

IMPROVING REGULATORY BENEFIT-COST ANALYSIS

Susan E. Dudley & Brian F. Mannix ♦

To paraphrase Winston Churchill, benefit-cost analysis may be the worst tool for policymaking, except for all the others that have been tried.¹

If regulatory interventions in market transactions are to have any hope of achieving desired outcomes, they must be based on an understanding of the tradeoffs associated with alternative actions. Every president since Jimmy Carter has recognized this and required regulatory agencies to analyze the benefits and costs of proposed regulations before they are issued. Across developed countries, benefit-cost analysis (BCA) is the principal public policy tool for laying out available information in a way that allows policy makers to make balanced, efficient regulatory decisions in the face of limited resources. However, BCA has limitations. Despite numerous advances in the field, a number of significant problems have arisen that challenge its legitimate use in informing and evaluating public policy decisions.

The barriers to improving BCA are both institutional and technical. Among the institutional factors constraining the sound application of BCA in regulatory matters are that (1) legislation is often either silent on its use, or explicitly prohibits it; (2) BCA is conducted by regulatory agencies who use it to advocate for, rather than objectively analyze, proposed new regulations; (3) efforts to counteract agencies' parochial perspective have not been as effective as they could be; and (4) incentives for *ex post* evaluation of *ex ante* estimates of the benefits and costs of regulatory actions are lacking.

Technical barriers stem from the way agencies conduct regulatory BCA, which tends to systematically bias the results. In particular, (1) analysts often start with a presumption that economic markets are fragile and prone to failure, but that their regulatory solutions will work exactly as planned, and that private decision makers are subject to cognitive biases that regulators somehow do not exhibit; (2) they identify co-benefits without

♦ Susan E. Dudley is director of the George Washington University Regulatory Studies Center; she served as Administrator of the White House Office of Information and Regulatory Affairs from 2007-2009. Brian F. Mannix is a research professor in the George Washington Regulatory Studies Center; he served as the Environmental Protection Agency's Associate Administrator for Policy, Economics and Innovation from 2005 to 2009. This article was prepared for the *Journal of Law & Politics* symposium, Reining in the Administrative State, on October 21, 2016.

¹ 444 Parl. Deb. HC (5th ser.) (1947) col. 206-07 (UK).

searching for corresponding co-costs; (3) they apply risk assessment methods that are fundamentally incompatible with BCA; and (4) retrospective review is analytically challenging.

This article briefly reviews the process by which regulations are developed in the United States and the role for BCA. It then examines the institutional and technical factors limiting the use of BCA as a tool for improving regulatory policy. It concludes with some recommendations.

I. U.S. REGULATORY PRACTICES

When issuing new regulations, federal agencies are constrained by their enabling legislation,² by the Administrative Procedure Act (APA), which requires agencies to provide public notice and seek comment before issuing new regulations,³ and by executive requirements for regulatory impact analysis (primarily BCA).⁴ Presidents from both parties for more than forty years have supported *ex ante* regulatory impact analysis to make agencies weigh the likely positive and negative consequences of regulations before they are issued.⁵

Executive Order (“E.O.”) 12,866,⁶ issued by President Bill Clinton in 1993, and reinforced by George W. Bush,⁷ Barack Obama,⁸ and Donald Trump,⁹ currently guides the development and review of regulations.¹⁰ It expresses the philosophy that regulations should (1) address a “compelling public need, such as material failures of private markets”; (2) be based on an assessment of “all costs and benefits of available regulatory alternatives, including the alternative of not regulating”; and (3) “maximize net benefits” to society unless otherwise constrained by law.¹¹ It also assigns

² See U.S. CONST. art. I, § 1 (“All legislative powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and a House of Representatives.”).

³ Administrative Procedure Act, Pub. L. No. 79-404, 60 Stat. 237 (1946) (codified as amended in scattered sections of 5 U.S.C.).

⁴ Exec. Order No. 12,866, 58 Fed. Reg. 51,735 (Oct. 4, 1993) (“In deciding whether and how to regulate, agencies should assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating.”).

⁵ See Susan E. Dudley, *Improving Regulatory Accountability: Lessons from the Past and Prospects for the Future*, 65 CASE W. RES. L. REV. 1027 (2015).

⁶ 58 Fed. Reg. at 51,735 (Sept. 30, 1993).

⁷ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, CIRCULAR A-4: REGULATORY ANALYSIS (2003).

⁸ Exec. Order No. 13,563, 76 Fed. Reg. 3821 (Jan. 18, 2011).

⁹ Exec. Order No. 13,777, 82 Fed. Reg. 12285 (Feb. 24, 2017).

¹⁰ President Trump has imposed additional procedures, including the requirement that every new regulation be offset by the removal of at least two existing regulations. See Exec. Order No. 13,771, 82 Fed. Reg. 9339 (Jan. 30, 2017).

¹¹ Exec. Order No. 12,866, 58 Fed. Reg. at 51,735.

the Office of Information & Regulatory Affairs (OIRA) within the Office of Management and Budget (OMB) responsibility for reviewing executive branch agency proposed and final regulations before they are issued, along with supporting analyses.

A. Role of BCA in Regulatory Development

In the U.S. and most developed countries, BCA is considered an important aspect of *ex ante* regulatory impact analysis.¹² OIRA explains that the purpose of the regulatory impact analysis (RIA)¹³ is to provide the public with a “careful and transparent analysis” of the effects of regulatory actions. It should include “an assessment and (to the extent feasible) a quantification and monetization of benefits and costs anticipated to result from the proposed action and from alternative regulatory actions.”¹⁴

The purpose of the RIA is to inform agency decisions in advance of regulatory actions and to ensure that regulatory choices are made after appropriate consideration of the likely consequences. To the extent permitted by law, agencies should proceed only on the basis of a reasoned determination that the benefits justify the costs (recognizing that some benefits and costs are difficult to quantify). Regulatory analysis also has an important democratic function; it promotes accountability and transparency and is a central part of open government.¹⁵

B. BCA in Practice

OIRA reports each year to Congress on the benefits and costs of the major rules it reviewed over the previous ten years. In its most recent final (2015) report, it states:

The estimated annual benefits of major Federal regulations reviewed by OMB from October 1, 2004, to September 30, 2014, for which agencies estimated and monetized both benefits and costs, are in the aggregate between \$216 billion and \$812 billion, while the estimated

¹² See ORG. FOR ECON. CO-OPERATION & DEV., INTRODUCTORY HANDBOOK FOR UNDERTAKING REGULATORY IMPACT ANALYSIS (RIA) (2008), <https://www.oecd.org/gov/regulatory-policy/44789472.pdf>.

¹³ Note that in the U.S., the regulatory impact analysis document is frequently abbreviated as RIA, while the act of doing the analysis is not.

