court case \textit{Ex parte Young}\textsuperscript{42} and its progeny.\textsuperscript{43} However, neither the Supreme Court nor the literature has offered a comprehensive account of this constitu-

\textsuperscript{33} See infra Part III.A.

\textsuperscript{34} See Valerie L. Starr, \textit{Should Pre-Enforcement Judicial Review of Administrative Compliance Orders Be Available Under the Clean Air Act?}, 38 SUFFOLK U. L. REV. 903, 915 (2005) (“By permitting pre-enforcement judicial review, formal adjudication in the court of appeals hinders the compliance order’s expediency.”); see also Quiggle, \textit{supra} note 12, at 329 (explaining that plaintiffs who are permitted to seek pre-enforcement review of compliance orders “will likely succeed in both slowing down the enforcement of environmental laws and delaying the cleanup of catastrophic contamination”).

\textsuperscript{35} See Christopher M. Wynn, \textit{Facing a Hobson’s Choice? The Constitutionality of the EPA’s Administrative Compliance Order Enforcement Scheme Under the Clean Air Act}, 62 WASH. & LEE L. REV. 1879, 1901 (2005) (noting that the majority view is that immediate review prevents floods of challenges); see also Quiggle, \textit{supra} note 12, at 328; Marin K. Levy, \textit{Judging the Flood of Litigation}, 80 U. CHI. L. REV. 1007, 1073–74 (2013) (noting that this argument is problematic because it would require empirical evidence of the number of cases that lead to a flood of challenges).

\textsuperscript{36} See infra Part III.A.1.a.

\textsuperscript{37} See infra Part III.A.1.a.

\textsuperscript{38} See infra Part III.A.1.a.

\textsuperscript{39} See infra Part III.A.1.a.

\textsuperscript{40} See infra Part III.A.2.

\textsuperscript{41} Given that other scholars have addressed the issue, \textit{see, e.g.}, Davis, \textit{supra} note 15, at 190; Starr, \textit{supra} note 34, at 918, this Article does not analyze in depth the administrative law doctrines that have led courts to reach inconsistent conclusions on the immediate reviewability of orders.

\textsuperscript{42} 209 U.S. 123 (1908).
tional doctrine. This Article fills this gap by providing a theoretical framework to evaluate the constitutionality of delayed review provisions. Two key factors to consider are the degree of coercion to comply that recipients of orders experience and whether they are able to meaningfully challenge the order in court after having previously complied with it.

This discussion focuses on a series of examples from environmental and worker safety areas, where the debate over delayed review of administrative orders has been the most animated. The analytical tools provided in both the public policy and constitutional sections of this Article, however, can be applied broadly. Similar questions have arisen under other statutes, such as the Federal Trade Commission Act, the Social Security Act, the National Labor Relations Act, and the Sarbanes-Oxley Act. Another area in which delayed review has been controversial is tax law. Specifically, scholars have been recently examining the contours of the bar on immediate review of the assessment and collection of taxes contemplated in the Anti-Injunction Act and the Declaratory Judgment Act. The conceptual frameworks proposed in this Article should also help illuminate analogous debates in these and other areas.

This Article proceeds in four parts. Part I provides concrete examples of environmental and worker safety statutes that illustrate the role of administrative orders as enforcement mechanisms and show why the timing of their judicial review is critical. While similar debates have also arisen in the context of challenges to regulations, this Article will focus specifically on administrative orders because the public policy and legal questions underlying the judicial review of these two forms of agency action do not overlap completely. Part II explores the constitutional and administrative law doctrines on which courts have relied to determine whether immediate review is permissible under different statutes. This analysis also reveals that courts have

---


46 Immediate or pre-enforcement review of regulations raises legal and public policy questions that do not perfectly overlap with those relevant to administrative orders. See McNeill, supra note 6, at 348–49 (noting the special importance—in the specific context of regulations—of determining whether this form of agency action is fit for review, as required by the doctrine of ripeness).
applied these principles inconsistently to statutes that are structurally very similar. Part III presents a vigorous criticism of the widely-accepted justifications for delayed review. These justifications are largely based on a distorted view of the effects that immediate review of administrative orders would have on enforcement and court caseloads. Finally, Part IV examines the constitutional constraints on delayed review. It starts by delving into the debate over whether there is a general constitutional right to judicial review. It then narrows its focus to provide an analytical framework to evaluate the constitutionality of different types of delayed review approaches under the doctrine of constitutionally intolerable choices.

I. ADMINISTRATIVE ORDERS: THE TENSIONS BETWEEN ENFORCEMENT AND JUDICIAL REVIEW

A. Administrative Orders As Enforcement Mechanisms

The notion of an order has very broad contours and, not surprisingly, the Administrative Procedure Act (“APA”) has defined it in the negative: an order is a decision by an agency that is not a rule. Administrative orders are used to ensure compliance with the mandates in federal statutes and regulations. Compliance orders—also known as citations in certain contexts—are a paradigmatic example of orders that serve this purpose. One commentator has defined them as “a directive from the [agency] requiring the recipient to comply with a particular statutory or regulatory requirement by a specified deadline.” Their main goal is to provide an alternative to the—usually more cumbersome—judicial enforcement option and offer a swift and flexible enforcement tool to address less serious violations. However, if the
recipient of the order does not comply with it voluntarily, agencies—which often lack authority to enforce these orders directly—will have to bring an action to seek enforcement of the order by a court.\(^{52}\) Despite these limitations, compliance orders are very widely used. In recent years, the EPA, for example, has issued between 800 and 900 compliance orders per year.\(^{53}\)

In order to illustrate the types of issues that delayed judicial review of administrative orders raise, the discussion in Parts II, III, and IV will focus on two broad areas of the law: environmental and worker safety law. For this reason, some additional background on these statutory frameworks is warranted.

The environmental statutes that the following sections will discuss are the Clean Air Act,\(^{54}\) Clean Water Act,\(^{55}\) Resource Conservation and Recovery Act (“RCRA”),\(^{56}\) and Comprehensive Environmental Response, Compensation, and Liability Act (“CERCLA”).\(^{57}\) These statutes share similar purposes. Congress enacted the Clean Air Act\(^{58}\) and the Clean Water Act\(^{59}\) to improve the quality of the nation’s air and waters, respectively. RCRA\(^{60}\) and CERCLA,\(^{61}\) on the other hand, have the goal of addressing the negative effects that hazardous substances can have on human health and the environment. These and other environmental statutes generally give the EPA four different possible courses of action after the agency identifies a potential violation: The EPA may issue a compliance order, assess civil penalties, seek the imposition of criminal penalties, or bring an enforcement action in court.\(^{62}\)

Administrative orders also play an important role in the area of worker safety. In particular, this Article focuses on two statutes: the Federal Mine

\(^{52}\) Id. (explaining that the EPA lacks contempt powers and that the only way of compelling compliance lies with the courts).


\(^{58}\) Id. § 7401(b) (2012).


\(^{61}\) Id. § 9604(a).

\(^{62}\) See, e.g., Tenn. Valley Auth. v. Whitman, 336 F.3d 1236, 1240–41 (11th Cir. 2003) (explaining the enforcement tools under the Clean Air Act); see also 33 U.S.C. § 1319(b)–(d) (2012) 42 U.S.C. § 6928(a) (RCRA) (Clean Water Act); id. § 7413(b) (2012) (Clean Air Act); id. §§ 9606(b), 9609(c) (CERCLA).
Safety and Health Amendments Act of 1977 (Mine Safety Act)\textsuperscript{63} and the Occupational Safety and Health Act of 1970 (OSH Act).\textsuperscript{64} These two statutory frameworks, which have the goal of ensuring the health and safety of workers and miners,\textsuperscript{65} are examined together due to the similarity of their enforcement mechanisms.\textsuperscript{66} Under both Acts, the Secretary of Labor may issue an administrative order—referred to as “citation”—directing the employer or mine operator to abate a violation within a certain time period.\textsuperscript{67} Other enforcement options include the possibility that the Secretary of Labor assess a civil penalty,\textsuperscript{68} seek the imposition of criminal penalties,\textsuperscript{69} or bring an enforcement action.\textsuperscript{70}

\textbf{B. The Importance Of Orders’ Timing Of Review}

The timing of review of administrative orders can have a very significant impact not only on the recipients of the orders, but also on those whom the different statutes intend to protect.\textsuperscript{71} As the examples below illustrate, disallowing immediate judicial review of administrative orders can raise very different types of concerns. In the interest of clarity, the following discussion considers compliance orders and environmental cleanup orders separately, even though the legal mechanisms that delay the judicial review of these two types of orders operate in a very similar fashion.


One could theoretically challenge administrative orders at three different points in time: (i) immediately after they are issued, (ii) after the agency brings an enforcement action,\textsuperscript{72} or, under some statutes, (iii) after an administrative commission rules on their validity.\textsuperscript{73} Under the first option, recipients of a compliance order would simply be able to challenge it in court as soon as they receive it, without having to wait for any further action from the agency. The other options adopt a delayed review approach, which requires the recipient of the order to wait—without bringing a challenge in court—

\begin{itemize}
  \item \textsuperscript{63} 30 U.S.C. §§ 801–964 (2012).
  \item \textsuperscript{64} 29 U.S.C. §§ 651–678 (2012).
  \item \textsuperscript{66} Ne. Erectors Ass’n v. Sec’y of Labor, 62 F.3d 37, 40 (1st Cir. 1995).
  \item \textsuperscript{68} 29 U.S.C. § 666(a)–(d) (2012); 30 U.S.C. § 820(a), (b) (2012).
  \item \textsuperscript{69} 29 U.S.C. § 666(e) (2012); 30 U.S.C. § 820(d) (2012).
  \item \textsuperscript{70} 29 U.S.C. § 660(b) (2012); 30 U.S.C. § 816(b) (2012).
  \item \textsuperscript{71} Most notably, the general public, workers, and wildlife. \textit{See supra Part I.A.} (explaining the goals of the Clean Air Act, Clean Water Act, RCRA, CERCLA, OSH Act, and Mine Safety Act).
  \item \textsuperscript{72} \textit{See infra} Part I.B.1; \textit{see also} Davis, \textit{supra} note 15, at 190.
  \item \textsuperscript{73} \textit{See infra} Part I.B.1.b.
\end{itemize}
until the government takes action. The type of governmental action that will provide access to court—i.e., bringing an enforcement action or having a commission rule on an administrative appeal—varies depending on the specific statute considered. However, the statutes examined in this Article adopt one of two approaches. One generally occurs with environmental statutes and the other alternative is more common in worker safety judicial review provisions.

i. The Need for an Enforcement Action: Delayed Review in the Context of Environmental Statutes. In the case of compliance orders issued under the main environmental statutes, the timing of judicial review can be very relevant to the recipient of the order. The four environmental statutes examined in this Article authorize the imposition of fines for failing to comply with an order’s requirements. These fines could reach astronomical amounts (e.g., up to $37,500 for each day of violation under the Clean Water Act). The daily nature of the fines is what makes the timing of review of compliance orders relevant. The greater the number of days without complying with an order, the larger the fine the potential violator may have to face.

With immediate review, the recipient of the order would be able to seek review of an administrative order as soon as the agency issues it. This could also include a request for a preliminary injunction. If the court grants the preliminary injunction, the accrual of daily penalties for failing to comply with the order will cease, at least while the court is examining the challenge.

Under a delayed review approach, however, the recipient of the order would only be able to question its validity in court when the agency brings an enforcement action. This leaves plaintiffs with a complicated choice: complying with the order (which could be improper) or refusing to do so at their own risk. In the latter case, the daily penalties can accumulate. If the agency, several months later—given the lack of compliance with the order—
Delayed Judicial Review

brings a judicial enforcement action, and the operator is not successful in challenging the order in court, the total amount of fines could be very high.79

ii. Prior Ruling by a Commission: OSH Act and Mine Safety Act. Under both the OSH Act and the Mine Safety Act, if the Secretary of Labor (Secretary) determines that the employer or mine operator has violated a statute, regulation, or order, the Secretary must issue a citation. As with compliance orders in the environmental context, a citation under the OSH Act and the Mine Safety Act serves the purpose of identifying the alleged violation and providing a reasonable timeframe for its abatement.80 Once the Secretary has issued a citation, the employer or operator may choose to comply with it voluntarily, contest it at the administrative level, or simply disregard it. If the employer or operator disregards the citation, the Secretary must send a notification with a proposed penalty and provide a period of fifteen or thirty days to contest the notification or proposed assessment of the penalty.81 If the employer or operator does not respond in the time given, the proposed penalty becomes final and unreviewable by any other administrative body or court.82

If the employer or operator decides to contest either the initial order or the notification and proposed penalty, however, a Commission will hear that challenge.83 After the appropriate hearing, the Commission will issue an order affirming, modifying, or vacating the Secretary’s citation and proposed penalty.84 The judicial review provisions of both statutes allow the operator or employer to challenge the order of the Commission in the appropriate court of appeals.85 The Secretary may also bring an action in federal court to either obtain review or seek enforcement of an order of the Commission.86

This framework also allows for an immediate or delayed review approach. The question that the judicial review provisions of the statute did not explicitly answer was the following: may employers and operators seek immediate judicial review of citations issued by the Secretary without having to go through the process before the Commission? Delayed review can have similar effects to those under the environmental statutes examined above because the OSH Act and the Mine Safety Act provide that operators or employers who fail to comply with a citation may be assessed a penalty of

79 See Law of Envir. Prot. § 9:262 (explaining the doubling of penalties and the criminal implications of the issuance of compliance orders).
80 30 U.S.C. § 814(a) (2012) (this provision of the Mine Safety Act provides that “[e]ach citation shall be in writing and shall describe with particularity the nature of the violation, including a reference to the provision of the chapter, standard, rule, regulation, or order alleged to have been violated. In addition, the citation shall fix a reasonable time for the abatement of the violation”). The OSH Act includes identical language. See 29 U.S.C. § 658(a) (2012).
up to $7,000\textsuperscript{87} and $5,000,\textsuperscript{88} respectively, per day of violation. This has created the following concern: if an operator or employer receives a citation with which it disagrees and fails to comply, penalties could accumulate throughout the administrative appeal process before the Commission and later, while the court of appeals is examining the challenge.\textsuperscript{89} In order to address this problem, operators and employers have urged the courts to allow them to seek immediate judicial review of the initial citation by the Secretary.\textsuperscript{90}

2. Futile Challenge: Delayed Review of Cleanup Orders

Delayed review has also been controversial in another context: cleanup actions issued under CERCLA and RCRA. CERCLA authorizes the EPA to clean up contaminated sites and recover, at a later time, cleanup costs from those who are responsible for the pollution.\textsuperscript{91} The statute also allows private parties to carry out the cleanup and other remedial activities at the site.\textsuperscript{92} In the latter case, the EPA may issue different types of orders—e.g., one directing a person to take action to remediate contamination that poses a threat to human health or the environment\textsuperscript{93} or an order to memorialize an agreement between the agency and a potentially responsible party.\textsuperscript{94}

Section 113(h) of CERCLA, however, deprives federal courts of jurisdiction to review challenges of cleanup actions that have been selected following the appropriate procedure, including the types of orders noted in the preceding paragraph.\textsuperscript{95} This affects not only actions under CERCLA but also those that could be brought under other federal statutes, such as RCRA.\textsuperscript{96}

\textsuperscript{87} 29 U.S.C. § 666(d).
\textsuperscript{88} 30 U.S.C. § 820(b)(1).
\textsuperscript{91} See 42 U.S.C. §§ 9604(a)(1), 9607(a) (2012). For an analysis of the parties from which the EPA may recover cleanup costs, see Luis Inaraja Vera, Compelled Costs Under CERCLA: Incompatible Remedies, Different Statutes of Limitations, and Tort Law, 17 VT. J. ENVTL. L. 394, 396 (2016).
\textsuperscript{92} See 42 U.S.C. § 9604(a)(1). CERCLA contemplates two types of actions to address contamination, i.e., removal and remedial actions. Their main difference is that removal actions are generally short term, while remedial actions are broader and aim to provide a permanent solution for the contamination at a particular site. Compare id. § 9601(23), with id. § 9601(24). While the definitions in the statute include some activities that go beyond cleanup of contaminants, such as monitoring or provision of alternative water supplies, this Article, for the sake of succinctness, will refer to removal and remedial actions collectively as “cleanup actions.”
\textsuperscript{93} See 42 U.S.C. § 9606.
\textsuperscript{94} See id. § 9622(d)(3).
\textsuperscript{95} See id. § 9613(h).
\textsuperscript{96} See El Paso Nat. Gas Co. v. United States, 750 F.3d 863, 880–82 (D.C. Cir. 2014) (barring RCRA claims); Razore v. Tulalip Tribes of Wash., 66 F.3d 236, 239–40 (9th Cir.
This has been controversial when the potential plaintiff is the recipient of the order and also when the challenging party is a citizen group concerned with the impacts that a particular cleanup action could have on the environment or human health. This can occur, for example, when the method used to clean up a piece of land results in some of the contaminants becoming airborne.

When the plaintiff is a citizen group, however, the jurisdictional bar is not absolute. It instead limits when these challenges may be brought. Before Congress enacted the amendment that incorporated the jurisdictional bar to the statute, judicial actions against these cleanup activities could, in theory, have been brought immediately after the agency formally approved them and issued the corresponding order. Under the current version of the statute, however, only delayed review is available. Any challenge of cleanup actions under the citizen suit provision can only be brought once the cleanup is finalized.

It is worth highlighting that the jurisdictional bar in CERCLA was enacted with the purpose of preventing recipients of orders—i.e., generally those who are potentially responsible parties under the statute—from resorting to litigation to delay cleanups. Taking this into account, the problem with applying it to citizen suits is twofold. First, some of these suits are brought precisely because the cleanups are paralyzed or moving forward at a very slow pace. Second, delayed review in cases where the cleanup itself will lead to harm to human health or the environment makes the challenge futile: by the time the action is brought, any harm will have already occurred.

1995) (barring citizen suit under the Clean Water Act and RCRA); McClellan Ecological Seepage Situation v. Perry, 47 F.3d 325, 331 (9th Cir. 1995) (barring claims under RCRA and the Clean Water Act).

97 See Healy, supra note 12, at 301 (noting the two types of scenarios in which citizen suits are brought to protect human health or the environment); Pollans, supra note 18, at 442.


100 See 42 U.S.C. § 9613(h)(4) (2012) (foreclosing citizen suits of cleanup actions that are "to be undertaken at the site"). Courts have explained that this language requires cleanups to be finalized before a challenge may be brought. See, e.g., Clinton Cty. Comm’rs v. EPA, 116 F.3d 1018, 1025 (3d Cir. 1997) ("a citizens’ suit challenging a ‘removal’ action may not be brought even after completion of that removal action."); Schalk v. Reilly, 900 F.2d 1091, 1093 (7th Cir. 1990) (interpreting the language in § 9613(h)(4) to bar citizen suits to cleanups that are not completed); Alabama v. EPA, 871 F.2d 1548, 1557 (11th Cir. 1989) (noting that actions may be brought "only after a remedial action is actually completed").


102 See Pollans, supra note 18, at 483.
II. THE LEGAL DEBATE OVER DELAYED REVIEW OF ADMINISTRATIVE ORDERS

A. Legal Doctrines In Play

While the scholarship on delayed review of administrative orders has addressed the different policy concerns associated with the two timing-of-review options discussed above, the case law has primarily focused on a limited number of legal issues. The two most debated questions have been whether a given statute precludes immediate review of administrative orders and if this form of agency action has the necessary attributes to be considered final.103 A smaller number of cases have also paid some attention to whether barring immediate review could violate the requirements of due process.104 This section provides a general overview of these legal doctrines and then examines how courts have ruled on the availability of immediate review under environmental and worker safety statutes.

1. Statutory Preclusion

Section 702 of the APA provides that “[a] person suffering legal wrong because of agency action, or adversely affected or aggrieved by agency action within the meaning of a relevant statute, is entitled to judicial review thereof.”105 As the Supreme Court has explained, this provision encapsulates “the basic presumption of judicial review.”106 Section 701 of the APA, however, limits this presumption of reviewability in two instances, the relevant one for the purposes of this discussion being when “statutes preclude judicial review.”107 While the presumption of reviewability is strong, the Supreme Court has noted that it is a rebuttable one.108 To overcome the

---

103 See infra Part II.B (examining the different conclusions that courts have reached applying these two doctrines to orders issued under the Clean Water Act, Clean Air Act, RCRA, CERCLA, the Mine Safety Act, and the OSH Act).
104 The Supreme Court has not addressed this argument with any of the environmental statutes. See infra Part II.B. However, some lower courts have examined this question. See, e.g., Gen. Elec. Co. v. Jackson, 610 F.3d 110, 113 (D.C. Cir. 2010); Solid State Circuits, Inc. v. EPA, 812 F.2d 383, 392 (8th Cir. 1987); Wagner Seed Co. v. Daggett, 800 F.2d 310, 316 (2d Cir. 1986).
107 5 U.S.C. § 701(a) (2018). As one author has explained, the Supreme Court has not taken a stance on whether Congress’s power to limit jurisdiction is limited by Article III or the Due Process Clause. Harold H. Bruff, Availability of Judicial Review, in A GUIDE TO JUDICIAL AND POLITICAL REVIEW OF FEDERAL AGENCIES 17 (Michael E. Hertz et al. eds., 2015).
108 Mach Mining, LLC v. EEOC, 135 S. Ct. 1645, 1651 (2015) (noting that the presumption “fails when a statute’s language or structure demonstrates that Congress wanted an agency to police its own conduct”).
presumption one of two things is necessary: specific statutory language or evidence of congressional intent.109

When considering preclusion in cases in which the language of the statute provides no clear guidance, the question becomes: how significant must the evidence of congressional intent be to preclude review? Those in favor of delayed review have often relied on Block v. Community Nutrition Institute.110 This decision provided a comprehensive list of the types of evidence that may support a finding of implied preclusion, namely, the legislative history, the “contemporaneous judicial construction barring review and the congressional acquiescence in it,” and the structure of the statute.111 In addition, while other decisions had explained that implied preclusion required “clear and convincing evidence,”112 the Court in Block reinterpreted this expression to mean that the intent to preclude only needed to be “fairly discernible.”113

More recent Supreme Court opinions, however, have used the “clear and convincing” language again, which suggests a higher standard.114 Moreover, one scholar has suggested that Block should not be afforded much weight because it possibly confuses the concepts of standing and preclusion.115 As explained below, the issue of implied preclusion is a critical piece in the immediate-versus-delayed-review analysis, given that most of the statutes discussed in this Article—with the exception of CERCLA, which incorporates a provision that expressly imposes timing-of-review limitations—do not explicitly preclude immediate review of orders.

2. The Doctrine of Finality

An additional limitation on judicial review of agency action that has been very relevant in cases dealing with administrative orders is the doctrine of finality. The logic behind it is that courts should not meddle in ongoing agency proceedings and should instead wait until the agency reaches a final decision.116 To that end, Section 704 of the APA contemplates the require-

110 467 U.S. 340.
111 Id. at 349–50. For a recent case applying these factors, see Cuozzo Speed Techs., 136 S. Ct. at 2140–42.
113 467 U.S. at 351–52.
114 Compare Cuozzo Speed Techs., 136 S. Ct. at 2140–42, with Elgin v. Dep’t of Treasury, 567 U.S. 1, 10 (2012).
115 WILLIAM F. FOX, UNDERSTANDING ADMINISTRATIVE LAW 281 (2012).
116 Limiting Judicial Intervention in Ongoing Administrative Proceedings, 129 U. Pa. L. Rev. 452, 467 (1980) (explaining that, in addition to the requirement of exhaustion of administrative remedies, the doctrine of finality provides “a complementary approach to the problem of judicial intervention in ongoing administrative proceedings”).
ment that agency action be final before a court may review it. 117 Except in cases in which the statute explicitly resolves this issue, courts will have to determine, when relevant to the case at hand, whether certain agency action is final.

Courts generally use a two-part test to make this determination. 118 First, in order to be final, the action must mark the “consummation of the agency’s decisionmaking process.” 119 The Supreme Court has interpreted this requirement to mean that the decision must present the agency’s final position on the matter and is not “merely tentative” or “interlocutory.” 120 Second, the agency’s action “must be one by which rights or obligations have been determined, or from which legal consequences will flow.” 121 This does not occur when the agency is merely providing a recommendation that is not binding on the relevant actors. 122 Similarly, actions that do not establish any obligations or prohibitions will not meet this standard. 123

While authors and courts generally focus on this two-part test when analyzing finality issues, some court decisions have taken into account additional factors. 124 Certain Supreme Court decisions, for example, have explained that a person seeking review under the APA must also establish that there is “no other adequate remedy in a court.” 125 This requirement, which comes directly from Section 704 of the APA, can be relevant in the context of administrative orders. The key question is whether the fact that the recipient of the order has the possibility of opposing an enforcement action brought by the agency provides an adequate remedy that justifies barring pre-enforcement—i.e., immediate—review. 126

3. Procedural Due Process

Even when an order is not fit for review based on the doctrines of preclusion or finality, there are cases in which judicial review is nevertheless constitutionally mandated. The Fifth and Fourteenth Amendments to the Constitution require that any deprivation of “life, liberty, or property” be

---

120 Alaska Dep’t of Envtl. Conservation v. EPA, 540 U.S. 461, 463 (2004); Bennett, 520 U.S. at 178.
121 Bennett, 520 U.S. at 178; see also Chi. & S. Air Lines, 333 U.S. at 113; Murray Energy Corp., 788 F.3d at 336.
122 See Bennett, 520 U.S. at 178.
123 Murray Energy Corp., 788 F.3d at 336.
124 See Murray Energy Corp., 788 F.3d at 336–37; Bruff, supra note 107, at 8.
126 See U.S. Army Corps of Eng’rs v. Hawkes Co., 136 S. Ct. 1807, 1815 (2016); Sackett, 566 U.S. at 120.
carried out following “due process of law.” While substantive due process focuses on whether the government has an appropriate reason to take a person’s “life, liberty, or property,” procedural due process is concerned with the procedures that the government follows when doing so. The key questions to guide the procedural due process analysis are: (i) is due process required?, (ii) when should the required process take place?, and (iii) what type of process is required?

To address these questions, it is useful to turn to the seminal court decision on procedural due process, *Mathews v. Eldridge.* In that case, the plaintiff had his disability benefits terminated after having had an opportunity to submit written documents to the Social Security Administration. The plaintiff’s claim was that the procedures in place before the termination of these types of benefits did not comport with due process. In order to answer the question of whether the procedures were constitutionally sufficient, the Court laid out what is regarded as the current test for procedural due process. Its three factors are the following: (i) “the private interest that will be affected by the official action,” (ii) “the risk of an erroneous deprivation of such interest through the procedures used,” and (iii) “the Government’s interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.”

It is critical, for the purposes of this Article, to differentiate this traditional procedural due process requirement from what some have referred to as the *Ex parte Young* doctrine, which also has a constitutional underpinning. As examined at length in Part IV, the gist of the doctrine adopted in *Ex parte Young*
Young is that recipients of orders should not be put in a position in which, in order to challenge the order, they have to risk the imposition of high fines or imprisonment. This would occur, for example, when one is forced to disobey a law in order to be able to challenge it. While some courts have treated this doctrine as a requirement of procedural due process, others have correctly observed that it constitutes a separate constitutional principle.

B. The State Of The Issue Under The Different Statutes

In light of the strong economic incentive to comply with administrative orders, the issue of whether the recipients of compliance orders may seek judicial review before the agency decides to bring an enforcement action in court—which could potentially take a very long time—can be critical. As discussed earlier, interest in obtaining judicial review of administrative orders can also arise in other contexts, such as when citizen groups seek judicial review of a cleanup order issued under CERCLA. Delayed review in those cases can make the later challenge futile. Although all the relevant statutes share similar goals to protect the environment, human health, and the safety of workers, courts have been inconsistent when answering the question of whether immediate judicial review of administrative orders is available. These discrepancies are, in many cases, not justified by the language of the enforcement provisions of the different statutes. This subsection examines how courts have decided this issue under the Clean Air Act, Clean Water Act, RCRA, CERCLA, the Mine Safety Act, and OSH Act.

1. Clean Water Act

Before the Supreme Court examined this question in 2012, the agreement among lower courts was that immediate review of administrative orders issued under the Clean Water Act was barred based on the doctrine of

---

137 See infra Part IV.B.1.
138 This issue, which the article discusses in Part IV, arises when a statute only allows a recipient of an order to question its validity in the context of an enforcement action brought by the agency. See infra Part IV.C.1.
139 See infra notes 295–303 and accompanying text (explaining that while procedural due process is mostly concerned about the process that is followed when a person is deprived of life, property, or liberty at the administrative or judicial levels, the Ex parte Young doctrine tries to ensure that recipients of orders will have an opportunity to question the validity of such orders in court).
140 See supra Part I.B.1.
141 See supra Part I.B.2.
142 See supra Part I.B.2.
143 See supra Part I.A.
144 Compare Acker v. EPA, 290 F.3d 892, 894 (7th Cir. 2002) (finding an order issued under the Clean Air Act to not be immediately reviewable), with Sackett v. EPA, 566 U.S. 120, 126–27, 129 (2012) (concluding that orders issued under the Clean Water Act are final and immediately reviewable by the courts).
Delayed Judicial Review

preclusion.\textsuperscript{145} In \textit{Sackett v. EPA}, however, the Supreme Court examined the government’s contentions that the order at issue was not final and that the Clean Water Act impliedly precluded its immediate review.\textsuperscript{146} The relevant facts of the case are as follows. Michael and Chantell Sackett (“the Sacketts”) had placed fill material on their property.\textsuperscript{147} The EPA considered this action to violate the Clean Water Act and issued an order directing the Sacketts to restore their property to its original state.\textsuperscript{148} The Sacketts sought immediate review of the order, but both the district court and the Ninth Circuit concluded that the Clean Water Act precluded that challenge.\textsuperscript{149}

On the finality question, the Supreme Court ruled against the government and found that the order was final.\textsuperscript{150} Focusing on the first finality prong, the Court noted that the order “determined rights or obligations” because it imposed on the Sacketts a legal duty to restore their land, and the failure to comply with the order exposed the Sacketts to penalties if the EPA decided to initiate an enforcement proceeding in the future.\textsuperscript{151} The government’s argument—based on the second prong of the finality test—that the order did not constitute the “consummation of the agency’s decisionmaking process” because it invited “informal discussion” ultimately failed.\textsuperscript{152} The Court considered that the mere possibility of talking to EPA officials did not offer the Sacketts a meaningful opportunity for agency review and that, therefore, the order was final.\textsuperscript{153}

As for the preclusion issue, the Court opined that the Clean Water Act did not preclude immediate review.\textsuperscript{154} Unlike other environmental statutes,\textsuperscript{155} the Clean Water Act contains no language precluding judicial review. For that reason, the Court focused its analysis on whether the Act should be interpreted to impliedly preclude immediate review of compliance orders.\textsuperscript{156} The EPA’s argument was that the purpose of the statute—improving the quality of the nation’s waters—as well as the fact that there are provisions in the statute expressly allowing judicial review of other agency actions, led to the conclusion that the immediate review of compliance orders was impliedly precluded.\textsuperscript{157} The Court disagreed and pointed out that these argu-

\textsuperscript{145} See Stein et al., \textit{supra} note 25, at 10,819.
\textsuperscript{146} See \textit{Sackett}, 566 U.S. at 126–27, 129.
\textsuperscript{147} \textit{Id.} at 125.
\textsuperscript{148} \textit{Id.} at 126.
\textsuperscript{149} \textit{Id.} at 125.
\textsuperscript{150} See \textit{id.} at 126–27.
\textsuperscript{151} \textit{Id.} at 126 (internal quotation marks omitted).
\textsuperscript{152} \textit{Id.} at 127 (internal quotation marks omitted).
\textsuperscript{153} \textit{Id.}
\textsuperscript{154} \textit{Id.} at 128–31.
\textsuperscript{155} CERCLA is an example of a statute that does explicitly preclude immediate review. \textit{See supra} Part 1B.II.
\textsuperscript{156} \textit{Sackett}, 566 U.S. at 129.
\textsuperscript{157} \textit{Id.} at 128–31.
ments were not sufficient to overcome the presumption of reviewability in the Administrative Procedure Act.\textsuperscript{158}

The \textit{Sackett} case provided a final answer to the question of whether administrative orders issued by the EPA under the Clean Water Act are immediately reviewable. More recent cases have reached a similar conclusion with respect to determinations made by the Corps of Engineers under this same statute.\textsuperscript{159} Because these challenges were resolved based on the doctrines of finality and preclusion, the Court did not have an opportunity to address the due process argument.\textsuperscript{160} Nevertheless, Justice Alito’s concurrence in \textit{Sackett} included a statement indicating that delayed review in that case would have led to results that are “unthinkable” in “a nation that values due process.”\textsuperscript{161}

2. Clean Air Act

While other environmental statutes such as the Clean Water Act and RCRA have also raised statutory preclusion questions, the controversy over whether the EPA’s orders issued under the Clean Air Act are subject to immediate review by courts has focused on finality. The Clean Air Act, however, is different from other statutes in this regard because, although it does not specify the timing of review of administrative orders, it does contain a general provision that allows for judicial review of \textit{any} final action of the Administrator.\textsuperscript{162} Thus, determining if administrative compliance orders are final has become the key part of the analysis.

Circuit courts of appeals have disagreed over whether compliance orders under the Clean Air Act are final. Some circuits have answered this question in the negative.\textsuperscript{163} In a Seventh Circuit case, for example, the court concluded that the plaintiff—i.e., the recipient of the order—had failed to meet either of the two prongs of the finality test.\textsuperscript{164} First, the court noted that the order did not meet the first prong because it did not represent the summation of the agency’s decisionmaking process.\textsuperscript{165} It viewed the order as simply notifying the plaintiff of the “potential” legal consequences that fur-

\textsuperscript{158} Id. at 129.

\textsuperscript{159} See, \textit{e.g.}, U.S. Army Corps of Eng’rs v. Hawkes Co., 136 S. Ct. 1807, 1813–16 (2016) (finding that a jurisdictional determination by the Corps of Engineers was immediately reviewable).

\textsuperscript{160} The due process argument would come into play when the plaintiff believes that an order deemed not final or whose review is precluded by the statute should nonetheless be immediately reviewable by a court. If a court concludes that the order is final and subject to judicial review, there is no need to examine the due process claim.

\textsuperscript{161} \textit{Sackett}, 566 U.S. at 132 (Alito, J., concurring).


\textsuperscript{163} Acker v. EPA, 290 F.3d 892, 894 (7th Cir. 2002); see also Solar Turbines, Inc. v. Seif, 879 F.2d 1073, 1076 (3d Cir. 1989); Asbestec Constr. Servs., Inc. v. EPA, 849 F.2d 765, 768 (2d Cir. 1988).

\textsuperscript{164} Acker, 290 F.3d at 894.

\textsuperscript{165} Id.
ther violation of its obligations under the Clean Air Act could entail. As for the second prong, that is, the requirement that legal consequences flow from the agency action, the Seventh Circuit interpreted it to mean that when an order does not impose a penalty, that order is generally not final. As a result, the court noted that, because the EPA’s order did not impose a sanction, it was not final and, thus, also not fit for judicial review.

Other circuits have reached the opposite conclusion. The Ninth Circuit determined that the two finality requirements had been met with respect to an order that the EPA had issued under the Clean Air Act. First, it concluded that the order was the agency’s last word on the matter and thus constituted the “consummation of the agency’s decision-making process.” Second, it explained that “rights or obligations” had been determined because the order had the effect of paralyzing the plaintiff’s facility, and that legal consequences could flow from the order—in the form of criminal and civil penalties—if the plaintiff failed to comply with it.

It is unclear if the Supreme Court has settled whether orders issued under the Clean Air Act are final. As one commentator has explained, the Supreme Court appears to have conceded in *Alaska Department of Environmental Conservation v. EPA* that compliance orders under the Clean Air Act are final and therefore reviewable. In that case, the Court explained that the order issued to the operator met the two finality prongs. First, it noted that the agency “had spoken its ‘last word’” on the matter and, second, the order had legal effect by potentially subjecting the operator to civil and criminal penalties if it defied it.

Other commentators, however, do not share the view that this case resolved the circuit split with regard to the finality of orders under the Clean Air Act. One possible reason to doubt that the Supreme Court settled this question is that, while the EPA had argued before the Ninth Circuit that the order was not final, it reconsidered its position when the case reached the Supreme Court. The practical consequence would be that the Court’s statements on the finality of the order were dicta. Interestingly, the EPA’s counsel

---

166 Id. (emphasis omitted).
167 Id. at 895.
168 *Acker*, 290 F.3d at 894–95.
169 See *Alaska Dep’t of Envtl. Conservation v. EPA*, 244 F.3d 748 (9th Cir. 2001); *All-steel, Inc. v. EPA*, 25 F.3d 312 (6th Cir. 1994).
170 *Alaska Dep’t of Envtl. Conservation*, 244 F.3d at 751.
171 Id. at 750 (internal quotation marks omitted).
172 Id.
173 See *Nichols*, supra note 30, at 205–06 (pointing out that this holding is applicable to other compliance orders under the Clean Air Act).
174 *Alaska Dep’t of Envtl. Conservation*, 540 U.S. at 482.
175 Id. at 483.
176 See *Stein et al.*, supra note 25, at 10,820–21 (noting that, years after the Supreme Court decision in *Alaska Dep’t of Envtl. Conservation* was published, there was a circuit split on this issue).
suggested in the oral argument that the agency’s position on finality applied only to that particular dispute, given that it presented an unusual fact pattern.\textsuperscript{178} In that case, the state had issued a permit to a mining facility operator, and then the EPA had ordered the operator to halt the construction of the plant.\textsuperscript{179} Therefore, the obligation to stop the construction only arose after the issuance of the order because the operator had a valid permit. This is different from the more common scenario in which the order merely directs the operator to comply with what a statute or regulation already required.