¹⁴ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, REGULATORY IMPACT ANALYSIS: A PRIMER 2 (2011), https://www.whitehouse.gov/sites/default/files/omb/inforeg/regpol/circular-a-4_regulatory-impact-analysis-a-primer.pdf

¹⁵ *Id.*

annual costs are in the aggregate between \$57 billion and \$85 billion, reported in 2001 dollars. . . . These ranges reflect uncertainty in the benefits and costs of each rule at the time that it was evaluated.¹⁶

Thus, according to this report, the benefits of regulations issued over the last ten years are almost a factor of ten higher than the costs. However, these estimates represent only a fraction of promulgated rules, and “a closer examination reveals that the benefit figures are highly dependent on a few assumptions and that the ranges presented are unlikely to reflect the true uncertainty surrounding them.”¹⁷ It is significant that the reported benefits and costs are based on *ex ante* estimates developed by the agencies themselves before the regulations went into effect. OMB cautions that its “reliance on those estimates in this Report should not necessarily be taken as an OMB endorsement of all the varied methodologies used by agencies to estimate benefits and costs.”¹⁸ OMB identifies several key uncertainties embedded in these estimates, including how regulations’ expected reduction in risks to life are valued and the numerous “assumptions used in projecting the health impact of reducing particulate matter.”¹⁹

Those caveats often get lost in public discourse, however, and the aggregate estimates are widely reported, without qualification, as evidence of the net benefits of federal regulatory activity.²⁰

Why, after decades of practice, has BCA not lived up to its potential? The barriers to improving regulatory BCA are both institutional and technical. These are discussed in the next two sections.

II. INSTITUTIONAL BARRIERS TO IMPROVING BCA

Several institutional barriers limit the extent to which robust BCA informs regulatory policy decisions. First, although presidents of both parties have long required agencies to base new regulations on BCA of

¹⁶ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, 2015 REPORT TO CONGRESS ON THE BENEFITS AND COSTS OF FEDERAL REGULATIONS AND AGENCY COMPLIANCE WITH THE UNFUNDED MANDATE REFORM ACT 1–2 (2015), https://obamawhitehouse.archives.gov/sites/default/files/omb/infocore/2015_cb/2015-cost-benefit-report.pdf (internal footnotes omitted).

¹⁷ Susan E. Dudley, *OMB’s Reported Benefits of Regulation: Too Good to Be True?*, REGULATION, Summer 2013, at 26.

¹⁸ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, *supra* note 14, at 14.

¹⁹ *Id.*

²⁰ See, e.g., Richard Pierce, Jr., *The Regulatory Budget Debate*, 19 N.Y.U. J. LEGIS. & PUB. POL’Y 249 (2016).

alternative approaches, Congress continues to pass laws that do not permit explicit consideration of tradeoffs. Second, BCA is conducted by the agencies wishing to issue new regulations, and as a result, is often used as a tool for advocacy, rather than a neutral tool for analysis. Third, efforts to counter agencies' parochial perspective, such as oversight and public comment, have not been fully effective. Fourth, the government rarely evaluates regulatory outcomes *ex post* to determine the accuracy of *ex ante* analyses. These are each discussed below.

A. Legislation Precludes BCA

Executive Order 12,866 requires agencies to apply BCA to examine alternative policy options, stating, “in choosing among alternative regulatory approaches, agencies should select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity), *unless a statute requires another regulatory approach.*”²¹

This last caveat is significant. Legislation delegating regulatory authority to executive-branch agencies rarely includes explicit requirements for agencies to base their regulatory decisions on BCA. The Safe Drinking Water Act is a notable exception. The Toxic Substances Control Act was another exception, but with support from industry and non-governmental organizations alike, Congress amended it in 2016 to remove most of the benefit-cost balancing language.²² Most statutes are silent on whether regulations should be based on BCA,²³ but some have been interpreted as precluding a weighing of costs against benefits.²⁴

As a result, although E.O. 12,866 requires agencies to estimate both benefits and costs of regulations, especially those “economically significant” rules expected to have impacts of \$100 million or more in a year,²⁵ a small fraction of regulations each year actually includes such analysis. In 2014, for example, executive branch agencies issued 53 economically significant rules, of which only 13 were accompanied by full BCA. This count does not include independent regulatory agencies, such as the Securities and Exchange Commission, the Federal Communications

²¹ Exec. Order No. 12,866, 58 Fed. Reg. 51,735 (Oct. 4, 1993) (emphasis added).

²² Frank R. Lautenberg Chemical Safety for the 21st Century Act, Pub. L. No. 114-182, 130 Stat. 448 (2016).

²³ Arthur Fraas, *The Role of Economic Analysis in Shaping Environmental Policy*, 54 L. & CONTEMPORARY PROBLEMS 113, 114 (1991).

²⁴ See *Whitman v. Am. Trucking Ass'ns*, 531 U.S. 457 (2001).

²⁵ Exec. Order No. 12,866, 58 Fed. Reg. at 51,735.

Commission, and the Consumer Financial Protection Bureau,²⁶ which are not subject to E.O. 12,866 or OIRA oversight and rarely conduct even rudimentary BCA.²⁷

B. BCA is Conducted by Regulatory Agencies

One important institutional barrier to better regulatory analysis is that RIAs are conducted by the agencies themselves, and agencies face incentives to demonstrate that the benefits of their desired actions exceed the costs.²⁸ RIAs are often developed after decisions are made and used to justify, rather than inform, regulations. As noted above and discussed in the next section on technical barriers to better analysis, regulatory benefit estimates, in particular, are highly uncertain, relying on hypothetical models and numerous assumptions, which are rarely subjected to *ex post* evaluation for accuracy.²⁹

C. Efforts to Counter Agencies' Parochial Perspective Have Been Ineffective

OIRA's role in reviewing agency regulations before they are published is an important one, and evidence suggests it contributes to higher quality analysis in executive-branch agencies compared to independent regulatory agencies.³⁰ Nevertheless, as a reliable check on agencies' analysis, it has drawbacks. First, OIRA has fewer than 50 analysts responsible for reviewing all of the significant regulations of the executive branch.³¹

²⁶ After the recent decision in *PHH Corp. v. CFPB*, the CFPB is no longer an independent agency, and may be subject to E.O. 12,866, although the president clearly has the discretion to exempt them. 839 F.3d 1 (D.C. Cir. 2016). See Susan Dudley, *CFPB Court Ruling is a Victory for Individual Liberty*, FORBES (Oct. 12, 2016, 11:51AM), <http://www.forbes.com/sites/susandudley/2016/10/12/cfpb-court-ruling-is-a-victory-for-individual-liberty/>.

²⁷ OMB's 2015 Report to Congress states: "Independent agencies still have challenges in providing monetized estimates of benefits and costs of regulation." OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, *supra* note 16, at 32.

²⁸ Stephen Breyer observed that "well-meaning, intelligent regulators, trying to carry out their regulatory tasks sensibly, can nonetheless bring about counterproductive results." Breyer attributes this problem to a combination of public perceptions, congressional actions, and uncertainties inherent in understanding and predicting risks. These external factors exacerbate the problem of "tunnel vision," a phrase he uses to describe how agencies single-mindedly pursue a particular goal to a point that "the regulatory action imposes high costs without achieving significant additional safety benefits." STEPHEN BREYER, *BREAKING THE VICIOUS CIRCLE: TOWARD EFFECTIVE RISK REGULATION* 11 (1995).

²⁹ Dudley, *supra* note 17, at 26–27.

³⁰ Arthur Fraas & Randall Lutter, *On the Economic Analysis of Regulations at Independent Regulatory Commissions*, 63 ADMIN. L. REV. (SPECIAL ISSUE) 213, 226–27 (2011).

³¹ Susan Dudley & Melinda Warren, *Regulators' Budget Reflects President Trump's Priorities: An Analysis of the U.S. Budget for Fiscal Years 1960 through 2018*, 39 REGULATORS' BUDGET REP. 1 (July 18, 2017),

Perhaps more importantly, as an office within the Executive Office of the President, in addition to offering “a dispassionate and analytical ‘second opinion’ on agency actions,”³² it is responsible for pursuing the president’s priorities. Often agencies eschew analysis and instead choose a politically popular option; the same political forces will likely have influence within the Executive Office of the President.

Public comment is a key accountability tool, but it often comes late in the regulatory development process after agencies have conducted analysis and crafted positions. Regulatory impact analyses are dense and complex documents, often running into the thousands of pages, making meaningful public comment and peer review difficult.

The legislature has not been effective at monitoring individual RIAs or holding agencies accountable for BCA; as noted above, while Congress debates the merits of proposed legislation that would require BCA, it continues to pass new legislation that precludes it.³³ Confronted with a silent statute, the Supreme Court has found that agency balancing of benefits and costs may be unavailable,³⁴ more recently, it has found it optional;³⁵ and more recently still, it has found that a decision may be arbitrary under the APA if it does *not* take costs into consideration.³⁶

D. Ex ante Estimates Are Not Evaluated ex post

Ex ante benefit and cost estimates are not verified with empirical data *ex post*, even though several executive orders (for example, 12,866; 13,563; and 13,610) direct agencies to evaluate existing regulations. These retrospective review guidelines have been met with limited success, largely because they did not change underlying incentives.³⁷ Unlike other government programs that are reassessed each time their funds are appropriated, regulations, once created, tend to exist in perpetuity.³⁸

https://regulatorystudies.columbian.gwu.edu/sites/regulatorystudies.columbian.gwu.edu/files/downloads/2018_Regulators_Budget-2017-07-18.pdf.

³² Fed. Regulatory Rev., 74 Fed. Reg. 8819 (Feb. 26, 2009).

³³ Frank R. Lautenberg Chemical Safety for the 21st Century Act, Pub. L. No. 114-182, 130 Stat. 448 (2016).

³⁴ See *Whitman v. Am. Trucking Ass’ns*, 531 U.S. 457 (2001).

³⁵ *Entergy Corp. v. Riverkeeper, Inc.*, 556 U.S. 208 (2009).

³⁶ *Michigan v. EPA*, 135 S. Ct. 2699 (2015).

³⁷ *Reducing Unnecessary and Costly Red Tape through Smarter Regulations: Hearing Before the Joint Econ. Comm.*, 113th Cong. 32–46 (2013) (statement of Susan E. Dudley, Director, The George Washington University Regulatory Studies Center).

³⁸ Susan E. Dudley, *Retrospective Evaluation of Chemical Regulations* (Org. for Econ. Cooperation & Dev. Env’t Working Papers, No. 118).

For new regulations, OIRA serves a gatekeeper role, which compels regulating agencies to present analysis consistent with executive order requirements if they wish to issue new rules. On the other hand, once a regulation is issued, the consequence of not conducting *ex post* analysis is less problematic from the agency's perspective, in that the regulation will remain on the books.³⁹

Compounding this asymmetric incentive structure is that regulated parties may be more motivated to prevent a potentially burdensome regulation from being implemented than to advocate for a regulation to be removed.⁴⁰ Once a regulation is in place, it confers a competitive advantage on some parties, especially those who have already invested in compliance.⁴¹ Incumbents and other beneficiaries are thus less likely to support evaluation that may lead to changes or repeal.⁴²

III. TECHNICAL BARRIERS TO IMPROVING BCA

BCA is a valuable tool for informing policy decisions when collective action is necessary, but it is necessarily a static exercise, dependent upon assumptions and models of how the world would look in the future with and without a regulatory intervention. It is not a replacement for market processes, which are dynamic and responsive to diverse preferences and changing circumstances.⁴³ Often, the numerous assumptions on which estimated benefits and costs depend are highly uncertain and not transparent to decision-makers or the public.⁴⁴ Technical barriers to high quality BCA include a lack of attention to the compelling public need for government intervention in markets (the "market failure"), lack of objectivity in identification of benefits and costs, incompatibility of underlying risk assessments, and challenges in conducting retrospective analysis to corroborate estimates after regulations are implemented.

³⁹ Sofie E. Miller & Susan E. Dudley, *Regulatory Accretion: Causes and Possible Remedies*, 67 ADMIN. L. REV. ACCORD 98, 111 (2016), http://www.administrativelawreview.org/wp-content/uploads/2016/03/MillerDudley_PublishedVersion-1.pdf.

⁴⁰ Dudley, *supra* note 38, at 9.

⁴¹ Miller & Dudley, *supra* note 39, at 109–10.

⁴² U.S. agencies suggest that "environmental advocacy groups have constituted a weak barrier to evaluation." Nicholas Hart, Evaluation at EPA: Determinants of the U.S. Environmental Protection Agency's Capacity to Supply Program Evaluation (Aug. 31, 2016) (unpublished Ph. D. dissertation, George Washington University), <https://pqdtopen.proquest.com/doc/1812333751.html?FMT=AI>.

⁴³ Friedrich A. von Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 530 (1945).

⁴⁴ Susan E. Dudley & Marcus Peacock, *Improving Regulatory Science: A Case Study of the National Ambient Air Quality Standards*, 24 SUP. CT. ECON. REV. 49 (2017).

A. Lack of Attention to Market Failure

Executive orders are explicit that agencies' RIAs should clearly articulate the core problem that requires regulatory action. EO 12,866 states:

Federal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling public need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people. In deciding whether and how to regulate, agencies should assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating.⁴⁵

The concept of “market failure” is an important one in regulation. The market exchange of goods and services between willing buyers and sellers efficiently relies on price signals to allocate scarce resources to their highest and best use.⁴⁶ Absent a clearly identified market failure, regulation and other forms of government intervention can disrupt those market forces and lead to misallocation of resources,⁴⁷ inevitably resulting in *negative* net benefits. Thus, targeting a systemic problem rather than relying on anecdotes to support regulation is important. If a regulation is not based on a “compelling public need,” it is more susceptible to special-interest pressures. As the OECD observes, “[a] basic aspect of RIA is that it must be conducted with this ‘whole of society’ view in mind, rather than paying undue attention to impacts on individual groups that may be lobbying for regulation.”⁴⁸

Yet many RIAs claim very large net benefits without identifying a compelling public need to justify the necessity of a collective solution imposed by the government.⁴⁹ In some cases, the RIA does not explain

⁴⁵ 58 Fed. Reg. 51,735 (Oct. 4, 1993).