It is important to note that the Supreme Court’s decision in \textit{Sackett} did not necessarily resolve the uncertainty over whether immediate review is available with respect to orders issued under the Clean Air Act. As some authors have pointed out, it is unclear whether the Court would extend its view that orders under the Clean Water Act are immediately reviewable to other statutes such as the Clean Air Act.\textsuperscript{180} Another aspect to consider is that the EPA published a memorandum in March of 2013 suggesting that several types of orders issued under the Clean Air Act are immediately reviewable.\textsuperscript{181} The EPA’s view, however, is subject to change.\textsuperscript{182}

Before moving on to the next statute, a note on due process is in order. It is worth highlighting that the Supreme Court cases on the Clean Air Act never addressed the due process argument—which only comes into play when, under the finality and preclusion doctrines, the order is deemed to not be immediately reviewable.\textsuperscript{183} As a result, a circuit split still exists on whether due process would require immediate review of orders under the Clean Air Act.\textsuperscript{184} Part IV will suggest a framework to analyze this question.

\textsuperscript{179} Id.
\textsuperscript{180} See Stein et al., supra note 25, at 10,822 (noting that \textit{Sackett} “may” affect RCRA); Schiff, supra note 25, at 134 (explaining that it is unclear whether this decision applies to all compliance orders). However, the language in the more recent Supreme Court case dealing with the orders issued by the Corps of Engineers under the Clean Water Act suggests that the Court may be willing to apply the conclusions it reached in \textit{Sackett} to other statutes. See \textit{U.S. Army Corps of Eng’rs v. Hawkes Co.}, 136 S. Ct. 1807, 1815 (2016) (“As we have long held, parties need not await enforcement proceedings before challenging final agency action where such proceedings carry the risk of ‘serious criminal and civil penalties.’”).
\textsuperscript{181} Susan Shinkman, Dir., Office of Civil Enf’t, EPA, Memorandum on the Language Regarding Judicial Review of Certain Administrative Enforcement Orders Following the Supreme Court Decision in \textit{Sackett v. EPA} 5 (Mar. 21, 2013) (requiring the inclusion of language in compliance orders indicating that the orders are subject to judicial review) https://www.epa.gov/sites/production/files/documents/linguageregarding-sackett032113.pdf [https://perma.cc/5UFW-AD9S].
\textsuperscript{183} Nichols, supra note 30, at 195.
\textsuperscript{184} Id.
3. RCRA

The enforcement frameworks in RCRA and the Clean Water Act are very similar in that they are both completely silent with respect to whether courts may review orders immediately.\(^{185}\) As a result, the litigation over whether compliance orders under RCRA may be challenged at the pre-enforcement stage has mostly focused on the issue of preclusion.\(^{186}\)

Most courts have agreed that RCRA impliedly precludes review of compliance orders until the EPA brings an enforcement action.\(^{187}\) To reach this conclusion, courts have articulated different arguments. Some courts have considered that Congress’s purpose was to free the “EPA from the burden of answering challenges at every phase of environmental clean-up.”\(^{188}\) Another common argument has focused on the structure of the enforcement mechanisms in the statute.\(^{189}\) According to a district court judge, when the EPA decides to take action against the potential release of hazardous waste, RCRA gives the agency a choice between initiating a judicial action immediately and issuing an administrative compliance order.\(^{190}\) Courts have explained that allowing the recipient of a compliance order to challenge it in court would effectively eliminate “Congress’s scheme of giving the EPA some choice on how to proceed.”\(^{191}\)

In short, courts have concluded that compliance orders issued under RCRA may not be challenged outside of an enforcement action initiated by the EPA—i.e., are not subject to immediate review. Although the arguments that these courts have relied on may be at odds with the Supreme Court’s position in the *Sackett decision*—which dealt with the Clean Water Act—it remains uncertain whether the Court would interpret RCRA and the Clean Water Act consistently in this regard.\(^{192}\) Moreover, it is important to note that, currently, the EPA’s 2013 memorandum requires the agency to include

---

\(^{185}\) Stein et al., *supra* note 25, at 10,822. As noted earlier, the Clean Air Act is different because it contains a provision allowing judicial review of any final action of the EPA. See *supra* Part II.B.2.


\(^{188}\) See, e.g., *Mobil Oil*, 1997 WL 1048911, at *5.

\(^{189}\) See *Ross Incineration Servs.*, 118 F. Supp. 2d at 846.

\(^{190}\) *Id.* at 843; see also 42 U.S.C. § 6928(h) (2012).

\(^{191}\) *Ross Incineration Servs.*, 118 F. Supp. 2d at 846.

\(^{192}\) See Stein et al., *supra* note 25, at 10,822 (noting that *Sackett* “may” affect RCRA); Schiff, *supra* note 25, at 134 (explaining that it is unclear whether this decision applies to all compliance orders).
language in compliance orders issued under RCRA stating that the order is subject to judicial review.\footnote{Shinkman, supra note 181, at 3–5.}

4. CERCLA

CERCLA is different from the other statutes considered above in that it is the only one that contains a provision expressly precluding immediate judicial review of administrative orders.\footnote{42 U.S.C. § 9613(h).} However, as explained earlier, this prohibition is structured differently for recipients of orders than for environmental groups who wish to challenge a cleanup order under the citizen-suit provision of the act.\footnote{See supra Part I.B.2.}

In the first case, the preclusion provision in CERCLA would not allow the recipient of the order to challenge it in court immediately.\footnote{See 42 U.S.C. § 9613(h).} However, if the EPA were to bring an enforcement action or one to recover a penalty for failure to comply with the order, the recipient of the order would then be able to challenge its validity in that judicial proceeding.\footnote{Id. § 9613(h)(2); see Gabison, supra note 13, at 191 (explaining that “[i]f the EPA elects to go forth with the suit, the private individual cannot influence the process until the lawsuit begins by joining the suit as a plaintiff”).} The second scenario deals with situations in which citizens and organizations have objections about the cleanup activities at a particular site. Those—other than recipients—willing to challenge the agency action approving the cleanup action to be undertaken at the site would be able to do so under the citizen-suit provisions of the Act.\footnote{See 42 U.S.C. § 9659.} However, courts have interpreted the preclusion provision of CERCLA\footnote{Id. § 9613(h)(4).} to not permit challenges to ongoing cleanup actions.\footnote{See, e.g., Clinton Cty. Comm’rs v. U.S. EPA, 116 F.3d 1018, 1025 (3d Cir. 1997); M.R. (Vega Alta), Inc. v. Caribe Gen. Elec. Prods., Inc., 31 F. Supp. 2d 226, 234 (D.P.R. 1998).} In other words, citizen groups are only able to challenge cleanups once they have finalized.\footnote{See, e.g., Gen. Elec. Co. v. Jackson, 610 F.3d 110, 113 (D.C. Cir. 2010); Solid State Circuits, Inc. v. EPA, 812 F.2d 383, 392 (8th Cir. 1987); Wagner Seed Co. v. Daggett, 800 F.2d 310, 316 (2d Cir. 1986).}

Given the explicit bar on immediate challenges, it is not surprising to see that many recipients of compliance orders have resorted to procedural due process arguments in an attempt to persuade courts to grant them immediate review of orders.\footnote{See Jackson, 610 F.3d at 113.} These challenges, however, have been generally unsuccessful. Before any penalties can be imposed on the recipient of an order, there will be a hearing in court and, thus, the due process requirements are met.\footnote{As examined in Part IV, courts have also rejected Ex parte}
Young-type arguments against CERCLA, which generally focus on the hurdles that recipients of orders face before being able to seek judicial review.204

5. The OSH Act and Mine Safety Act

The concern under the OSH Act and Mine Safety Act is that, if the Secretary of Labor issues a citation to an operator or employer, and the latter files an administrative appeal, the penalties for non-compliance with this order could accumulate while the Commission is evaluating the validity of the citation. For that reason, operators and employers have tried to obtain immediate review of the Secretary’s citations.205

However, courts have refused to allow such challenges. To justify this outcome, courts have mostly relied on the doctrine of preclusion.206 Although the provisions of the OSH Act and Mine Safety Act do not expressly preclude judicial review of citations issued by the Secretary, courts have found these claims to be impliedly precluded.207 Two reasons have supported this finding. The first one is the structure of these Acts. While the statutes are silent on the reviewability of orders issued by the Secretary, they expressly allow challenges of orders issued by the Commission.208 The second argument that courts have used to justify implied preclusion under the Acts is that recent amendments to these two statutes suggest that the legislators had the intent of strengthening the two Acts’ enforcement provisions.209

Those seeking immediate review of citations have also relied on the Ex parte Young doctrine. Their claim has been that the penalties that may accumulate while the Commission is examining the administrative appeal create an impermissible coercion to comply, and compliance can lead to harm that will not necessarily be repaired with later judicial review.210 Courts have generally rejected this argument, especially after the Supreme Court ruled on a similar issue in Thunder Basin. In that case, a mine operator was seeking immediate review of a letter sent by the manager of the Mine Safety and Health Administration.211 The Supreme Court’s response to this argument incorporated references to both the requirement of procedural due process, when the Court used the expression “pre-deprivation hearing,” and to the Ex

---

204 See Solid State Circuits, 812 F.2d at 389–92; Wagner Seed, 800 F.2d at 317.
206 See Thunder Basin Coal Co., 510 U.S. at 207–12; Big Ridge, 715 F.3d at 652–54; E. Bridge, LLC v. Chao, 320 F.3d 84, 88–91 (1st Cir. 2003); Sturm, Ruger & Co. v. Chao, 300 F.3d 867, 871–76 (D.C. Cir. 2002) (applying Thunder Basin Coal Co.’s test for preclusion to decide whether the court had jurisdiction to hear an immediate challenge to a warrant issued under the OSH Act).
207 See Thunder Basin Coal Co., 510 U.S. at 207–12; Big Ridge, 715 F.3d at 652–54; E. Bridge, 320 F.3d at 88–91; Sturm, Ruger & Co., 300 F.3d at 871–76.
208 See supra note 207.
209 See supra note 207.
210 See Thunder Basin Coal Co., 510 U.S. at 216; Big Ridge, 715 F.3d at 653.
211 Thunder Basin Coal Co., 510 U.S. at 204.
parte Young doctrine, which the Court explicitly alluded to. 212 Similarly to what lower courts had ruled on the procedural due process claim under environmental statutes, the Court noted that, because the penalties would only be payable after the court reviewed them, there was no pre-deprivation hearing. 213 With regard to the Ex parte Young argument, the Court explained that, given that mine operators can obtain temporary relief from certain citations—meaning that, in these instances, penalties would not accumulate while the Commission is examining their validity—there was no Ex parte Young violation either. 214

While many courts have relied on Thunder Basin to decide similar challenges to the OSH Act and Mine Safety Act, it is important to note how narrow this ruling actually was. 215 First, this was not a facial challenge to the constitutional validity of the enforcement and judicial review provisions of the Mine Safety Act. On the contrary, the Court specifically stated that the petitioner’s claim was not an “abstract challenge to the Mine Act’s statutory review scheme,” but was limited to the “present situation.” 216 Second, while the Court explained that an Ex parte Young challenge would fail in cases in which temporary relief from the accumulation of penalties was available, it never clarified what would happen in the not uncommon case of citations for which the Act does not provide this option. 217

In short, the Thunder Basin ruling on whether the Mine Safety Act could raise constitutional concerns under Ex parte Young was very narrow. In addition, the Supreme Court has not considered these types of challenges in the context of the OSH Act or any of the environmental statutes discussed above. A framework that allows courts and legislatures to evaluate the constitutionality of delayed review provisions is long overdue. Part IV of this Article will take on this task after a critical analysis, in Part III, of the public policy arguments in support of delayed review.

212 Id. at 218.
213 Id.
214 Id.
215 See, e.g., Big Ridge, 715 F.3d at 653–54; Reich v. Manganas Painting Co., 104 F.3d 801, 802–03 (6th Cir. 1997).
216 Thunder Basin Coal Co., 510 U.S. at 218 n.22.
217 Id. at 218. The constitutional problems that arise under these other types of citations for which no temporary relief is available were thoroughly examined by a district court many years before the Court decided Thunder Basin Coal Co. See Lucas v. Morton, 358 F. Supp. 900, 903–05 (W.D. Pa. 1973) (explaining how, without an avenue to obtain immediate judicial review in the case of irreparable harm, barring temporary relief of certain orders could be unconstitutional).
Delayed Judicial Review

III. THE QUESTIONABLE PUBLIC POLICY FOUNDATION OF DELAYED JUDICIAL REVIEW

A. Arguments in Support of Delayed Judicial Review

Delayed judicial review has an unquestionable intuitive appeal. If it increases the pressure on recipients of orders to comply with what an agency demands, then this form of review is preferable because, despite the fairness concerns that it may raise, it will help further the goals of the particular statute under which an order is issued.\footnote{See David Montgomery Moore, Pre-Enforcement Review of Administrative Orders to Abate Environmental Hazards, 9 PACE ENVTL. L. REV. 675, 678 (1992) (explaining that immediate review of orders has been perceived to be in tension with the fulfillment of the statute’s purpose); Starr, supra note 34, at 918 (“Pre-enforcement judicial review of administrative compliance orders hinders the purpose of the statute.”).} Those who support delayed judicial review tend to also rely on a series of generic arguments about the dangers of immediate review, namely that it leads to delays in enforcement\footnote{See Healy, supra note 12, at 325.} and to a significant increase in the number of court challenges of compliance orders.\footnote{See Quiggle, supra note 12, at 328–29.} The more serious these effects, the easier it is to brush aside the fairness concerns associated with delayed review. The following analysis shows that these arguments are not as persuasive as a superficial look would suggest and that they can only justify delayed review under a very limited set of circumstances.

1. The Argument That Allowing Immediate Review Would Cause Important Delays in Enforcement

One of the main arguments against immediate judicial review of administrative orders is that it delays enforcement and that this, in turn, could lead to harm to the environment or human health.\footnote{See Starr, supra note 34, at 913 (explaining that formal adjudication in court of appeals hinders the compliance order’s expediency.”); see also Quiggle, supra note 12, at 329 (explaining that plaintiffs who are permitted to seek pre-enforcement review of compliance orders “will likely succeed in both slowing down the enforcement of environmental laws and delaying the cleanup of catastrophic contamination”).} There are two theories that scholars use to support this idea. The first one is that allowing a challenge in court causes, by itself, a direct delay in the enforcement of the order.\footnote{See Healy, supra note 12, at 326 (explaining the EPA’s reluctance to continue with cleanups after the response action is challenged).} The second theory is that the delay is of an indirect nature, brought about, for example, by the way an agency reacts to a challenge of one of its orders.\footnote{See Healy, supra note 12, at 328–29.} The Delays Directly Resulting From a Challenge. The first theory is the least convincing one. The legal effects of compliance orders are not inter-
rupted merely as a result of the plaintiff’s bringing a challenge in court.\textsuperscript{224} The order will only become unenforceable while the court is examining the merits of the challenge if the recipient of the order seeks a preliminary injunction and the court decides to grant it.\textsuperscript{225} However, the likelihood of this happening in cases in which the order is both legal and necessary to prevent the occurrence of a substantial harm to the environment is low.

To be successful, the plaintiff seeking a preliminary injunction must generally show a likelihood of success on the merits, irreparable harm, that the harm to the moving party is higher than the harm to the non-moving party, and that the balance between the public and private interests favors the public interest.\textsuperscript{226} Before examining the four prongs of the test, it is important to highlight an implied assumption of the argument that immediate review inevitably leads to delay, i.e., that the order in question is appropriate. In other words, the premise is that the courts would uphold the order, if challenged. This is because, by identifying the issue as ‘delay,’ one is assuming that the challenge itself would ultimately be unsuccessful and that the problem is that the order would end up being enforced later than it would otherwise have been.

With this assumption in mind, if we are focusing on appropriate orders, by definition, the likelihood that plaintiffs will succeed on the merits is going to be low. As courts have put it, plaintiffs must show “[m]ore than a mere possibility of relief.”\textsuperscript{227} If the order that is being challenged is proper, the plaintiffs’ ability to demonstrate to the court that they are likely to prevail at the end of the proceeding will be low in most cases.

As for the second factor—i.e., the existence of irreparable harm—the Supreme Court has set a high bar by noting that movants do not meet their burden by merely “showing some possibility of irreparable injury.”\textsuperscript{228} Plaintiffs only meet their burden if they show that they would be “likely” to suffer an irreparable injury.\textsuperscript{229} However, in the majority of cases involving the types of plaintiffs that those concerned with immediate review tend to focus on the most—i.e., manufacturing companies\textsuperscript{230}—the harm experienced

\textsuperscript{224} See Davis, supra note 15, at 204.

\textsuperscript{225} See id.


\textsuperscript{227} Nken v. Holder, 556 U.S. 418, 434 (2009) (alteration in original) (internal quotation marks omitted).

\textsuperscript{228} Id. (internal quotation marks omitted).


\textsuperscript{230} See Quiggle, supra note 12, at 329 (focusing on “large polluters like General Electric”).
by the recipient of the order will be of an economic nature, which is often deemed not “irreparable” by courts.\footnote{See, e.g., Davis v. Pension Benefit Guar. Corp., 571 F.3d 1288, 1295 (D.C. Cir. 2009), abrogated on other grounds by Winter v. Nat. Res. Def. Council, Inc., 555 U.S. 7 (2008); see also Ace Am. Ins. Co. v. Wachovia Ins. Agency Inc., 306 Fed. App’x 727, 731 (3d Cir. 2009) (“[A] preliminary injunction must be the only way of protecting the plaintiff from harm and may not be granted to relieve purely economic harm.” (citing Frank’s GMC Truck Ctr., Inc. v. Gen. Motors Corp., 487 F.2d 100, 102 (3d Cir. 1988)); Daily Caller v. U.S. Dep’t of State, 152 F. Supp. 3d 1, 6 (D.D.C. 2015) (“[W]e see no reason to depart from the general rule that economic harm does not constitute irreparable injury.”); Safari Club Int’l v. Jewell, 47 F. Supp. 3d 29, 36 (D.D.C. 2014) (“[T]he general rule in this Circuit is that economic harm does not constitute irreparable injury . . . the fact that economic losses may be unrecoverable does not, in and of itself, compel a finding of irreparable harm.” (internal quotation marks omitted))).}

The third and fourth factors, which merge when the government is the defendant,\footnote{Nken, 556 U.S. at 435.} also favor the agency. The cases in which granting a preliminary injunction most damaging for the government and the general public—i.e., those where the lack of compliance with the order would cause an important environmental or health impact—present a balance between public and private interest in favor of the private interest. This supports a denial of the preliminary injunction, especially in light of the Supreme Court’s view that environmental harm—i.e., what the court would balance against the economic injury that the plaintiff would allege—can rarely be remedied with monetary damages and tends to be irreparable.\footnote{Amoco Prod. Co. v. Vill. of Gambell, 480 U.S. 531, 545 (1987).}

In short, this analysis of the factors that guide whether the court will grant a preliminary injunction shows that such an outcome is unlikely in cases involving adequate orders. As a result, the probability that these challenges would directly cause delays in the enforcement process is also low. Moreover, given the balancing of harms that the test demands, it is even less likely that a preliminary injunction would be granted in cases in which doing so could result in serious harm to the environment or human health.

\textit{ii. Indirect Sources of Delay.} The second delay theory, which is based on the \textit{indirect} effects of litigation, is more persuasive, but it is very limited in scope. The claim is that litigation tends to divert agency resources, thereby reducing the agency’s capacity to attend to enforcement activities.\footnote{See Richard L. Revesz, \textit{Federalism and Environmental Regulation: A Public Choice Analysis}, 115 Harv. L. Rev. 553, 598 (2001) (noting that “not spending time in the litigation of enforcement action” may increase the number of sites that are remediated); see also Healy, supra note 12, at 525–26 (explaining these indirect sources of delay); Levy, supra note 34, at 1061.}

While there may be some noticeable effects of litigation on how swiftly certain compliance activities are performed, it is important to highlight three points.

First, agency resources—at least, the time and work of some government employees—will be expended every time that an agency contests or is otherwise involved in a challenge of an administrative action. This fact, by itself, does not automatically justify limiting judicial review or explain why

\begin{references}
\begin{itemize}
\item \footnote{Nken, 556 U.S. at 435.}
\item \footnote{Amoco Prod. Co. v. Vill. of Gambell, 480 U.S. 531, 545 (1987).}
\item \footnote{See Richard L. Revesz, \textit{Federalism and Environmental Regulation: A Public Choice Analysis}, 115 Harv. L. Rev. 553, 598 (2001) (noting that “not spending time in the litigation of enforcement action” may increase the number of sites that are remediated); see also Healy, supra note 12, at 525–26 (explaining these indirect sources of delay); Levy, supra note 34, at 1061.}
\end{itemize}
\end{references}
judicial review should be delayed in some instances—e.g., in environmental cases—but not in others. Second, these claims are documented with respect to one statute. In particular, these types of delays seem to arise in the context of CERCLA and only when the agency has to develop a factual record to aid the court in ruling on the claim that a particular party should not be held liable for the cleanup costs associated with a contaminated site. The need to spend resources on a factual record, however, will also depend on whether the defense raised is essentially a legal one or one that requires a thorough analysis of the industrial activities carried out by the potentially responsible party. Thus, while the argument may be valid in some cases, it is only persuasive under a very limited set of circumstances. Thus, it cannot be reasonably used to support a broader claim that substantial enforcement delays are an inherent feature of immediate judicial review.

Third, even in the CERCLA context, part of the delay seems to result from the agency’s reluctance to move forward with the cleanup process when a court is examining a challenge, even in the absence of a preliminary injunction. This can occur when the agency perceives that not continuing with the cleanup action will be strategically advantageous in court. This type of delay could be avoided by the agency and, therefore, does not provide an adequate basis to support the need to bar immediate challenges either.

2. Allowing Immediate Judicial Review and the Alleged Flood of Challenges

The other argument often offered in support of delayed judicial review is that, with immediate review, the agency would “quite possibl[y] face[ ] judicial review of every administrative order it issues.” The idea behind this view is that, because the order generally requires its recipient to act in a way that will be economically costly, a simple cost-benefit analysis would always lead the operator to challenge the order regardless of its merit in order to avoid these compliance costs.

235 In fact, Professor Healy, who lays out this argument in his scholarship, does not state that this claim is valid in other contexts. See Healy, supra note 12, at 325–26; see also Healy, supra note 99, at 12 (explaining the different types of settings under which potentially responsible parties under CERCLA may challenge their liability in court.).

236 For example, the question of a party’s ownership of a facility is essentially a legal question that can affect whether that party is potentially liable under the statute or not. See 42 U.S.C. § 9607(a)(1).

237 Healy, supra note 12, at 326.

238 See id.

239 Quiggle, supra note 12, at 328; see also Wynn, supra note 35, at 1901 (noting that the majority view is that immediate review prevents floods of challenges).

240 See Quiggle, supra note 12, at 363–64. It is important to note that there is no empirical support for this claim, as some scholars have deemed necessary for those making similar “flood of challenges” arguments. See Levy, supra note 35, at 1073.
The problem with this argument is that it overlooks the costs that recipients of orders face if they decide to bring a challenge. Under the traditional economic model for litigation, parties make their decision on whether to bring a court action taking into account the “expected value” of their claim and its potential costs. To do so, the potential plaintiff will ponder the following key factors: the likelihood of success of the claim (P), the gain in the event of success (G), and the costs associated with litigating the case (C). This process involves “attaching probabilities and payoffs” to the different options. In the typical case involving the challenge of an order, cost is the key issue—i.e., costs incurred with the challenge and compliance costs potentially avoided if the challenge is successful. The following table lays out the different possible scenarios:

<table>
<thead>
<tr>
<th>Cost of Bringing the Action</th>
<th>Cost of Compliance</th>
<th>Fines for Non-Compliance</th>
<th>Gain from Delayed Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Challenge</td>
<td>N/A</td>
<td>Full Cost</td>
<td>N/A</td>
</tr>
<tr>
<td>Challenge</td>
<td>Full Cost</td>
<td>(1 – P) x Cost</td>
<td>(1 – Pi) x Cost Pi x Cost</td>
</tr>
</tbody>
</table>

Under a “no challenge” scenario, the recipient of the order must pay the full cost of compliance. If the plaintiff brings a challenge, the costs associated with that decision will vary depending on whether the plaintiff wins or loses in court and also be based on the court’s decision on whether to grant a preliminary injunction. “P” represents the likelihood of prevailing in the challenge and “Pi” is the likelihood of obtaining a preliminary injunction. Under this “challenge” scenario, there is always a cost to bringing the action, but the evaluation of the other costs and gains will depend on P and Pi. How likely the recipient of the order is to have to ultimately pay compliance costs depends on the probability of prevailing in the suit (P), whereas the likelihood of having to pay fines for non-compliance that accrue while the court is examining the merits of the case depends on the probability of obtaining a preliminary injunction (Pi). Similarly, the savings from not having to comply with the order until after the date of the court’s ruling are tied to the likelihood of securing a preliminary injunction (Pi).

244 These costs include legal fees and experts, and also reputational costs and the time dedicated to the lawsuit. See Robbennolt, *supra* note 241, at 2.
245 Where (1 – P) is the likelihood of losing and (1 – Pi) is the likelihood of not being granted the preliminary injunction.
246 This, of course, only applies to a situation in which the court grants a preliminary injunction and dismisses the challenge to the order.
What this model shows is that the decision of whether to challenge an order may point in different directions depending on the factors outlined above. In other words, the cost-benefit analysis does not always justify a challenge. When the order is perceived to be proper and, thus, likely to be upheld, its recipient has a higher probability of having to pay the cost of compliance and the full cost of bringing the claim. For the reasons explained above, the likelihood of obtaining a preliminary injunction in these cases is also low. If the plaintiff does not obtain a preliminary injunction, continuing with the challenge without complying with the order will also create a serious risk of having to pay substantial daily fines if the court upholds the order. Therefore, these actions would be brought primarily when the costs of compliance are extremely high, or when the order is perceived to be so flawed that the chances of a successful challenge are high.

The notion that introducing immediate review would not cause recipients of compliance orders to systematically challenge them in court has also been confirmed in practice with the Clean Water Act. Before 2012, courts only permitted delayed review of compliance orders issued under the Clean Water Act. This changed in 2012 as a result of the Supreme Court decision in Sackett. More than five years later, attorneys have not observed “the flood of litigation that many expected.” On the contrary, the number of challenges of compliance orders under the Clean Water Act has been very low.

B. Evaluating The Convenience Of Delayed Review In Specific Cases

The purpose of the previous analysis was to challenge the notion that the arguments about delay and unmanageable amounts of litigation can support delayed review in all circumstances. The conclusion that delayed review is not necessarily the superior alternative in many instances begs the question: how can we evaluate, from a public policy standpoint, whether delayed review is better than immediate review? Part III.B provides a basic frame-

---

247 McMahon, supra note 44, at 1384–85 (explaining how the expenses associated with litigation can be higher than the advantage conferred by winning in court and how this operates as a disincentive to sue).

248 See supra Part III.A.1.a. As also explained above, the assumption is that those making these arguments in support of delayed review are primarily concerned about challenges to perfectly appropriate orders, which courts would be very likely to ultimately uphold.

249 As the D.C. Circuit pointed out, many recipients of these orders comply with what the agency requires from them because they believe that the order is “accurate and would withstand judicial review.” Gen. Elec. Co. v. Jackson, 610 F.3d 110, 128 (D.C. Cir. 2010).


252 Even a broad search on Westlaw for cases decided by U.S. district courts after Sackett including the terms “compliance order” and “Clean Water Act” (which includes cases in which administrative orders were not being challenged at all) yields an average of ten cases per year in the entire United States.
Delayed Judicial Review

work to make this type of determination. This discussion examines a set of environmental statutes that, as explained above, share a similar enforcement structure: the Clean Air Act, the Clean Water Act, and RCRA.\textsuperscript{253}

The first step in the analysis is to divide the universe of situations in which compliance orders may be issued into four categories. The first two categories differentiate between orders that are appropriate—defined as those that would be expected to withstand judicial review—and those that are deficient—i.e., those likely to be set aside by a court. The other distinction focuses on whether the recipient of the order would be willing to challenge it in court if immediate judicial review were available. The following matrix shows the four possible scenarios.

<table>
<thead>
<tr>
<th></th>
<th>Would Challenge</th>
<th>Would Not Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appropriate Order</strong></td>
<td>Appropriate WC</td>
<td>Appropriate WNC</td>
</tr>
<tr>
<td><strong>Deficient Order</strong></td>
<td>Deficient WC</td>
<td>Deficient WNC</td>
</tr>
</tbody>
</table>

To evaluate whether delayed review is the preferable enforcement/review option, this discussion will analyze each category and compare two different scenarios: one where there is immediate judicial review with one in which only delayed review is available. It is worth noting that most of the criticisms of immediate judicial review have focused on the first category—appropriate order/plaintiff would challenge.\textsuperscript{254} With respect to the second, third, and fourth categories, the notion that immediate judicial review creates net benefits should be less controversial.

1. **Appropriate Order/Plaintiff Would Challenge**

This category includes compliance orders that courts would uphold but that their recipients would challenge, given the opportunity. Before revisiting the arguments in favor or against allowing immediate review of these types of orders, it is important to examine the different scenarios that this category deals with. The following table synthesizes the different expected outcomes with delayed and immediate review of compliance orders.

\textsuperscript{253} See \textit{supra} Part I.B.

\textsuperscript{254} See \textit{supra} Part III.A.
With Delayed Review | With Immediate Review
---|---
The operator complies with it to avoid risk of penalties. | The operator complies with it to avoid risk of penalties.
The operator complies after the EPA brings a successful enforcement action. | The operator complies after the EPA brings a successful enforcement action.
The operator does not comply with the order and the EPA does not bring an enforcement action. | The operator does not comply with the order and the EPA does not bring an enforcement action.

The operator complies after the EPA brings a successful enforcement action.

As the table shows, introducing the possibility of immediately challenging a compliance order would only add one additional scenario to this category of orders: the recipient of the order complies with it after an unsuccessful challenge in court. Those who oppose immediate judicial review of compliance orders focus on this particular scenario and claim that this form of judicial review would (i) lead to delays in the enforcement of environmental laws and regulations that could result in harm to human health and the environment, and (ii) cause courts to be flooded with challenges of compliance orders. As the discussion above explained, none of these arguments is persuasive.

With regard to the argument that there would be substantial delays in enforcement, a mere challenge without a preliminary injunction would not render the order unenforceable, and the penalties for non-compliance with it would continue to accrue. These penalties only contribute to incentivizing compliance. Moreover, applying the current test for preliminary injunctions, the likelihood that a plaintiff would be able to obtain this form of temporary relief in the case of appropriate orders is low, and especially of those orders trying to prevent serious harm to human health or the environment.

As for the argument relating to the potential flood of challenges, this is also unlikely to occur. The economic model discussed in detail above suggests that the likelihood of success is an important factor in the decision of whether to challenge an order or not. With appropriate orders, the

---

255 Quiggle, supra note 12, at 329 (explaining that plaintiffs who are permitted to seek pre-enforcement review of compliance orders “will likely succeed in both slowing down the enforcement of environmental laws and delaying the cleanup of catastrophic contamination”); Starr, supra note 34, at 918 (“By permitting pre-enforcement judicial review, formal adjudication in the court of appeals hinders the compliance order’s expediency.”).

256 See Wynn, supra note 35, at 1901 (noting that the majority view is that immediate review prevents floods of challenges); see also Quiggle, supra note 12, at 328.

257 See supra Part III.A.1.

258 See supra Part III.A.1.

259 See supra Part III.A.1.

260 See supra Part III.A.2.
probability of success is low by definition. Thus, unless the cost of compliance is very high, the assumption that plaintiffs would systematically challenge compliance orders is unconvincing. Further, the experience with the Clean Water Act confirms that immediate review does not necessarily lead to a significant increase in litigation.

2. Appropriate Order/Plaintiff Would Not Challenge

This is the most desirable category because orders are appropriate, and there is no unnecessary litigation. Introducing immediate review would not have a negative impact on this subset of orders, given that it would provide their recipients with an option—challenging the order—that they are not interested in using.

However, there would be global positive effects associated with immediate judicial review: it would increase the share that this category represents in the total number of orders. Scholars have noted that more extensive judicial review creates a powerful incentive for the executive to avoid acting in an irrational or lawless fashion, increasing the quality of the final product—in this case, of administrative orders. Therefore, allowing immediate review of orders should increase the overall share of appropriate orders.

3. Deficient Order/Plaintiff Would Not Challenge

There are at least two scenarios in which, if immediate review were available, a deficient order would remain unchallenged. First, if an operator realizes that the cost of challenging the order would be higher than that of complying with it, it may choose to not challenge the order. The cost of complying with the order, however, could still be significant. Therefore, reducing the number of cases in which this occurs would still be important from both a fairness and an efficiency standpoint. While allowing immediate review would not translate into an increase in judicial review of orders in the ‘deficient/would not challenge’ category, this form of review may still contribute, as noted above, to an overall increase in the quality of orders. This, in turn, would make this undesirable category of orders smaller.

Second, the recipient of a compliance order may comply voluntarily with it because the order is very lenient and potentially more advantageous to the owner/operator than the statute or regulations allow for. This has been an issue, for example, in the context of CERCLA/RCRA cleanup actions,

261 Of course, there could still be a small number of cases in which a plaintiff is able to bring a colorable challenge of an appropriate order.
262 See supra Part III.A.2.
263 See supra Part III.A.2.
265 See supra Part III.B.2. (explaining how more extensive judicial review is perceived to lead to orders of a higher quality).
where the type/level of cleanup that the order directs the recipient to undertake/reach does not meet appropriate standards, or where the time provided to complete the cleanup is unreasonably long. In these cases, permitting immediate judicial review would provide citizen groups the opportunity to challenge these orders before the cleanup is finalized.

4. Deficient Order/Plaintiff Would Challenge

This is the most problematic situation and the one that could be dramatically reduced with immediate review. Allowing immediate judicial review of compliance orders would solve the problem of deficient orders that have an economic impact on their recipients that is sufficiently great to justify their challenge in court, but for which, either by statute or judicial interpretation, prompt judicial review is not available. As explained earlier, the option of seeking injunctive relief is critical to this type of challenge—so that the potential penalties do not accumulate. However, unlike with the first category—i.e., appropriate orders that their recipient would challenge—the likelihood of success in the case of a deficient order would be higher, which would make the issuance of the preliminary injunction more likely. Moreover, with deficient orders, the assumption that their enforcement would help prevent damage to the environment or human health does not necessarily hold true.

5. Summing Up: Overall Appropriateness of Immediate Review of Environmental Compliance Orders

There are several conclusions to extract from the preceding analysis. First, criticisms of immediate review of environmental compliance orders apply to only one of the four categories of orders: i.e., the first one—appropriate orders that would be challenged by their recipients. Second, the arguments against immediate review of these types of orders—based on the purported dramatic increase in the number of challenges and delays in enforcement—are not very persuasive. Third, it can hardly be controversial that, under categories two, three, and four, immediate review would be beneficial or, at least, not harmful. Last, the economic analysis described in Part

266 See, e.g., Petition Alleging Violations Of The Human Rights Of Various Residents Of Vieques, Puerto Rico By The United States Of America at 33 (Inter-Am. Comm’n H.R. Sept. 23, 2013), http://www.nlginternational.org/report/Vieques_IACHR_petition-FINAL.pdf [https://perma.cc/6GUM-62FG] (claiming that the polluter “engaged in substandard cleanup practices, such as the open-air detonation of munitions and open burning of vegetation, which the Environmental Protection Agency determined to be hazardous to the environment.”).