⁴⁶ Susan E. Dudley et al., *Consumer's Guide to Regulatory Impact Analysis: Ten Tips for Being an Informed Policymaker*, 8 J. BENEFIT-COST ANALYSIS 187, 191 (2017).

⁴⁷ As the OECD observes: “Identifying one or more significant sources of market failure provides evidence of a potential case for regulation. However, regulation frequently fails to address the identified market failure effectively and efficiently. There is a risk that market failure may be supplanted (or compounded) by regulatory failure.” ORG. FOR ECON. CO-OPERATION AND DEV., *supra* note 12, at 7.

⁴⁸ *Id.* at 6.

⁴⁹ Sofie Miller & Brian Mannix, *One Standard to Rule Them All: The Disparate Impact of Energy Efficiency Regulations*, in NUDGE THEORY IN ACTION 251, 251–55 (Sherzod Abdukadirov ed., 2016).

why private markets are unable to reach solutions superior to government action. For example, the Environmental Protection Agency (EPA) and the Department of Transportation (DOT) estimate their joint fuel economy rules will have large negative costs (which suggests that they would be justified even if they caused environmental harm), because, according to their calculations, the fuel savings consumers will derive from driving more fuel-efficient vehicles will outweigh the increased purchase price.⁵⁰

At the same time that RIAs are cavalier about articulating why markets are unable to respond to the problem identified, they blindly assume that regulators are unbiased and knowledgeable and that the regulation will work exactly as planned.⁵¹

For example, while DOT and EPA do not identify a material failure of private markets that would prevent consumers from reaping the huge cost savings described above absent government regulation, they do make heroic assumptions to arrive at those estimates. Their results depend heavily on assumptions about future energy prices and the choice of discount rate—a rate significantly lower than consumers reveal they use when making personal decisions. Their RIAs do not appear to appreciate other vehicle attributes which consumers might value. By looking at average prices and usage patterns and by applying a low discount rate, the regulators paradoxically conclude that by taking away consumers' choices, they can make them better off.⁵²

This appears to be a classic case of the “planner’s paradox,” where planned solutions always look better on paper than unplanned solutions because the planner sees only his “data, assumptions, biases, and understandings of the way the world works . . . All of the unseen difficulties with the planned solution—the data, assumptions, biases, and understandings of the world that turn out to be wrong—are invisible to the analyst because the data he considers are his own.”⁵³

⁵⁰ According to EPA & DOT, under the recently proposed standards for heavy-duty trucks, “the buyer of a new long-haul truck in 2027 would recoup the investment in fuel-efficient technology in less than two years through fuel savings.” Press Release, Environmental Protection Agency, EPA, DOT Propose Greenhouse Gas and Fuel Efficiency Standards for Heavy-Duty Trucks (June 19, 2015), <https://www.epa.gov/newsreleases/epadot-propose-greenhouse-gas-and-fuel-efficiency-standards-heavy-duty-trucks>.

⁵¹ See Adam Smith, Utilizing Behavioral Insights (without Romance): An Inquiry into the Choice Architecture of Public Decision-Making, (Oct. 16, 2016) (Geo. Wash. Univ. Regulatory Studies Ctr. unnumbered working paper), <https://regulatorystudies.columbian.gwu.edu/utilizing-behavioral-insights-without-romance-inquiry-choice-architecture-public-decision-making>.

⁵² Dudley, *supra* note 17, at 26–30.

⁵³ Brian Mannix, *The Planner's Paradox*, REGULATION, Summer 2003 at 8, 8–9.

This problem is accentuated by the increasingly prevalent application of behavioral insights to justify regulation based on the cognitive biases and fallacies of individuals acting on their own behalf.⁵⁴ RIAs rarely consider that government decision-makers may suffer from similar, if not more, problematic biases.⁵⁵

Yet, almost by definition, regulatory policies substitute the judgment of government regulators for those of individuals, and it is easy to succumb to what Nobel laureate Friedrich Hayek called the “fatal conceit.”⁵⁶ When agencies calculate large net benefits without being able to identify a material failure of private markets, and must depend instead on assumptions about consumer irrationality such that consumers cannot be trusted to make decisions in their own self-interest, those benefits should be viewed with skepticism.⁵⁷

B. Lack of Objectivity in Identification of Benefits and Costs

As observed above, agencies have incentives to demonstrate that proposed regulations will have net benefits to society, which can lead to RIAs that look more like advocacy tools than neutral fact-finding documents. This is perhaps most evident in EPA air-quality regulations, where so-called “co-benefits” play an important role, while the concept of “co-costs” is never used. OMB reports that rules that reduce reductions in fine air particles (PM_{2.5}) contribute between 61 and 80 percent of the estimated benefits of all federal regulations.⁵⁸ Many of these benefits derive from ancillary reductions in PM_{2.5} that EPA expects will occur coincidentally from controls aimed at other pollutants. For example, in 2010, four regulations claimed 100 percent of their quantified benefits from ancillary reductions in PM_{2.5}.⁵⁹ Three of those regulations targeted emissions of toxic air pollutants and the fourth established standards for sulfur dioxide. “In 2012, 99 percent of the reported benefits from EPA’s

⁵⁴ Brian F. Mannix & Susan E. Dudley, *The Limits of Irrationality as a Rationale for Regulation*, 34 J. POL’Y ANALYSIS & MGMT. 705, 707, 711 (2015); see also Brian F. Mannix & Susan E. Dudley, *Please Don’t Regulate My Internalities*, 34 J. POL’Y ANALYSIS & MGMT. 715, 716 (2015).

⁵⁵ See Niclas Berggren, *Time for Behavioral Political Economy? An Analysis of Articles in Behavioral Economics*, REV. AUSTRIAN ECON. 199 (2012); Smith, *supra* note 51.

⁵⁶ See von Hayek, *supra* note 43, at 530.

⁵⁷ Susan Dudley, *Perpetuating Puffery: An Analysis of the Composition of OMB’s Reported Benefits of Regulation*, 47 BUS. ECON. 165 (2012).

⁵⁸ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, 2012 REPORT TO CONGRESS ON THE BENEFITS AND COSTS OF FEDERAL REGULATIONS AND UNFUNDED MANDATES ON STATE, LOCAL, AND TRIBAL ENTITIES 12 (2012).