267 See supra Part III.A.1.a.

268 A classic example, in the worker safety context, that this assumption may not be accurate, especially when dealing with a deficient order, can be found in Ne. Erectors v. Sec’y of Labor, 62 F.3d 37 (1st Cir. 1995). There, an employer was trying to obtain immediate review of administrative action that had the effect of compromising workers’ safety. See Buchman, supra note 89, at 201–03.
2019]  

Delayed Judicial Review  

239

III.A suggests that the first category of orders does not, as defenders of delayed review would argue, include the vast majority of the orders that agencies issue.\(^{269}\) One of the features of this category is that the recipient of the order would challenge it. However, the economic model shows that there can be many scenarios under which these potential plaintiffs may be unwilling to challenge the orders because there is an economic incentive to not do so unless the costs of compliance are very high.\(^{270}\)

**IV. CONSTITUTIONAL CONCERNS WITH DELAYED JUDICIAL REVIEW**

Most of the statutes described in part II of this Article are silent on whether immediate judicial review of orders is permitted.\(^{271}\) This has led some courts to conclude that Congress nonetheless intended to foreclose such review.\(^{272}\) With respect to CERCLA, courts have been virtually unanimous in reaching the same conclusion in light of the statute’s explicit bar on immediate review of cleanup orders.\(^{273}\) Because the majority of these cases have focused on the doctrines of preclusion and finality, the question of whether Congress is constitutionally permitted to limit judicial review in this manner is still not conclusively answered.

The extent to which Congress may limit judicial review of agency action generally has been amply debated. Some authors see judicial review of agency’s orders and regulations as a basic principle of the American legal system that can be traced back to *Marbury v. Madison*.\(^{274}\) Others deny the notion that there is a general constitutional right to judicial review of agency action.\(^{275}\) These two positions represent the two sides of a spectrum: one that would require judicial review in all cases and another one that would be comfortable with denying it completely in certain instances.

Delayed judicial review falls somewhere between these two extreme positions: it is placing an important limitation on judicial review, but it is not eliminating it altogether. Some could claim that leaving open the possibility of review under certain conditions—which delayed review generally does—complies with any constitutional requirement of access to courts. Others

\(^{269}\) The argument relating to the “flood of challenges” is premised on the notion that most recipients of orders would challenge them. The economic argument laid out above questions that assumption. The consequence of this is that the first category of orders—in which the recipient would challenge the order—would not integrate the vast majority of appropriate orders. In other words, many appropriate orders would fall into the “appropriate/recipient would not challenge” category. In this same vein, see Gen. Elec. Co. v. Jackson, 610 F.3d 110, 128 (D.C. Cir. 2010) (explaining that many recipients of these orders comply with what the agency requires from them because they believe that the order is “accurate and would withstand judicial review”).

\(^{270}\) Id.

\(^{271}\) See supra Part II.B.1, 2, 3, 5.

\(^{272}\) See supra Part II.B.2, 3, 5.

\(^{273}\) See supra Part II.B.4.

\(^{274}\) Berger, supra note 3, at 89 n.180.

\(^{275}\) Bagley, supra note 4, at 1309.
would counter, however, that those who are coerced to comply with an administrative order under the threat of increasing daily penalties are seeing their right to judicial review impermissibly curtailed.276

To address this issue, Part IV starts by examining whether there is a general constitutional right to obtain judicial review of agency action. The second part of the analysis focuses on a more limited theory of judicial review—the doctrine of constitutionally intolerable choices—and provides a comprehensive framework to evaluate the constitutionality of different delayed review provisions.

A. Is There A Constitutional Right To Judicial Review?

If delayed judicial review is, in fact, problematic from a constitutional standpoint, there is a threshold question that must be addressed first. And that is whether the Constitution expressly or impliedly recognizes a general right to challenge administrative action in court. Given the constitutional dimension of this hypothetical right, it would include not only situations where the statute is silent—i.e., implied preclusion—but also those in which Congress has explicitly barred judicial review of a certain form of agency action—i.e., express preclusion.277

Scholars have been debating, and disagreeing on, this issue for a long time.278 However, these conversations in the literature have, at least, contributed to delineating the potential sources of such a constitutional requirement. Two doctrines that scholars have frequently cited as constitutional hooks to justify the existence of a universal right to judicial review are separation of powers and the non-delegation doctrine.279

276 See Ex parte Young, 209 U.S. 123, 148 (1908); John Harrison, Ex Parte Young, 60 Stan. L. Rev. 989, 991–93 (2008) (explaining the legal framework in Ex parte Young and how the Court concluded that it was unconstitutional because, in practice, it inhibited judicial review).

277 For a discussion of the difference between these two forms of preclusion and their treatment by courts, see supra Part II.A.1.

278 See, for example, the back and forth between Professors Raoul Berger and Kenneth Culp Davis that started in the 1960s. Berger, supra note 3, at 57–58 (noting that the right to judicial review, as that to be protected from arbitrariness, is mandated by the Constitution); Kenneth Culp Davis, Administrative Arbitrariness—A Postscript, 114 U. Pa. L. Rev. 823, 831 (1966) [hereinafter Davis, Postscript] (explaining that “Mr. Berger’s position is that all discretion must be reviewable for arbitrariness or abuse, and that position is not at all supported by constitutional doctrine.”). Further see Berger, Sequel, supra note 47, at 603–04; Kenneth Culp Davis, Administrative Arbitrariness Is Not Always Reviewable, 51 Minn. L. Rev. 643, 646 (1967) [hereinafter Davis, Reviewable] (questioning whether Prof. Berger’s conclusion follows from the cases he cites); see also Berger, Synthesis, supra note 46, at 980–81 (listing—in response to Professor Davis’s claim—the cases on which he supports his position that judicial review is constitutionally required).

279 Berger, supra note 3, at 88–89 (stating that one of the justifications for the availability of judicial review “as of right” has, as one of its sources, the constitutional restrictions on delegation of powers); Congressional Preclusion of Judicial Review of Federal Benefit Disbursement: Reasserting Separation of Powers, 97 Harv. L. Rev. 778, 785–87 (1984) (tying
2019] Delayed Judicial Review 241

the scholarship on this matter is often unclear when tying arguments in support of universal judicial review to specific constitutional doctrines.280

Supporters of the notion that there is a universal right to judicial review often view it as a necessary structural element of our legal system. The argument starts with the premise that the Constitution does not allow the government to exercise its power arbitrarily, which is a notion the Supreme Court recognized and later codified in the Administrative Procedure Act.281 With that understanding in mind—which has generally not been questioned by the critics of this theory282—some scholars have then argued that, if these prohibitions are to have any effect, courts must have the ability to review acts of the executive in order to determine whether the government has acted in compliance with the law and the Constitution.283

Some scholars have questioned the validity of this conclusion vehemently.284 One common criticism has been that the notion that judicial review must be generally available does not follow from the fact that the Constitution proscribes arbitrary action by the executive.285 In other words, an agency may be prohibited from acting arbitrarily, but this does not necessarily mean that courts have authority to review these actions in the absence of a congressional grant of jurisdiction.286

Be that as it may, most scholars have been able to agree on one thing: the extent to which the Constitution requires judicial review of agency action is uncertain. One scholar explained in the 1970s that the question of whether there is a right to judicial review mandated by the Constitution is “unresolved and obscure.”287 More recently, another scholar has stated that de-

280 See Jaffe, supra note 3, at 420 (stating that “in our system of remedies,” an individual has the right to secure judicial review, but without specifying the particular constitutional doctrine that supports this conclusion).

281 See Berger, supra note 3, at 57.

282 Professor Kenneth Culp Davis is very critical of the idea that judicial review is always available but agrees that the Constitution does not permit arbitrary behavior on the part of the executive. See Davis, Reviewable, supra note 278, at 646 (“Mr. Berger presents an impressive collection of Supreme Court statements that our Constitution and our institutions leave no place for arbitrary exercise of power. I agree with those statements.”).

283 See Jaffe, supra note 3, at 403 (“[T]here is in our society a profound, tradition-taught reliance on the courts as the ultimate guardian and assurance of the limits set upon executive power by the constitutions and legislatures.”); see also Berger, Synthesis, supra note 47, at 980–81 (explaining that it would not be reasonable for the Supreme Court to state that the executive must not act arbitrarily if there was no way of enforcing this prohibition).

284 See Bagley, supra note 4, at 1314 n.166 (arguing that the claim that the Constitution requires judicial review of arbitrariness “finds little to recommend it in either historical or modern practice.”); Davis, Reviewable, supra note 278 (noting that arbitrariness and abuse are not always reviewable).

285 Davis, Reviewable, supra note 278, at 646 (noting that it is a “mistake[ ] is to equate lack of authority to act arbitrarily with judicial reviewability”).

286 See id.

terminating how much the Constitution limits Congress’s authority to limit the jurisdiction of federal courts is “maddeningly hard to say.”

The main reason why this question eludes resolution is that the Supreme Court has not provided a clear framework delineating Congress’s power to restrict the jurisdiction of federal courts. The Court has never explicitly stated that there is a general right to judicial review of administrative action. However, it has expressed that a categorical bar on judicial review can raise serious constitutional concerns.

Regardless of where one falls on this debate, the uncertainty surrounding this doctrine of ‘universal judicial review’ makes the doctrine unfit to answer the question of whether it is constitutionally proper for Congress to enact delayed review provisions. The next subsections explore a narrower doctrine that can be useful in determining the constitutional boundaries of legal provisions placing limitations on immediate judicial review of agency action.

**B. The Doctrine Of Constitutionally Intolerable Choices**

Even in the absence of a general right to obtain judicial review of agency action, there is one constitutional doctrine that restricts the types of delayed judicial review frameworks that Congress may enact. The so-called doctrine of constitutionally intolerable choices limits Congress’s ability to draft judicial review provisions in a way that allows an agency to coerce the recipients of its orders or regulations into compliance while also making access to courts excessively burdensome. The scenario in which this typically occurs is one where the recipient of a compliance order (i) can only challenge its validity after failing to comply with and (ii) the statute provides very severe penalties for the violation of the order.

---

288 Bagley, supra note 4, at 1310.

289 See Bruff, supra note 4, at 17 (noting that “[t]he Court was avoiding [the] lurking constitutional issue.”); Congressional Preclusion of Judicial Review of Federal Benefit Disbursement: Reasserting Separation of Powers, supra note 279, at 791 (pointing out that courts have been dodging the constitutional question).

290 See Johnson v. Robison, 415 U.S. 361, 366–67 (1974) (explaining that a statutory bar that bars courts from examining the constitutionality of the legislation would be constitutionally troublesome); Congressional Preclusion of Judicial Review of Federal Benefit Disbursement: Reasserting Separation of Powers, supra note 279, at 791 (indicating that “in these cases, [courts’] apprehensions have been evident”).

291 See Thunder Basin Coal Co. v. Reich, 510 U.S. 200, 218 (1994) (describing the situation in which a “constitutionally intolerable choice” exists as one where the recipient of the order can be coerced into compliance by severe penalties).

292 See, e.g., Reisman v. Caplin, 375 U.S. 445, 447 n.7, 449 (1964) (noting that the addressee of the summons—i.e., the witness—may make “objections at the hearing before the court,” which occurs after the witness refuses to comply with the summons and the Secretary brings an action before a judge or United States commissioner); Okla. Operating Co. v. Love, 252 U.S. 331, 336 (1920) (pointing out that the only possible judicial review is one in the context of “proceedings to punish for contempt”); Ex parte Young, 209 U.S. 123, 145–46 (1908) (explaining that the validity of the orders at issue can only be tested if they are disobeyed first).
These situations can raise different concerns that can be tied to a variety of more general constitutional principles. Courts have often looked to due process as one of the constitutional hooks for the doctrine of constitutionally intolerable choices.\textsuperscript{293} Another possibility is to tie this doctrine to the principle of separation of powers, given that it prevents the legislative branch from insulating certain acts of the executive from the scrutiny of the courts. Some courts have even relied on other constitutional requirements such as the right to equal protection of the laws.\textsuperscript{294}

As mentioned earlier, however, it is important to differentiate the doctrine of constitutionally intolerable choices from the requirement of procedural due process.\textsuperscript{295} Procedural due process generally addresses the need for a hearing, and it is satisfied even if it takes place at the administrative level.\textsuperscript{296} The doctrine of constitutionally intolerable choices, on the other hand, is mostly concerned with access to courts.\textsuperscript{297} For this reason, providing further opportunity to be heard at the administrative level can resolve procedural due process issues. Violations of the doctrine of constitutionally intolerable choices, on the other hand, can only be addressed by reducing the coerciveness of the agency action\textsuperscript{298} or by granting the potential plaintiff with swift access to court.\textsuperscript{299}

One of the first Supreme Court decisions that articulated this doctrine—at the beginning of the 20th century—was \textit{Ex parte Young}\textsuperscript{300}. While this case is generally known for the Court’s holding on sovereign immunity and federalism,\textsuperscript{301} it is also a landmark case on access-to-courts issues.\textsuperscript{302} Later cases have helped clarify some of the questions that \textit{Ex parte Young} did not fully

\textsuperscript{293} See, e.g., Wadley Southern Co. v. Georgia, 235 U.S. 651, 661 (1914) (“any party affected . . . is entitled, by the due process clause, to a judicial review); Thunder Basin Coal Co., 510 U.S. at 218 (examining the issue under Part IV of the opinion, which responds to plaintiff’s due process claim).

\textsuperscript{294} See supra Part IIA.3.

\textsuperscript{295} See, e.g., Wadley Southern Co. v. Georgia, 235 U.S. 651, 661 (1914) (“any party affected . . . is entitled, by the due process clause, to a judicial review); Thunder Basin Coal Co., 510 U.S. at 218 (examining the issue under Part IV of the opinion, which responds to plaintiff’s due process claim).

\textsuperscript{296} See supra Part II.A.3.

\textsuperscript{297} See Mathews v. Eldridge, 424 U.S. 319, 349 (1976) (concluding that the administrative procedures in question were sufficient to comport with the requirement of due process).

\textsuperscript{298} See, e.g., Reisman, 375 U.S. at 447 n.7, 449; Okla. Operating Co., 252 U.S. at 336; Ex parte Young, 209 U.S. at 145–46.

\textsuperscript{299} See infra Parts IV.B.1, 2, 3.

\textsuperscript{300} Ex parte Young, 209 U.S. 123, 145 (1908). For an explanation of the basic facts underlying that case, as well as the multiple legal proceedings involved, see Harrison, supra note 276, at 991–94.

\textsuperscript{301} See, e.g., Note, Congressional Intent to Preclude Equitable Relief — Ex Parte Young After Armstrong, 131 HARV. L. REV. 828, 844 (2008) (“[T]he Ex parte Young doctrine . . . has focused on federalism concerns about federal courts’ interference with state action.”); Vicki C. Jackson, Seminole Tribe, the Eleventh Amendment, and the Potential Evisceration of Ex Parte Young, 72 N.Y.U. L. REV. 495, 510–11 (1997) (pointing out that the principle that “suits against state officers are not regarded as suits against the State for purposes of the Eleventh Amendment” is commonly referred to as the “Ex Parte Young” doctrine”).

\textsuperscript{302} See Harrison, supra note 276, at 993 (explaining that the Court “concluded that [the statutory penalties] were inconsistent with the requirements of due process because of the burden they placed on access to the courts”).
address.\textsuperscript{303} None of these court decisions, however, offer a complete picture of the doctrine of constitutionally intolerable choices and its different moving parts. The following subsections address this problem by proposing a three-factor framework to assess the constitutionality of delayed judicial review provisions.

1. Severity of the Punishment

As noted above, coercion is one of the central elements of the doctrine of constitutionally intolerable choices. If recipients of an order are coerced to comply with it to avoid very severe penalties and doing so leads them to lose the right to seek judicial review, there is a high likelihood that the statute is unconstitutional. Therefore, the severity of the punishment for violating the order is a key element in this analysis.\textsuperscript{304} This first factor focuses on the punishment that the statute authorizes. In contrast, the next subsection will examine legal mechanisms that may temper the actual penalties that a court would be likely to impose.

\textit{Ex parte Young} provides an example of when the punishment is so great that it could easily deter potential plaintiffs from resorting to courts. In that case, the punishment for failing to comply with the orders issued by the Railroad and Warehouse Commission was particularly severe: daily fines of up to $5000 (over $130,000 in current dollars)\textsuperscript{305} and imprisonment for a period not exceeding five years.\textsuperscript{306} Moreover, courts had interpreted the statute not to allow challenges to these orders unless the plaintiff had violated them first and the state had brought an enforcement action.\textsuperscript{307} The Supreme Court concluded that this framework was unconstitutional.\textsuperscript{308}

This case raises two questions with regard to the severity of the punishment that is required to violate this doctrine. The first one is whether a statute that does not include imprisonment as a potential punishment for those who violate the orders could still be deemed coercive under the doctrine of constitutionally intolerable choices. This question should be answered in the affirmative. Monetary fines can generate a very powerful incentive to comply with what an agency demands.\textsuperscript{309} In extreme cases, for example, severe fines may threaten the continuation of a business.


\textsuperscript{304} See \textit{Ex parte Young}, 209 U.S. at 147 (explaining that when fines are enormous, they can intimidate the potential plaintiff from taking the issue to court).


\textsuperscript{306} \textit{Ex parte Young}, 209 U.S. at 145.

\textsuperscript{307} Id. at 145–47.

\textsuperscript{308} Id. at 148.

\textsuperscript{309} There are various examples of decisions in which the Supreme Court has examined, under the doctrine of constitutionally intolerable choices, statutes that did not contemplate penalties of imprisonment, without concluding that this circumstance made them automatically compliant with this constitutional doctrine. See Thunder Basin Coal Co v. Reich, 510 U.S.
Delayed Judicial Review

The second question relates to the amount of the fines. When the statute authorizes the imposition of a very high one-time fine, the deterrent effect to forego judicial review can be substantial. That is what occurred in Ex parte Young, where the total amount of the fine could have reached approximately $134,000 in today’s dollars.310 However, lower daily penalties, if allowed to accumulate, can have an even more powerful coercive effect. This was illustrated in Okla. Operating Co. v. Love, where an order of the Oklahoma Corporation Commission prohibited the plaintiff from requesting payments for laundry services that exceeded certain maximum rates.311 As in Ex parte Young, the orders were initially treated as not subject to immediate review.312 The individual fines that the statute allowed for the failure to comply with the order were ten times lower than in Ex parte Young, but the Court did not hesitate to find them unconstitutionally coercive because “each day’s continuance of the refusal after service of the orders [was] ‘a separate offense.’”313

Last, it is important to note that the doctrine of constitutionally intolerable choices does not apply to those who are not recipients of compliance orders. While these potential plaintiffs may wish to bring a challenge, the agency has not directed its request at them, and thus, they cannot be fined for failing to comply with the order.314 Therefore, when considering a third-party plaintiff, there is no coercion of the type that the doctrine of constitutionally intolerable choices intended to prevent.315

2. Is There a Reasonable and Predictable Possibility of Not Being Punished if a Challenge is Unsuccessful?

While the previous factor looked at the maximum punishment authorized by the statute, this factor examines something different. It focuses on

---

200, 218 (1994) (focusing its rejection of the claim on the availability, for the plaintiff, of temporary relief that would reduce the amount of the fines, without mentioning the fact that the statute did not authorize imprisonment of non-compliant parties); Okla. Operating Co. v. Love, 252 U.S. 331, 336 (1920) (ruling for plaintiff in a case involving a statute that authorized monetary fines only); Wadley Southern Co. v. Georgia, 235 U.S. 651, 659 (1914) (examining the claim that the statute was for the imposition of high monetary fines, without reference to the fact that imprisonment was not authorized in the statute).

310 Ex parte Young, 209 U.S. at 145.
312 Id. at 336.
314 As explained above, this can arise under CERCLA when a citizen group is dissatisfied with a cleanup order directed to a potentially responsible party. See supra Part II.B.4.
315 This idea echoes the Supreme Court’s inclination, in the context of some implied preclusion claims, to interpret statutory bars to judicial review literally when the plaintiff is deemed a third party. See, e.g., Block v. Cnty. Nutrition Inst., 467 U.S. 340, 341–48 (1984) (finding implied preclusion where the plaintiffs were consumers and non-profit organizations challenging an order that set minimum milk prices between handlers of dairy products and producers).
other features of the statute—or its judicial interpretation—that may provide some reassurance to potential plaintiffs that they will not be severely punished if they decide to violate the order with the goal of challenging it, but ultimately lose in court. In other words, this subsection analyzes whether there are any legal mechanisms that may reduce the coercive effect of the threatened penalties. These mechanisms typically adopt one of two forms: the so-called “good faith” defense or the temporary relief approach.

Under a good faith defense, the plaintiff may fail to comply with the order and question its validity in court without risking serious penalties, as long as the violation of the order was in good faith, i.e., reasonably believing that there was a valid defense justifying non-compliance. While this approach can provide some peace of mind to some plaintiffs and is therefore preferable to a statute that does not contemplate this exception, there are two reasons why it is not as effective in preventing coercion as a statute that allows for immediate review to begin with.

First, it may be difficult, in close cases, to determine the types of arguments that the court may deem “reasonable.” One way to address this problem is to put the burden of showing the plaintiff’s unreasonableness on the agency, unless there are detailed regulations that can assist the plaintiff in making this determination. Second, a statute may be ambiguous with respect to whether a good faith defense is available.

It is important to distinguish the good faith defense from cases in which a statute merely confers discretion to the court to determine the amount of the fines or whether they should be imposed at all. While a good faith defense can save a statute from violating the doctrine of constitutionally intolerable choices, it does not follow that giving discretion to the court justifies that same result. While some courts have adopted that position, as explained above, the key consideration under the doctrine of constitutionally intolerable choices is the extent to which recipients of orders feel coerced to comply even when they believe that the order is improper. Statutes that


317 For an example of this type of approach, see Solid State Circuits, Inc., 812 F.2d at 392 (discussing that, absent regulations, the burden should be placed on the agency).

318 See, e.g., Reisman, 375 U.S. at 447–49 (examining at length whether the statutes at issue in that case should be interpreted to only punish “contumacious” violations).

319 See Gen. Elec. Co. v. Jackson, 610 F.3d 110, 118 (D.C. Cir. 2010) (citing other decisions for the proposition that “there is no constitutional violation if the imposition of penalties is subject to judicial discretion,” none of which were issued by the Supreme Court); Big Ridge, Inc. v. Fed. Mine Safety and Health Comm’n, 715 F.3d 631, 653–54 (7th Cir. 2013) (listing discretion as one factor that supported the constitutionality of the statute that the court was considering in that particular case, even though the court’s citation to Thunder Basin Coal Co. v. Reich does not support that point).

320 See supra Part IV.B.1.
simply provide that the court “may” impose those penalties, as opposed to indicating that it “shall” do so, leave some uncertainty as to what the punishment will be. This is unlikely to appease the plaintiffs’ concerns and eliminate the coercion they experience to comply with the order.

The other way of reducing the coercive effect of the threatened penalties is to provide some form of temporary relief during an administrative appeal. This applies when a statute allows the imposition of daily penalties for non-compliance with the order but also contains an exhaustion requirement that forces the recipient of the order to go through an administrative appeal process before seeking judicial review. The OSH Act and Mine Safety Act discussed above are an example of this type of framework.321 If the agency grants the temporary relief, the fines will not accumulate during the administrative review period. This approach may help reduce the pressure that recipients of orders experience under some delayed review statutory schemes.322 However, if the criteria to grant or reject the request for temporary relief are not carefully drafted, the statute may leave the door open for the agency to deny these types of requests in order to encourage immediate compliance with the order.

3. Is It Possible to Comply with the Order and Still Challenge It in Court?

The typical case to which the doctrine of constitutionally intolerable choices applies is one where the recipient of the order is able to bring a challenge only after failing to comply with it.323 However, the doctrine has also been invoked in the context of statutes that require those who receive an order to first go through an administrative appeal process before they can access courts. In the latter scenario, the recipients of orders can bring a challenge regardless of whether they comply with the order or not. In these cases, the question then becomes: does the fact that recipients of orders can challenge them in court—albeit not immediately—regardless of whether they comply remove this type of framework from the doctrine of constitutionally intolerable choices? In other words, does the fact that compliance with the order does not prevent a future court challenge make the statute automatically constitutional? There are two main views on whether that should be the case.

Under the first one, if the recipient of the order has the opportunity to comply with the order and then challenge it in court, there is no constitu-

321 Examples of this type of framework can be found in 30 U.S.C. § 815(b)(1)(A) (2012); 29 U.S.C. § 659(b) (2012). For an explanation of how the process works, see supra Part II.B.5.
322 See, e.g., Thunder Basin Coal Co., 510 U.S. at 218 (explaining how temporary relief takes the statute out of the scope of the Ex parte Young doctrine). Other courts have followed suit. See, e.g., Big Ridge, Inc., 715 F.3d at 653 (7th Cir. 2013) (noting that the recipient of the order could “request that the Secretary delay imposing the penalties until further review”).
323 See supra Part IV.B.1.
However, this interpretation of the doctrine leaves an important problem unaddressed. If the agency’s request is not very onerous, or the harm that it may cause to the recipient of the order can be reversed, then the operator may have a relatively safe avenue to seek and obtain judicial review of the order, even if it comes at a later time. The operator may just comply with the order and challenge it later, after the agency has ruled on the appeal.

However, if complying with the order may cause irreparable harm, later judicial review may be futile. This can occur, for example, if the agency requests that an operator share sensitive information or make an important investment to modify a facility. In these types of cases, while there is a theoretical right to judicial review after compliance, it may be useless to the recipient of the order, who has decided to meet the agency’s demands to avoid the risk of penalties—the harm is already done. This situation ends up being very similar to what the Supreme Court was trying to avoid in cases such as *Ex parte Young*: the operator is coerced into compliance by the threat of severe fines, while the role that courts play is minimized. 325

To address this concern, the second and better view on whether these types of frameworks comply with the doctrine of constitutionally intolerable choices is more nuanced. The statute is constitutional as long as compliance with the order does not cause its recipient to suffer irreparable harm. 326 This can be prevented either by avoiding delayed judicial review altogether or by allowing courts to make a preliminary finding on whether immediate judicial review is necessary in that particular case to avoid irreparable harm. The latter option, however, presents certain risks. For example, as has happened under the OSH Act, courts may misinterpret existing precedent and neglect to do an irreparable-harm analysis when recipients of orders bring immediate challenges. 327

In short, statutory provisions that would otherwise be unconstitutional under the previous two factors can avoid that result if they meet two conditions. First, they must allow recipients of orders to bring a challenge even if they comply with the orders, and second, they must still provide an opportunity to obtain immediate judicial review when the harm associated with compliance is serious and irreparable.
The framework laid out in Part IV.B provides a set of tools to assess the constitutionality of enforcement and judicial review provisions. The first two factors—severity of the fines and reasonable expectation of not being punished for mere good faith disobedience—must be present in order for the statute to be unconstitutionally coercive. The third factor—dealing with the possibility of complying with the order and challenging it later—on the other hand, may prevent a statute that meets the first two factors from reaching unconstitutional results. The following discussion uses this framework to examine the constitutional issues raised by several federal statutes.

1. Clean Water Act, Clean Air Act, RCRA, and CERCLA

As explained in more detail earlier,\(^\text{328}\) the enforcement approaches in the Clean Air Act, the Clean Water Act, RCRA, and CERCLA have many similarities. All four statutes allow the EPA to issue compliance orders.\(^\text{329}\) There is, however, a critical difference in how courts have interpreted their judicial review provisions. After Sackett,\(^\text{330}\) the Clean Water Act has been interpreted to allow immediate review of compliance orders, and thus, access to courts is not delayed or otherwise significantly limited.\(^\text{331}\) This interpretation, therefore, has removed the judicial review provisions of the Act from the scope of the doctrine of constitutionally intolerable choices. Because the plaintiff can request a preliminary injunction from the court as soon as it receives the order, there is no coercion to comply built around a delayed access to courts.

In contrast, under some enforcement provisions of the Clean Air Act,\(^\text{332}\) CERCLA, and RCRA, the recipient of the order has more limited options. It is important to note that the analysis below applies to the Clean Air Act and RCRA insofar as they are interpreted to allow delayed judicial review only, as many courts have done.\(^\text{333}\) Under a delayed review approach, the plaintiff has two options after receiving a compliance order. The first option is to comply with it, which also operates as a relinquishment of any future right to judicial review.\(^\text{334}\) Alternatively, the recipient of the order may refuse to

\(^{328}\) See supra Part I.B.1.
\(^{329}\) See, e.g., 42 U.S.C. § 6928(c) (2012) (RCRA); id. §§ 7413(c), 7477 (Clean Air Act); For a more exhaustive list of the provisions providing the EPA authority to issue compliance orders, see Nichols, supra note 29, at 196–200.
\(^{331}\) See supra Part II.B.1.
\(^{332}\) This assumes they are still interpreted, as some circuit courts have done, to bar immediate judicial review. See supra Part II.B.2.
\(^{333}\) See supra Part II.B.2, II.B.3.
\(^{334}\) See supra Part I.B.1.
abide by the order’s terms. In this second case, the EPA has the authority to bring a judicial action requesting the court to enforce the administrative order and/or impose penalties on its recipient. Under certain interpretations of the Clean Air Act, and under RCRA and CERCLA, this is the only avenue that will allow the recipient of the order to obtain judicial review.

The constitutional framework laid out above strongly suggests that some of the enforcement provisions in the Clean Air Act and RCRA present serious constitutional problems if they continue to be interpreted to bar immediate judicial review of orders. CERCLA, on the other hand, has some features that justify a different conclusion. The following discussion analyzes these statutes under the three-factor framework introduced in the previous subsection.

First, the three statutes allow courts to impose fines of tens of thousands of dollars per day on those who fail to comply with an order. These daily penalties accrue while the agency decides whether it will bring an enforcement action in court, but during this time, the recipient of the order may not obtain judicial review. The final dollar amounts of the fines can therefore easily reach six figures. As explained above, the Supreme Court has considered lower daily penalties to be coercive. A fine that today would approximate $6000 was deemed to “obviously” meet the coerciveness standard.

Second, based on the language of the statutes, the recipient of an order issued under certain provisions of the Clean Air Act or RCRA has no reasonable expectation of not being severely punished. Section 3008 of RCRA provides that the violation of a compliance order can lead to the imposition of fines of up to $25,000 per day. This section, however, contains no language suggesting that those violating a compliance order can...
availing themselves of a good faith defense.\footnote{Id. § 6928(b) (RCRA). With orders issued under other provisions of RCRA, however, the conclusion is different. Section 7003(b) authorizes the imposition of penalties to “[a]ny person who willfully violates” an order issued under section 7003(a). Id. § 6973. Consistent with this language, courts have been more receptive to recognizing a good faith defense in the case of § 7003(a) orders. See, e.g., United States v. Valentine, 885 F. Supp. 1506, 1514–16 (D. Wyo. 1995) (deeming it necessary to examine whether the recipient of the order violated the order in good faith).} Similarly, while this type of defense is available under section 113(e) of the Clean Air Act, it only applies to some types of orders.\footnote{42 U.S.C. § 7413(e) (limiting the “sufficient cause” defense to the failure to comply with certain administrative subpoenas and orders under § 7414, but leaving out, for example, orders issued under § 7413).} In contrast, section 106 of CERCLA provides that fines can be imposed on those who “without sufficient cause, willfully” violate a compliance order.\footnote{Id. § 9606(b).} As some courts have recognized, this language creates a good faith defense and therefore appeases the constitutional concerns under the doctrine of constitutionally intolerable choices.\footnote{See Gen. Elec. Co. v. Jackson, 610 F.3d 110, 118–19 (D.C. Cir. 2010) (“CERCLA’s ‘willfulness’ and ‘sufficient cause’ requirements are quite similar to the good faith and reasonable grounds defenses that the Supreme Court has found sufficient to satisfy due process.”); Solid State Circuits v. EPA, 812 F.2d 383, 391 (8th Cir. 1987) (explaining that the language “sufficient cause” meets the constitutional standard); Wagner Seed Co. v. Daggett, 800 F.2d 310, 316 (2d Cir. 1986) (“Assuming the inclusion of the willfulness standard, a good faith defense may be read into § 9606(b).”).} CERCLA avoids this problem altogether by providing a good faith defense.

The last factor focuses on the possibility of complying with an order and then challenging it in court. When the Clean Air Act,\footnote{See supra Part I.B.1., II.B.1.} RCRA, and CERCLA are interpreted to only allow delayed review of orders, compliance operates as a relinquishment of the right to obtain judicial review.\footnote{The Eleventh Circuit concluded that this enforcement framework was unconstitutional in Tennessee Valley Auth. v. Whitman, 336 F.3d 1256, 1258–59 (11th Cir. 2003).} Therefore, this factor does not cure the statutes’ coercion issues identified under the first two factors.

In conclusion, these three factors strongly suggest that some of the enforcement and judicial review provisions in the Clean Air Act and RCRA, if interpreted to allow delayed access to courts only, are unconstitutional.\footnote{29 U.S.C. § 660(a) (2012); 30 U.S.C. § 816(a)(1) (2012).} CERCLA avoids this problem altogether by providing a good faith defense.

2. The OSH Act and Mine Safety Acts

The OSH Act and Mine Safety Act have very similar enforcement and judicial review structures. As explained earlier, most courts have interpreted these statutes to allow only delayed review of administrative orders.\footnote{See supra Part II.B.5.} Once an employer or operator receives a citation, it must go through an administrative appeal process before it will gain access to a court.\footnote{See supra Part II.B.2. While an ad-
ministrative commission examines the appeal, the daily penalties authorized by the statute may accumulate. Although courts have overwhelmingly supported the constitutionality of these enforcement and judicial review provisions, there are some cases in which the Mine Safety Act could lead to constitutionally impermissible results.

As for the first factor in the constitutional framework proposed in Part IV.B, the daily fines that both statutes authorize for the failure to comply with a citation are coercive under *Okla. Operating Co. v. Love*, where the Supreme Court considered a $6000 fine to be clearly excessive. The OSH Act and Mine Safety Act allow courts to impose penalties of up to $7000 and $5000, respectively, per day of violation of a citation—the same order of magnitude that the Supreme Court has deemed to have a significant coercive power to induce compliance with an order.

The second factor—which examines if the recipient of an order may have a reasonable expectation of not being punished if the later challenge of the order fails—requires a more extensive discussion. This part of the analysis focuses on whether there is a good faith defense to the penalties or if there is ample opportunity to minimize their effect by obtaining temporary relief at the administrative level. Only one of the two statutes contemplates a good faith defense to the penalties that accrue while the Commission examines a challenge to the citation. The OSH Act incorporates clear language in that regard, and courts have interpreted it so that the default rule is that non-compliance penalties do not accrue during the administrative appeal process unless there is a finding of bad faith. This alleviates the constitutional concerns with the statute, given that it reduces the coercion that the daily penalties could generate. There is no similar provision creating a good faith presumption or defense in the Mine Safety Act.

The possibility of obtaining temporary relief from the accumulation of fines while the Commission is hearing a challenge is also different in the OSH Act and Mine Safety Act. The good faith provision in the OSH Act operates as temporary relief because it has the effect of preventing the accrual of daily fines before the Commission has ruled on the administrative challenge of the citation. The Mine Safety Act, on the other hand, has a more intricate mechanism by which recipients of orders may obtain temporary relief. The operator has the opportunity to submit a request for temporary relief, and the Commission decides based on the likelihood that the operator will succeed in the appeal and the probability of harm to miners if such relief is granted.

---

355 See supra Part IV.B.1.
360 See id. § 815(b)(2)(A), (B).
However, there is an important caveat: the opportunity to obtain temporary relief under the Mine Safety Act does not apply to orders issued under two subsections of the main enforcement provision of the statute. These are not minor exclusions. One of these subsections refers to orders that the Secretary may issue for any violation of any health or safety standard, rule, order, or regulation. Given that the Mine Safety Act does not include a good faith defense, the provision eliminating an operator’s opportunity to seek temporary relief from certain orders is problematic from a constitutional standpoint.

This conclusion may seem to be at odds with the Supreme Court’s decision in *Thunder Basin*. The Court treated the possibility of obtaining temporary relief as a feature of the Mine Safety Act that supported its constitutionality. However, the Supreme Court’s position cannot be reasonably interpreted to mean that all the enforcement and judicial review provisions of the Act are constitutional. First, the case that the Court was examining did not deal with one of the types of orders for which there is no temporary relief, so this language should be considered dicta with respect to the excluded orders. Second, the plaintiff did not bring a facial challenge of the statute and, therefore, the decision was also not intended to provide an abstract assessment of the constitutionality of the Act, as the Court itself pointed out.

The third factor of the analysis is identical for both statutes. Neither the OSH Act nor the Mine Safety Act limits the ability of a recipient of a citation to challenge it in court after having complied with it. This is an important advantage compared to the enforcement and judicial review provisions in the environmental statutes discussed above. However, as explained earlier, this only solves the constitutional problem if, in cases in which compliance would cause serious irreparable harm, the recipient of the order has the opportunity to obtain immediate review of the order. Otherwise, the operator could end up in a situation in which it has to choose between the immediate harm caused by compliance or the risk of very serious fines in the future if it decides to not comply with the order. None of the statutes expressly recognize the possibility of allowing immediate review in these cases. Some courts, however, have expressed their inclination to permit these types of

---

361 See id. § 815(b)(2) (providing that “[n]o temporary relief shall be granted in the case of a citation issued under subsection (a) or (f) of section 814”); see also Lucas v. Morton, 358 F. Supp. 900, 904 (W.D. Pa. 1973) (quoting the language from a decision of the Board of Mine Operations echoing this interpretation of sections 105(d) and 104(b) and (f), which now correspond to sections 815(b) and 814 (a) and (f) of the Act).