⁵⁹ Dudley, *supra* note 57.

mercury and air toxics rule were co-benefits.”⁶⁰ The mercury rule was vacated by the Supreme Court,⁶¹ and the agency’s reliance on co-benefits was questioned in the course of the litigation.⁶²

In principle, a benefit-cost analysis should be “complete.” It should include all of the significant consequences of a policy decision: direct and indirect, intended and unintended, beneficial and harmful. In practice, all such analyses must to some degree fall short of completeness. The problem here is that agencies do not appear to be undertaking the search for benefits and costs objectively. On the benefit side of the equation, they quantify or list every conceivable good thing that they can attribute to a decision to issue new regulations, while on the cost side they only consider the most obvious direct and intended costs of complying with the regulation. Thus, in setting stringent utility-emissions standards, EPA dismisses risks associated with reduced electric reliability, the competitiveness of the U.S. economy in international trade, or the effect that higher electricity prices will have on the family budget. In establishing new fuel economy standards, EPA, DOT, and the Department of Energy use unrealistic assumptions to estimate consumer energy and fuel savings, without considering all the other complex factors that go into individual decisions about which car or appliance to buy.⁶³

C. Risk Assessments Are Incompatible with BCA

When regulations are intended to reduce risks to human health or the environment, the BCAs will rely upon scientific risk assessments for critical inputs. In some applications, such as the Federal Emergency Management Administration’s (FEMA’s) evaluation of projects that mitigate damage from natural hazards, risk assessment and BCA are compatible. FEMA uses benefit-cost modules that have probabilistic risk assessments built in, complete with damage functions, to ensure that hazard-mitigation funds are spent cost-effectively.⁶⁴

⁶⁰ Dudley, *supra* note 14, at 28.

⁶¹ See *Michigan v. EPA*, 135 S. Ct. 2699, 2711–22 (2015).

⁶² Susan E. Dudley, *Supreme Court’s EPA Mercury Ruling is a Victory for Common Sense Regulation*, THE CONVERSATION (June 30, 2015, 1:32PM), <http://theconversation.com/supreme-courts-epa-mercury-ruling-is-a-victory-for-common-sense-regulation-44073>.

⁶³ Dudley, *supra* note 17, at 26–30.

⁶⁴ Brian Mannix, *Are Chemical Risk Assessment and Benefit Cost Analysis Compatible?*, GEO. WASH. UNIV. REGULATORY STUDIES CTR. (Jan. 6, 2016), <https://regulatorystudies.columbian.gwu.edu/are-chemical-risk-assessment-and-benefit-cost-analysis-compatible>.

However, practices for developing *chemical* risk assessments generally are not compatible with BCA because they explicitly strive to err on the side of precaution by overstating risk.⁶⁵ Rather than presenting probabilities and a range of outcomes that reflect uncertainties, chemical risk assessments often generate precise-sounding predictions that hide not only considerable uncertainty about the actual risk, but the reliance on deliberately biased inferences and assumptions for handling that uncertainty.⁶⁶

Though these practices are intended to be precautionary and health-protective, when used in BCA they may have the opposite effect. That is due to the “health-wealth effect”: when consumers’ incomes go down, they will buy less of everything, including risk reduction.⁶⁷ In the economy as a whole, if we spend a billion dollars’ worth of resources on anything, that is a billion dollars less that will be spent on everything else in the consumers’ market basket – including risk reduction. Estimates vary, but some empirical research suggests that a \$15 million decrease in income is associated with the loss of an additional statistical life.⁶⁸

The use of exaggerated risk estimates in BCA leads to policies that spend too much money to reduce those risks, which *increases* other risks because of this “health-wealth” effect. In fact, if a particular mortality risk is overestimated by a factor of ten or more, as is often the case with chemical risk assessments, then the income effect will not only diminish the expected benefits: it will end up killing more people than it saves.⁶⁹

Whenever a predictive risk assessment relies on a long chain of inferences, there will be many opportunities to introduce a bias in the prediction of risk. Sometimes risk assessors may simply be designing their model to minimize false negatives, while paying a bit less attention to false positives. At every step, people may think they are just being cautious – and erring on the side of safety. But the cumulative result is just the

⁶⁵ For example, EPA’s “Risk Assessment Principles and Practices” document states: “[s]ince EPA is a health and environmental protective agency, EPA’s policy is that risk assessments should not knowingly underestimate or grossly overestimate risks. This policy position prompts risk assessments to take a more ‘protective’ stance given the underlying uncertainty with the risk estimates generated.” ENVTL. PROT. AGENCY, AN EXAMINATION OF EPA RISK ASSESSMENT PRINCIPLES AND PRACTICES 13 (2004).

⁶⁶ See Dudley & Peacock, *supra* note 44, at 5.

⁶⁷ Mannix, *supra* note 64.

⁶⁸ Randall Lutter et al., *The Cost-Per-Life-Saved Cutoff for Safety-Enhancing Regulations*, 37 ECON. INQUIRY 599 (1999).

⁶⁹ Mannix, *supra* note 64.

opposite of what they intended. The exaggeration of chemical risks actually kills people.⁷⁰

D. Ex post BCA Challenging

As noted above, agencies rarely conduct *ex post* evaluation to verify the accuracy of their *ex ante* analysis and assumptions. This is not only due to the poor incentives discussed above; retrospective review poses analytical challenges. Once a regulation is in place, it is difficult to accurately say what the outcome would have been without it.⁷¹ For example, would air emissions have increased directly with economic and population growth, or would technological change and citizen preferences have driven emissions lower?⁷² Measuring opportunity costs (what activities or innovations were foregone to achieve regulatory goals?) is difficult, and measuring regulatory benefits is often harder.⁷³ We have no way of knowing, for example, how many important pharmaceuticals have *not* been discovered, because of the barriers to innovation presented by drug regulations.

IV. RECOMMENDATIONS

This section offers some modest recommendations to address the institutional and technical problems facing BCA.

A. Institutional

1. Legislation requiring BCA

Congress should legislatively require BCA and establish a standard that new regulations consider a range of reasonable alternatives and attempt to maximize net benefits to society. Several bipartisan bills in recent Congresses would have codified the language of President Clinton's Executive Order 12,866 and President Obama's Executive Order 13,563.⁷⁴ This would lend congressional support to the orders' nonpartisan principles and guiding philosophy that before issuing regulations, agencies should identify a compelling public need, evaluate the likely effects of different regulatory approaches, and select the approach that provides the greatest

⁷⁰ *Id.*

⁷¹ Dudley, *supra* note 38, at 7.

⁷² Jonathan Adler, *The Fable of Federal Regulation: No, States Didn't Ignore Environmental Problems*, PERC REP. (Dec. 10, 2004), <http://www.perc.org/articles/fable-federal-regulation>.

⁷³ Dudley, *supra* note 38, at 7.

⁷⁴ Principled Rulemaking Act, S. 1818, 114th Cong. (2015).

net benefit for the country.⁷⁵ Ideally, such a requirement would override authorizing statutes that ignore or explicitly prohibit analysis of tradeoffs.