364 See id. at 216 (“[T]he company sued before a citation was issued.”).

365 See id. at 218 n.22 (explaining that the petitioner expressly stated that it was not bringing an abstract challenge to the statute).

366 See supra Part IV.B.3.
immediate challenges when the facts support a finding of irreparable harm.\(^{367}\)

In sum, the penalties that both the OSH Act and the Mine Safety Act authorize are susceptible of generating a powerful incentive to comply. This coercion, however, is significantly lower in the case of the OSH Act due to the provision in the statute creating a good faith defense and temporary relief from the accumulation of penalties. The constitutional propriety of the Mine Safety Act, on the other hand, points in different directions depending on the type of order at issue. The coercion that a recipient will experience is likely to be low when temporary relief is available for the particular type of order that the operator received. This is not so, however, with orders for which there is no such option. In these cases, because the level of coercion to comply is high, the third factor is decisive. The enforcement and judicial review provisions of the statute will only be constitutional if they are interpreted to allow immediate judicial review of orders that can cause serious irreparable harm to their recipients.

**CONCLUSION**

Delayed judicial review of agency action has attracted the attention of scholars and practitioners for decades. Surprisingly, there are two aspects of the immediate/delayed review dichotomy that had not been studied in depth. This Article addresses them both.

First, many scholars have argued that administrative compliance orders in the environmental context should not be subject to immediate judicial review because these challenges would cause important delays in enforcement and overburden courts. As this Article shows, these criticisms are misguided. Enforcement would not be delayed by simply allowing a plaintiff to challenge an administrative order. This would only occur if courts grant preliminary injunctions that render the orders unenforceable while they are evaluating the merits of the plaintiffs’ claims. Under the current preliminary injunction jurisprudence, courts are very unlikely to grant this form of injunctive relief, especially in cases in which substantial harm could follow from non-compliance with the order. Moreover, courts would not be flooded with immediate challenges of orders because, as the available empirical evidence supports, litigation costs and the risk of higher penalties are likely to deter frivolous challenges.

Second, this Article argues that the doctrine of constitutionally intolerable choices places important limitations on the types of delayed review provisions that legislatures can enact and provides a framework to evaluate the

\(^{367}\) See, e.g., Thunder Basin Coal Co., 510 U.S. at 212–13 (dismissing the constitutional claim, in part, because the petitioner had not proved that it would be suffering irreparable harm in that particular case, which suggests that a different outcome would be possible if an operator were able to prove a harm of that nature); Lucas v. Morton, 358 F. Supp. 900, 905 (W.D. Pa. 1973).
2019] 

*Delayed Judicial Review*

constitutionality of delayed review provisions under this doctrine. Examined under this framework, there are provisions in the Clean Air Act, RCRA, and Mine Safety Act that present very serious constitutional issues, if interpreted, as many courts have done, to deny immediate judicial review.
ARTICLE

DECODING PANDORA’S BOX:
ALL WRITS ACT AND SEPARATION
OF POWERS

JENNIFER X. LUO*

Table of Contents

I. INTRODUCTION .............................................. 257
II. BACKGROUND OF THE ALL WRITS ACT ................... 261
III. APPLICATION OF THE ALL WRITS ACT TO PRIVACY .... 265
   A. How New York Telephone Co. Opened Pandora’s Box ... 266
   B. The Role of the All Writs Act as a Gap-Filling
      Jurisdictional Device ...................................... 268
   C. Application of the All Writs Act to Data Decryption ... 269
IV. THE ROLE AND JURISDICTIONAL BASIS FOR JUDICIAL REVIEW
    UNDER THE ALL WRITS ACT ............................. 273
    A. Whose Right Is the All Writs Act Enforcing? .......... 273
    B. Jurisdictional Issues as Applied to Private Entities ... 275
       1. Ancillary Jurisdiction Analysis—Lack of Statutory
          Duty on Technology Companies .......................... 276
       2. Inherent Jurisdiction Analysis—Is the Order Really
          Enforcing the Court’s Jurisdiction and Duties? .... 277
V. ACCESS TO COURT: PRACTICAL AND STANDING ISSUES
    UNDER THE NEW YORK TELEPHONE CO. FRAMEWORK ...... 279
    A. Practical Issues with Applying the New York Telephone
       Co. Test ...................................................... 279
    B. Standing of Individual Defendants ........................ 280
VI. ALL WRITS ACT AND THE SEPARATION OF POWERS:
    COMMUNICATIONS ASSISTANCE FOR LAW ENFORCEMENT
    ACT ............................................................ 282
    A. CALEA and Device Decryption ............................ 282
    B. A Modified New York Telephone Co. Test ............... 284
VII. CONCLUSION ................................................ 285

I. INTRODUCTION

For much of history, locks have existed to guard our homes and personal belongings. We display enormous ingenuity in the design of the most

* B.A., Rice University; J.D., Yale Law School.
complex locks, all for a sense of security in our physical belongings and by extension, ourselves. In the twenty-first century, physical locks and the items they guard are increasingly digitalized and stored in electronic devices and digital clouds. The companies that create these devices have in turn become the modern equivalents of locksmiths. And sometimes, even these “locksmiths” do not have the key to such locks. This phenomenon raises some interesting questions, such as—what are the duties of these “locksmiths” to assist the government in decrypting the locks that they and their consumers have created? Is Congress, the Judiciary, or the Executive Branch best positioned to decide this question? And as non-democratic states implement draconian and purposely ambiguous laws that require tech companies to hand over encryption keys to monitor their citizens,1 how can we develop solutions that will stay true to the values of our democracy?

The Executive Branch’s answer, at the moment, is the All Writs Act. Since 2008, the government has filed more than sixty applications under the All Writs Act to compel private companies such as Apple, Google, and even a major bank, to decode consumer login information.2 While recent publicity in the San Bernardino shooting may have brought the All Writs Act to the national spotlight,3 the fact is that “the alleged crimes in these cases have ranged from drugs to counterfeiting.”4 As the frequency of such orders increased since 2008, the government has also become bolder in the scope of requests sought.5

One explanation for this aggressive use of the Act is Congress’s recent failure to mandate encryption backdoors.6 More alarmingly, many such orders were issued on the same day that the government sought them, were

---


3 Segal, supra note 2.

4 Id.; see also Letter to Court at *1, In re Apple Inc., 149 F. Supp. 3d 341, 346 (E.D.N.Y. Feb. 17, 2016) (No. 1:15-MC-1902) (stating that the government has filed multiple applications for similar orders, some of which are pending).

5 See Segal, supra note 2. Compare In re Order Requiring Apple, Inc. to Assist in the Execution of A Search Warrant Issued by This Court, No. 14-1760-WGC, at *1 (D. Md. Aug. 29, 2014) (specifying that Apple would not be required to “enable law enforcement’s attempts to access any encrypted data”), with Government’s Ex Parte Application for Order Compelling Apple Inc. to Assist Agents in Search, In re Search of an Apple iPhone Seized During the Execution of a Search Warrant on a Black Lexus IS300, No. 15-0451M (C.D. Cal. Feb. 15, 2016).

6 See Segal, supra note 2; see also Erin Kelly, Congress Wades into Encryption Debate with Bill to Create Expert Panel, USA TODAY (Jan. 11, 2016, 4:10 PM), https://www.usatoday
sealed, and received no objections from the companies the orders targeted, shielding these cases from public discourse. While the Legislative Branch has failed to mandate decryption by private technology companies, the Executive Branch has resorted to a little-known statute from the eighteenth century to accomplish the same end.

With its roots going back to two sections of the Judiciary Act of 1789, the All Writs Act is perhaps one of the most vaguely worded statutes in existence. The text of the Act, as amended, provides in its entirety:

(a) The Supreme Court and all courts established by Act of Congress may issue all writs necessary or appropriate in aid of their respective jurisdictions and agreeable to the usages and principles of law.
(b) An alternative writ or rule nisi may be issued by a justice or judge of a court which has jurisdiction.

Until the recent trend to invoke the Act to compel the decryption of personal devices, the government used the Act sparingly and often referred to it as a “statute of last resort.” For example, the Act has been used to compel a telephone company to install a pen register where no legislation specifically gave the government such powers. In recent years, prosecutors have increasingly drawn on the Act’s authority to seek judicial orders to compel private technology companies to decode consumer devices. Meanwhile, Congress has yet to provide a legislative solution that addresses whether private entities can be compelled to decode consumer devices in

7 See Segal, supra note 2.
8 See Judiciary Act of 1789, ch. 20, §§ 13–14, 1 Stat. 73, 80–82; see also United States v. N.Y. Tel. Co., 434 U.S. 159, 179 (1977); Lonny S. Hoffman, Removal Jurisdiction and the All Writs Act, 148 U. Pa. L. Rev. 401, 433 (1999). Sections 13 and 14 of the First Judiciary Act provided the antecedents for 28 U.S.C. § 1651(a). See Hoffman, supra, at 401 n.4. In 1940, the two provisions were essentially consolidated into what is now referred to as the All Writs Act, promulgated in its current form in § 1651(a), which substantially mirrors the language of section 14. See Michael D. Sousa, A Casus Omissus in Preventing Bankruptcy Fraud: Ordering a Search of a Debtor’s Home, 73 Ohio St. L.J. 93, 111–12 (2012); see also Pa. Bureau of Corr. v. U.S. Marshals Serv., 474 U.S. 34, 40–42 (1985) (discussing the legislative history of § 1651 as consolidating section 14 and various provisions without substantial change in meaning or purpose). As a result of the 1948 revisions, the All Writs Act is the sole statutory authority on which a court may base its issuance of an extraordinary writ, with the exception of 28 U.S.C. § 2241(a) concerning writs of habeas corpus. See Hoffman, supra, at 434–35.
9 28 U.S.C. § 1651 (2012). A “writ” is a court’s written order commanding the addressee to do or refrain from doing some specified act in the name of a state or other competent legal authority. Writ, BLACK’S LAW DICTIONARY (10th ed. 2014).
10 See, e.g., Mongelli v. Mongelli, 849 F. Supp. 215, 220 (S.D.N.Y. 1993) (“[T]he All Writs Act must be invoked only as a last resort—it is not a ‘catch-all’ statute granting jurisdiction when all else fails.”).
11 See N.Y. Tel. Co., 434 U.S. at 159.
certain situations. This is perhaps not surprising given the ultra-sensitive nature of data privacy and consumer protection.

Case law applying the Act also offers little guidance. Thus far, the Court has only discussed application of the Act in the pre-internet era. Various courts and legal scholars have derived the following requirements for the usage of the All Writs Act:

1. no other law applies;
2. the issuing court has jurisdiction over the underlying matter on an independent basis and the order is “in aid of” that jurisdiction;
3. exceptional circumstances are present that make issuance under the Act necessary or appropriate; and
4. the issuance of relief is done in conformity with the “usages and principles of law.”

However, the latter three prongs are intrinsically vague and ambiguous. Their amorphous qualities offer little guidance to courts in the decryption context and can be easily used to justify either the issuance or denial of an order.

Yet the question of whether the Act in fact provides such authority implicates jurisdictional and separation of powers concerns. While most critics of the Act as applied to decryption orders object based on privacy concerns, I argue that courts have little jurisdictional basis to compel private entities that have no public or statutory duty to provide law enforcement assistance. Additionally, the existing framework of analysis fails to take into account areas where a gap is not necessarily intentional or inadvertent, but rather the result of legislative battles and Congress’s inability to arrive at a consensus.

This Article argues that the issuance of decryption orders under the All Writs Act does not merely enforce the issuing court’s existing jurisdiction but threatens to violate the separation of powers doctrine by enabling the executive branch to bypass legislative consensus. In doing so, the court also violates the freedom of third-party corporations not to assist. To safeguard the original function of the All Writs Act as a gap-filler, courts should take into account whether a pre-existing gap in legislation is the result of unresolved legislative battle rather than an unintended gap in legislation. Only in this way can courts address the issue of compelled decryption orders without undermining the safeguards of our democracy.

13 See, e.g., N.Y. Tel. Co., 434 U.S. at 172.
To set the stage, Part II discusses the history of the All Writs Act, including the predecessors to the Act and their common usage by federal courts before 1948—when the Act was amended to take its current form. In part, it offers a brief overview of how the Act’s predecessors were limited to the issuance of writs of mandamus and injunction. Part III then examines applications of the Act in Supreme Court jurisprudence and discusses the role of the Act as a statutory gap-filling device. It studies how the Court first applied the Act to the area of privacy in United States v. New York Telephone Co.,15 where the Court compelled a telephone company to install a pen register in order to wiretap criminal suspects. It then analyzes how the Court subsequently drew the boundaries around application of the Act as a gap-filler. Finally, it summarizes recent lower courts’ application of the Act to compel decryption by private third-party corporations.

Part IV questions whether application of the Act in the privacy context thus far truly enforces the court’s existing jurisdiction, or if it furthers the agenda of the government and undermines the separation of powers doctrine. In particular, it considers the tenuous grounds for ancillary and inherent jurisdiction under the Act to adjudicate decryption cases. Part V explains the practical issues that defendants face to effectively challenge an All Writs Act order, the lack of incentives for third-party corporations to challenge such orders, and the implications of establishing a new judicial norm that assesses the cost of decryption on a case-by-case basis. Finally, Part VI analyzes the legislative history of the Communications Assistance for Law Enforcement Act ("CALEA")16 and proposes a modified New York Telephone Co. test that takes into account the reason and nature of a gap. Here, I suggest that where a gap is the result of legislative gridlock, courts lack jurisdiction to fill such a gap under the All Writs Act.

II. BACKGROUND OF THE ALL WRITS ACT

The All Writs Act provides that, “[t]he Supreme Court and all courts established by Act of Congress may issue all writs necessary or appropriate in aid of their respective jurisdictions and agreeable to the usages and principles of law.”17 The Act traces its lineage back to the Judiciary Act of 1789.18

18 See Hoffman, supra note 8, at 433–39; see also 16 CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 3933, at 535 (2d ed. 1996) (explaining that “§ 1651(a) is a combination of two provisions that trace back to sections 13 and 14 of the First Judiciary Act”). In fact, the Act was entangled with controversy and confusion even from its birth. “[C]hief Justice Marshall declared section 13 unconstitutional to the extent it sought to enlarge the Court’s original jurisdiction.” Id.; see Marbury v. Madison, 5 U.S. (1 Cranch) 137, 175–80 (1803) (noting that the Constitution allows Congress to extend the Supreme Court’s appellate, but not original, jurisdiction).
Its exact predecessors are sections 13 and 14 of the Judiciary Act of 1789,\textsuperscript{19} which would later be consolidated into section 1651(a).\textsuperscript{20} Specifically, section 13 authorized the Supreme Court:

\begin{quote}
[T]o issue writs of prohibition to the district courts, when proceeding as courts of admiralty and maritime jurisdiction, and writs of mandamus, in cases warranted by the principles and usages of law, to any courts appointed, or persons holding office, under the authority of the United States.\textsuperscript{21}
\end{quote}

Section 13 thus provided district courts with two types of power: (1) to issue writs of prohibition in courts of admiralty and maritime jurisdiction; and (2) to issue writs of mandamus “in cases warranted by the principles and usages of law.”\textsuperscript{22}

Similarly, section 14 authorized federal courts to “issue writs of scire facias, habeas corpus, and all other writs not specially provided for by statute, which may be necessary for the exercise of their respective jurisdictions, and agreeable to the principles and usages of law.”\textsuperscript{23} The operative words are “all other writs not specially provided for by statute,” which are then limited by the qualifiers that they must be “necessary” for the court’s exercise of jurisdiction and “agreeable” to the ends of the law. Thus, while section 13 gave courts the power to issue writs of mandamus, section 14 provided broader power to issue any writ not already provided for by statute. However, both sections also limited courts’ power to issue such writs by requiring that they be agreeable to the “principles and usages of law.”\textsuperscript{24} Section 14 additionally requires that any writs issued must be “necessary” for the courts’ existing jurisdiction.\textsuperscript{25}

The current All Writs Act was codified in the Judicial Code in 1948 when provisions of sections 13 and 14 were consolidated into section 1651 without substantive amendment.\textsuperscript{26} The original phrase “not specifically provided for by statute” was removed in this consolidation.\textsuperscript{27} The legislative history appears to be scant and Congress made “necessary changes in phra-

\begin{footnotesize}
\begin{itemize}
\item[19] Hoffman, supra note 8, at 433; see also Judiciary Act of 1789, ch. 20, §§ 13–14, 1 Stat. 73, 80–82.
\item[21] Judiciary Act of 1789, ch. 20, § 13, 1 Stat. 73, 81.
\item[22] Id.
\item[23] Id. at 81–82.
\item[24] Id.
\item[25] Id.
\item[26] 28 U.S.C. § 1651(a) (1948) (Reviser’s Note); see also Hoffman, supra note 8, at 434, n. 146 (“The 1948 codification consolidated sections 342, 376, and 377 of the 1940 version of the Code which, in turn, were derived from sections 234, 261, and 262 of the Judicial Code of 1911.”) (citing Act of March 3, 1911, ch. 231, 36 Stat. 1156, 1162).
\item[27] Compare id. § 1651(a) (1948) with 28 U.S.C. § 377 (1940).
\end{itemize}
\end{footnotesize}
seology” without substantive amendment. However, the legislative history did provide that the new section was “expressive of the construction recently placed upon [the all writs provision] by the Supreme Court in U.S. Alkali Export Assn. v. U.S. [sic] . . . .”28 In U.S. Alkali Export Association,29 the petitioner argued that a statute stripped the district court of jurisdiction and sought cert review from the Supreme Court after the district court denied the motion to dismiss.30 The Court affirmed the district court’s jurisdiction and found that interlocutory review under section 262 of the Judicial Code was appropriate where the underlying question involved the propriety of the court’s equity jurisdiction.31 Nevertheless, the Court stated the traditional use of writs under section 262 was “to confine inferior courts to the exercise of their prescribed jurisdiction or to compel them to exercise their authority when it is their duty to do so.”32 Chief Justice Stone wrote:

The writs may not be used as a substitute for an authorized appeal; and where, as here, the statutory scheme permits appellate review of interlocutory orders only on appeal from the final judgment, review by certiorari or other extraordinary writ is not permissible in the face of the plain indication of the legislative purpose to avoid piecemeal reviews.34

Indeed, the Supreme Court reiterated in Pennsylvania Bureau of Correction v. U.S. Marshals Service that Congress’s consolidation of sections 13 and 14 into the All Writs Act was “intended to leave the all writs provision substantially unchanged” and “that the 1948 changes in phraseology do not mark a congressional expansion of the powers of federal courts to authorize issuance of any ‘appropriate’ writ.”35 Thus, Congress’s consolidation of sections 13 and 14 into section 1651 has been interpreted to maintain, rather than expand or limit, the powers of the federal courts under the All Writs Act.36 Section 1651 also became the only existing statutory authority on which a court may base its issuance of

29 H.R. REP. NO. 80-308, at A145 (1947); see also Pa. Bureau of Corr., 474 U.S. at 40 ( noting that the legislative history of the 1948 codification of § 1651 indicates that the new section should be interpreted in line with the Alkali holding).
31 Id. at 202–04.
32 See id. at 203 (“The questions now presented involve the propriety of the exercise, by the district court, of its equity jurisdiction, and an asserted conflict between its jurisdiction and that of an agency of Congress said to be charged with the duty of enforcing the antitrust laws . . . .”).
33 Id. at 202.
34 Id. at 203.
36 See id. at 41–42 (stating that Congress intended to maintain the powers of the federal courts under the All Writs Act, not to expand them).
extraordinary writs (aside from writs of habeas corpus codified in section 2241(a)).

Prior to 1948, the Supreme Court frequently had to police lower courts’ unauthorized use of section 14 to issue writs of mandamus. For example, in *Roche v. Evaporated Milk Association,* the Supreme Court reversed the circuit court’s issuance of a writ of mandamus directing the district court to reinstate guilty pleas in abatement based on violations of the final judgment rule. Indeed, circuit courts sometimes issued writs of mandamus to district courts even though appellate jurisdiction was only available after a “final” judgment. In addition, parties often sought to enforce state court judgments by seeking writs of mandamus from federal courts, which would unwittingly grant them despite a lack of jurisdiction. As an example, the Supreme Court in *McIntire v. Wood* invalidated a district court’s issuance of a writ of mandamus under section 14 to a state office. Finding the lower court lacked original jurisdiction in the first place, the Court concluded that the lower court’s issuance of a writ of mandamus was improper because the court’s authority was “confined exclusively to those cases in which it may be necessary to the exercise of their jurisdiction.” In doing so, the Court emphasized that section 14 was not intended to be an independent source of jurisdiction. Prior to 1948, the Supreme Court frequently had to reverse the lower courts’ issuance of mandamus under section 14.

---

37 See, e.g., *De Beers Consol. Mines v. United States*, 325 U.S. 212, 225 (1945) (invalidating injunction issued by lower court based on lack of jurisdiction); *Rosenbaum v. Bauer*, 120 U.S. 450, 459 (1887) (affirming an order remanding the case for lack of jurisdiction); *Bath Cty. v. Amy*, 80 U.S. 244, 247 (1871) (“The Circuit Courts of the United States have no power to issue writs of mandamus to State courts, by way of original proceeding, and where such writ is neither necessary nor ancillary to any jurisdiction which the court then had.”); *Riggs v. Johnson Cty.*, 73 U.S. (6 Wall.) 166, 197–98 (1867) (describing when a lower court would lack jurisdiction to issue writ of mandamus).

38 319 U.S. 21 (1943).

39 *Id.* at 32; see also *Stephen I. Vladeck, Military Courts and the All Writs Act, 17 GREEN BAG 191, 193, n.6 (2014)* (citing *Roche v. Evaporated Milk Ass’n*, 319 U.S. 21, 25–31 (1943)).

40 See *Vladeck, supra* note 39, at 193.


42 11 U.S. (7 Cranch) 504 (1813).

43 *Id.* at 505–06 (holding that the Court lacked jurisdiction because of the absence of a statute specifically conferring jurisdiction).

44 *Id.* at 506.

45 Similarly, in *McClung v. Stillman*, 19 U.S. (6 Wheat.) 598 (1821), the Court reiterated that Section 14 was not intended to be an independent source of jurisdiction. *Id.* at 600–01 (1821). There, the state court issued a writ of mandamus citing section 14, and the Supreme Court invalidated the writ based on the court’s lack of jurisdiction. *Id.* at 605; see also *Amar, supra* note 20, at 457–58 (pointing out that sections 14, 15, and 17 of the First Judiciary Act reveal a distinction between the word “power” and “jurisdiction”). Thus, under Section 14, federal courts have the “power” to issue writs which may be necessary for the exercise of their respective “jurisdictions.” *Amar, supra* note 20, at 457–58. Parsing the Act’s text in this manner, Professor Amar concludes that “it is clear from context that the Act is investing courts with certain authority if and when they have independently founded jurisdiction . . . . ‘Jurisdiction’ must be established first, and independently; ‘power’ then follows, derivatively.” *Id.*
2019] Decoding Pandora’s Box 265

After 1948, “orders under the All Writs Act typically took the form of injunctions.”

Injunctions issued under the Act were relatively rare because they could “only be issued when necessary to protect a court’s underlying subject matter jurisdiction.” Thus, they had to aid a court’s existing jurisdiction on some independent ground such as original jurisdiction, diversity jurisdiction, or statutory jurisdiction. Furthermore, these injunctions were typically issued “to safeguard ongoing proceedings or to effectuate already-issued orders.” Specifically, where a party’s conduct violates a previously issued court order, the court may use the Act to enjoin such conduct. For example, the Act had been used to issue injunctions to enjoin a proposed merger so as to maintain the status quo. It was also used to issue injunctions against expelled members of a union in order to enforce a consent decree under the Racketeer Influenced and Corrupt Organizations Act.

III. APPLICATION OF THE ALL WRITS ACT TO PRIVACY

The application of the All Writs Act to the area of privacy is perhaps unexpected given its historical use as a mandamus- or injunction-issuing mechanism. But its relevance to privacy is also not surprising given its vague and amorphous qualities as a jurisdiction-preserving tool. This Section discusses in depth the first case that applied the All Writs Act to the area of privacy—United States v. New York Telephone Co.—the Supreme Court’s subsequent readings of the All Writs Act, and recent applications of the Act in the area of privacy.

46 See 1-10A MOORE’S MANUAL: FEDERAL PRACTICE AND PROCEDURE § 10A.05 [hereinafter 1-10A MOORE’S MANUAL] (discussing ancillary injunctions under the All Writs Act, which are used to prevent conduct that could frustrate the court’s jurisdiction); Potapchuk, supra note 14, n.99, at 1421 (citing Klay v. United Healthgroup, Inc., 376 F.3d 1092, 1099–1100 (11th Cir. 2004)).

47 See Potapchuk, supra note 14, at 1421.


49 See Potapchuk, supra note 14, at 1420–21, n.101 (citing Klay, 376 F.3d at 1099; Barton, 569 F.2d at 1359–60) (noting that the All Writs Act permits a district court to issue any order “necessary to enable the court to try the issue [in a pending case] to final judgment” and “develop the material issues and to bring them to a complete resolution”); see also United States v. N.Y. Tel. Co., 434 U.S. 159, 172 (1977) (noting the long-recognized authority of the federal courts to issue orders under the All Writs Act to “effectuate and prevent the frustration of orders it has previously issued in its exercise of jurisdiction otherwise obtained”).


52 See United States v. Int’l Bhd. of Teamsters, 266 F.3d 45, 50 (2d Cir. 2001).
A. How New York Telephone Co. Opened Pandora’s Box

The seminal case that provides the doctrinal support for forced decryption took place in the pre-digital era in the 1970s. In New York Telephone Co., the government sought an order under the All Writs Act to compel a public telephone company to install a pen register on two telephone lines in order to wiretap suspects in a gambling operation. The government argued that such an order was necessary in order to carry out a prior warrant to wiretap the suspects’ phone conversations. The district court authorized the All Writs Act order directing the New York Telephone Co. (the “Company”) to furnish the FBI with “all information, facilities and technical assistance necessary to employ the pen registers unobtrusively.” The Company declined to install the pen register, though it provided the information necessary for the government to do so. The government determined that it was not feasible for it to install the pen register without alerting the suspects and insisted that the Company install unused telephone lines to monitor the suspects.

The Company moved in the Southern District of New York:

[T]o vacate that portion of the pen register order directing it to furnish facilities and technical assistance to the FBI in connection with the use of the pen register on the ground that such a directive could be issued only in connection with a wiretap order conforming to the requirements of Title III of the Omnibus Crime Control and Safe Streets Act of 1968.

The district court ruled that the Wiretap Act did not govern pen registers but that it had jurisdiction to authorize the installation of the pen registers based on a showing of probable cause and the court’s inherent power under the All Writs Act. The Second Circuit subsequently held that the district court abused its discretion in ordering the Company to assist in the installation and

53 434 U.S. at 174–78. A “pen register” is a device that “records the numbers dialed on a telephone by monitoring the electronic impulses caused when the dial of the telephone is released” and “does not overhear oral communications and does not indicate whether calls are actually completed.” Id. at 161 n.1.; see also Hoffman, supra note 8, at 416–19 (discussing N.Y. Tel. in the context of All Writs Act); Ian J. McCarthy, IOS Fear the Government: Closing the Back Door on Governmental Access, 49 U. Tol. L. Rev. 179, 194–96 (2017) (same); Stephen J. Otto, Whether the Department of Justice Should Have the Authority to Compel Apple Inc. to Breach Its iPhone Security Measures, 85 U. Cin. L. Rev. 877, 880 (2017) (same); Potapchuk, supra note 14, at 1428–30 (same).

54 434 U.S. at 175–76.
55 Id. at 161.
56 Id. at 162–63.
57 Id. at 163.
58 Id.
59 Id.
operation of pen registers “in the absence of specific and properly limited Congressional action.” It expressed concerns that:

[S]uch an order could establish a most undesirable, if not dangerous and unwise, precedent for the authority of federal courts to impress unwilling aid on private third parties,” and that “there is no assurance that the court will always be able to protect [third parties] from excessive or overzealous Government activity or compulsion.

The Supreme Court reversed the Second Circuit. It held the Company could be compelled to install pen registers under the All Writs Act. First, it analyzed the removability of the Company from the underlying controversy. Because there was probable cause that the Company’s pen registers were being used to facilitate a criminal enterprise, the Company was not removed from the underlying matter. Second, the Court took into account the nature of the Company as a highly regulated public utility. Finding that such an entity has a duty to serve the public, the Court held that the Company lacked any “substantial interest in not providing assistance.” Third, the order was not burdensome for the Company because the Company regularly employed pen registers to check billing operations and detect fraud. Furthermore, the order required that the government fully reimburse the Company at prevailing rates. Thus, compliance with the order would not disrupt the Company’s business operations. Fourth, without the Company’s compelled assistance, the authorized surveillance could not have been accomplished. Finally, the order was not inconsistent with “recent congressional actions” since Congress, in the Wiretap Act, commanded assistance necessary to accomplish an electronic interception, even though it did not specifically provide for the installation of pen registers. Thus, the Court affirmed the use of the All Writs Act after examining the removability and nature of the third party, the burdensomeness of the order, whether the order could be accom-

60 Id. at 164 (quoting Application of the U.S. in the Matter of an Order Authorizing the Use of a Pen Register or Similar Mech. Device, 538 F.2d 956, 961 (2d Cir. 1976) (No. 1068, Docket 76-1155)).
61 Id. (alterations in original) (quoting Application of the U.S. in the Matter of an Order Authorizing the Use of a Pen Register or Similar Mech. Device, 538 F.2d 962–63 (No. 1068, Docket 76-1155)).
62 Id. at 178.
63 Id. at 174–78.
64 Id. at 174.
65 Id.
66 Id. at 175.
67 Id.
68 Id.
69 Id. at 176.
plished without the third party’s assistance, and whether the order was consistent with congressional intent.\textsuperscript{70}

\textbf{B. The Role of the All Writs Act as a Gap-Filling Jurisdictional Device}

Eight years after \textit{New York Telephone Co.}, the Supreme Court decided another landmark case involving the All Writs Act in \textit{Pennsylvania Bureau of Correction v. U.S. Marshals Service}. There, the Court clarified the role of the All Writs Act as a gap-filling device and warned against the use of the Act to expand jurisdiction.\textsuperscript{71} A lower court issued an order under the All Writs Act directing the United States Marshals Service to transport state inmates from a county facility to the federal court to testify as witnesses.\textsuperscript{72} After the Marshals objected to the order, the Third Circuit reversed in part and held that “the All Writs Act did not confer power upon the District Court ‘to compel non-custodians to bear the expense of [the production of witnesses] simply because they have access to a deeper pocket.’”\textsuperscript{73} Recognizing that the All Writs Act may be “necessary or appropriate” to enforce the court’s exercise of jurisdiction, the Court emphasized the limited role of the Act to “fill[] the interstices of federal judicial power when those gaps threatened to thwart the otherwise proper exercise of federal courts’ jurisdiction.”\textsuperscript{74} Because the Act is a “residual source of authority to issue writs,” where a statute specifically addresses the issue at hand, that statute is the controlling authority.\textsuperscript{75} Unlike in \textit{New York Telephone Co.}, where there was “a gap in federal statutes,” here the habeas corpus statute already provided for the issuance of a writ.\textsuperscript{76} Thus, the use of the All Writs Act was inappropriate because it supplanted a federal statute rather than fill a gap in jurisdiction.

While \textit{Pennsylvania Bureau of Correction} is an instance where the Court underscored the limited role of the All Writs Act, another case decided a decade before \textit{New York Telephone Co.}\textemdash\textit{FTC v. Dean Foods Co.}\textsuperscript{77}\textemdash stands for the view that the All Writs Act should be used to issue orders where Congress failed to provide a solution. There, the newly created Fed-

\footnotesize
\textsuperscript{70} Justice Stevens, joined by Justice Brennan and Justice Marshall, dissented from the opinion. Writing for the minority, Justice Stevens pointed out that the Court’s usage of the All Writs Act was unprecedented both in terms of “purpose” and “means.” \textit{Id.} at 187 (Stevens, J., dissenting in part). He argued that the purpose of the order was not in aid of the court’s jurisdiction and duties, but rather in aid of the government’s interest. \textit{Id.} Moreover, the means exceeded the court’s historical usage of a common-law writ. \textit{Id.} at 188–89. Finally, the minority pointed out that the order in such a form was similar to a writ of assistance under colonial rule, which was specifically outlawed in the nation’s history. \textit{Id.} at 190.

\textsuperscript{71} 474 U.S. 34, 43 (1985).

\textsuperscript{72} \textit{Id.} at 35–36.

\textsuperscript{73} \textit{Id.} at 36 (alteration in original) (citation omitted).

\textsuperscript{74} \textit{Id.} at 41.

\textsuperscript{75} \textit{Id.} at 43.

\textsuperscript{76} \textit{Id.} at 42 n.7.

\textsuperscript{77} 384 U.S. 597 (1966).
eral Trade Commission sought a preliminary injunction under the All Writs Act to stop the respondents, Dean Foods Company and Bowman Dairy Company, from merging until it reviewed the legality of the merger.\textsuperscript{78} The FTC argued that if the merger was subsequently ruled illegal after its consummation, it would be impossible to restore the merged entities’ separate status, which meant the court of appeals would thus be deprived of its appellate jurisdiction.\textsuperscript{79} However, respondents argued the fact that Congress did not give the FTC express statutory authority to request preliminary relief under the All Writs Act meant no such relief would be available.\textsuperscript{80} The Court, on the other hand, reasoned that Congress could not have entrusted the enforcement of the Clayton Act to the FTC without allowing the court of appeals to exercise its derivative power under the All Writs Act.\textsuperscript{81} Thus, in the absence of explicit congressional direction, courts may exercise their authority under the All Writs Act to ensure effective judicial review of administrative agencies.\textsuperscript{82}

Although Pennsylvania Bureau of Correction and Dean Foods do not concern the subject of privacy, they both underscore the role of the All Writs Act as a gap-filling device while drawing boundaries around the Act’s usage. At one end of the spectrum, Dean Foods and New York Telephone Co. may stand for the idea that the All Writs Act should apply as long as Congress is silent on a particular legislative issue and where there is a gap in jurisdiction. At the other end, Pennsylvania Bureau of Correction states that courts should not use the All Writs Act where Congress has already provided a legislative solution. What these cases have not addressed, however, is a scenario where Congress has engaged in significant legislative debate on an existing “gap” but has failed to take action. Such a scenario is the current situation in the data decryption context.

\textbf{C. Application of the All Writs Act to Data Decryption}

Although the significance of New York Telephone Co. dissipated after Congress passed the Electronic Communications Privacy Act (“ECPA”)\textsuperscript{83} to regulate surveillance involving pen registers, its framework can still apply to data privacy cases.\textsuperscript{84} Indeed, the government has increasingly sought All Writs Act orders compelling third-party device manufacturers like Apple to

\textsuperscript{78} Id. at 599.
\textsuperscript{79} See id. at 599–600.
\textsuperscript{80} See id. at 605–06.
\textsuperscript{81} Id. at 606–07.
\textsuperscript{82} See id. at 607–08.
\textsuperscript{84} See Potapchuk, supra note 14, at 1429 n.138 (citing Smith v. Maryland, 442 U.S. 735, 741 (1979); In re Application of the U.S. for an Order Authorizing (1) Installation & Use of a Pen Register & Trap & Trace Device or Process, (2) Access to Customer Records, & (3) Cell Phone Tracking, 441 F. Supp. 2d 816, 831 (S.D. Tex. 2006)).
decode defendants’ password-protected consumer devices.85 One of the few cases to have addressed this issue in depth is In re Order Requiring Apple, Inc. to Assist in the Execution of a Search Warrant Issued by this Court (“In re Apple, Inc.”).86 There, the government secured search warrants for the defendant’s residence and seized an iPhone 5 that used Apple’s software for its operating system.87 It then sought a warrant to search the device but was unable to bypass the iPhone’s passcode security.88 The government requested that Apple unlock the device. Apple’s response—consistent with its past practice in at least seventy instances89—was that it would unlock the iPhone only if a court issued a lawful order requiring it to do so.90 Subsequently, the government filed an application to compel Apple to decode the iPhone under the All Writs Act, which the district court denied.

The district court analyzed the order application with a two-tiered test. First, the warrant had to meet the statutory requirements of the All Writs Act, which the court broke down into three elements: the issuance of the writ must be (1) “in aid of” the issuing court’s jurisdiction; (2) “necessary or appropriate” to provide such aid to the issuing court’s jurisdiction; and (3) “agreeable to the usages and principles of law.”91 Second, if the warrant satisfied these threshold requirements, the court could issue the writ after considering three discretionary factors: (1) the closeness of the relationship between the directed entity and the matter; (2) the reasonableness of the burden imposed on the writ’s subject; and (3) the necessity of the writ to aid the court’s jurisdiction.92 The first three elements came from the statutory text of the All Writs Act, and the next three factors arose out of the Court’s holding in New York Telephone Co.93

Analyzing the order under this framework, the district court concluded that although the application was “in aid of” and “necessary or appropriate” to the court’s jurisdiction, it was not “agreeable to the usages and principles of law.”94 Additionally, all three discretionary factors weighed against the government.95 Two primary concerns drove the court’s denial of the All

85 See Sorkin, supra note 12; see also In re Apple Inc., 149 F. Supp. 3d 341, 348 (E.D.N.Y. 2016) (citation omitted) (“Apple alluded to ‘additional requests similar to the one underlying the case before this Court’ and the fact that it has ‘been advised that the government intends to continue to invoke the All Writs Act in this and other districts in an attempt to require Apple to assist in bypassing the security of other Apple devices in the government’s possession.’”).
87 Id. at 345–46.
88 Id.
89 Id. at 346.
90 Id.
91 Id. at 350.
92 Id. at 351; see also id. (recognizing that the third discretionary factor replicates the second statutory element).
93 149 F. Supp. 3d at 350–51.
94 Id. at 351–54.
95 Id. at 350–51.
Writs Act order. First, it was arguable that another statute, the CALEA, exempted Apple from forced decryption and implicitly prohibited compelled decryption by private entities. Second, the burdens imposed on Apple were significantly higher than those imposed on the Company in *New York Telephone Co.*

The district court’s application of the *New York Telephone Co.* framework is interesting because instead of treating the *New York Telephone Co.* test as a three-prong test that interprets the All Writs Act, the court created a six-prong test, with three mandatory factors based on the text of the All Writs Act, and three discretionary factors derived from the *New York Telephone Co.* framework. This six-prong test is therefore a little redundant because the three discretionary factors were originally used to help the *New York Telephone Co.* Court assess whether a writ is “necessary or appropriate” and “agreeable to the usages and principles of law.” Nevertheless, the court’s analysis of CALEA and the concerns it raised pose valid questions for future decryption cases under the All Writs Act. Whether the Act provides means to compel decryption is a pressing issue that courts around the country are increasingly facing.

After *In re Apple, Inc.*, the first and only appellate decision to address this issue in depth thus far was *United States v. Blake*. The Eleventh Circuit applied a five-prong test derived from *New York Telephone Co.* and *Pennsylvania Bureau of Correction*, examining whether the use of the All Writs Act was (1) necessary or appropriate to carry out an issued order, (2) not otherwise covered by statute, (3) not inconsistent with the intent of Congress, (4) whether the third party was too far removed from the underlying case, and (5) whether any burden imposed on the compelled party was reasonable. The court held that the use of the All Writs Act was appropriate. There was no other way for the FBI to execute the district court’s warrant to search the contents of the iPad, no statute expressly permitted or prohibited it, compelling decryption was not inconsistent with the intent of Congress, Apple was not too far removed from the case as a non-party because its

---

97 *In re Apple, Inc.*, 149 F. Supp. 3d at 354.
98 Id. at 368–72.
99 Id. at 350–51.
101 Since *In re Apple, Inc.*, a federal magistrate court judge granted an All Writs Act order compelling Apple to decode the iPhone of a defendant in the 2015 San Bernardino attack. *See In re Search of an Apple iPhone Seized During the Execution of a Search Warrant on a Black Lexus IS300, No. ED 15-0451M, at *1 (C.D. Cal. Feb. 16, 2016).* However, the judge did not issue any opinion explaining the rationale for granting such an order and no further adjudication on the matter took place. Id.
102 868 F.3d 960 (11th Cir. 2017).
technology provided the means by which the iPad was locked, and complying with the order was not costly for Apple. In applying this test, the court devoted most of the discussion to the third prong, which analyzed whether the use of the Act as a gap-filler would be inconsistent with the intent of Congress. It examined the CALEA, which required “‘telecommunication carrier[s]’ to provide certain forms of assistance to law enforcement, while exempting ‘information services’ companies—a category that includes Apple—from those same requirements.” The court reasoned this exemption did not run counter to using the All Writs Act to compel decryption because section 1002 merely evinced Congress’s intent to exempt device manufacturers from designing a back-end decryption channel for law enforcement. In doing so, the court described section 1002 to be “all about design choices” and created a “distinction between initial design and later access” not discussed in the statute. Only after creating this artificial distinction could the court justify why the exemption in section 1002 did not prohibit forced decryption under the All Writs Act.

The court’s interpretation of section 1002 requirements as solely concerning “design choices” at the pre-manufacturing stage lacks evidentiary support in the text of the statute. First, section 1002(a) makes clear it is about “capability requirements,” which broadly include “enabling the government, pursuant to a court order or other lawful authorization, to intercept, to the exclusion of any other communications, all wire and electronic communications.” Nowhere does this provision limit its applications solely to the design stage. In fact, this provision requires broad cooperation by telecommunication carriers to supply law enforcement agencies with “authorized communications interceptions,” to “deliver[] intercepted communications and call-identifying information to the government,” and to “enabl[e] the government, pursuant to a court order or other lawful authorization, to access call-identifying information.” Additionally, section 1002(b) lays out three exemptions to section 1002(a), one of which has a paragraph on “[d]esign of features and systems configurations.” This exempts any tele-

---

104 See id. at 971–72.
105 Id. at 971–72.
106 Id. at 972 (alteration in original).
107 Id.
108 Id.
109 47 U.S.C. § 1002(a)(1) (2012). Additional provisions require that carriers provide specific information such as “call-identifying information” and “intercepted communications,” id. § 1002(a)(3), pursuant to court orders “unobtrusively and with a minimum of interference with any subscriber’s telecommunications services and in a matter that protects—
(A) the privacy and security of communications and call-identifying information not authorized to be intercepted; and
(B) information regarding the government’s interception of communications and access to call-identifying information,” id. § 1002(a)(4).
111 Id. § 1002(a)(3).
112 Id. § 1002(a)(2).
113 Id. § 1002(b)(1)(A).
Decoding Pandora’s Box 273

communications equipment provider from section 1002(a)’s requirements to “require any specific design of equipment, facilities, services, features, or system configurations to be adopted.”114 If section 1002(a) is solely about design choices, then section 1002(b)(1)’s exemption from the provision applicable to the design stage of equipment for telecommunication carriers would defeat the purpose of section 1002(a). Therefore, it cannot be the case that section 1002(a) requirements are all about design choices.115

IV. THE ROLE AND JURISDICTIONAL BASIS FOR JUDICIAL REVIEW UNDER THE ALL WRITS ACT

It is the role of courts to function as neutral arbiters of power. But this role can be complicated in the context of the All Writs Act. This section argues that it is a legal fiction to justify decryption orders under the All Writs Act as mere instruments to enforce the court’s jurisdiction. This legal fiction treats third-party private corporations as an extension of the government with a duty to assist law enforcement despite the lack of any statutory duty. Additionally, it is unclear what jurisdictional framework applies in the decryption context and whether those frameworks can provide any limits at all.

A. Whose Right Is the All Writs Act Enforcing?

A decryption order under the All Writs Act does not just involve the simple question of whether the order enforces the court’s prior search warrant, but also whether it imposes additional responsibilities on third parties that have no obligations to the court or the underlying matter. Although the All Writs Act is ambiguously worded, its text nonetheless provides some limitations: it can only be used to issue “writs” and such writs must be “in aid of [the courts’] respective jurisdictions.”116 As Section II shows, the predecessors to the All Writs Act—sections 13 and 14 of the Judiciary Act—were primarily used to issue writs of mandamus or prohibition that enforced or invalidated a pre-existing court order.117 In this sense, the All Writs Act helped enforce the court’s underlying jurisdiction by making sure that a pre-existing order was carried out within the court’s jurisdiction. How-

114 Id. (emphasis added).
115 See In re Apple, Inc., 149 F. Supp. 3d 341, 354, 354 n.12 (E.D.N.Y. 2016) (interpreting section 1002(a) to prescribe general, not design-specific, responsibilities on telecommunications service providers).
117 See supra Section II; see also Ex parte Bollman, 8 U.S. (4 Cranch) 75, 93 (1807) (“[C]ourts which are created by written law, and whose jurisdiction is defined by written law, cannot transcend that jurisdiction. It is unnecessary to state the reasoning on which this opinion is founded, because it has been repeatedly given by this court; and with the decisions heretofore rendered on this point, no member of the bench has, even for an instant, been dissatisfied.”).
ever, in the forced decryption context, the use of the All Writs Act to carry out a pre-existing search warrant goes beyond enforcing the court’s rights because it violates the rights of third-party providers in a way not provided for by the original search warrant.

An order compelling a third party to decode a defendant’s device arguably functions rather like a writ of mandamus in that it compels a party to carry out a specific performance. However, whereas a writ of mandamus compels actions from another government entity, a decryption order directs action from a private entity that is otherwise free from any similar public or statutory duty. The validity underlying a writ of mandamus rests on the petitioner’s judicially enforceable and legally protected right. For example, in *New York Telephone Co.*, the subject of the All Writs Act order was a “highly regulated public utility” and thus subject to “a duty to serve the public.”118 By not providing a pen register, the Company deprived the government of a legal right under public law. This refusal explains why the government had a valid grievance when it petitioned the Court to order the Company’s compliance with the All Writs Act order. Compelled decryption cases, on the other hand, involve private entities that are otherwise not subject to any similar statutory or public duty. Although companies like Apple and other technology companies provide services to vast members of the public, whether they should be subject to public and statutory duties is a normative question for the Legislative Branch. After all, at the beginning of the twentieth century, privately owned utility companies were not considered governmental agencies until various state and federal legislative bodies began regulating them.119 Whether technology companies should have similar duties as public utilities is a question for the Legislative Branch and, ultimately, the electorate. Until such laws are passed, companies like Apple do not have any statutory duty to comply with decryption orders. A petitioner can only be aggrieved when she is denied a legal right by someone who has a legal duty to either perform or refrain from performing. Here, the petitioner is the government, which, without any statutory law, lacks a legal right to demand any action from private entities in a courtroom setting. It is therefore questionable that any writ can be issued under the All Writs Act to a private entity at all.

In this regard, a writ directing a private corporation with no public or statutory duty to specific performance is closer to the historically abhorred writ of assistance rather than a writ of mandamus. The Fourth Amendment’s prohibition of general warrants was a direct reaction to the oppressive British practice of allowing courts to issue “writs of assistance” in the colonial period.120 The same principle explains why our Constitution requires that Congress must pass specific legislation to allow federal courts the power to

120 See *N.Y. Tel. Co.*, 434 U.S. at 180 n.3–4 (Stevens, J., dissenting).
issue search warrants. Such legislation must also specify the grant of authority and limitations on the places to be searched, the objects of the search, and the requirements for the issuance of a warrant.

### B. Jurisdictional Issues as Applied to Private Entities

The text of the All Writs Act requires that the writ must be “necessary or appropriate in aid of . . . jurisdiction[],” but it does not specify the nature of this jurisdiction. However, the All Writs Act does not provide federal courts with an independent grant of jurisdiction. Thus two types of jurisdiction may support an All Writs Act order: (1) ancillary jurisdiction, which allows the court to assert jurisdiction over a subject matter that is different from, but related to, a matter properly before the court; and (2) inherent jurisdiction, which is the inherent power of a court to issue remedies to effectuate its existing jurisdiction. The rationale for ancillary jurisdiction rests on the argument that an All Writs Act order is only issued when necessary to protect the court’s underlying jurisdiction. Similarly, the rationale for inherent jurisdiction stems from a court’s inherent power to issue extraordinary writs. The Supreme Court in *New York Telephone Co.* did not specify what jurisdiction the All Writs Act order was premised on. Since an All Writs Act order in the decryption context ostensibly seeks to

---

121 See id. at 180–81 (Stevens, J., dissenting); see also Harris v. United States, 331 U.S. 145, 158 (1947) (Frankfurter, J., dissenting).


125 See Dugas v. Am. Sur. Co., 300 U.S. 414 (1937); Labette Cty. Comm’rs v. United States ex rel. Moulton, 112 U.S. 217, 223–24 (1884); Board of Education v. York, 429 F.2d 66, 69 (10th Cir. 1970) (affirming use of the All Writs Act to order appellants to send their son to a specific public school was ‘necessary and appropriate’ in aid of the court’s jurisdiction over the underlying segregation problems and that “[t]he equitable powers of the federal courts exist independently of the 1964 Act.”); Morrow v. District of Columbia, 417 F.2d 728, 737 (D.C. Cir. 1969) (holding that All Writs Act order to compel a judge to vacate an existing order is premised on the court’s “ancillary jurisdiction”); see also N.Y. Tel. Co., 434 U.S. at 187 n.18 (citing 9 J. MOORE, ET AL., MOORE’S FEDERAL PRACTICE ¶¶ 110.27–110.28 (2d ed. Supp. 1975)) (“Here, we are faced with an order that must be necessary or appropriate in the exercise of a district court’s original jurisdiction.”)

126 See Morrow, 417 F.2d at 738 n.36 (“It has been a traditional common law power of appellate courts to use extraordinary writs like writs of mandamus in aid of their appellate jurisdiction.”) (citing State ex rel. Gillette v. Niblack, 53 N.E.2d 542, 542 (Ind. 1944); People ex rel. Earle v. Circuit Court, 48 N.E. 717, 719 (Ill. 1897)).

127 See N.Y. Tel. Co., 434 U.S. at 251 n.19 (Stevens, J., dissenting) (“The Court never explains on what basis the District Court had jurisdiction to enter this order. Possibly, the
compel a private company to decode a consumer device in connection with a pre-existing warrant, both ancillary and inherent jurisdiction may exist as potential grounds for jurisdiction. However, there are also various issues with applying either jurisdictional framework to the decryption context.

1. Ancillary Jurisdiction Analysis—Lack of Statutory Duty on Technology Companies

The theory of ancillary jurisdiction springs from the equitable doctrine that a court may consider a subject matter over which it has no independent jurisdiction when it is necessary in order to dispose of the principal case or do complete justice in the principal case. Because ancillary jurisdiction is a common law doctrine that was created out of necessity, there is no “general, all-encompassing jurisdictional rule.” Thus, it is unclear what the limits of ancillary jurisdiction are to aid the court’s original jurisdiction, and whether orders to compel decryption fall within those limits. Ancillary jurisdiction does not “turn[] on whether the court has... original equity powers.” However, the Supreme Court has found ancillary jurisdiction where “an additional party has a claim upon contested assets within the court’s exclusive control, or when necessary to give effect to the court’s judgment.” In delineating the boundaries of ancillary jurisdiction, courts have looked at whether:

(1) the ancillary matter arises from the same transaction which was the basis of the main proceeding, or arises during the course of the main matter, or is an integral part of the main matter; (2) the ancillary matter can be determined without a substantial new fact-finding proceeding; (3) determination of the ancillary matter through an ancillary order would not deprive a party of a substantial procedural or substantive right; and (4) the ancillary matter must be set-
tled to protect the integrity of the main proceeding or to insure that the disposition in the main proceeding will not be frustrated. 133

Therefore, one should analyze the limits of the court’s ancillary jurisdiction based on the “relationship between the original [warrant] and the ancillary order.” 134 Under this framework, grounds for ancillary jurisdiction are tenuous because a third-party private entity has no duty or relationship to the underlying crime or warrant. Even if a private third party’s assistance may be integral to effectuating the court’s enforcement of a prior warrant, it is not the company’s action that led to the obstruction. The third-party company was never an accomplice to the defendant in the underlying crime, nor did it place any obstruction to the execution of the warrant. Indeed, the closest analogy would be an order to a lock manufacturing company to open a lock that only the defendant has the keys to. The third-party entity has no legal duty to perform the required act. Nor does the petitioner have a legal right to the third party’s performance by virtue of the warrant. The third-party entity is therefore not bound under any legal duty to any of the parties, because it does not owe such a duty to any parties to the litigation. Unlike the Company in New York Telephone Co., which is a public utility company that is both subject to public utility law and has a statutory duty to serve the public, no statutory law imposes any legal duty on the third-party entity in compelled decryption cases. Therefore, a critical piece of the puzzle is missing for the court to establish ancillary jurisdiction.

2. Inherent Jurisdiction Analysis—Is the Order Really Enforcing the Court’s Jurisdiction and Duties?

Courts have also based the All Writs Act’s jurisdiction on the “inherent powers of a court of equity to prevent the defeat or impairment of its jurisdiction.” 135 Indeed, “[t]he power to enter judgment and, when necessary, to enforce it by appropriate process, is inherent in the Court’s appellate jurisdiction.” 136 Of course, implicit in the courts’ exercise of inherent powers is that such power must be used solely to advance the status quo and duties of the court. As the Supreme Court stated in FTC v. Dean Foods Co., an injunction under the All Writs Act was upheld because it was necessary to “preserve the status quo while administrative proceedings are in progress and prevent...
impairment of the effective exercise of appellate jurisdiction.” 137 Maintaining the status quo of the courts means not only preventing the impairment of the court’s judgment but also preserving the court’s role as a neutral protector of the parties’ rights. Yet orders in aid of a court’s own duties and jurisdiction inevitably protects the legal rights of one of the parties. In New York Telephone Co., for example, the requirement that the Company install a pen register enabled the government’s right to effectively investigate an alleged crime. Similarly, in Harris v. Nelson, 138 the court’s ordering of discovery in connection with a habeas corpus proceeding protected the rights of a prisoner. 139 Thus, any time a court issues an order under its inherent jurisdiction, it is also enforcing the rights of a party at the expense of the opposing party. Such exercise of inherent jurisdiction enforces the warrant at the expense not only of the defendant’s privacy rights but also of a third and unrelated party. Supporters of compelled decryption might argue that a third party’s encryption enabled the defendant to lock the defendant’s device. Thus, the third party has prevented the execution of a search warrant and should be subject to the inherent jurisdiction of the court. However, such a third party did not intentionally impede the warrant’s execution; it merely provided the means by which to do so. It is true that Congress imposes criminal penalties on any person who “forcibly assaults, resists, opposes, prevents, impedes, intimidates, or interferes with any person authorized to serve or execute search warrants” in the Omnibus Diplomatic Security and Antiterrorism Act of 1986. 140 But this statutory provision does not impose a positive duty to assist with the execution of a warrant but rather prohibits conduct such as “assaults” and “intimidat[ions]” that would constitute a tort or crime against any individual in the first place. Such criminal acts likely require intent as an element, 141 whereas one cannot impute intent to the simple act of providing an option to encrypt a consumer device. In this sense, companies like Apple are more akin to being a lock manufacturer or a security company. That Congress did not criminalize or prohibit a third party from merely providing the mechanism to lock a consumer device means that courts cannot overstep their jurisdictional boundaries to punish third parties for the same act.

139 See id. at 290.
140 18 U.S.C. § 2231 (2012); see also United States v. N.Y. Tel. Co., 434 U.S. 159, 190 n.21 (Stevens, J., dissenting) (explaining that “[t]his section was originally enacted as part of the Espionage Act of 1917”).
141 See, e.g., United States v. Norton, 808 F.2d 908, 909 (1st Cir. 1987) (interpreting a similar statute that contained the word “intimidat[ion]” to suggest “that it was not designed to punish pure ‘frightening’ without any element of intent to injure . . . .”)
Decoding Pandora’s Box

2019] 279

V. ACCESS TO COURT: PRACTICAL AND STANDING ISSUES UNDER THE NEW YORK TELEPHONE CO. FRAMEWORK

Application of the New York Telephone Co. framework in the privacy context has focused on a balancing test that weighs the government’s need for assistance against the non-party’s interest in its own autonomy. However, this application faces several practical hurdles: (1) the test assumes that a non-party will challenge the order when the reality is that most nonparties are unlikely to do so; and (2) standing and other issues can exist where defendants, rather than the non-parties, are the ones challenging the order. These difficulties in turn mean that few such cases will face judicial scrutiny.

A. Practical Issues with Applying the New York Telephone Co. Test

The New York Telephone Co. test is not ideal for analyzing cases in the privacy context because of three practical issues: (1) private entities tend not to challenge decryption orders for various reasons, so it is often individual defendants who challenge All Writs Act orders; (2) individual defendants who challenge such orders lack access to proprietary information that demonstrates the cost of such orders; and (3) the courts are not well-positioned to evaluate the burden of such orders on technology companies since the cost of compliance will only rise over time as decryption becomes a judicial norm.

Technology companies are not as incentivized as defendants to challenge All Writs Act orders. Thus far, individual defendants—not third-party entities—have brought the majority of challenges to decryption orders under the All Writs Act. In fact, Apple did not object to most of the All Writs Act orders it faced until recently. Indeed, third-party entities may be unwilling to incur significant legal costs to object to an All Writs Act order unless the cost of decoding is prohibitive to their business operations. They may not want to jeopardize their relationship with the government by turning down such orders. Furthermore, corporations have little to lose in privacy terms by complying with an All Writs Act order because disclosure of their consumers’ data would only implicate the privacy of their users’ information. Thus, the unfortunate real-

142 See, e.g., United States v. Blake, 868 F.3d 960 (11th Cir. 2017).
143 See In re Apple, Inc., 149 F. Supp. 3d 341, 346 (E.D.N.Y. 2016) (citation omitted) (“Apple’s response, consistent with its past practice in at least 70 instances, was that it could and would unlock [Defendant]’s phone for the agents, but only if a court issued a lawful order requiring it to do so. Also consistent with past practice, Apple provided the agents with the specific technical language it deemed sufficient to make clear its obligation to provide the services that would allow the agents to gain access to the iPhone’s passcode-protected data.”).
ity is that most third parties are unlikely to object to an All Writs Act order unless the costs of compliance becomes astronomical in the long run. But by then, a judicial norm may already be so entrenched that it becomes difficult to challenge such orders.

The fact that third-party corporations tend not to challenge All Writs Act orders means that defendants face mounting practical issues that hinder them from effectively challenging such orders. For one, defendants often lack access to proprietary information and insider corporate information such as the administrative costs and technical burdens to decrypt devices.\(^{145}\) Private corporations control and closely monitor such information, and they have no incentive to share it with criminal defendants. Meanwhile, two prongs of the *New York Telephone Co.* test require that parties demonstrate the burdens of such orders on nonparties.\(^{146}\) Defendants are therefore rarely equipped to meet this burden given their lack of access to such information. Thus, the practical reality of applying the *New York Telephone Co.* test to All Writs Act orders in forced decryption cases is that defendants face an uphill battle.

Finally, courts are not well-positioned to assess the burdens imposed on nonparties because the *New York Telephone Co.* test only evaluates the cost of decryptions in a single case, whereas the cumulative cost of such orders is much higher. Aside from the logistical problem of estimating the burden of a single decoding procedure, the real question is the cost of complying with such orders in the long-term. Because the request to decode an iPad involves the same procedure in every investigation, a court’s decision inevitably impacts future cases. Even though a court decides whether an order is reasonable in a single case, other courts in subsequent cases may lean on the finding that it would not be burdensome to decode an iPad. A single case that seemingly incurs little cost in fact sets important precedent and invites future accumulation of such costs, for which courts cannot account. Furthermore, such analysis cannot take into account the reputational costs\(^{147}\) of complying with such orders. Thus, the indirect consequences of such orders may also result in costs that are difficult for the court to quantify.

### B. Standing of Individual Defendants

Additionally, the government has frequently raised standing issues in cases where individual defendants rather than non-party entities challenge

\(^{145}\) See, e.g., *Blake*, 868 F.3d at 972–73 (finding that defendant challenging the All Writs Act order could not demonstrate unreasonable technical burden imposed on Apple).

\(^{146}\) See supra Section III.A.

\(^{147}\) See Apple Inc.’s Response to Court’s October 9, 2015 Memorandum and Order at *4, *In re Apple, Inc.*, 149 F. Supp. 3d 341 (E.D.N.Y. Oct. 19, 2015) (No. 15 MISC 1902 (JO)) (“Forcing Apple to extract data in this case, absent clear legal authority to do so, could threaten the trust between Apple and its customers and substantially tarnish the Apple brand. This reputational harm could have a longer term economic impact beyond the mere cost of performing the single extraction at issue.”).
All Writs Act orders. For example, in Blake, the government argued that defendants challenging the All Writs Act lacked Fourth Amendment standing because the order only implicated the Fourth Amendment rights of Apple, not of individual defendants.148 It did not implicate the defendants’ rights because the order was directed at Apple, not at the defendants. Thus, only Apple had standing to challenge the All Writs Act order. In doing so, the government relied on United States v. Padilla,149 and In re Order Requiring [XXX], Inc. to Assist in the Execution of a Search Warrant Issued by This Court by Unlocking a Cellphone (“In re Cellphone”).150 Both cases, however, are inapposite in this context. Padilla does not address the underlying question of whether the defendants’ Fourth Amendment rights were violated.151 In re Cellphone only addresses the question of whether a third party such as a cellphone provider had standing to challenge the search warrant.152 It did not address whether standing was limited to that third party.153 The court’s rationale was that a third party such as a cellphone manufacturer may incur “unduly burdensome” expenses in order to comply with the warrant and did not discuss any limitation of standing as applied to the defendants.154

In fact, the defendants likely do have standing because the All Writs Act clearly implicates their Fourth Amendment rights. “Standing to invoke the exclusionary rule . . . exist[s] where the Government attempts to use illegally obtained evidence to incriminate the victim of the illegal search.”155 The Fourth Amendment requires that the party invoking standing “must have objectively reasonable expectation of privacy in the place searched or the item seized.”156 As owners of password-protected devices, defendants have a reasonable expectation of privacy in the contents of those devices. It would be preposterous if a citizen could lose Fourth Amendment standing to protect her right to privacy simply because of a corporation’s failure to challenge the order.

The government argues the All Writs Act merely enforces a pre-existing warrant. But this argument ignores the fact that the inquiry is whether the All Writs Act order exceeded the scope of the search warrant and thus

148 See Blake, 868 F.3d at 969 n.4.
149 508 U.S. 77 (1993); see Brief for United States at 56, United States v. Blake, 868 F.3d 960 (11th Cir. 2017) (No. 15-13395); see also Padilla, 508 U.S. at 81 (“It has long been the rule that a defendant can urge the suppression of evidence obtained in violation of the Fourth Amendment only if that defendant demonstrates that his Fourth Amendment rights were violated by the challenged search or seizure.”).
150 In re Order Requiring [XXX], Inc. to Assist in the Execution of a Search Warrant Issued by This Court by Unlocking a Cellphone, No. 14 Mag. 2258, 2014 WL 5510865, at *2 (S.D.N.Y. Oct. 31, 2014) [hereinafter “In re Cellphone”]; see Brief for United States, supra note 149, at 56, 61.
153 See id. at *2.
154 Id.
156 Rehberg v. Pauli, 611 F.3d 828, 842 (11th Cir. 2010).
violated the defendants’ Fourth Amendment rights, not whether the pre-existing warrant is valid. Even if the pre-existing warrant is valid, the defendants likely do have a reasonable expectation of privacy in the contents of their iPads that do not fall within the scope of the warrant. Therefore, the All Writs Act order runs the risk of violating the defendants’ privacy rights by exceeding the scope of the search warrant. If it exceeded the scope of the search warrant, then the defendants’ Fourth Amendment rights would be violated.157

VI. ALL WRITS ACT AND THE SEPARATION OF POWERS: COMMUNICATIONS ASSISTANCE FOR LAW ENFORCEMENT ACT

Because the All Writs Act only applies to inadvertent gaps in legislation, its application to privacy raises two issues: first, an exemption under the CALEA arguably addresses this gap; second, application of the Act may undermine separation of powers in light of legislative proposals to address this issue. Here, I argue that a gap indeed exists under the CALEA but that the New York Telephone Co. framework fails to take into account the reason for that gap. The framework is insufficient because it does not consider scenarios where the gap is not the result of an unintended oversight but rather the result of legislative failure to reach a consensus. Where it is clear that a gap was the result of legislative gridlock, the use of the All Writs Act as a gap-filler is inappropriate. Such usage undermines separation of powers and judicial independence because it turns courts into an expedient way to circumvent legislative consensus. It undermines both the role of the court as a neutral decision-maker and the functioning of a healthy democracy.

A. CALEA and Device Decryption

The CALEA was designed primarily to preserve the ability of law enforcement agencies to engage in lawful electronic surveillance. However, it “addressed the responsibilities of private companies to preserve and allow access to records relating to wire and electronic communications.”158 The text of the CALEA provides three exemptions that limit its application to specific circumstances.159 One of these limitations, section 1002(b)(2), exempts equipment manufacturers from providing certain assistance to law enforcement listed in section 1002(a), which requires that telecommunications carriers be capable of “enabling the government, pursuant to a court order or other lawful authorization, to intercept, to the exclusion of any other com-

157 See Horton v. California, 496 U.S. 128, 140 (1990) (“If the scope of the search exceeds that permitted by the terms of a validly issued warrant . . . the subsequent seizure is unconstitutional without more.”).
munications, all wire and electronic communications.” However, “information services” and “equipment, facilities, or services that support the transport or switching of communications for private networks or for the sole purpose of interconnecting telecommunications carriers” are exempt from such requirements under section 1002(b)(2). “Information services” is defined as the “offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information via telecommunications.” This category includes companies like Apple since products like the iPad and iPhone provide information services.

The fact that Congress exempted information services from the CALEA’s obligations is significant—the legislative history of the CALEA explained that one of the considerations of doing so was “to avoid impeding the development of new communications services and technologies.” Indeed, given that information services were still in their infant stage in the 1990s, when the CALEA was passed, regulations on how information services companies could design their products would have stunted innovation.

Since its passing, a robust legislative history shows that Congress repeatedly considered amending the CALEA to either require or prohibit compelled decryption by private information services. In 2015, for example, a proposed bill attempted to insulate companies like Apple from decryption orders. Meanwhile, other legislative proposals tried to achieve the opposite outcome. In 1991, then Senator Joe Biden (D-Del.) introduced the Comprehensive Counter-Terrorism Act of 1991 that sought to expand the CALEA to require third-party assistance to decode consumer devices. Additionally, in July 2015, the Senate held hearings in an effort to craft an approach to balance privacy and law enforcement interests with respect to smart phone encryption. Rather than an unintentional “gap” in legislation, the issue of whether private entities can be compelled to decrypt consumer products for law enforcement purposes is in fact a quagmire of frequent legislative debates.

---

160 Id. § 1002(a)(1).
161 Id. § 1002(b)(2).
162 Id. § 1001(6)(A).
163 Moreover, a second exemption broadly frees telecommunication carriers from “decrypting, or ensuring the government’s ability to decrypt” communications. Id. § 1002(b)(3). This provision specifically contemplates whether a “telecommunications carrier” can be required to decrypt communication encrypted by a customer. It narrowly provides that such decryption is only possible where the (1) encryption was provided by the carrier, and (2) the carrier possess the information necessary to decrypt the communication. Id.
This section proposes that courts should adopt a modified New York Telephone Co. test that examines the legislative history of a statutory gap. The Supreme Court has long prescribed the All Writs Act to function as a gap-filler that “renders it unnecessary for Congress to anticipate every circumstance in which a federal court might properly act to vindicate the rights of the parties before it.” It has also drawn the boundaries of the Act at two extremes: at one end, the Act cannot be interpreted to empower courts to do something already specifically authorized by another statute. At the other end, it cannot be interpreted to authorize something specifically prohibited by an existing statute. This results in an incredible amount of latitude.

The gulf between these extremes means a variety of gaps exist that the All Writs Act may fill. Therefore, courts should take into account the legislative history of a particular gap so as to understand whether it was the result of intentional deliberation, unintentional inadvertence, or political quagmire. The use of the Act in the third scenario is the most dangerous. Where a gap is the result of a lack of political consensus, the Executive Branch, if successful in seeking the writ, would achieve a legislative goal that Congress has considered but failed to address. Such a result would undermine the separation of powers laid out in our Constitution by transforming the All Writs Act from a jurisdiction-preserving device meant to protect the integrity of the judiciary into an executive tool to circumvent legislative powers that only Congress possesses. It is therefore inconsistent with the “usages and principles” of the Act. Hence, a court applying the All Writs Act should consider whether the use of the Act exceeds its original function by examining the nature of a legislative gap when applying the New York Telephone Co. test.

Furthermore, the necessity prong of the All Writs Act also means that the government should exhaust other remedies before seeking a decryption order. However, too often, the government rushes into court without seeking these other remedies first. A grand jury subpoena to the defendant to release the password to a personal device or cloud account is the most direct method possible. In fact, in some cases, defendants have voluntarily surrendered their passwords upon being asked by the police. Where the defendant refuses to cooperate with a subpoena, the district court may place the defendant


170 See, e.g., United States v. Apple Mac Pro Computer, 851 F.3d 238, 248 (3d Cir. 2017) (“Forensic analysts also found an additional 2,015 videos and photographs in an encrypted application on Doe’s phone, which Doe had opened for the police by entering a password.”).
under contempt. However, these remedies are not without limitations, since for some defendants, the threat of imprisonment for contempt may far outweigh much harsher sentences. Yet the government should be required to show that it has at least attempted these options before seeking an All Writs Act order, especially since the former are much less onerous and implicate fewer constitutional and privacy concerns. Therefore, a minimum requirement in applying the *New York Telephone Co.* test should be for the government to exhaust these options before invoking the All Writs Act.

### VII. Conclusion

This Article addresses an emerging issue that will likely become increasingly relevant as long as no legislative solution is provided to address personal device decryption. As the frequency and scope of such decryption orders increase, it is only a matter of time before the issue of whether the All Writs Act is a valid means of forcing decryption by third-party providers must be addressed by either the Supreme Court or Congress. How this issue is resolved will not only affect how criminal investigations proceed in the future but also shape the jurisdictional limits of the judiciary.

The existing framework from *New York Telephone Co.* is no longer sufficient to delineate when the All Writs Act should apply to compel decryption of personal devices. This framework fails to take into account situations where substantial legislative gridlock causes a statutory gap such that the use of the Act would undermine the separation of powers doctrine. The All Writs Act was only meant to preserve the courts’ existing jurisdiction, not to fulfill the Executive Branch’s wishes where Congress has failed to provide an answer. This improper use of the Act in the decryption context is even more apparent when one examines the jurisdictional basis, or lack thereof, that the courts have over third-party and private entities like Apple.

The use of the All Writs Act as a blunt tool for the Executive Branch is also particularly dangerous when legitimized by the Judiciary. Private entities, whose privacy interests are not at stake, are not as incentivized as individual defendants to challenge such orders. Meanwhile, individual defendants lack the resources and access to proprietary information to adequately challenge such orders in court.

From its early days as a mere mechanism for preserving courts’ jurisdiction to its recent role in the privacy context, judicial grant of All Writs Act orders may become dangerously close to judicial activism. Left unaddressed, this trend of using the All Writs Act to order private entities to decrypt private consumer devices for government investigations may become an insidious judicial norm that chips away not only at the rights of defendants and the economic freedom of corporations, but also at the very foundation of our democracy.
NOTE
THE DISTORTIVE EFFECT OF THE NATIONAL PRACTITIONER DATA BANK ON MEDICAL MALPRACTICE LITIGATION AND SETTLEMENT

GEORGE MALIHA*

Table of Contents

I. INTRODUCTION ................................................. 287
II. LIABILITY INSURERS, PHYSICIANS, AND THE NPDB .......... 290
   A. The Basics of the Interaction Between the Insured And the Insurance Company .......................... 290
   B. An Overview of the NPDB .................................... 294
   C. The Pernicious Influence of the NPDB on Physician Behavior in Medical Malpractice Litigation .......... 297
III. PHYSICIANS’ FAILED ATTEMPTS TO CHALLENGE AN INACCURATE OR MISLEADING NPDB REPORT .............. 300
IV. ANOTHER PARADIGM OF A GOVERNMENT-SUPERVISED “BLACKLIST”: CREDIT REPORTING ......................... 305
V. RECOMMENDATIONS ............................................. 308
VI. CONCLUSION ..................................................... 309

I. INTRODUCTION

Congress created the National Practitioner Data Bank (“NPDB”) in 1986 to address a concern that medical liability cases were increasing throughout the nation.\(^1\) In order to prevent physicians from moving from state to state in order to escape a poor outcome, the NPDB was supposed to provide a central clearinghouse of information for every physician in the country—regardless of where they practiced.\(^2\) The idea was quite simple: whenever a physician was involved in a malpractice case or a professional disciplinary action, it would be reported to the NPDB and form a record for that physician.\(^3\) New practices considering working with that physician

---


2. Id.

3. See id.
could request the information as part of evaluating him or her for credentialing.

However, the NPDB distorts medical malpractice litigation and settlement—harming defendant-physicians, plaintiff-patients, and insurers. Hospitals considering credentialing or renewing privileges for a physician effectively must request the NPDB report—they will be penalized in a malpractice action if they do not. Further, the NPDB plays an important role in determining whether to allow the physician to practice at the institution. Thus, the fear of an NPDB entry from malpractice litigation or settlement is driving physicians to press their insurers to mount a complete defense and litigate claims to judgment. In some cases, operating under laws conceived to prevent insurers from leaving their clients with judgments over policy limits, insurers have incentives and the legal right to settle against a physician’s wishes and interests. In other cases, physicians and insurers prolong litigation where both sides would be open to settlement with the plaintiff had the NPDB not provided a barrier.