2. Stronger legislative and judicial checks on agency decisions

Legislation could apply the Executive Order requirements to independent agencies and make compliance with them judicially reviewable.⁷⁶ Judicial review could be valuable because agencies tend to take more seriously aspects of their missions that are subject to litigation, and might be particularly important for independent regulatory commissions, whose actions are not subject to OIRA review.⁷⁷ “Like executive and congressional oversight, judicial oversight would likely make regulatory agencies more accountable for better decisions based on better analysis.”⁷⁸

3. More vigorous and meaningful public input

Another way to shine more light on agency decisions and the supporting analysis would be through publication of advance notices of proposed rulemaking (ANPR) for high-impact rules. As noted above, regulatory-impact analyses are often developed after decisions are made and are used to justify, rather than inform, regulations. ANPRs could be valuable for soliciting input from knowledgeable parties on a range of possible approaches, data, models, etc., before particular policy options have been selected.⁷⁹ These might include “back of the envelope” analyses that consider the effects of a wide range of alternatives.⁸⁰ “Successful reforms might involve pre-rulemaking disclosure of risk-assessment information to engage broad public comment on the proper choice of studies, models, assumptions, etc. long before any policy decisions are framed, or ‘positions’ established.”⁸¹

⁷⁵ Dudley, *supra* note 5, at 1044.

⁷⁶ Dudley, *supra* note 5, at 1045.

⁷⁷ Catherine M. Sharkey, State Farm “with Teeth”: Heightened Judicial Review in the Absence of Executive Oversight, 89 N.Y.U. L. REV. 1589, 1620 (2014).

⁷⁸ Reducing Unnecessary and Costly Red Tape through Smarter Regulations, *supra* note 37, at 8.

⁷⁹ Susan E. Dudley & Kai Wegrich, Regulatory Policy and Practice in the United States and European Union (Mar. 10, 2015) (Geo. Wash. Univ. Regulatory Studies Ctr. unnumbered working paper), <http://regulatorystudies.columbian.gwu.edu/achieving-regulatory-policy-objectives-overview-and-comparison-us-and-eu-procedures>.

⁸⁰ Christopher Carrigan & Stuart Shapiro, *What's Wrong with the Back of the Envelope? A Call for Simple (and Timely) Benefit-Cost Analysis*, 11 REGULATION & GOVERNANCE 203 (2017).

⁸¹ Dudley, *supra* note 44.

4. *Require agencies to evaluate existing regulations before issuing new ones*

While *ex post* evaluation has a long tradition in other areas (particularly in programs financed through the fiscal budget),⁸² it has received little attention in the regulatory arena, despite government guidelines requiring it.⁸³ In essence, *ex ante* analyses are hypotheses of the effects of regulatory actions. Better regulatory evaluation would allow agencies and others to test those hypotheses against actual outcomes.⁸⁴ This would not only inform decisions related to the benefits and costs of existing policy, but would provide feedback that would improve future *ex ante* analyses and future policies.⁸⁵ To incentivize more robust evaluation of regulations once they are in effect, agencies could be required to test the validity of previous BCA predictions before commencing new regulation. As a condition for issuing new regulations, agencies could be required to present a robust framework for later evaluation and a commitment to gather necessary data.

5. *An independent body may offer more objective retrospective review*

Rather than leaving the responsibility for retrospective evaluation with regulatory agencies, Congress could assign an independent body responsibility for reviewing the accumulated stock of regulations and making recommendations to repeal rules or sets of rules.⁸⁶ This model has the potential to address some of the accumulated regulatory burden⁸⁷ and to improve regulatory evaluation. An independent third-party review would offer an objectivity that past efforts (which depend on regulatory agencies themselves to identify outmoded regulations) lacked,⁸⁸ and “would likely

⁸² Stijn Smismans, *Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal*, 6 EUR. J. RISK REG. 6, 12 (2015).

⁸³ Joseph Aldy, *Learning from Experience: An Assessment of the Retrospective Reviews of Agency Rules and the Evidence for Improving the Design and Implementation of Regulatory Policy*, ADMIN. CONF. U.S. (Nov. 18, 2014), <https://www.acus.gov/report/retrospective-review-report>. Aldy writes that federal regulatory agencies have a mixed record on *ex post* review, despite their “long track record of prospective analysis of proposed regulations that can address these questions.” *Id.* at 4.

⁸⁴ Hypothesis testing is an essential element of the scientific method, and provides the feedback necessary to inform predictions. See Jose Wudka, *What Is the “Scientific Method”?*, U.C. RIVERSIDE (Sept. 24, 1998), http://physics.ucr.edu/~wudka/Physics7/Notes_www/node6.html.

⁸⁵ Dudley *supra* note 38, at 27.

⁸⁶ See, e.g., S. 708, 114th Cong. (2015); S. 1683, 114th Cong. (2015).

⁸⁷ Michael Mandel & Diana Carew, *Regulator Improvement Commission: A Politically-Viable Approach to U.S. Regulatory Reform*, PROGRESSIVE POL’Y INSTITUTE (May 2013), http://www.progressivepolicy.org/wp-content/uploads/2013/05/05.2013-Mandel-Carew_Regulatory-Improvement-Commission_A-Politically-Viable-Approach-to-US-Regulatory-Reform.pdf.

⁸⁸ As Greenstone observed, “the process of self-evaluation is challenging for all organizations, as it requires complete objectivity. Indeed, history is unkind to organizations that fail to get outside reviews of their work.” *Examining Practical Solutions to Improve the Federal Regulatory Process: Roundtable*

identify reform opportunities agencies would miss.”⁸⁹ Perhaps most importantly, institutionalizing a third-party review could improve review by motivating better data collection and more rigorous evaluation of whether risk management regulation is actually achieving its desired effect.⁹⁰

6. Offset requirements could provide incentives for better BCA

To motivate retrospective evaluation of regulations, several countries have “initiated programs that require new regulatory costs to be offset by removal of existing regulatory burdens.”⁹¹ President Trump issued E.O. 13,771 in January 2017 requiring that “for every one new regulation issued, at least two prior regulations be identified for elimination, and that the cost of planned regulations be prudently managed and controlled through a budgeting process.”⁹²

While such approaches do not explicitly balance the benefits of regulations against the costs⁹³ and may not be appropriate for all types of regulations, they could motivate governmental and non-governmental agents to develop approaches to quantify the benefits and costs of regulations so they can trade off less-cost effective rules and retain those that are achieving their goals. Along these lines, sunset provisions could provide incentives for evaluation of regulations’ effects.