Webb v. Witt illustrates this conflict and its harm. Empress Nazirah Ashanti Webb was delivered with an injury to her brachial plexus, causing permanent paralysis of her right arm. In the resulting medical malpractice suit against the physicians and hospital, the insurer recommended settlement. One of the physicians, Dr. Susan Kaufman, D.O., resisted this move.

4 See id. at A-5–7.
6 42 U.S.C. § 11135(a)–(b).
8 Whether a physician can press his or her insurer to litigate depends on their contractual arrangement. Some states require insurers to provide “no consent” to settle insurance policy options. See, e.g., N.Y. COMP. CODES R. & REGS. tit. 11, § 70.2 (2016); see also Waters, supra note 7. However, some states require insurers to not give physicians the right to veto a settlement within policy limits. See, e.g., FLA. STAT. § 627.4147(b)(1) (2016).
9 Medical malpractice litigation can be quite lengthy already, especially if settlement negotiations break down and a case goes to trial. See Neil Vidmar, Medical Malpractice and the American Jury: Confronting the Myths about Jury Incompetence, Deep Pockets, and Outrageous Damage Awards 49–67 (1995).
12 Webb, 876 A.2d at 861–62.
13 See id.
14 Id. at 862. The role of the hospital as the insurer or acquirer of insurance adds yet another conflict to the picture, but it does not change the fundamental distorting role of the NPDB. It will not be analyzed in depth in this paper. In this case, the hospital was the only named insured on the malpractice policy, but the insurance policy at issue stated that the insurer must make “a reasonable attempt to consult” Kaufman before settling. Id. at 861. This wrinkle recapitulates the fact that many malpractice insurance policies do not require the physician-insured to consent to settlement (so-called “pride provisions”). Kent D. Syverud, The Duty to Settle, 76 VA. L. REV. 1113, 1175–76 (1990). Another variation, “hammer” clauses, require consent not to be withheld unreasonably. Jonathan L. Schwartz & Seth L. Laver, The
to settle on her behalf, arguing that the settlement would reduce her ability to practice medicine by preventing her from entering insurer networks and obtaining hospital privileges.\textsuperscript{15} She specifically alleged that reporting a settlement to the NPDB would prevent her from securing future medical liability insurance.\textsuperscript{16} The court rejected her claims and allowed the insurer to settle on the hospital and physicians’ behalf.\textsuperscript{17} Yet, that same court noted “that Kaufman’s main concern does not appear to be the settlement itself, but rather how it is reported to the federal and state databanks. Clearly, her interests are served by a settlement of the claims, so long as the reported percentage of her responsibility is zero.”\textsuperscript{18} Thus, the physician was unable to litigate her claims, the insurer chose to settle a claim against the interest of the insured, and the resulting litigation delayed payment to the plaintiff—a payment that all parties appeared to agree was reasonable.

The NPDB’s brooding shadow over medical malpractice has led many litigants and commentators to term it a “blacklist.”\textsuperscript{19} Part II explores whether this term is appropriate by describing the NPDB in the context of insurer-physician relations. This discussion will connect the well-described model of insurer-insured relations to the prescient concerns raised about the NPDB’s potential distortive effect on litigation and settlement as the data bank was being enacted in the late 1980s. Part III will explore mechanisms to alter reports—and place a physician’s “side” into the record kept by the NPDB. Attempts to alter reports have triggered litigation against reporting entities and the NPDB itself, and although they have largely failed, these suits illustrate the unique problems that the NPDB causes physicians. In Part IV, these unsuccessful suits will be contrasted against a body of law surrounding the accuracy of another putative “blacklist”—credit scores.\textsuperscript{20} Part V will begin to sketch out some basic policy recommendations.


\textsuperscript{15} Id. at 862–63.

\textsuperscript{16} Webb, 876 A.2d at 863. Dr. Kaufman also claimed that similar effects would arise from reporting to New Jersey’s data bank. See id. at 865–66; see also N.J. STAT. ANN. § 45:9–22.22–25 (West 2016).

\textsuperscript{17} The court rejected Dr. Kaufman’s arguments on procedural and substantive insurance law theories. See Webb, 876 A.2d at 863.

\textsuperscript{18} Id.


Liability insurance overshadows the conduct of malpractice litigation. The terms of agreements between insurers and physicians materially affect the conduct of a given malpractice litigation. And, those same terms can lead to yet more conflict after a malpractice suit has concluded.\textsuperscript{21}

This part will focus on the decision to continue to litigate or to settle. As such, this section is divided into three parts: Part A will provide a basic overview of the complex—and sometimes counterintuitive—dynamics that connect a liability insurer, insured, defense counsel, plaintiff, plaintiff’s counsel, and juries in a medical malpractice situation. Part B will explain what the NPDB actually is and briefly outline its operation. This part will omit a lengthy discussion of physician peer review boards and antitrust immunity, though that part of the law has also been criticized recently.\textsuperscript{22} Part C will bring Parts A and B together, demonstrating how the NPDB distorts medical malpractice litigation and settlement to the detriment of physicians and plaintiffs. Although the subject of this paper is the federal NPDB, state malpractice databases should, presumably, produce similar effects on litigation and settlement within their respective boundaries.\textsuperscript{23}

\textbf{A. The Basics of the Interaction Between the Insured and the Insurance Company}

Medical malpractice is not merely the conflict between a physician and his or her former patient. “Tort litigation and liability insurance are symbiotic institutions . . . .”\textsuperscript{24} Courts and legislatures have imposed obligations on insurers—such as the duty to defend or settle—in order to try to align the incentives of parties.\textsuperscript{25} Nonetheless, a typical medical malpractice suit in-

\textsuperscript{21} Even the quintessential 1L contracts and malpractice case, \textit{Hawkins v. McGee}, 146 A. 641 (N.H. 1929), spawned a conflict between Dr. McGee and his insurer to cover those now-famous damages described in the case. \textit{See} McGee v. U.S. Fid. & Guar. Co., 53 F.2d 953, 956 (1st Cir. 1931) (holding that the insurance company was under no obligation to cover the judgment against Dr. McGee).

\textsuperscript{22} See 42 U.S.C. §§ 11111–11115; \textit{see also} Van Tassel, supra note 19, at 2052–57.

\textsuperscript{23} \textit{See generally} Jeffrey P. Donohue, Note and Comment, \textit{Developing Issues Under the Massachusetts ‘Physician Profile’ Act}, 23 AM. J.L. & MED. 115 (1997). Several states have their own malpractice reporting requirements or authorize their state medical boards to obtain such information from courts. \textit{See, e.g.}, ALA. CODE § 34-24-56 (2016); ALASKA STAT. ANN. § 08.64.345 (West 2016); ARK. CODE ANN. § 17-95-103 (West 2016); MASS. GEN. LAWS ANN. ch. 112, § 5C (West 2016); MICH. COMP. LAWS ANN. § 333.16231 (West 2016); NEV. REV. STAT. ANN. § 630.173 (West 2016); OHIO REV. CODE ANN. § 4731.281 (West 2016); P.R. LAWS ANN. tit. 20, § 134a (2013).

\textsuperscript{24} Syverud, supra note 14, at 1114.

\textsuperscript{25} This paper does not purport to cover all the nuances of this area of law, but a good review can be found at \textit{Tom Baker, Insurance Law and Policy} 524–636 (2d ed. 2008). Further, regardless of whether an insurance company may or may not settle over the objections of its insured, it must inform the insured as to its action. \textit{See, e.g.}, Rogers v. Bobson, 407 N.E.2d 47, 49 (Ill. 1980).
volves four major players: 1) the plaintiff (often represented by a lawyer working for a contingent fee), 2) the defendant-physician, 3) the medical malpractice liability insurer, and 4) the lawyer for the physician hired by the insurer.26 All four players have distinct interests that may or may not align depending on the posture and merits of the litigation. However, to simplify the analysis for our purposes, this paper concentrates upon the physician and his or her liability insurer.27

Liability insurance law, however, did not develop specifically to meet the concerns of physicians and address medical malpractice—but to protect consumers who could not afford to pay judgments above their liability insurance caps. Crisci v. Security Insurance Co. of New Haven, Connecticut28 provides a paradigmatic example. Rosina Crisci owned an apartment building where June DiMare was injured when a step broke.29 The DiMares sued Ms. Crisci for $400,000. Ms. Crisci had liability insurance for $10,000, and she triggered the policy provisions obligating Security Insurance to defend her and “make any settlement it deemed expedient.”30 In its investigation, Security Insurance determined that if the jury believed the DiMares’ claims, the verdict would be no less than $100,000. However, Security Insurance refused to settle the case for $6500.31 (The insurance company would only be on the hook for $10,000 anyway.) The jury returned a verdict of $101,000, leaving Ms. Crisci with liability over nine times her policy limit.32 Nonetheless, the court found the insurer liable for the excess over the policy limit.33 While physicians are protected by the same rule,34 physicians sued for medical malpractice often suffer from a different problem: they frequently want to risk the expense and unpredictability of discovery, depositions, trials, and juries to get a vindicating verdict—with their liability insurance potentially footing the bill.35 The issue for the physician, then, is not only the

27 Physicians and insurers also conflict with the appointed counsel. See generally Richard H. Underwood, The Doctor and His Lawyer: Conflicts of Interest, 30 KAN. L. REV. 385 (1982).
29 Id. at 175.
30 Id.
31 See id. at 175–76. Ms. Crisci was willing to pay $2,500 of a $9,000 settlement offer by plaintiff. Id.
32 Id. at 176.
33 Id. at 178. The Crisci or “reasonable offer” rule, with some modifications, is now the majority rule defining the insurer’s duty to settle. “Under the reasonable-offer test, the question is whether an insurer under a policy without limits would have accepted the offer.” KENNETH S. ABRAHAM & DANIEL SCHWARCZ, INSURANCE LAW AND REGULATION 615 (6th ed., 2015).
34 See, e.g., Garner v. Am. Mut. Liab. Ins. Co., 107 Cal. Rptr. 604, 608–09 (Ct. App. 1973) (holding that even when physicians’ attorneys testify that they would not expect the jury to return a verdict above the policy limits, an insurance company must come to its own independent judgment as to merits of a physician’s case and cannot delegate that responsibility to a board that does not bear the risk and liability for a jury award above the policy limit).
money but also the reputational and long-term costs of even a settled medical malpractice suit. Beyond tangible costs, medical malpractice can trigger deep emotional turmoil and uncertainty in the physician, patient, and patient’s family. Magnifying these issues is the fact that the language of many malpractice insurance contracts gives control of settlement to the insurer—who does not have to contend with the reputational and other long-term costs of a lawsuit. The next several cases provide examples of courts ignoring these concerns. The issue of whether the physician committed malpractice is less relevant—but the courts’ refusal to engage in any more nuanced analysis of the costs of a malpractice lawsuit to a physician is.

For instance, some courts have refused to recognize situations where the duty to settle does not fit the circumstances well. *Harrison v. Long* illustrates this problem. Dr. Long was sued for malpractice by the parents of Bradley Harrison. Accordingly, Dr. Long activated his malpractice liability insurance from St. Paul Fire and Marine Insurance Co. St. Paul found it expedient to settle the case over the policy limits, so the insurer notified the Kansas Health Care Stabilization Fund, which would cover the settlement amount over the liability cap. Dr. Long objected to settlement on due process and property rights grounds as depriving his “right” to defend himself against accusations of malpractice. But, the court rejected his arguments as he had contracted with the insurance company to defend him and did not a physician perceives that a malpractice claim is legitimate, he or she will often push to settle to minimize the publicity and expense of the mistake. See *id.* at 546.  

See *West Wake Price & Co v. Ching* [1957] 1 W.L.R. 45, 53 (QB) (noting a lawsuit may damage a professional’s reputation even if the suit is not successful).  


Medical malpractice policy language is often not public, but it does enter public records in litigation over the meaning of the insurance contract. See, e.g., Mitchum v. Hudgens, 533 So. 2d 194, 196 (Ala. 1988) (emphasis added) (“‘We’ll defend any suit brought against you for damages covered under this agreement. We’ll do this even if any of the allegations of the suit are groundless, false or fraudulent. The company shall have the right and duty to defend any suit against the named insured seeking such damages, even if any of the allegations of the suit are groundless, false or fraudulent. The company may make such investigation and such settlement of any claim or suit as it deems expedient.’”). But see *Aquilina v. O’Connor*, 399 N.Y.S.2d 919, 921 (App. Div. 1977) (emphasis added) (“‘[E]ven if such suit is groundless, false or fraudulent; but the company may make such investigation, negotiation and settlement of any claim or suit as it deems expedient.’”). See also *KAN. STAT. ANN.* § 40-3403 (2016). Physicians paid a premium for the additional protection by the state. *Harrison*, 734 P.2d at 1157.  

Kansas had enacted the Fund to cover such damages “as a partial response to the increasing pressure of what was termed a national medical malpractice crisis.” *Id.; see also*
2019] National Practitioner Data Bank 293

have an overriding constitutional right to litigate.\textsuperscript{44} So, even though the liability insurer could still go to the Stabilization Fund after trial, St. Paul chose to settle the case over a physician objection.\textsuperscript{45}

Further, courts do not compute the devastating reputational effects that a settlement can inflict on physicians. Feliberty v. Damon\textsuperscript{46} provides one example of this type of conflict. Thomas Michaels, an iron worker, visited Dr. Feliberty, a Buffalo physician, complaining of “recurring throat pain and lumps in his neck.”\textsuperscript{47} After a brief exam, Dr. Feliberty recommended Michaels get an x-ray of his chest.\textsuperscript{48} Michaels did not return to see the physician.\textsuperscript{49} Two-and-a-half years later, Michaels sued Dr. Feliberty for failure to diagnose lymphoma.\textsuperscript{50} A medical malpractice panel unanimously found negligence, and the case proceeded to trial.\textsuperscript{51} The final judgment against Dr. Feliberty was $743,000.\textsuperscript{52} Although the physician wished to appeal, the insurer settled for $700,000—within Dr. Feliberty’s liability limits.\textsuperscript{53} Despite the physician alleging that his medical practice had been ruined and that he had been forced to leave his practice location, the court found no problem with the insurer’s conduct because the insurance contract had given the insurer an exclusive right to settle.\textsuperscript{54} The physician could have purchased more expensive insurance if he wished to have a “consent to settle” clause.\textsuperscript{55}

Even though a physician’s loss in court heightens the conflict between an insurer who wishes to settle and a physician who wishes to appeal, an outcome favorable to a physician presents a similar conflict. In Aquilina v. O’Connor,\textsuperscript{56} Dr. Aquilina, a neurosurgeon, was sued by the estate of a deceased patient.\textsuperscript{57} Dr. Aquilina “moved for summary judgment upon the grounds that he had not been the treating physician, only an emergency con-

\textsuperscript{44} Id. at 1160–61. Dr. Long’s required contribution to the Stabilization Fund increased 400 percent. Id. at 1158.
\textsuperscript{45} Presumably, the insurer was attempting to save money on lawyers’ fees and litigation costs, for the Stabilization Fund did cover both judgments and settlements over liability insurance limits, see KAN. STAT. ANN. § 40-3403(c)(1) (“Any amount due from a judgment or settlement which is in excess of the basic coverage liability of all liable resident health care providers or resident self-insurers for any personal injury or death arising out of the rendering of or the failure to render professional services within or without this state . . . .”).
\textsuperscript{46} 527 N.E.2d 261 (N.Y. 1988).
\textsuperscript{47} Id. at 261.
\textsuperscript{48} Id.
\textsuperscript{49} Id. at 261–62.
\textsuperscript{50} Feliberty, 527 N.E.2d at 262; see also N.Y. JUD. LAW § 148-a (McKinney 1983) (repealed 1991). New York no longer uses mandatory medical malpractice panels.
\textsuperscript{51} Feliberty, 527 N.E.2d at 262.
\textsuperscript{52} Id.
\textsuperscript{53} Id. (“While the settlement was within policy limits and plaintiff therefore technically suffered no out-of-pocket loss, he is understandably concerned about . . . a different interest—his professional reputation. This . . . contract, however, specifies that the ‘company may make such investigation and such settlement of any claim or suit as it deems expedient.’”).
\textsuperscript{54} Id.
\textsuperscript{56} Id. at 920.
sultant, and that his diagnosis had been correct and proper.” 58 Plaintiff wished to discontinue the action and Aquilina’s insurance-retained attorney negotiated a settlement dismissing the action with prejudice. However, Dr. Aquilina desired to sue plaintiff and her attorney for malicious prosecution and abuse of process—but was unable due to the settlement. 59 Despite a “consent to settle” clause in the insurance policy, the court held that dismissal of a suit negotiated between the two parties does not qualify as a settlement and that Dr. Aquilina had no cause of action. 60

Thus, even before the NPDB was enacted, physicians and insurers suffered a potentially severe conflict of interest. 61 However, the NPDB would provide an even stronger incentive to physicians not to settle, worsening the conflicts illustrated above. Before turning to the NPDB’s effect on litigation, this Article turns to the question of what the NPDB is and its basic operation.

B. An Overview of the NPDB

The NPDB arose out of a well-founded concern about physicians escaping liability for malpractice. Congress found that “[t]he increasing occurrence of medical malpractice and the need to improve the quality of medical care have become nationwide problems . . . ” and that “[t]here is a . . . need to restrict the ability of incompetent physicians to move from State to State . . . .” 62 As part of a program to fix this problem, Congress mandated that any insurer making a settlement on a malpractice claim report it to the NPDB. 63 Other reportable incidents to the NPDB include sanctions by state medical boards, 64 disciplinary actions (and some investigations) by hospital peer-review boards, 65 and criminal convictions or exclusion from government healthcare programs. 66 NPDB reports should include basic demographic information, the amount of a settlement, a description of settlement or judgment terms, and a “description of the acts or omissions and injuries

58 Id.
59 Id.
60 Id. at 921.
63 Id. § 11131(a) (“Each entity (including an insurance company) which makes payment under a policy of insurance, self-insurance, or otherwise in settlement (or partial settlement) of, or in satisfaction of a judgment in, a medical malpractice action or claim shall report . . . information respecting the payment and circumstances thereof.”); see also 45 C.F.R. §§ 60.7, .14, .16 (2016).
64 42 U.S.C. § 11132; see also 45 C.F.R. §§ 60.8–10.
65 42 U.S.C. § 11133; see also 45 C.F.R. §§ 60.11–12.
National Practitioner Data Bank

or illness upon which the action or claim was based . . . .” The narrative portion of the report—the critical description of the acts or omissions—is capped at 4000 characters, including punctuation and spaces.

The law also imposes obligations on hospitals and health care systems. Hospitals are required to request information from the NPDB whenever they are considering whether to grant privileges to a physician and every two years thereafter. Although the law presumes that “a payment in settlement of a medical malpractice action or claim shall not be construed as creating a presumption that medical malpractice has occurred,” some courts have found that the NPDB reports are discoverable in corporate liability litigation, such as negligent credentialing cases. Further, the law provides some modest penalties for not reporting, though there is little evidence that the penalties are being enforced.

Although the NPDB embodies laudable goals, Congress was made aware of potential issues with collecting and reporting medical malpractice settlements. First, settlements, especially low-value settlements, reflect mere nuisance value and not the merits of a case, for “a malpractice suit in and of itself . . . may have nothing whatsoever to do with quality of care . . . .”

---

67 Id. § 60.7; see also 42 U.S.C. § 11131(b).
68 NPDB GUIDEBOOK, supra note 1, at E-11.
69 45 C.F.R. § 60.17; see 42 U.S.C. § 11135(a).
70 42 U.S.C. § 11137(d) (emphasis added).
71 See, e.g., Klaine v. S. Ill. Hosp. Servs., 47 N.E.3d 966, 975 (Ill. 2016) (“[W]e believe it is clear that information reported to the NPDB, though confidential, is not privileged from discovery in instances where, as here, a lawsuit has been filed against the hospital and the hospital’s knowledge of information regarding the physician’s competence is at issue.”).
72 See 42 U.S.C. § 11131(c).
73 Teninbaum, supra note 19, at 90 (“History reveals that the NPDB has been unable or unwilling to enforce the rules: in response to a Freedom of Information Act request made during the drafting of this article, HHS has never . . . levied a single fine against any person or entity for failure to report a malpractice claim.”).
74 Medical Malpractice: Hearing on H.R. 5110 Before the Subcomm. On Health & the Env’t of the H. Comm. on Energy and Commerce, 99th Cong. 305 (1986) [hereinafter Medical Malpractice Hearing] (statement of John Horty, President, National Council of Community Hospitals) (“Cases often are settled merely for their nuisance value, and it would be misleading to include such a settlement in the reporting mechanism. In theory the persons evaluating the information reported to the Secretary will be professionals and will be able to evaluate the significance of settlements. Nevertheless, we do not think it is in anyone’s interest that information about small settlements be maintained and circulated, however much they be confidential in theory.”); id. at 328–29 (statement of Dr. Harry S. Jonas, President, American College of Obstetrics & Gynecology) (“Malpractice settlements frequently reflect economic considerations on the part of the insurer and the practitioner involved. It is often cheaper and less traumatic to settle a claim rather than go to court and fight the case on its merits. These decisions are totally outside the scope of any peer review process and may or may not reflect upon a physician’s competence. Likewise, claims filed against a physician represent mere allegations not conclusive proof of malpractice.”); id. at 442 (statement of Dr. Raymond Scalettar, Member of Board of Trustees, American Medical Association) (“Such data will include an extremely large amount of complex and possibly misleading information that will necessitate careful interpretation to be meaningful. . . . Other suits may be settled for small, and even relatively substantial amounts. These amounts may reflect the nuisance value, costs of litigation and the possibility—particularly in severe injury cases—that juries will find for the plaintiff more out of sympathy than a conviction that true negligence occurred.”); id. at 465.
Settlements, in fact, “are very often notoriously misleading guides to anything that happened and it is not clear . . . that the probative value of a settlement would be the same as that of the other information to be reported.” This concern led to a proposal to provide a minimum threshold for reporting, around $10,000 to $25,000. Critically, one witness (a member of a hospital association) testified that physicians may well choose to fight a nuisance claim rather than settle, be reported to the NPDB, and face a record when he or she chooses to apply for future hospital privileges. The American Medical Association agreed.

A second problem was that medical malpractice is not uniform across medical specialties and practice settings. Some physicians can expect multiple lawsuits against them, others not so much. This is not to say that the number of filed, settled, or otherwise resolved suits is irrelevant, but the number and value of claims must be understood in the context of the physician’s specialty, practice, and location. For instance, surgical specialists are sued more, while pediatricians are sued less often but tend to have larger

---

75 Id. at 350 (statement of Dr. Harry L. Metcalf, Chairman, Board of Directors, American Academy of Family Physicians). He continued: “Because this legislation appears to require reporting of . . . every payment under an insurance policy resulting from a malpractice judgment or settlement we question whether the reporting requirement will have the desired result.” Id.


77 See, e.g., Medical Malpractice Hearing, supra note 74, at 305, 313–14 (statement of John Harty, President, National Council of Community Hospitals); id. at 316 (statement of Richard P. Kusserow, Inspector General, United States Department of Health and Human Services) (“A lot of different States have been wrestling with this problem. I will mention Georgia set it at $10,000; New Jersey, $25,000; and California is at $30,000.”).

78 See id. at 313–14 (statement of John Harty, President, National Council of Community Hospitals) (“I guess my concern is that a lot of nuisance claims are filed, a lot of nuisance claims are settled rather than being fought when perhaps they should be fought, and you end up with people having a certain record that might get into other hands when it got into the hospitals and have an effect on the physician.”). It is important to note Mr. Harty was not representing the medical profession, but a hospital association, when he broached the concern.

79 See id. at 442 (statement of Dr. Raymond Scalettar, Member of Board of Trustees, American Medical Association) (“The number of cases going to litigation also would increase as the reporting requirements would make defendants less willing to settle cases.”).

80 Id. at 29 (statement of Dr. Harry S. Jonas, President, American College of Obstetrics & Gynecology) (“The significance of a certain number of claims depends on a number of factors, including the number of patients seen or procedures performed by the physician, the type of subspecialty practice in which the physician is engaged, and even the fact that the physician is chairman of a department and thus, likely to be named automatically in a great number of suits. As the Dean of a medical school, and as a former head of the obstetrics-gynecology department, I have been named to a number of suits for which I had no direct involvement.”); id. at 442 (statement of Dr. Raymond Scalettar, Member of Board of Trustees, American Medical Association) (“An average neurosurgeon, for example, has been sued four times. Some of the suits may involve errors of judgement or mistakes under the extraordinary circumstances of their practice, which says nothing about the physician’s competence.”).
2019] National Practitioner Data Bank 297

damages since children can need a lifetime of care.\textsuperscript{81} Third, even by the early 1980s, many physicians had ceded control of the litigation strategy to their medical malpractice liability carriers, which could make decisions “without regard to the merit of the claim.”\textsuperscript{82}

In the end, Congress valued disclosure of all malpractice settlements over these concerns expressed by physicians, attorneys, and hospital administrators. There is no indication that the courts perceived or Congress intended the NPDB to be such a high-stakes endeavor for physicians. Indeed, Congress’s findings of fact indicate that the law is supposed to “restrict the ability of incompetent physicians to move from State to State without disclosure or discovery of the physician’s previous damaging or incompetent performance.”\textsuperscript{83} The courts continue to downplay the results of a report to the NPDB. One described the NPDB “by itself” as not quite a rebuke but merely an informational disclosure, stressing that “[t]he official purpose of the report is to disclose information, not to reprimand.”\textsuperscript{84} Another minimized its role: “Information in the [NPDB] is intended ‘only to alert . . . health care entities that there may be a problem with a particular practitioner’s professional competence or conduct’ because the practitioner has been the subject of a disciplinary action.”\textsuperscript{85}

However, the NPDB’s actual results on litigation and settlement are different.

C. The Pernicious Influence of the NPDB on Physician Behavior in Medical Malpractice Litigation

The NPDB weighted the scales further against physicians settling cases. As discussed in Part I.A., physicians were already less likely than a typical

\textsuperscript{81} See id. at 365 (statement of Dr. Harry S. Jonas, President, American College of Obstetrics & Gynecology) (“It is sometimes a long time before settlements or final judgments come down. It could be a matter of years. If there is a physician who has 26 suits filed against him or her, that information should alert entities to inquire further. That physician should not be unduly judged because of the point I made about malpractice suits not necessarily meaning incompetence. Nevertheless, the suits filed could be relevant to the medical staff considering privileges for that individual.”). See generally Anupam B. Jena et al., Malpractice Risk According to Physician Specialty, 365 New Eng. J. Med. 629 (2011).

\textsuperscript{82} Medical Malpractice Hearing, supra note 74, at 286 (statement of Jack Owen, Executive Vice-President, American Hospital Association) (emphasis added) (“It is well known that the majority of small malpractice payments represent the disposition of ‘nuisance’ claims—actions in which the costs of defense would exceed the settlement payment. Most decisions in this regard are made by a provider’s insurance carrier or fund, without regard to the merit of the claim. It is also recognized that the vast majority of these claims are in fact meritless and often are made in the anticipation that a ‘nuisance’ settlement will be offered.”).

\textsuperscript{83} 42 U.S.C. § 11101(2); see also H.R. Rep. No. 99-903, at 2 (1986), reprinted in 1986 U.S.C.C.A.N. 6384 (“The purpose of this legislation is to improve the quality of medical care by encouraging physicians to identify and discipline other physicians who are incompetent or who engage in unprofessional behavior.”).

\textsuperscript{84} Rochling v. Dep’t of Veterans Affairs, 725 F.3d 927, 932 (8th Cir. 2013).

\textsuperscript{85} Leal v. Sec’y, U.S. Dep’t of Health & Human Servs., 620 F.3d 1280, 1283 (11th Cir. 2010) (emphasis added) (citing NPDB GUIDEBOOK, supra note 1, at A-3).
defendant to settle, but the NPDB enhanced that penchant further. A 2003 study of the claims processed by several medical malpractice carriers showed that the number of malpractice suits resulting in payment has fallen and defense costs have risen.86 While these changes are concentrated in small-claim medical malpractice—where suits are less likely to be meritorious anyway—increased defense costs often translate to higher malpractice premiums.87

Although plaintiff’s lawyers may screen claims more effectively in order to try to prevent protracted litigation with a physician attempting to avoid reporting,88 NPDB obligations are also driving a spate of litigation between physicians and insurance companies on whether to settle and how precisely to apportion blame.89 The issue in these cases is not the amount of the settlement. Rather, it is whether the physicians involved will be reported or whether a hospital insurance plan or defendant will take the entire blame and not report to the NPDB. For instance, in *Melendez v. Hospital for Joint Diseases Orthopedic Institute*,90 an orthopedic resident was apportioned a quarter of the blame for an incident that the hospital insurance plan had decided to settle.91 The resident challenged the apportionment to prevent reporting to the NPDB at such an early stage of his career.92 Although the court decided against the resident on procedural grounds,93 it failed to consider that residents are supposed to operate under the supervision of an attending physician.94

Another case illustrates how the NPDB can punish physicians for performing emergency care and correcting others’ mistakes. In *Doe v. South Carolina Medical Malpractice Liability Joint Underwriting Association*,95 Dr. Doe was on duty in an emergency room when he responded to a "may-
day” in the hospital’s intensive care unit. After some delay in reestablishing
airway access, the patient was found to be brain dead. The same carrier
insured all the physicians that had responded, but Dr. Doe received separate
counsel. The insurer decided to settle with the patient’s estate, prompting
Dr. Doe to request that the settlement not be made in his name. Based upon
the insurer’s attorney’s opinion that South Carolina’s Good Samaritan statute
would not apply—and not on an examination of Dr. Doe’s medical
judgment—the insurer made the settlement in Dr. Doe’s name and reported
him to the NPDB—along with six other defendants. The court found no
breach of a duty of good faith in the insurer’s apportionment of the blame.

Accordingly, physicians and their counsel have searched for ways to
“dodge” reporting, defeating the disclosure-driven focus of the NPDB and
damaging other dispute-resolution systems in medicine. Several techniques
to escape reporting to the NPDB have been developed and reported in the
literature, including 1) corporate shielding; 2) paying verbal demands for
malpractice injury; 3) paying out-of-pocket without liability insurance;
4) refunding money paid for medical care; 5) pre-complaint mediation; and 6) some other statutory schemes. It is hard to determine how wide-
spread these practices are since they escape public reporting and litigation
systems. Nonetheless, when faced with a potential report to the NPDB, a
physician and his or her counsel may resort to these methods to avoid report-
ing. Further, requiring reporting appears to interfere with health-system and

96 Id. at 672.
97 Id. at 672–73.
98 Id. at 673.
99 S.C. CODE ANN. § 15-1-310 (2016) (“Any person, who in good faith gratuitously ren-
ders emergency care at the scene of an accident or emergency to the victim thereof, shall not
be liable for any civil damages for any personal injury as a result of any act or omission by
such person in rendering the emergency care or as a result of any act or failure to act to provide
or arrange for further medical treatment or care for the injured person, except acts or omissions
amounting to gross negligence or willful [sic] or wanton misconduct.”). 100 Id. at 672, 673 n.4.
101 Id. at 676.
102 Teninbaum, supra note 19, at 105–06 (corporate entities do not report their corporate
liability to the NPDB).
103 Id. at 106–07; see also 42 U.S.C. § 11151(7) (2012) (“The term ‘medical malpractice action or claim’ means a written claim or demand for payment based on a health care pro-
vider’s furnishing (or failure to furnish) health care services, and includes the filing of a cause
of action, based on the law of tort, brought in any court of any State or the United States
seeking monetary damages.”).
104 Teninbaum, supra note 19, at 107; see also Am. Dental Ass’n v. Shalala, 3 F.3d 445,
446 (D.C. Cir. 1993) (“Although the Act does not define the term ‘entity,’ its language and
structure indicate clearly that Congress did not intend the statutory term ‘entity’ to include
individual practitioners.”).
105 Teninbaum, supra note 19, at 107–08; see also 45 C.F.R. § 60.7 (2016) (“For purposes
of this section, the waiver of an outstanding debt is not construed as a ‘payment’ and is not
required to be reported.”).
106 Teninbaum, supra note 19, at 108.
107 Id. at 108–09; see also Or. REV. STAT. § 31.250 (2016) (requiring that parties in a
medical malpractice suit enter into “dispute resolution” after commencing litigation).
hospital programs to encourage physician apology, since such disclosures can end up in the NPDB—while a corporate apology does not. These developments defeat the purpose of the NPDB. However, the difficulties that physicians have faced in trying to change or “put their side of the story” into often legitimate reports makes the dodging techniques more appealing. This paper now turns to physicians’ attempts to place their own statements into the NPDB record.

III. PHYSICIANS’ FAILED ATTEMPTS TO CHALLENGE AN INACCURATE OR MISLEADING NPDB REPORT

Current NPDB procedures to challenge reports are inadequate for many medical malpractice claims, especially those that do not reach formal adjudication. As discussed above, the mere existence of the NPDB disincentivizes settlement and heightens conflicts between physicians and insurers. However, compounding this issue are the poor safeguards against inaccurate or misleading reports. A poorly drafted or potentially disputed report can ripple through the duration of a physician’s career. All else being equal, NPDB reports that better reflect the reality of a situation and that afford more protections to physicians would lessen these incentives to prolong medical malpractice litigation. The next several cases illustrate the difficulties providers face when they attempt to change or dispute a report. They will then provide examples of NPDB report language (the NPDB does not provide sample reports for physicians or lawyers to review).

As a preliminary matter, there are two major mechanisms to change a report: 1) a “subject statement” and 2) dispute resolution. The first—the subject statement—is essentially a 4000-character response to the entity’s 4000-character narrative report. It does not obligate a reporting entity to change their description of the event. To change an entity report, a physician must enter the dispute process. This process notifies the filer of the report as well as anyone who has queried the NPDB for the last three years that a report is disputed. At that point, the filing entity has an opportunity to alter the report on its own. If not, providers may request that the NPDB “review the report for accu-


\[110\] See NPDB GUIDEBOOK, supra note 1, at F-2–3.

\[111\] There is actually “dispute status,” whereby the report is described as in “dispute,” and “dispute resolution,” where the NPDB can take action in certain limited circumstances described below. See id. at F-3.

\[112\] Id.

\[113\] Id.
National Practitioner Data Bank

racy,"114 but the NPDB “will only review the accuracy of the reported information, and will not consider the merits . . . of the action . . . .”115 In listing examples of pertinent documentation for a malpractice settlement, the NPDB lists merely the written claim or settlement and explicitly labels as “unrelated” any items relating to the facts and merits of the underlying malpractice case.116

This system means that sometimes the NPDB becomes a repository of resolved or transient concerns about a physician, turning relatively minor incidents into life-changing events. For instance, in Doe v. Thompson,117 a physician struggled to erase a temporary suspension in which the physician was completely cleared. A St. Louis hospital filed a report with NPDB “regarding its summary suspension of [Dr. Doe] for an indefinite period.”118 After lifting its suspension, the hospital filed the following report: “This report revises the adverse action report filed 2/17/94, which informed the [NPDB] of the summary suspension of Dr. Doe pending psychiatric evaluation. Based on the receipt of a positive evaluation, the summary suspension of Dr. Doe has been revoked, effective April 25, 1994.”119

Dr. Doe “objected to describing the psychiatric evaluation as ’positive’”—or any mention of the evaluation at all—claiming that the inclusion was preventing him from seeking new privileges at any hospital.120 The NPDB partially agreed and amended the report:

This report revises the adverse action report filed February 17, 1994, which informed the [NPDB] of the summary suspension of Dr. Doe pending an appropriate evaluation. Based on the receipt of an evaluation that found that Dr. Doe was not suffering from any type of psychiatric disorder, the summary suspension of Dr. Doe has been revoked effective April 25, 1994.121

While the court found that he had a right to change his report using statutory authority outside the NPDB, it took Dr. Doe two years to reach that point.122 Ultimately, he was barred by the statute of limitations.123

In fact, the NPDB goes out of its way to stay out of disputes between a physician, hospital, or insurer, which means that the NPDB can hold records

---

114 45 C.F.R. § 60.21(c)(1) (2016); see also 42 U.S.C. § 11136 (2012).
115 45 C.F.R. § 60.21(c)(1).
116 See NPDB GUIDEBOOK, supra note 1, at F-6 tbl.1.
118 Id. at 126.
119 Id. (alteration in original) (emphasis added).
120 Id. at 126–27.
121 Id. at 127 (alteration in original) (emphasis added).
122 The court concluded that the heightened review standards of the Privacy Act applied. See id. at 129–33; see also 5 U.S.C. § 552a (2012).
123 Thompson, 332 F. Supp. 2d at 134. The statute of limitations under the Privacy Act is two years, except where there is material and willful misrepresentation. See 5 U.S.C. § 552a(g)(5).
that do not tell the whole story of a particular episode. The data bank has often reaffirmed its stay-out-of-the-merits position in litigation against physicians and other health practitioners seeking to mitigate damage from a potentially inaccurate report. This position is reasonable when the NPDB serves as a clearinghouse for criminal convictions or other situations that have strong due process protections. For instance, few would dispute the accuracy of the following partially reproduced NPDB report for a physician who pleaded guilty to a count of mail fraud for developing a fraudulent medical claim based on a fictitious car accident:

DATE OF ACTION: 02/13/92
ADVERSE ACTION CODE: 406.00 REPRIMAND
LENGTH OF ACTION: PERMANENT
EFFECTIVE DATE OF ACTION: 02/13/92
ACT/OMISSIONS DESCRIPTION: CIVIL PENALTY $1,000.00
RESPONDENT PLED GUILTY TO ONE COUNT OF MAIL FRAUD

However, when the NPDB records some of the details of a complex medical malpractice case—one that may not ever reach a final disposition—the record can simply be incomplete. For instance, the 4000-character entity narrative and subject response cannot cover the complexity of a case like Doe v. Rogers. Although there is no information whether the incident matured into a medical malpractice claim, the limited record indicates that it well could. Dr. Doe performed a late-night, emergency laparoscopic appendectomy on a fourteen-year old girl suffering from acute appendicitis. During the procedure, Dr. Doe removed tissue—described by him as an “inflamed band” but characterized by the anesthesiologist as the patient’s

124 See NPDB GUIDEBOOK, supra note 1, at F-16.
125 See, e.g., Leaf v. Sec’y, U.S. Dep’t of Health & Human Servs., 620 F.3d 1280, 1284 (11th Cir. 2010) (“Dr. Leaf’s affidavits disputed to some extent the Hospital’s version of his conduct. That dispute is outside the scope of the Secretary’s review.”); Doe v. Rogers, 139 F. Supp. 3d 120, 148–49 (D.D.C. 2015) (agreeing with Leaf); Simpkins v. Shalala, 999 F. Supp. 106, 111 (D.D.C. 1998) (“Moreover, this court is convinced that regardless of what HHS’s obligations may or may not be to review whether a reporting entity acted correctly in a peer review action or an entity’s investigation of a doctor, it has a duty to determine whether the events that transpired should have resulted in a Data Bank Report . . . . Surely there is a difference between requiring HHS to review whether the health care entity acted correctly in suspending the doctor and requiring HHS to review whether the entity in fact suspended the doctor. In this case, HHS had a duty to review whether Dr. Simpkins’ resignation and D.C. General’s actions with regard to Dr. Simpkins triggered D.C. General’s reporting responsibilities under the HCQI Act.”).
127 Id. at 814.
fallopian tube.\textsuperscript{129} Pathology confirmed that the removed tissue was indeed the patient’s inflamed fallopian tube, and Dr. Doe did not dispute that he failed to recognize the anatomy properly.\textsuperscript{130} However, the resulting NPDB report failed to include any indication of the elements of a medical malpractice claim, especially a deviation from the standard of care (i.e., what steps the physician should have taken):

In June 2009, the physician commenced practice at the Hospital in thoracic and general surgery. On Friday, October 2, 2009, the physician performed a laparoscopic appendectomy on a 14–year–old female. In the course of performing the procedure, the physician inadvertently removed part of one of the patient’s fallopian tubes.\textsuperscript{131}

Although the case concentrates on the propriety of reporting after Dr. Doe voluntarily suspended his practice at the hospital,\textsuperscript{132} the NPDB’s description of the potential malpractice event that triggered the remainder of the case is paltry at best. It provides little context and prevents healthcare institutions from assessing whether a physician is likely to repeat a mistake of such magnitude again.

While one may respond that the example was taken from a case more about suspension of privileges than about medical malpractice, poor practice can be the predicate for a suspension of privileges or medical malpractice.\textsuperscript{133} Further, the NPDB challenges—whatever their specific facts—indicate the frustration and anxiety physicians feel towards a listing.\textsuperscript{134} Of course, there are incompetent physicians that the system must detect and potentially remediate, and many of the cases may well involve those “bad apples” in the profession. Indeed, the NPDB “is not designed to provide protection to phy-


\textsuperscript{130} See Rogers, 139 F. Supp. 3d at 129–30.

\textsuperscript{131} Id. at 131. The remainder of the report describes the disciplinary action taken by the hospital and not the actual care of the patient herself.

\textsuperscript{132} Id. at 131–32.

\textsuperscript{133} See, e.g., Elam v. College Park Hosp., 183 Cal. Rptr. 156, 165–66 (Ct. App. 1982) (“In light of our conclusion Hospital owes generally a duty to insure the competency of its medical staff and to evaluate the quality of medical treatment rendered on its premises, the filed papers pertaining to the motion for summary judgment are replete with triable issues of fact. For example, whether Hospital should have conducted an investigation through its peer review committee upon notice of the \textit{Bailey} case? Whether the committee had conducted its periodic reviews of Schur in a nonnegligent manner? Assuming a review was made after notice of the \textit{Bailey} case, was it performed in a nonnegligent manner? If it had been made in a careful and proper manner, would the committee have recommended revocation or suspension of Schur’s staff privileges?”). Unfortunately, these reported cases do not reproduce the exact language used in the NPDB reports, and the reports are otherwise confidential to the public.
Physicians at all costs . . ."135 But, the dramatic consequences of one mistake—especially when the factual circumstances appear less than crystal clear—indicate that the NPDB skews too far the other way and captures good physicians who make mistakes.

Indeed, sometimes the NPDB can simply record misleading information. Rochling v. Department of Veterans Affairs136 demonstrates how potentially competent physicians can be subsumed by the NPDB—or at least do not have the opportunity to present their case. In Rochling, a patient with liver disease was treated at an Oklahoma VA hospital, where he underwent a laparoscopic cholecystectomy137 and liver biopsy.138 A gastroenterologist attempted to perform an endoscopic retrograde cholangiopancreatogram (“ERCP”)139 but was unable to do so, prompting a transfer to a VA Hospital in Little Rock, Arkansas with a recommendation to attempt the ERCP again.140 Dr. Rochling assumed care of the patient at this point.141 He judged that an ERCP was unnecessary based on the patient’s clinical condition.142 The patient subsequently died.143

In the resulting malpractice suit against the VA, the patient’s family placed the blame on the VA surgeon, not Dr. Rochling:

The VA surgeon placed clips on the patient’s common bile duct and left them there when the surgery was completed, which was below applicable standards of care. Following the surgery the patient began suffering [sic], signs and findings indicative of biliary obstruction. Nevertheless, the VA staff failed to timely recognize these signs symptoms and findings [sic] and failed to follow-up with the appropriate imaging studies and corrective surgery. As a proximate result of the foregoing the patient died.144

The VA settled, but assessed the entire settlement against Dr. Rochling, not the surgeon who failed to remove the clips, and reported accordingly to the NPDB.145 Dr. Rochling exhausted his administrative remedies against the VA, and the NPDB indicated “that the arguments [Dr.] Rochling made were beyond the scope of proper HHS review.”146 The reviewing court failed to

---

135 Leal v. Sec’y, U.S. Dep’t of Health & Human Servs., 620 F.3d 1280, 1285 (11th Cir. 2010).
136 725 F.3d 927 (8th Cir. 2013).
137 This is a surgical removal of the gallbladder. See Thomas Lathrop Stedman, Stedman’s Medical Dictionary 365 (28th ed. 2006).
138 See Rochling, 725 F.3d at 929.
139 This is a method of visualizing pancreatic, hepatic, and common bile ducts. See Stedman, supra note 137, at 364.
140 See Rochling, 725 F.3d at 929.
141 See id.
142 See id.
143 See id.
144 Id. (alteration in original) (emphasis added).
145 See id. at 930.
146 Id.
recognize Dr. Rochling’s due process claims, leaving him with a mark on his record for the remainder of his career—based on the actions of another physician.\footnote{Id. at 932–33.}

These various cases highlight a critical shortcoming of NPDB: by staying out of assessing the accuracy of its data bank, NPDB perpetuates inaccuracies and disputed episodes over a physician’s career. Its limited space prevents the development of context and nuance, and its restrictive dispute process leaves providers with few avenues to correct a damaging report. Further, the courts have not been able to provide much relief from this scheme.\footnote{An additional note: In one case, a physician filed a claim of libel in an action in regards to a NPDB report. See Sobel v. United States, 571 F. Supp. 2d 1222. 1226 (D. Kan. 2008). However, the libel claim failed in Sobel, as it will every time, for two technical reasons: First, a defamation claim is not available under the Federal Tort Claims Act. See id.; see also 28 U.S.C. § 2680(h) (2012). Second, libel requires public disclosure, and the NPDB is a confidential database. Cf. Restatement (Second) of Torts § 558 (Am. Law Inst. 1965) (emphasis added) (“To create liability for defamation there must be . . . an unprivileged publication to a third party . . . .”).}

This system, though, is not the only way to administer a putative “blacklist.” Congress and the courts have provided other examples of—not perfect—but better systems in the form of credit reporting.

\textbf{IV. Another Paradigm of a Government-Supervised “Blacklist”: Credit Reporting}

The NPDB is not alone in the ranks of lists that purport to collect data but can dramatically impact careers and livelihood.\footnote{149 Elizabeth Doyle O’Brien, Comment, Minimizing the Risk of the Undeserved Scarlet Letter: An Urgent Call to Amend § 1681e(b) of the Fair Credit Reporting Act, 57 Cath. U.L. Rev. 1217, 1220–22 (2008).} Credit scores and reports represent a similar make-or-break type of institution for everyone requiring credit. One legislator has even noted that a “poor credit history is the ‘Scarlet Letter’ of 20th century America.”\footnote{136 Cong. Rec. 18766 (1990) (statement of Rep. Annunzio (D-Ill.)).} One key difference between the NPDB and credit reports, however, is that a federal agency compiles the former while private entities develop the latter. This difference has led Congress to increasingly regulate private credit rating agencies, providing consumers with some opportunities to challenge the factual content of their credit scores—namely through the Fair Credit Reporting Act (“FCRA”).\footnote{151 Pub. L. No. 91-508, 84 Stat. 1114 (1970) (codified at 15 U.S.C. §§ 1681-81x (2012)). The FCRA was written as an amendment to the Consumer Credit Protection Act passed two years earlier. See Pub. L. No. 90-501, 84 Stat. 146 (1968) (codified at 15 U.S.C. § 1601).}

The FCRA provides consumers the right to request a “reasonable reinvestigation” of a particular entry on a credit report.\footnote{152 15 U.S.C. § 1681i.} “If . . . an item of the information is found to be inaccurate or incomplete or cannot be verified,” the reporting agency will “promptly delete that item of informa-
tion . . . ”153 If a deletion is made, previous requestors of the credit report can be notified.154 Like with the NPDB, there have been disputes as to whether the system is working.155 However, at the very least, “the consumer may file a brief statement setting forth the nature of the dispute” if no resolution is found.156 Furthermore, unlike with the NPDB, furnishers of credit information (banks, credit card companies, etc.) are obligated to investigate consumer complaints.157 Unlike the NPDB’s stay-out-of-factual-disputes attitude, the FCRA does give consumers—the subjects of the collected information—the opportunity to dispute the content of reports and insert their side of the story and use the resources of the reporting agency to investigate issues.

Some courts have been willing to provide teeth to these statutory commands.158 Accordingly, the failure of a credit reporting agency to investigate a consumer’s dispute of an entry on the report has led to successful claims against the reporting agencies.159 Likewise, creditors themselves can be held responsible for failing to report to a credit agency that a particular debt is in dispute.160 However, courts have been more hesitant to penalize credit agencies and credit information furnishers when an inaccuracy is not apparent on

---

153 Id. § 1681i(a)(5)(A).
154 See id. § 1681i(d).
155 See O’Brien, supra note 149, at 1224 n.41.
156 15 U.S.C. § 1681i(b). Similar to the NPDB, the FCRA further provides that “[t]he consumer reporting agency may limit such statements to not more than one hundred words if it provides the consumer with assistance in writing a clear summary of the dispute.” Id.
158 However, the Supreme Court has recently emphasized that the traditional standing requirements must be met, even with a specific statutory scheme. See Spokeo, Inc. v. Robins, 136 S. Ct. 1540, 1550 (2016) (“On the one hand, Congress plainly sought to curb the dissemination of false information by adopting procedures designed to decrease that risk. On the other hand, Robins cannot satisfy the demands of Article III by alleging a bare procedural violation. A violation of one of the FCRA’s procedural requirements may result in no harm.”).
159 See, e.g., Dennis v. BEH-I, LLC, 520 F.3d 1066, 1070 (9th Cir. 2008) (holding a credit agency liable for failing to properly investigate a consumer dispute when the agency reported that consumer had a judgment entered against him, when a simple check of a civil register would show the opposite is true); Morris v. Equifax Info. Servs., 457 F.3d 460, 467–68 (5th Cir. 2006) (holding a credit agency liable for failing to investigate consumer dispute even though the agency did not own consumer’s record); Lazarre v. JPMorgan Chase Bank, N.A., 780 F. Supp. 2d 1330, 1335 (S.D. Fla. 2011) (holding a boilerplate response by an “early warning” credit fraud detection system could indicate to a reasonable jury that the credit agency was not providing the consumer a reasonable investigation).
160 See, e.g., Drew v. Equifax Info. Servs., LLC, 690 F.3d 1100, 1106–07 (9th Cir. 2012) (holding a “fraud block” letter from a credit reporting agency is sufficient to trigger a re-investigation by the information furnisher, even though the entity expected a different type of notification); Saunders v. Branch Banking & Tr. Co. of Va., 526 F.3d 142, 150–51 (4th Cir. 2008) (holding a jury could conclude that a bank’s failure to disclose to a credit agency the disputed nature of a debt could trigger liability); Stroman v. Bank of Am. Corp., 852 F. Supp. 2d 1366, 1374–75 (N.D. Ga. 2012) (holding once a credit reporting agency notifies a creditor of a dispute, that creditor is obligated to undertake a reasonable investigation); Mazza v. Verizon Washington DC, Inc., 852 F. Supp. 2d 28, 35 (D.D.C. 2012) (holding consumer had pleaded sufficiently to proceed against a credit information furnisher on an alleged failure to investigate a dispute).
the face of a legal instrument\textsuperscript{161} or when a consumer disputes the legal status of an obligation.\textsuperscript{162}

Nonetheless, some courts have interpreted the FCRA to provide extensive and substantive protections to consumers—especially after they have exhausted reexamination procedures. For instance, in \textit{Henson v. CSC Credit Services},\textsuperscript{163} a state court clerk erroneously listed the Hensons as having a judgment entered against them.\textsuperscript{164} After noting that “a consumer reporting agency is required to follow ‘reasonable procedures to assure maximum possible accuracy’ of the information contained in a consumer’s credit report,”\textsuperscript{165} the court reasoned that the agency—absent any notification—could not be held liable as a matter of law for relying on a judgment docket.\textsuperscript{166} However, once the Hensons notified the credit agency that there was a problem, the duty of reinvestigation required the credit agency to go beyond mere reliance on the docket.\textsuperscript{167} With notice, an agency could better direct its resources and more individually investigate the nature and accuracy of the complaint.\textsuperscript{168}

“Accordingly, a credit reporting agency may be required, in certain circumstances, to verify the accuracy of its initial source of information, in this case the Judgment Docket.”\textsuperscript{169}

Although by no means perfect, the high-stakes “credit data bank” might provide some basis for reform from the NPDB.

\textsuperscript{161} See, e.g., Wright v. Experian Info. Sols., Inc., 805 F.3d 1232, 1241 (10th Cir. 2015) (holding a credit agency need not employ a tax lawyer to verify lien documents and may rely on a LexisNexis report, even if inaccurate); Childress v. Experian Info. Sols., Inc., 790 F.3d 745, 747 (7th Cir. 2015) (“What the plaintiff wants would thus require a live human being, with at least a little legal training, to review every bankruptcy dismissal and classify it as either voluntary or involuntary. That’s a lot to ask—to too much when one considers the alternative, which is for the agency to act only upon receiving information from the bankruptcy petitioner indicating that the petition has indeed been voluntarily dismissed.”); Stroud v. Bank of Am., 886 F. Supp. 2d 1308, 1314 (S.D. Fla. 2012) (holding a bank had discharged its obligation by verifying personal information on purchase information and bank records).

\textsuperscript{162} See, e.g., DeAndrade v. Trans Union LLC, 523 F.3d 61, 68 (1st Cir. 2008) (holding when a consumer is currently in litigation with a bank on the status of a loan, a credit agency is not obligated to report such a dispute when the loan appears valid on its face).

\textsuperscript{163} 29 F.3d 280 (7th Cir. 1994).

\textsuperscript{164} Id. at 282–83. The confusion is perhaps understandable. Jeff Henson (the brother of the party in the case) purchased a car from his brother Greg Henson (the party of the case) and financed his purchase with a loan. Id. at 282. The car was stolen, and Jeff stopped making payments. Id. The creditor sued both Jeff and Greg, though the suit did note that Greg had no interest in the car any longer. Id. Accordingly, judgment was rendered against Jeff alone. Id. at 282–83.

\textsuperscript{165} Id. at 284 (quoting 15 U.S.C. § 1681e (1994)).

\textsuperscript{166} Id. at 285.

\textsuperscript{167} Id. at 286–87.

\textsuperscript{168} Id.

\textsuperscript{169} Id. at 287. “On remand, the Hensons will have the burden of showing that they brought the alleged error in Greg’s credit report to Trans Union’s attention. If they meet their burden, the trier of fact must weigh the above-mentioned factors in deciding whether Trans Union violated the provisions of section 1681i.” Id.
Harvard Journal on Legislation

V. RECOMMENDATIONS

The NPDB embodies a good objective—disclosure of malpractice data—that has not been implemented well over the last three decades. There is certainly a need to prevent physicians from moving between hospitals, health systems, or jurisdictions to avoid the consequences or disclosure of their professional performance. However, the registry cannot simply become an unfair and undeserved way to raise malpractice premiums and prevent physicians from gaining privileges at hospitals. Likewise, the registry cannot be yet another incentive for physicians to fight malpractice lawsuits that they or their insurance companies would be willing to settle otherwise. Yet, continuing under the present system promises to perpetuate the conflicts between defendant-physicians and plaintiff-patients, physicians and their liability insurers, and physicians and the NPDB itself.

Some of this litigation arises from the dreaded report to the NPDB. A core issue driving physicians is that every medical malpractice settlement—or even judgment—does not represent a systemic reflection on a physician’s ability and competence. This proposition is trivial outside of medical malpractice: Every judgment against a pharmaceutical company, manufacturer, or other type of defendant does not mean that the defendant is not fit to participate in the market. Further complicating this issue is a general reluctance to have mistakes discussed in a public forum.170

As such, a preliminary proposal for reform would be to separate reporting and disclosure. Although issues of confidentiality would have to be resolved,171 if the NPDB did not report all the data it received, some of the pressure on the medical malpractice tort system would be relieved. Assuming that there is not a systemic pattern of underreporting in a particular specialty, locality, or practice-type, the now-thirty years of data in the NPDB provides a glimpse into the “typical” malpractice profile of physicians in the country across specialty, practice type, and geographic location.172 Physicians who deviate from that “typical” profile—whether by number of reports, dollar amount of settlements, time between suits, or some other


171 Cf. Medical Malpractice Hearing, supra note 74, at 457 (testimony of Dr. Raymond Scalettar, Member Board of Trustees, American Medical Association) (“We don’t think [the data] should be collected on a Federal level. We think that the information that would be placed in a Federal repository might not have all the necessary privacy features, . . . would not be necessarily verifiable, [and] could not be analyzed.”).

172 Cf. Michelle M. Mello & David M. Studdert, Building a National Surveillance System for Malpractice Claims, 51 HEALTH SERVS. RES. 2642, 2645 (2016) (“Consideration should be given to whether the NPDB might be restructured to serve as a more robust malpractice claims surveillance system, offering researchers a wider range of variables and data linkage opportunities.”).
“trigger” — will have their reports disclosed to hospitals when they apply for privileges.

Further, the NPDB — in a public or private form — could take a more active role in investigating disputed reports. Like credit agencies and creditors in the context of the FCRA, the NPDB can look into disputed cases, especially those that have not had the benefit of full discovery and adjudication. The government has already developed competency in investigating medical injuries in the process of defending federally employed physicians in Federal Tort Claims Act malpractice cases as well as assessing compensation for vaccine173 and radiation-testing174 injuries. Lessons from these institutions can be translated over to assessing the circumstances and context of malpractice reports.

VI. Conclusion

However the NPDB is reformed — if it is reformed — the Data Bank must shed the legalistic formalism that it is purely informational. From the beginning, many recognized that the NPDB drives the players in the medical malpractice system. And, how it decides to collect, verify, and report data has tangible effects throughout medical malpractice. Although by no means a silver bullet, reforming the NPDB so that it functions as a true information warehouse will help align the incentives of physicians, insurers, and patients as well as provide vital data on the quality of physicians and other providers.


NOTE

RETHINKING FDA’S REGULATION OF COSMETICS

Grace Wallack*

The Food, Drug, and Cosmetic Act’s treatment of cosmetics has been largely unchanged since it was first passed in 1938. Now, with the growth of a billion-dollar “natural” cosmetics market, as well as interest from legislators in revitalizing the FDCA, the safety of conventional cosmetics is once again in the public consciousness. Academic interest has come back and many papers call for more stringent cosmetic regulation,1 but largely under the existing legal framework. This article accomplishes three things. First, it gives a modern overview of how the FDA actually regulates potentially harmful chemicals in cosmetic products, which is done through a largely voluntary and industry-run system. Second, the article explores the difference between the current individualized ingredient review and the cumulative exposure problem. Finally, the article proposes a new framework for regulating cosmetic ingredients that takes into consideration an individual’s cumulative exposure from total cosmetic use, as well as long-term, rather than immediate, effects.

Table of Contents

I. INTRODUCTION ........................................ 312 R
II. A HISTORY OF FDA COSMETIC REGULATION .......... 313 R
   A. Current Tools for Evaluating Cosmetic Safety ...... 317 R
III. THE NEW MARKET FOR “NATURAL COSMETICS” ........ 321 R
IV. INGREDIENTS OF CONCERN ........................... 323 R
   A. BHA and BHT ........................................ 324 R
   B. Coal Tar Dye ......................................... 324 R
   C. Dioxane .............................................. 325 R
   D. Ethanolamines (DEA, MEA, TEA and Related Chemicals) ........................................... 325 R
   E. Formaldehyde-releasing Preservatives .......... 326 R
   F. Parabens ............................................... 326 R
   G. Phthalates .............................................. 327 R
   H. Triclosan ............................................... 327 R
   I. Talc .................................................. 328 R
V. POTENTIAL FIXES ........................................ 331 R

* B.A.; Georgetown University, 2013; J.D. Candidate, Harvard Law School, 2019. The author thanks Professor Jacob Gersen and the students in Food Law Lab for feedback on the early versions of this article. The author also sincerely thanks Josh Silver, Marjorie McLean, Winston Shi, Cory Trio, and all the members of the Harvard Journal on Legislation editing team for their valuable assistance and insights.

As consumers have grown more conscious about healthy living overall, many have turned this attention to their beauty routines as well. The natural and/or organic beauty market has been growing rapidly and is estimated to reach $25.11 billion by 2025. Consumers are increasingly concerned about preservatives and other ingredients in traditional cosmetics that are linked to cancer and reproductive toxicity.

Concern about natural and organic ingredients has prompted multiple consumer safety organizations to push for tighter regulations of chemicals in cosmetics. Senators Dianne Feinstein (D-Cal.) and Susan Collins (R-Me.) introduced the Personal Care Safety Act (“PCPA”) in 2017 to strengthen the FDA’s regulatory powers over the cosmetics industry.

Independent organizations and non-profit groups have also stepped in to fill the need for more information about cosmetic ingredients. The Environmental Working Group, for example, maintains a database of cosmetic products on the market and their ingredients. They state “Americans’ frequent exposures to cosmetics and personal care products raise questions about the potential health risks from the myriad of un-assessed ingredients in them. These ingredients migrate into the bodies of nearly every American.”

This interest in chemical exposure raises a new challenge for regulators: how to evaluate long-term exposure from a repeated use of multiple cosmetics, rather than immediate harm from individual products. This Article will explore this problem, looking first at the current state of cosmetic regulation in the United States, followed by a description of the cumulative exposure.

---


People are generally aware of the Food and Drug Administration’s ("FDA") framework for regulating food and drugs. But cosmetics regulation receives much less attention. The FDA has authority to regulate cosmetics under the Food, Drug, and Cosmetic Act ("FDCA")6 and the Fair Packaging and Labeling Act ("FPLA`).7 However, cosmetics are subject to significantly less stringent requirements than their food and drug counterparts.

This lack of regulation creates two problems. First, because the FDA is limited in its ability to regulate cosmetics, the agency is less able to positively assure consumers that their cosmetics are safe. A lack of clear guidelines over "clean" or "natural" beauty leaves consumers at risk for falling for faulty claims. This Article will explore this problem by looking at the ingredients of concern to clean-beauty advocates and examining how the FDA currently regulates them.

Second, current FDA regulations do not take into account that exposure to toxins—from both beauty products and the surrounding environment—is not uniform along race, class, and gender lines. Women use more products than men,8 and products marketed towards women of color contain more hazardous materials than those marketed towards white women.9 Additionally, data shows the level of certain harmful toxins that are found in both cosmetic and non-cosmetic products vary by race.10

The Article will detail these issues and evaluate other models for cosmetic regulation. This Article argues that a stronger regulatory regime would better address harmful ingredients in cosmetics in the context of overall usage and give consumers more confidence in the safety of cosmetics overall.

II. A History of FDA Cosmetic Regulation

Cosmetics were not always under the FDA’s control. The 1906 Pure Food and Drug Act prohibited the sale of foods or drugs that were misbranded or adulterated but did not include the pre-market approval mecha-

---

nism for drugs that exists today.\textsuperscript{11} Although regulators attempted to enforce the act aggressively, the courts quickly limited their powers by requiring the government to prove the company intended to defraud consumers.\textsuperscript{12} By the 1930s, the public was increasingly outraged about the sale of dangerous (and sometimes entirely useless) drugs, and, along with the FDA itself, pushed Congress to update the old law. Some of the most egregious products on the market—which the old law kept out of FDA’s reach—included a radium-based tonic, worthless “cures” for tuberculosis, and an eyelash dye that caused permanent blindness.\textsuperscript{13} The FDA compiled a list of such products to demonstrate why new regulations were needed. But the law was not passed until after a true “therapeutic disaster” occurred. In 1937, one company marketed a new “wonder drug” to treat strep infections (usually in children), which included a chemical compound similar to antifreeze.\textsuperscript{14} Over 100 people died from the medicine, and the public outcry ultimately pushed the new FDCA through Congress.\textsuperscript{15}

The FDCA defines a cosmetic by its intended use. A cosmetic is an “article[ ] intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body. . . for cleansing, beautifying, promoting attractiveness, or altering the appearance.”\textsuperscript{16} This definition covers moisturizers, perfumes, lipsticks, nail polish, most makeup, and shampoos.\textsuperscript{17}

However, some products that we traditionally think of as cosmetics can actually be regulated as drugs. The FDCA can define a drug by its intended use, among other characteristics.\textsuperscript{18} For example, a regular shampoo would be classified as a cosmetic, but an anti-dandruff shampoo would be a drug,
because it is intended to treat dandruff. Other cosmetic/drug combinations include fluoride toothpaste, antiperspirants, acne washes, and moisturizers or makeup that double as sunscreen. Products like these must meet the requirements for both cosmetics and drugs. The exact same product can even become a drug, rather than a cosmetic, based on marketing and intended use alone. For instance, “a fragrance marketed for promoting attractiveness” would be a cosmetic. But the same exact fragrance marketed with certain “aromatherapy” claims, such as “assertions that the scent will help the consumer sleep or quit smoking,” would qualify as a drug because of its intended use. These cosmetic/drug combinations must list the active ingredients before the other ingredients.

In the 1970s and 80s, the FDA increased attention on these so-called “cosmecutical” products that straddled the line between cosmetics and drugs. Anti-wrinkle claims drew the most attention, both because of the potential skin irritants within them (AHAs, BHAs, and retinoids) and because of manufacturers’ over-hyped claims. Regulation of “cosmecuticals” has been a topic of scholarly interest for the last twenty-plus years. Products like these are considered cosmetics in the European Union (“EU”) but “have been classified by the US FDA as Over-the-Counter (“OTC”) drugs,

---

20 Id.
21 Id.
22 Id.
24 Alpha Hydroxy Acids (AHAs) and Beta Hydroxy Acids (BHAs) are both chemical exfoliants and common ingredients in skincare products. See Beta Hydroxy Acids, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/cosmetics/productsingredients/ingredients/ucm107943.htm [https://perma.cc/A3Y2-4G99] (last visited Nov. 13, 2018). The most common BHA used is salicylic acid, a common treatment for acne. Id. “Products containing AHAs are marketed for a variety of purposes, such as smoothing fine lines and surface wrinkles, improving skin texture and tone, unblocking and cleansing pores, and improving skin condition in general.” Alpha Hydroxy Acids, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/Cosmetics/ProductsIngredients/Ingredients/ucm107940.htm [https://perma.cc/GD38-PLK7] (last visited Nov. 13, 2018). Glycolic acid and lactic acid are common AHAs. Id. Retinoids, of which retinol is a common variety, are vitamin-A derivatives used in skincare products for anti-aging and anti-acne benefits. See Corrie Pikul, The One Thing Dermatologists Agree On (Other Than Sunscreen), HUFFINGTON POST (Sept. 29, 2014), https://www.huffingtonpost.com/2014/09/29/benefits-of-retinoids-retinol_n_5845764.html [https://perma.cc/BFS8-W292].
25 Liang & Hartman, supra note 23, at 265.
including sunscreen products, anticavity toothpastes, antiperspirants, antidandruff preparations, skin protectants and hair restorers.\footnote{27 See Gerhard J. Noyheck et al., \textit{Safety Assessment of Personal Care Products/Cosmetics and Their Ingredients}, 243 \textit{TOXICOLOGY & APPLIED PHARMACOLOGY} 239, 241 (2010).}

Regulation of cosmetics bears some similarity to food and drugs but also differs in important ways. Unlike drugs, cosmetic products and ingredients do not need pre-market approval, except for certain color additives.\footnote{28 \textit{Cosmetic or Drug, supra note 19.}} Cosmetic manufacturers are also not required to register or file their product formulations with the FDA, and no registration number is required to import cosmetics into the United States. However, products intended for retail sale must list their ingredients, much like food products.\footnote{29 21 C.F.R. § 701.3(c)(1) (2011). However, fragrances and other trade secrets are permitted to be excluded from the label, and companies can just list “fragrance.” \textit{Id.}} Like food, the ingredient label on cosmetics must “bear a declaration of the name of each ingredient in descending order of predominance, except that fragrance or flavor may be listed as fragrance or flavor.”\footnote{30 \textit{Id.}} Ingredients must also go by a specified name or, if a name is not specified, the ingredient should be labeled as the name adopted by the Cosmetic, Toiletry and Fragrance Association, Inc. (“CTFA”).\footnote{31 \textit{Id.}} Labeling is not required for products intended solely for professional use.\footnote{32 The use of chemicals in professional settings is nevertheless worrisome, as beauty technicians are often exposed to the products more frequently and at higher levels than their customers. See Sarah M. Nir, \textit{Perfect Nails, Poisoned Workers}, \textit{N.Y. Times} (May 11, 2015), https://www.nytimes.com/2015/05/11/nyregion/nail-salon-workers-in-nyc-face-hazardous-chemicals.html [https://perma.cc/ZE7V-48PF].} Furthermore, the FDA does not have the authority to issue product recalls, although it can request that companies recall a product.\footnote{33 \textit{Cosmetics, AM. CANCER SOC’Y} (May 28, 2014), https://www.cancer.org/cancer/cancer-causes/cosmetics.html [https://perma.cc/P5CV-SJV5].}

In addition to the FPLA—which requires cosmetic labels to not be “misleading” in the label, packaging, or fill of the bottle—cosmetics cannot be “adulterated.” Adultered products are those that are poisonous, putrid, unsanitary or contaminated, or otherwise injurious to health.\footnote{34 \textit{FDA Authority Over Cosmetics: How Cosmetics Are Not FDA-Approved, But Are FDA-Regulated}, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/Cosmetics/GuidanceRegulation/LawsRegulations/ucm074162.htm [https://perma.cc/FB7J-NA46] (last updated July 24, 2018) [hereinafter \textit{FDA Authority}].} The FDA has only banned a few ingredients outright in the formulation of cosmetics; using these in any cosmetic automatically makes a product “adulterated”: 

- Bithionol (21 C.F.R. § 701.11)  
- Mercury compounds (§ 700.13)  
- Vinyl chloride (§ 700.14) (formerly used in aerosols)  
- Halogenated salicylanilides (§ 700.15)  
- Zirconium (when used in aerosols) (§ 700.16)  
- Chloroform (§ 700.18)
Rethinking FDA’s Regulation of Cosmetics

Methylene chloride (§ 700.19)
Chlorofluorocarbon propellants (§ 700.23)
Prohibited cattle materials (§ 700.27)

Some of the items listed above actually reference a few related substances. For example, the section banning “halogenated salicylanilides” actually lists four related chemical compounds, all of which are banned. But the list is still strikingly short and perhaps under-inclusive. Formaldehyde, a known carcinogen, is not banned from use in cosmetics in the United States. Some studies suggest one in five cosmetics in the U.S. market either contain or release formaldehyde. This relatively short list has raised alarm for those who are concerned about harmful chemicals in cosmetics, especially in relation to stricter regulation abroad.

A. Current Tools for Evaluating Cosmetic Safety

Because the FDA does not pre-approve cosmetic ingredients, the agency has other tools for ensuring that cosmetics in the market are safe to use. Most notably, the law requires cosmetic companies to sell safe products and to label them correctly. If the FDA determines a cosmetic is not safe, it can take regulatory action against a manufacturer, regardless of whether the product contains one of the ingredients listed above. Usually, the first action the agency takes is issuing a warning letter to the manufacturer, which details the violation and states that if the company does not remedy the situation, it could be subject to enforcement actions. The FDA also issues “untitled letters” for potential violations not significant enough for a warning letter. These letters suggest that a company should remedy some defect but do not state that the company would be subject to enforcement if they do not.

36 Exposing The Cosmetics Cover-Up: Is Cancer-Causing Formaldehyde In Your Cosmetics?, ENVTL. WORKING GRP., http://www.ewg.org/research/exposing-cosmetics-cover/formaldehyde-releasers#.We61_RNSwXo [https://perma.cc/49QR-L9BY] (last visited Nov. 4, 2018); see also infra Section IV.5.
37 See infra Section V.A.
38 See 21 C.F.R. § 700; id. § 701.
42 See id.
The vast majority of the warning letters on the FDA’s website concern cosmetic companies making improper drug claims on their products. Such letters highlight how, as discussed above, certain marketing claims can change the status of a product from a cosmetic to a drug. For example, some of these cited companies made claims for “removing wrinkles instantly” by “stimulating regeneration of cell tissues” or using essential oils “to help control eczema or psoriasis and heal the scalp.” These types of warning letters are common for products designed to treat acne, reduce cellulite, remove wrinkles, and restore hair growth.

If a company makes drug-like claims, the product is considered a new drug and must go through the proper drug-approval process. The FDA suggests that the company review its advertising claims in order to bring the product into compliance—but these types of violations have nothing to do with the underlying safety of the product. If a warning letter does not bring the company into compliance, the FDA can also request that the Department of Justice bring an action, and the FDA can also request a restraining order in court against the shipment of the product. However, the FDA does not have the authority to recall products, although it can ask that a company recall a product voluntarily. If a company decides to recall a product, the FDA is involved in monitoring the recall, and the FDA may issue public notice about the recall.

The FDA has various mechanisms for monitoring the safety of products. The agency has the authority to conduct inspections of cosmetic manufacturers without prior notice. The FDA does not have binding regulations on cosmetic manufacturing but does have a list of Good Manufacturing

---


46 Hollywood Skincare, supra note 44.

47 Healing Scents, supra note 45.

48 Warning Letters for Products Marketed as Cosmetics, supra note 43.

49 See, e.g., Healing Scents, supra note 45; Hollywood Skincare, supra note 44.

50 Id.


52 Id.

Rethinking FDA’s Regulation of Cosmetics 319

Practice Guidelines, which it encourages companies to follow. The FDA also collects “Adverse Event Reports” filed by consumers or health care professionals who experience a problem using cosmetics. However, manufacturers are not required to report complaints that they receive directly to the FDA.

The FDA also maintains a Voluntary Cosmetics Reporting Program (“VCRP”) for manufacturers and distributors to register their facilities and report products. Owners and operators of cosmetic manufacturing or packing facilities can register their facilities. In addition, any manufacturer or distributor can file a Cosmetic Product Ingredient Statement (“CPIS”) for each product the company distributes in the United States. This program allows the FDA to keep a list of manufacturers and product ingredients, which can be helpful for inspections or in the event of a recall. However, companies are not required to register. As a result, the FDA estimates this program covers only a fraction of the market.

So, how does the FDA evaluate the safety of cosmetic ingredients? Because manufacturers must ensure their products are safe, companies often test their products internally to ensure consumers will not have adverse reactions. However, this type of testing is probably not sufficient for long-term health hazards like cancer. The FDA suggests companies use available peer-reviewed data as well in evaluating the safety of their cosmetics.