B. Technical

1. Agencies should consider how regulation will affect competition

Agencies should be required to present evidence that they have identified a material failure of competitive markets or public institutions that requires a federal regulatory solution, and provide an objective evaluation of alternatives (including the alternative of not regulating) and

Before the Subcomm. on Regulatory Affairs & Fed. Mgmt. of the S. Comm. on Homeland Sec. and Gov’t Affairs, 114th Cong. (2015) (statement of Michael Greenstone, Milton Friedman Professor of Economics, University of Chicago, Director, Energy Policy Institute at Chicago).

⁸⁹ *A Review of Regulatory Reform Proposals: Hearing Before the S. Comm. on Homeland Sec. & Gov’t Affairs*, 114th Cong. (2015) (statement of Susan E. Dudley, Director, The George Washington University Regulatory Studies Center).

⁹⁰ Dudley, *supra* note 38, at 26.

⁹¹ Susan E. Dudley, *Can Fiscal Budget Concepts Improve Regulation?*, 19 N.Y.U. J. LEGIS. & POL’Y 259, 270 (2016).

⁹² Exec. Order No. 13,771, 82 Fed. Reg. 9339 (Jan. 30, 2017).

⁹³ Richard J. Pierce, Jr., *The Regulatory Budget Debate*, 19 N.Y.U. J. LEGIS. & POL’Y, 249, 251 (2016).

of the competitive and distributional impacts of different approaches.⁹⁴ Whatever their particular mission, regulators need to be mindful that competition is the most important regulator of our economy.⁹⁵

2. Agencies should recognize that behavioral insights apply to regulators as well as those being regulated

Regulations that derive most of their benefits from providing private monetary gains that individuals can achieve without government intervention, such as fuel savings from driving energy efficient cars, should require a particularly demanding burden of proof.⁹⁶ The analysis should be required to provide evidence that individuals behave irrationally (and do not learn) in the specific situation covered by the proposed regulation.⁹⁷ It should also provide evidence that *regulators* are not subject to biases that may color their judgment of consumer welfare, and explain why they are better able to judge individuals' preferences or be more faithful agents of individuals' interests than the individuals themselves.⁹⁸

3. Agencies should examine benefits and costs in a consistent manner

OMB recommends⁹⁹ that “[b]enefits and costs of a regulation should be assessed in a consistent manner,” yet agencies routinely count “co-benefits” without considering co-costs. As noted above, in principle, a benefit-cost analysis should account for all of the effects of a regulatory decision: indirect as well as direct, delayed as well as immediate, improbable as well as probable, unintentional as well as intentional. In practice, the analyst must define reasonable bounds on what to include. Agencies should define those boundaries in a way that not only produces a reasonably accurate and complete analysis, but also one that remains

⁹⁴ Exec. Order No. 12,866, 58 Fed. Reg. 51,735 (Oct. 4, 1993).

⁹⁵ Sofie E. Miller et al., *Steps to Increase Competition and Better Inform Consumers and Workers to Support Continued Growth of the American Economy*, GEO. WASH. UNIV. REG. STUD. CTR. (May 12, 2016), <https://regulatorystudies.columbia.gwu.edu/public-comment-national-economic-council-president%E2%80%99s-executive-order-13725-steps-increase>.

⁹⁶ Ted Gayer & W. Kip Viscusi, *Overriding Consumer Preferences with Energy Regulations*, 43 J. REG. ECON. 248, 264 (2013).

⁹⁷ Dudley et al. *supra* 46, at 192.

⁹⁸ Brian F. Mannix & Susan E. Dudley, *The Limits of Irrationality as a Rationale for Regulation*, 34 J. POL'Y ANALYSIS & MGMT. 705, 707, 711 (2015); *see also* Brian F. Mannix & Susan E. Dudley, *Please Don't Regulate My Internalities*, 34 J. POL'Y ANALYSIS & MGMT. 715, 716 (2015).

⁹⁹ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, *supra* note 16, at 54.

unbiased.¹⁰⁰ The problem with ‘co-benefits’ is not that ancillary [effects] are being included, but that they are being included selectively.”¹⁰¹

4. *Risk assessments should provide expected values and ranges based on probabilistic analysis*

Because risk assessment necessarily involves assumptions and judgments as well as pure scientific inputs, the government should establish procedures and incentives to make more transparent the effect different credible risk-assessment inputs and assumptions have on the range of plausible outcomes. This would make risk assessment more compatible with BCA, which is supposed to inform decision-makers of the expected value and range of the benefits and costs of different interventions.

5. *Regulations should be designed to facilitate natural experimentation and learning*

Ex ante RIAs necessarily depend on unverifiable assumptions and models of how the world would be had the regulation never been implemented, and how responses to regulatory requirements will alter those conditions.¹⁰² To test those hypotheses, agencies should be required to design regulations from the outset in ways that allow variations in compliance.¹⁰³ For example, EPA is required to review the National Ambient Air Quality Standards (NAAQS) every five years. Before it issued a new standard, it could be required to apply quasi-experimental techniques to gather and analyze epidemiology data and health outcome trends in different regions of the country and compare them against predictions.¹⁰⁴ Regulations such as these lend themselves to quasi-experimental techniques to examine regulatory benefits and costs, because

¹⁰⁰ Dudley et al. *supra* note 46, at 200.

¹⁰¹ Sofie E. Miller et al., *Public Comment on OMB’s 2015 Draft Report to Congress on the Benefits and Costs of Federal Regulation*, GEO. WASH. UNIV. REG. STUD. CTR. (Dec. 15, 2015), <https://regulatorystudies.columbian.gwu.edu/public-comment-omb%E2%80%99s-2015-draft-report-congress-benefits-and-costs-federal-regulations>.

¹⁰² Dudley, *supra* note 38, at 5.

¹⁰³ Dudley, *supra* note 38.

¹⁰⁴ See Louis A. Cox, *Public Interest Comment on The Environmental Protection Agency’s Proposed Rule: National Ambient Air Quality Standards for Ozone*, GEO. WASH. UNIV. REG. STUD. CTR. (Mar. 17, 2015), <https://regulatorystudies.columbian.gwu.edu/sites/regulatorystudies.columbian.gwu.edu/files/downloads/TCox-NAAQS-ozone-2015.pdf>; see also Francesca Dominici et al., *Particulate Matter Matters*, 344 SCIENCE 257 (2014).

different areas of the U.S. must respond differently depending on their attainment status.

QUOTAS

Shay Lavie [♦]

ABSTRACT

The Article addresses simple yet surprisingly overlooked questions—could numerical caps on legal rights be a valuable regulatory mechanism? In which circumstances should we employ them? It is the first to discuss numerical caps—quotas—as a distinct regulatory instrument, and the lessons it provides are pertinent to numerous legal settings.

The Article first sets out the theoretical framework for using quotas. It does so by synthesizing real-world examples and fleshing out the reasons for choosing quotas, especially non-tradable, over other regulatory alternatives, such as prices. Armed with the theoretical insights, the Article then suggests practical implications. In particular, capping the right to access courts through quotas can be valuable in balancing some of the conflicts that the American legal system faces. Such quotas restrict over-use of courts and push litigants to carefully invoke their rights, and they simultaneously guarantee a wide access to courts without imposing fees. Accordingly, the Article analyzes several litigation contexts—such as interlocutory appeals and pleading standards—in which policymakers can benefit from quotas.