The FDA also relies on, and suggests that companies consult, the Cosmetic Ingredient Review (“CIR”) panel on ingredient safety. The CIR is part of an industry trade association funded by the Personal Care Products Council that was established in 1976. The panel meets quarterly to assess the safety of ingredients in cosmetics based on manufacturer’s information.
and peer-reviewed data.\textsuperscript{65} The FDA takes these findings into consideration, but it neither votes on the panel nor is bound by the results. CIR publishes its findings in the \textit{International Journal of Toxicology} and on the CIR website.

The CIR prepares an annual list of substances for review, based upon the FDA’s voluntary reporting database.\textsuperscript{66} The panel then decides whether the ingredient is safe, not safe, or whether it needs more information to be reviewed.\textsuperscript{67} The panel also keeps a list of ingredients that are “safe for use in cosmetics, with qualifications”—usually meaning the ingredient is safe in a particular formulation or at a specific level.\textsuperscript{68} The panel also has published reports on background information about aerosols in products\textsuperscript{69} and potential endocrine-disrupting ingredients.\textsuperscript{70}

However, some scholars argue the CIR process is incomplete. As of 2005, the CIR had only evaluated an estimated 11 percent of ingredients used in cosmetics, leaving 89 percent completely untested.\textsuperscript{71} Furthermore, as some scholars note, “when the CIR does review ingredients, it generally focuses on the ingredients’ potential to cause short-term dermatological reactions such as rashes and eye irritation, not their potential to cause long term health problems such as cancer or reproductive harm.”\textsuperscript{72} As of April 2018, the CIR had only listed eleven ingredients as unsafe, three of which are considered safe in certain levels but unsafe at higher ones.\textsuperscript{73}


\textsuperscript{67} \textit{Id.} at § 42.


\textsuperscript{72} Shah & Taylor, supra note 1, at 204 (citing \textit{Stacy Malkan, Not Just A Pretty Face - The Ugly Side Of The Beauty Industry}, 11–12 (2007)).

\textsuperscript{73} \textit{Cosmetic Ingredient Rev.}, \textit{Unsafe: Quick Reference Table} (Apr. 2018), https://www.cir-safety.org/sites/default/files/U-122017posted042018.pdf [https://perma.cc/2KZ6-XGUU]. Note that this does not include the fourteen ingredients that the FDA bans, on which the CIR does not conduct review. \textit{See Cosmetic Ingredient Rev.}, supra note 35.
III. The New Market for “Natural Cosmetics”

Perhaps unsurprisingly, the market has reacted faster to the concern about harmful chemicals than regulators have. The market for “clean” beauty has been growing rapidly. All-natural beauty companies often point to the lack of regulation in the United States and the more stringent restrictions in the EU as evidence that “clean” beauty is important—and worth the premium price. Clean cosmetic advocates often focus on avoiding chemicals that are considered harmful in the long run because they cause cancer, birth defects, or other reproductive toxicity.

But “clean” beauty has no legal definition. “Natural” beauty products can range from those simply excluding certain harmful chemicals (i.e., phthalates or petrochemicals) to lipstick made entirely of food-grade materials. But there is virtually no regulation in the U.S. for what can be sold or marketed as a “natural” beauty product. In addition, the FDA does not regulate the term “organic” as applied to cosmetics, but a cosmetic can display the U.S. Department of Agriculture (“USDA”) organic label if ninety-five percent of it is made from organic agricultural products. However, if the product is not made up of mostly agricultural ingredients, there is simply no equivalent labeling scheme.

This has understandably caused both confusion and safety hazards. The FDA routinely cautions that organic or natural products, even if properly labeled as such, are not necessarily safe. A perusal of FDA’s recent recalls and alerts provides some notable examples. The FDA investigated a product called Wen’s Cleansing Conditioner—a product that now holds the title for most consumer complaints ever filed about any hair cleanser. Nevertheless, customers have complained that the

---

76 See infra Section IV.
Wen product has caused hair loss, breakage, balding, and itching. The FDA has received many complaints directly, and is investigating another 21,000 negative reports made to the company.

What if the “natural” cosmetic isn’t harmful, but is nevertheless not quite as “clean” as consumers think it is? A recent survey done by the Federal Trade Commission and USDA found that consumers believe cosmetics with organic claims conform to higher standards than is actually the case. As discussed above, consumers have little recourse with the FDA as long as the product and its ingredients are generally considered safe. Instead, consumers have turned to deceptive advertising class actions against companies. For example, one class action alleged that Jason and Avalon Organics products (both sold at Whole Foods) were marketed as either “Natural & Organic” or “Organic,” despite being made up of largely non-organic ingredients. Multiple lawsuits have been filed against Jessica Alba’s The Honest Company, which sells a wide variety of home goods, cleaning supplies, diapers, and cosmetics. One such lawsuit alleges the company’s shampoos and body washes are falsely labeled “natural and plant based”; another alleges the company’s baby foods and soaps are incorrectly labeled “organic.” A previous lawsuit alleged The Honest Company’s sunscreen both contained synthetic ingredients and was ineffective. Finally, a drugstore shampoo brand called Organix faces a lawsuit over deceptive labeling. Despite the name of the company and claims that their shampoos “contain or-

---

80 Wen by Chaz Dean, supra note 78.
81 Id.
ganic ingredients,” virtually none of the ingredients in the product are organic.87

Plaintiffs in these lawsuits typically do not allege that they were hurt by the non-natural ingredients. Rather they paid more for the product, believing it was organic or plant-based, than they otherwise would. But there are many consumers who believe that many of the chemicals in conventional cosmetics are unsafe.88 Studies show that many common ingredients in conventional cosmetics (phthalates, sulfates, parabens, etc.) are linked to cancer, interfere with reproductive organs, hasten the growth of skin lesions, or contribute to antibiotic resistance.89 Customers are likely purchasing products with all-natural or organic branding to avoid some of these chemicals, even though the FDA does not consider them so dangerous to ban outright. One study showed that fifty-five percent of pregnant women considered cosmetic use a risk during pregnancy.90

Not only is there a consumer interest for “clean cosmetics,” but new consumer groups have focused on removing chemicals that are found in many cosmetics and personal care products that can be harmful.91 Historically, the FDA has focused regulatory action on harmful active ingredients. These consumer groups seek to shift the regulatory focus away from the safety and efficacy of the active ingredient to the rest of the product formulation, and from individually harmful cosmetics to overall exposure.

IV. INGREDIENTS OF CONCERN

Many of the ingredients consumers are currently concerned about are used in a variety of products. These chemicals are often added as preservatives to extend the shelf-life of the product or to change the consistency of the product (into a lather, for example), or to preserve the scent of a product, rather than as the active ingredient. Parabens and phthalates are two commonly-cited chemicals present in a wide variety of products—from lotions and creams to plastic food packaging—and act as preservatives (parabens) and make plastics pliable (phthalates).92 This section will detail some of the

87 Complaint at 3, Golloher, 2014 U.S. Dist. LEXIS 91942, supra note 83.
88 See Why This Matters, supra note 5.
92 See Susan Matthews, Please Don’t Panic Over the Chemicals in Your Mac and Cheese, SLATE (July 14, 2017), http://www.slate.com/articles/health_and_science/medical_examiner/
most common chemicals listed as concerning by consumer advocates and
where they are commonly found to illustrate the problem. The list is not
exhaustive.93

A. BHA and BHT

BHA (butylated hydroxyanisole) and BHT (butylated hydroxytoluene)
are synthetic antioxidants that are used to extend the shelf-life of products.
They are usually found in lipsticks, moisturizers, and other creams.94 They
are also widely used in food products as preservatives.95 BHA and BHT can
cause skin irritations and allergic reactions. In addition, the International
Agency for Research on Cancer classifies BHA as a possible human carcino-
gen.96 Some evidence also suggests that BHA and BHT can mimic estrogen,
causing reproductive harm. BHA is currently banned in Europe, and in Cali-
ifornia, products containing BHA must contain a label stating that the ingre-
dient may cause cancer.97

Note that BHA (butylated hydroxyanisole) is not the same as the ingre-
dients called BHAs that are commonly used in exfoliants and anti-aging
products.98 Also known as salicylic acid, the FDA considers these products
safe to use, even though they do cause increased sensitivity to sunlight and,
therefore, increased risk for sunburn.99 However, studies on the long-term
effects are still being evaluated, and in the meantime the FDA recommends
people use sun protection as well.100

B. Coal Tar Dye

Coal tar dyes are found in many products, but are most prevalent in hair
dye and shampoos. Coal tar dyes are produced as a byproduct of coal manu-
facturing. Although coal tar is a known carcinogen, studies are mixed as to
whether, when used as a hair dye, it has any harmful effects. In Canada,

93 See, e.g., Chemicals of Concern, SAFE COSMETICS.ORG, http://www.safecosmetics.org/
get-the-facts/chem-of-concern/ [https://perma.cc/SX7W-ZQKP] (last visited May 9, 2018)
(listing over 25 categories of harmful ingredients).
94 See The Never List, BEAUTY COUNTER.ORG, https://www.beautycounter.com/the-never-
list/ [https://perma.cc/VC6S-KUNE] (last visited May 9, 2018).
95 DAV I D S UZUKI F OUN D., T H E “D IRTY D OZEN” I NGR E D I EN T S I NVESTIG AT ED I N T H E D AV I D
w p content/uploads/2010/10/dirty-dozen-BACKGROUNDER.pdf [https://perma.cc/VN8V-
HA7D].
96 Id.
97 Id.
Products/Ingredients/Ingredients/ucm107943.htm [https://perma.cc/5GV9-7JQH] (last updated
Nov. 15, 2017).
99 Id.
100 Id.
some coal tar dyes are prohibited in products that are used in the eye (like eyeshadow and mascara) but are allowed in hair dyes. The United States has a separate rule for coal tar used in hair dyes, and such products require a warning label.

C. Dioxane

Although not usually listed as an ingredient, dioxane is found in many liquid products that create suds (like shampoo, hand soap, etc.). Dioxane is created as a byproduct of manufacturing when ethylene oxide is added to other chemicals to make them less harsh. Research suggests dioxane is carcinogenic, especially to breast tissue, and is listed as a carcinogen by the EPA. However, the FDA does not require this to be listed as an ingredient on cosmetics, because it is technically a manufacturing byproduct. Despite this, surveys suggest that dioxane is quite prevalent: “Environmental Working Group’s analysis suggests that 97 percent of hair relaxers, 57 percent of baby soaps[,] and 22 percent of all products in [the] Skin Deep database contain dioxane.” Dioxane is listed as “Reasonably Anticipated To Be [A] Human Carcinogen” by the National Toxicology Program.

D. Ethanolamines (DEA, MEA, TEA and Related Chemicals)

DEA is used in cosmetics to make them creamy or sudsy. DEA can be harmful on its own, but it is more concerning when it reacts with nitrites also found in cosmetic products. When combined, the DEA can react and create nitrosamines, a known human carcinogen. “The European Union Cosmetics Directive restricts the concentration and use of cocamide and lauramide DEA in cosmetics, and limits the maximum nitrosamine concentration in products containing these ingredients.”

---

101 DAVID SUZUKI FOUND., supra note 95, at 3.
102 Prohibited & Restricted Ingredients, supra note 39.
103 1,4-Dioxane, SafeCosmetics.org, http://www.safecosmetics.org/get-the-facts/chemicals-of-concern/14-dioxane/ [https://perma.cc/8E38-ZSGF] (last visited May 9, 2017); see also DAVID SUZUKI FOUND., supra note 95, at 10.
104 1,4-Dioxane, supra note 103.
105 Id.
107 DAVID SUZUKI FOUND., supra note 95, at 4.
E. Formaldehyde-releasing Preservatives

Formaldehyde is another known human carcinogen. Many cosmetics use formaldehyde-releasing preservatives, which give off a small amount of formaldehyde over time to preserve the product (known as “off-gassing”).109 This is found in many cosmetics but most commonly in nail polishes. Off-gassing of formaldehyde from products can be a concern, since the consumer can inhale it.110 Plus, these ingredients may be used commonly elsewhere in the home, such as wood resins and plastics, so the consumer might be inhaling formaldehyde from multiple sources.111 It is worth noting that preservatives, which inhibit the growth of harmful molds, fungus, or bacteria, can be very important for the health of consumers.

F. Parabens

Parabens are widely used as preservatives in cosmetics. Studies suggest that somewhere between 75 and 90 percent of cosmetics include parabens, albeit at low levels.112 Parabens can penetrate the skin, interfere with hormone regulation, affect the male reproductive system, and have been linked to breast cancer.113 “It has been estimated that women are exposed to fifty mg per day of parabens from cosmetics.”114 Other urine studies have found concentrations of parabens to be higher in teenage girls than in boys.115

Paraben concentrations are regulated in Europe.116 “While the Cosmetic Ingredient Review recommends concentration limits for single (up to 0.4 percent) and total paraben concentrations (up to 0.8 percent) in a single product, these recommendations do not account for exposure to parabens from several products by a single individual.”117 Parabens have been linked to breast cancer in women. In addition, daily topical application of parabens (specifically methylparaben) might also lead to skin damage and potentially skin cancer.118

109 Such chemicals include: DMDM hydantoin, diazolidinyl urea, imidazolidinyl urea, methenamine, quaternium-15 and sodium hydroxymethylglycinate. DAVID SUZUKI FOUND., supra note 95, at 6.
110 Id.
111 Id.
112 Id. at 7; see also Parabens, SAFE COSMETICS.ORG, http://www.safecosmetics.org/get-the-facts/chemicals-of-concern/parabens/ [https://perma.cc/6W4K-HQRU] (last visited May 9, 2018).
113 DAVID SUZUKI FOUND., supra note 95, at 7.
114 Id. (citing Philippa D. Darbre & Philip W. Harvey, Paraben Esters: Review of Recent Studies of Endocrine Toxicity, Absorption, Esterase and Human Exposure, and Discussion of Potential Human Health Risks, 28 J. APPLIED TOXICOLOGY 561 (2008)).
116 Id.
117 Id.
118 Id.
However, like formaldehyde, preservatives can help prevent the growth of mold and fungus in cosmetic products. Because consumer preferences are shifting away from parabens, the FDA currently prioritizes enforcement against cosmetics that claim to use “natural preservatives” or “no preservatives” over traditional cosmetics because of the concern about the growth of bacteria.\textsuperscript{119}

\textbf{G. Phthalates}

Phthalates are used in a host of consumer goods, not just cosmetics.\textsuperscript{120} Typically used in manufacturing to make plastics pliable, they can show up in packaged foods and shower curtains. Because of this, most people in the United States show some level of phthalates in their system.\textsuperscript{121} Like parabens, exposure is not uniform, and “[w]hile levels of DEP have declined over time, disparities in exposure persist. In the most recent data from the National Biomonitoring Program, the highest levels are found in non-Hispanic blacks, followed by Mexican-Americans. Non-Hispanic whites have the lowest levels.”\textsuperscript{122} In cosmetics, some phthalates are used in nail polishes, others are commonly found as ingredients in perfumes.\textsuperscript{123} However, because they are commonly present in manufacturing processes and plastics, they may incidentally show up in other cosmetics.

Phthalates are thought to be endocrine disruptors. Some studies show that large doses can interfere with fertility or cause birth defects.\textsuperscript{124} As a result, the EU classifies them as potentially harmful and bans them from all cosmetics.\textsuperscript{125} The FDA does not currently have evidence suggesting that phthalates are harmful in cosmetics.\textsuperscript{126}

\textbf{H. Triclosan}

Triclosan is an antibacterial chemical usually found in hand soaps and hand sanitizers. Beyond cosmetics, it “can be found in a wide range of household products, including garage bags, toys, linens, mattresses, toilet fixtures, clothing, furniture fabric, paints, laundry detergent.”\textsuperscript{127} Triclosan can be absorbed through the skin, and “U.S. Centers for Disease Control and


\textsuperscript{120} Matthews, supra note 92.

\textsuperscript{121} DAVID SUZUKI FOUND., supra note 95, at 7.

\textsuperscript{122} Phthalates, supra note 10.

\textsuperscript{123} DAVID SUZUKI FOUND., supra note 95, at 5.

\textsuperscript{124} Id.

\textsuperscript{125} Phthalates, supra note 10.


\textsuperscript{127} DAVID SUZUKI FOUND., supra note 95, at 14.
Prevention scientists detected triclosan in the urine of nearly seventy-five percent of those tested (2,517 people ages six years and older).128 Studies suggest the chemical is an endocrine disruptor for humans and can also irritate the skin.129 Triclosan is very toxic for aquatic organisms and persists in the environment without breaking down for a long time.130

The FDA recently issued regulations finding that triclosan and triclo-carbon, along with twenty-three other chemicals, were no longer generally recognized as safe and effective when used in over-the-counter hand washes. In fact, no data existed showing their use was more effective than plain soap and water.131

I. Talc

Talc is a naturally occurring mineral used most commonly in baby powder, although it does appear in other cosmetic powder products, such as blush or eyeshadow.132 The IARC notes that there is limited evidence available and that talc might be carcinogenic to humans.133 However, recent anecdotal evidence suggests that routine application of baby powder to the genitals may increase risk of ovarian cancer in women.134 For example, Johnson & Johnson currently faces upwards of 6000 lawsuits claiming that lifelong use of baby powder caused ovarian cancer in some women. Some of the lawsuits claim the talc powder contained trace amounts of asbestos, even though asbestos has been banned since the 1970s.135 Other lawsuits argue the talc itself was the problem.

128 Id.
130 Id.
Current data is not available on whether talc could be to blame.136 And of course, cancer development can be a combination of many risk factors, including genetics and other lifestyle differences, and so the talc powder could be just a contributing factor.137 However, multiple studies conducted since the 1990s have found no association between talc use and ovarian cancer.138 The FDA is currently reviewing whether talc powder, applied to the genital area, causes greater cancer risk.139

Although the above list may sound alarming, it is not currently known whether repeated, long-term exposure to these chemicals from multiple sources is harmful. Some scientific articles suggest that the current method of testing ingredients (high dose studies on simulated skin or animals and creating a dose-response curve) is insufficient to show the true risk from lower levels of exposure.140 Others suggest that, because the chemical compounds from different products likely act differently or are metabolized differently, it is very unlikely that using multiple products would create cumulative harm.141 Furthermore, many scientists think it is unlikely that long-term use of cosmetics can have a harmful effect because the dosages are so small and usually applied topically (rather than ingested or injected, as with food or drugs).142 However, little evidence is available concerning how
the body actually absorbs cosmetics under conditions of regular use. Additionally, population-based studies have other limitations. It is hard to tell which exposure in a person’s regular activities (between home, work, ambient air pollution, etc.) caused a cancer, and studies must track people for years. The American Cancer Society explains that, because it is difficult to do long-term studies, there is virtually no information available about the long-term health effects of cosmetic use.

Furthermore, the effect from these ingredients, if there is one, is not distributed equally. Women use twice as many personal care products as men (twelve products versus six products, respectively), exposing them to more chemical compounds (168 vs. eighty-five, respectively). Many of the organic/natural products (those that are actually organic, not just those which claim to be) are more expensive, and may not be within the budget of a low-income family. Low-income families might also be exposed to more toxins overall, if they are unable to remove lead from their homes, update outdated furniture (coated in harmful flame-retardant chemicals), or purchase organic foods. Lead exposure is particularly acute in African-American neighborhoods. And although the FDA has evaluated common cosmetic products like eyeliner and lipstick and has found relatively low levels of lead, the FDA does not have the regulatory tools to consider lead content in the context of these broader economic and racial disparities.

Lead is one illustration of how the larger problem of the burden of chemical exposure from conventional personal care products may fall disproportionately on non-white, non-affluent people. A recent study of beauty products conducted by the Environmental Working Group (a consumer advocacy organization) found that, overall beauty products marketed towards African-American women were more hazardous than products marketed to

---

143 See id.
144 Id.; see also John A. Bukowski, Review of the Epidemiological Evidence Relating Toluene to Reproductive Outcomes, 33 REG. TOXICOLOGY & PHARMACOLOGY 147 (2001) (discussing limitations of drawing conclusions from occupational studies).
the general public. And African-American women also spend significantly more money on beauty products than women of other races.

The current framework thus poses two problems: First, the FDA is not well-equipped to evaluate or regulate harms from low-level, continuous exposure to the types of chemicals listed above. The concerns presented by consumer groups suggest we should at least consider changing how product safety is evaluated to test ingredients as they are used in context of a person’s total potential exposure, rather than individually.

Second, because the FDA is limited in how it evaluates cosmetic ingredients, the agency is not fully capable of assuring that their cosmetics are safe when they are in fact safe. A lack of clear guidelines over “clean” or “natural” beauty leaves consumers at risk for falling for faulty claims, as described above. Overall, a stronger regulatory system would allow the agency to both investigate actual hazards and give consumers more confidence that their other products are safe to use.

V. POTENTIAL FIXES

Right now, the FDA can only take action if a cosmetic is adulterated or misbranded. For some of the reasons discussed above, it is entirely possible that none of the chemicals of concern will be shown to cause long-term harms when used in cosmetics. However, the current structure of the FDCA does not enable the FDA to evaluate these potential harms, due to the limited information the FDA has on cosmetic ingredients and the current focus on reacting to complaints from consumers in the marketplace.

Congress should pass new legislation giving the FDA broader authority to regulate harmful ingredients in cosmetics. Recently proposed reforms, such as the Personal Care Products Safety Act, would improve reporting by making registration mandatory and would put the FDA in a better position to review safety of all the cosmetics in the market. More substantive changes—such as allowing the FDA to set limits on ingredients below the safety threshold based on the precautionary principle—would be necessary to really deal with the cumulative effect of consumer exposure. This Section details a few current regulatory options that have been considered and discusses other potential avenues for reform.

149 Pestano, supra note 8. EWG collects data on individual ingredients in cosmetics, and gives them a hazardous rating based on the potential concern and the reliability of the scientific evidence. It then rates personal care products from low to high hazard based on the score of the individual ingredients. See About Environmental Working Group’s Skin Deep, Envtl. Working Grp., https://www.ewg.org/skindeep/site/about.php#3 [https://perma.cc/7989-5NZ5] (last visited Oct. 14, 2018). Some think the rating system is largely overblown, since the concentrations of most ingredients is so low, and little data exists on whether, as used, the formulation actually poses a danger to humans. See Wischhover, supra note 148.

150 Shah & Taylor, supra note 1, at 212.

151 See Letter from Dayle Cristinzio, supra note 65.

152 See Shah & Taylor, supra note 1, at 240.
Harvard Journal on Legislation

A. European Model

Many advocates point to the European system as being more stringent than the U.S. system because the EU has banned over 1300 ingredients from use in cosmetics. However, many of these ingredients are not currently used in cosmetics, either in the EU or the United States. If a cosmetic in the EU does not contain any of those ingredients, or one of the restricted ingredients above an approved concentration, it is considered safe if the manufacturer has the appropriate safety data.

The EU does not technically require products to be pre-approved, but it does mandate that manufacturers conduct a much more stringent safety assessment than does the United States. A person with a university degree in pharmacology or another similar discipline must conduct the safety assessment. The safety assessment must consider:

- The general toxicological profile of each ingredient used;
- The chemical structure of each ingredient;
- The level of exposure of each ingredient;
- The specific exposure characteristics of the areas on which the cosmetic product will be applied; and
- The specific exposure characteristics of the class of individuals for whom the cosmetic product is intended.

In addition, the EU requires all products to be registered in the Cosmetic Products Notification Portal—an EU-wide database—before they can be marketed in the EU.

Because the EU safety assessment focuses more on systemic harms from use of the product rather than individual ingredients, it might be better suited to dealing with the problem of repeated, long-term use. But as of right now, it is not clear whether long-term effects, such as cancer, are being studied more in the EU than in the United States. Furthermore, it is hard to say for sure whether the EU is better off than the United States because of these regulations. The United States ranked seventh in the world in cancer

---

155 See Confidence in Cosmetics, COSMETIC, TOILETRY & PERFUMERY ASS’N, http://www.thefactsabout.co.uk/confidence-in-cosmetics/content/128#safety
156 2009 O.J. (L 342/46), 67.
157 Id. at 76; accord Assessing Safety, COSMETIC, TOILETRY & PERFUMERY ASS’N, http://www.ctpa.org.uk/content.aspx?pageid=426
158 2009 O.J. (L 342/46), 67.
159 Id. at 76; accord Assessing Safety, COSMETIC, TOILETRY & PERFUMERY ASS’N, http://www.ctpa.org.uk/content.aspx?pageid=426
incidents in 2011,\textsuperscript{158} faring better than some European countries, including Denmark and France, but much worse than many others.\textsuperscript{159} Separating harms from cosmetic use and other factors, such as genetics and lifestyle choices (i.e., smoking), make it virtually impossible to say for certain that the EU cosmetic regulations keep consumers safer.

\textbf{B. Personal Care Products Act}

In 2017, Senators Dianne Feinstein and Susan Collins introduced the PCPA, a law designed to update FDA regulation of cosmetics and protect consumers.\textsuperscript{160} The bill would have required the FDA to review the safety of at least five cosmetic ingredients each year. The bill also would have allowed the FDA to issue mandatory recalls of cosmetics with hazardous substances. As mentioned earlier, right now such recalls are voluntary.\textsuperscript{161}

The law would make some noteworthy changes to the FDA’s authority. First, the law would require mandatory registration for cosmetic manufacturers and mandatory submission of cosmetic ingredients.\textsuperscript{162} Second, by requiring the FDA to review the safety of at least five different chemicals or chemical categories each year, the law creates a positive mandate to review chemicals that is currently lacking.\textsuperscript{163} Three of the chemicals required for review in the first year are propyl paraben, formaldehyde, and lead acetate.\textsuperscript{164} Finally, the law would require companies to submit adverse health reports to the FDA when they occur.\textsuperscript{165} Each of these changes would improve the FDA’s ability to respond to consumer reports of adverse reactions and independently evaluate ingredients for their safety.


\textsuperscript{159} Not only is it difficult to draw conclusions from these data due to the host of lifestyle, cultural, and health-care related causes of cancer, but also there is reason to think the numbers are influenced mostly by rates of lung cancer, which is still the most prevalent form of cancer in the U.S. (and curiously, lung cancer is more deadly in the United States than in Europe despite lower rates of smoking in the United States). See John Horgan, Cancer Spending Higher in U.S. Than in Europe-and So Is Cancer Mortality, SCI. AM.: CROSS-CHECK (June 8, 2015), https://blogs.scientificamerican.com/cross-check/cancer-spending-higher-in-u-s-than-in-europe-and-so-is-cancer-mortality/ [https://perma.cc/KF2B-5MFZ]. For more detailed statistics, see generally Cancer Fact Sheets, WORLD HEALTH ORG., INT’L AGENCY FOR RESEARCH ON CANCER, http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx [https://perma.cc/KE8Q-ZB9D].


\textsuperscript{162} Id. at sec. 605–606.

\textsuperscript{163} Id. at sec. 608.

\textsuperscript{164} Id.

\textsuperscript{165} Id. at sec. 611.
The law also changes the scope of safety concerns that the FDA can investigate. Section 608 prescribes that, when the FDA does the annual review of the ingredients, it should consider:

The probable cumulative and aggregate effect in humans of relevant exposure to the ingredient or non-functional constituent or to any chemically or pharmacologically related substances from use in cosmetics or other products with similar routes of exposure under recommended or suggested conditions of use or their customary use, to the extent adequate data is available for analysis. In appropriate cases, the [FDA] may consider available information on the total exposure to an ingredient or non-functional constituent from all sources.\footnote{167}

This is a significant change from the FDA’s current focus on merely “adulterated” ingredients to considering aggregate effects. The scope of what the FDA can deem unsafe under the definition of “adulterated” has not been fully tested, but the prior legislative history of the FDCA suggests a legislative amendment might be needed in order to give the FDA such authority. The first version of the FDCA, when introduced in 1933, stated that “a cosmetic would be deemed adulterated if it was or could be “injurious” to the user under the usual or prescribed conditions of use, or if it contained any ‘poisonous or deleterious ingredient.’”\footnote{168} However, after opposition, this version of the bill was amended to remove the “injurious” provision, thus focusing “FDA regulation on the composition of the cosmetic as the source of injury.”\footnote{169} Because of this language, the FDA can currently only regulate cosmetics as adulterated if the ingredient in the product is harmful in isolation.\footnote{170} Amending the act to allow the FDA to consider cumulative effects from all sources would broaden the FDA’s ability to protect consumers.

This legislation would give the FDA many more tools to evaluate the safety of cosmetic ingredients. However, the law does have gaps. Because of the paucity of studies on cosmetics as applied and the difficulties in conducting such studies, there may not be enough data for the FDA to state reliably that a product is safe or unsafe within the review process. As described above, studies on cancer development take ten to twenty years, and so the FDA may not be in a position, after reviewing the evidence, to say whether a product is or is not safe.

\footnote{166}{Id. at sec. 608.}
\footnote{167}{Id.}
\footnote{168}{See Heymann, supra note 26, at 362.}
\footnote{169}{Id.}
\footnote{170}{See supra Section II.}
C. California Ingredient Registration System

In 2005, the California legislature passed the California Safe Cosmetics Act, which set up a state reporting system for cosmetics that use ingredients of concern. Under that Act, the California Department of Health Services creates a list of cosmetic ingredients that are considered potential toxins by any of the following:

1. A substance listed as known or reasonably anticipated to be a human carcinogen in a National Toxicology Report on carcinogens.
2. A substance given an overall carcinogenicity evaluation of Group 1, Group 2A, or Group 2B by the International Agency for Research on Cancer.
3. A substance identified as a Group A, Group B1, or Group B2 carcinogen, or as a known or likely carcinogen by the United States Environmental Protection Agency.
4. A substance identified as having some or clear evidence of adverse developmental, male reproductive, or female reproductive toxicity effects in a report by an expert panel of the National Toxicology Program’s Center for the Evaluation of Risks to Human Reproduction.

The department then requires any cosmetics company that sells cosmetics in the state to submit a list of its products that contain any of the listed ingredients. The department has also created a website listing both all of the ingredients of concern, as well as cosmetics that have been reported to the department. When a consumer goes to the database, they can search for a product, and when they click on it, the page shows the ingredients reported. The page states that the ingredients were reported pursuant to the California Safe Cosmetics Program, but that the products are not necessarily harmful. Many of the resources on the website are geared towards salon workers as much as they are to consumers. The program is designed to inform consumers and cosmetic industry workers about the products. However, the California system does not require any companies to change their formulations or display additional warnings about the product.


Id.


Id.
D. Expanding the FDA’s Risk Assessment Authority

Each of the policies described above have important benefits. The PCPA gives the FDA more tools to investigate the safety of cosmetics and ensure compliance. The California system gives more information to consumers so they can make more informed purchases. However, neither of these laws fully deal with the problem that data on long-term exposure, specifically for cancer and reproductive toxicity, can take years to develop. Nor do these proposals deal with overlapping exposures from multiple ingredients. In order to bridge the gap, the FDA should be given additional authority in the following ways.

1. Expanded Ingredient Database and Labeling

The FDA should establish a list—similar to the California list—of ingredients that are potentially harmful to human health. In addition to a mandatory registration like the one under the PCPA, companies should also be required to disclose any cosmetics containing ingredients on the list to the FDA. This would help the FDA understand how many companies are using certain ingredients and in which products, information the agency currently lacks.176

The FDA should also be given the authority to require additional product labeling for when cosmetics contain such ingredients, if such a label is considered appropriate and necessary.177 For example, if the company uses an ingredient, the FDA can require companies to use labels such as “may cause cancer,” “potential for reproductive toxicity,” or perhaps “this product contains ingredients that have not been evaluated by the FDA for safety.” Some warning labels like this are already required for certain coal tar hair dyes.178 However, the warning label would not be required for every ingredient, rather, only those the FDA evaluates as both potentially harmful and harmful given the method and frequency of exposure.

These labels would accomplish multiple goals. First, companies generally prefer not to have warning labels on their packaging, so this system would encourage companies to develop formulas without such ingredients—a trend that some private actors have already successfully achieved in some measure.179 Second, the labels would indicate to consumers that first, the

176 See Letter from Dayle Cristinzio, supra note 65.
177 For a more detailed look at the administrative burden under substantial evidence review for determinations such as this one, see Shah & Taylor, supra note 1, at 235–39.
179 The Environmental Working Group, the non-profit group behind the Skin Deep Database, has produced an “EWG Verified” seal that private brands can display on their packaging if the product meets EWG’s standards. See EWG Verified: A New Standard for Your Health, ENVT'L WORKING GRP., https://www.ewg.org/ewgverified/about-the-mark.php [https://
ingredients may not be safe, and second, it would change the perception that the FDA pre-approves ingredients in cosmetics, a commonly-held misconception.\(^{180}\)

Finally, the labels would bridge the gap between no regulation and a total ban, where scientific evidence suggests there could be harm but more information is needed. If the product is eventually considered too dangerous, it will already have been phased out of the market. If further testing shows there is no potential harm, the ingredient could be removed from the list. Because of the difficult nature of testing in this area, it is entirely possible that no study could conclusively determine whether a product definitely causes cancer. In such scenarios, following this method would adopt some measure of extra precaution, but without removing the ingredient entirely from the market.

2. Expanded Risk Assessments

Instead of testing individual products and relying on manufacturer’s safety tests, the FDA should be given the authority—similar to the EPA—to do human risk-assessment studies of overall cosmetic use.\(^{181}\) Risk assessments are used in many EPA regulatory decisions and include a consideration of both scientific information about safety and economic, social, and other factors.\(^{182}\) The EPA does both human health and ecological risk assessments.

Building on the list of potentially dangerous ingredients as outlined above, the FDA should be empowered to conduct human health risk assessments of those ingredients considered potentially harmful, taking into account information about dose amounts, all sources of exposure (including

\[^{180}\text{See }\text{Amy E. Newburger, Cosmeceuticals: Myths and Misconceptions, 27 Clinics in Dermatology} 446, 446 (2009).\]


non-cosmetic sources), and timing of exposure under normal use. This would also include collecting information on what products consumers typically use and how frequently consumers use them. The FDA would also consider issues like potential interactions between ingredients and whether the product is used more by at-risk demographics (such as children, adolescents, or pregnant and breastfeeding women). Similarly, the FDA can take into account differences in exposure across racial and socio-economic groups, to the extent they existed for that chemical.

The purpose of a risk assessment is not to determine an “optimal” level of exposure, but to guide decision makers in their ultimate actions. Thus, in order to make use of risk assessments, the FDA should be given the authority to set limits on chemical concentrations after considering the risk assessment, the economic costs, and the potential benefits but without finding that any use of the ingredient necessarily makes the product “adulterated.” These limits, similar to workplace exposure limits under the Toxic Substances Control Act, do not outlaw the use of a particular substance but rather set a technologically achievable limit based on both safety and potential cost.

Under this proposal, the FDA would also be allowed to limit the concentration of a potentially harmful ingredient below what is shown to be safe based on the evidence that people use multiple products daily containing the same ingredient. For example, if the risk assessment shows that consumers typically use a shampoo, conditioner, and moisturizer daily that all contain the same harmful ingredient, the FDA would be justified in limiting the maximum concentration in those three types of products below what it would otherwise be. Therefore, a consumer who uses all of the products would not be subject to more than the amount considered safe. Of course, the potential number of cosmetics a person could use is limitless, and so at some point, a consumer may be exposed to more than the overall limit. But they would be exposed to less under this regime than the current one.

The FDA could also use the risk assessment information to set company-wide limitations on concentrations of chemicals. With mandatory reporting, the FDA will have more information on which products typically include certain ingredients, and which products are typically used together. If a company markets a “line” of products that are intended to be used together, the FDA could limit the concentration of harmful ingredients on the entire line of products. For example, if a company sells a basecoat, nail polish color, and topcoat, the concentration across all three products of formaldehyde (or formaldehyde-releasing preservatives) would need to be less than the safe level for a product typically used alone. This would target products that consumers are most likely to use together, and thus reduce the overall level of exposure.

Finally, the FDA, in conjunction with other government agencies, needs more funding to do long-term studies of potential cancer-causing ingredients when used in cosmetics and applied repeatedly. These reforms would allow the FDA to independently evaluate the safety of cosmetic ingredients in the manner they are actually used by consumers. Expanded authority is not only necessary to evaluate whether more cosmetic ingredients are harmful, but also to ensure the FDA has the resources to say confidently that some ingredients are safe as currently used.

VI. Conclusion

With consumers becoming ever more conscious of health effects from cosmetics, it is possible to see the FDA as not keeping up with industry. This is not necessarily the case—conclusive data does not exist on whether these chemicals are harmful. But as more information comes forth, the FDA will need new tools to manage the risks, especially risks that take years or decades to develop. Expanding the FDA’s tools both to compel more disclosure and to investigate harms from long-term repeated use of products would help achieve these aims. Because research on cancer and reproductive risk is often indeterminate, a more balanced approach is needed. Creating a list of potentially risky ingredients is a good start and increasing the FDA’s ability to research such ingredients and assess their overall risk would provide consumers with more safe products and more information about what is currently unregulated.