[♦] Assistant Professor, Tel Aviv University Faculty of Law. I would like to thank Ben Alarie, Ronen Avraham, Zach Clopton, Kevin Clermont, Scott Dodson, Helen Irving, Yotam Kaplan, Amalia Kessler, Jacob Nussim, Moran Ofir, Liav Orgad, Oren Perez, Ariel Porat, Roy Shapira, Dov Solomon, Adam Shinar, Jennifer Shkabatur, and participants in the Stanford International Junior Faculty Forum, Annual American Civil Procedure Workshop, the American Law and Economics Association Annual Conference, the Italian Society of Law and Economics, Renmin University International Virtual Workshop, the Hebrew University of Jerusalem Law and Economics Workshop, and the faculty seminars in Cornell Law School, Tel Aviv Law School, and the College of Law and Business—Ramat Gan Law School for helpful discussions and comments. Aviv Avni, Ofir Givaty, Daniel Kaczelnik, Amit Kotiger, Sharon Salinas, and Shir Shrem provided excellent research assistance.

TABLE OF CONTENTS

INTRODUCTION.....	22
I. REGULATORY ROADMAP	27
A. Curbing Externalities	28
B. Price and Quantity.....	31
II. THE CASE FOR A QUOTA REGIME.....	33
A. Commodification	33
B. Willingness-to-Pay.....	36
C. Difficulties in Monetizing.....	38
III. LIMITATIONS	40
A. One-Size-Fits-All.....	41
B. Information Problems.....	42
1. <i>Policymakers</i>	42
2. <i>Right-holders</i>	43
IV. APPLICATION: LEGAL PROCEDURE.....	46
A. To Price or not to Price	46
B. Regulating Litigation Behavior.....	50
C. Implementation.....	53
1. <i>Adjudication behavior</i>	54
2. <i>Filing behavior</i>	57
CONCLUSION	60
A. Public Services.....	61
B. Government Bodies.....	64

INTRODUCTION

In July 2017, the Senate voted to approve sweeping sanctions against Russia, contrary to President Trump’s views.¹ In response, the President threatened to use his constitutional authority to veto the sanctions bill.² While President Trump did not exercise his veto power in this case,³ many

¹ Matt Flegenheimer, *Senate Forces Trump’s Hand on Russia Sanctions*, N.Y. TIMES, July 28, 2017, at A16.

² Michael Qazvini, *Trump Threatens to Veto Russian Sanctions Bill, Negotiate “Tougher” Deal*, DAILY WIRE (July 27, 2017), <http://www.dailywire.com/news/19067/trump-threatens-veto-russian-sanctions-bill-michael-qazvini#>.

³ Vivian Salana & Richard Lardner, *Trump Grudgingly Signs Russia Sanctions Bill*, ASSOCIATED PRESS, Aug. 3, 2017, <https://apnews.com/21e03f4267a84eeca4e6750ecf3128e9/Trump-grudgingly-signs-Russia-sanctions-bill>. In fact, as the bill passed by overwhelming margins, *id.*, and the Constitution allows Congress to override a presidential veto by a two-thirds majority of each chamber, U.S. CONST. art. I, § 7, cl. 2, any veto would have likely been overridden.

other past presidents have vetoed important bills.⁴ In fact, the use of the presidential veto power has significantly increased in the last two centuries,⁵ raising over-use concerns.⁶ These concerns have led scholars to consider ways to curb and regulate the veto power.⁷ The core challenge is promoting valuable vetoes and simultaneously discouraging inappropriate ones.⁸ A similar tradeoff arises in other contexts in which veto rights are used.⁹

A possible way to regulate the over-use of vetoes is imposing a pre-determined limitation on the number of times a veto could be employed. Consider a hypothetical rule that limits the President to several, say, five, vetoes per year. This cap ensures that the President would not use the veto right too often. More fundamentally, it motivates the President to think carefully before invoking her veto rights, presumably in the most important circumstances. This Article introduces and discusses the idea of such numerical limitations on the use of legal entitlements—“quotas” or “caps.”

The veto context is admittedly unique. But the challenge that veto powers pose—designing a mechanism to encourage useful behavior and minimize harmful acts—is a general one. A common tool to handle this challenge is forcing right-holders to pay for exercising their rights. In torts, for instance, strict liability regimes are designed to achieve this very goal: knowing that they would pay for the social harm they caused, right-holders are incentivized to engage only in socially desirable activities.¹⁰ Fines

⁴ See, e.g., J. Richard Broughton, *Rethinking the Presidential Veto*, 42 HARV. J. ON LEGIS. 91, 91–92 (2005) (discussing Andrew Jackson’s veto of the Bank Rechartering Bill of 1832).

⁵ *Id.* at 118; Charles L. Black, Jr., *Some Thoughts on the Veto*, 40 LAW & CONTEMP. PROBS. 87, 89–92 (1976).

⁶ Abner S. Greene, *Checks and Balances in an Era of Presidential Lawmaking*, 61 U. CHI. L. REV. 123, 126 (1994) (“the presidential veto looms as a structural threat”); Cass R. Sunstein, *An Eighteenth Century Presidency in a Twenty-First Century World*, 48 ARK. L. REV. 1, 10 (1994) (“The founders certainly did not anticipate the current situation, in which the veto power . . . implies a large and continuous presidential role in lawmaking itself.”).

⁷ See, e.g., Broughton, *supra* note 4, at 117 (suggesting prudential limits on the veto power); Greene, *supra* note 6, at 196 (concluding that “[i]n certain limited situations, we should accept a majority vote in both Houses of Congress as sufficient to block presidential [vetoes]”). These attempts are necessarily limited, as “[i]t is generally conceded today that the President may veto on any ground.” Joel K. Goldstein, *The Presidency and the Rule of Law: Some Preliminary Explorations*, 43 ST. LOUIS U. L.J. 791, 831 (1999).

⁸ Goldstein, *supra* note 7, at 830 (“[T]he framers . . . viewed the risk of vetoing and thereby ‘defeating a few good laws . . . amply compensated by the advantage of preventing a number of bad ones.’”) (quoting THE FEDERALIST No. 73, at 444 (Alexander Hamilton) (Clinton Rossiter ed., 1961)).

⁹ See, e.g., Eric A. Posner & Alan O. Sykes, *Voting Rules in International Organizations*, 15 CHI. J. INT’L L. 195, 204–06 (2014) (discussing the veto right of the permanent members of the United Nations Security Council).

¹⁰ See, e.g., Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055 (1972) (advocating, for this and other reasons, a rule that places the costs of accidents

fulfill a similar role.¹¹ Obviously, the power of these traditional regulatory mechanisms is limited. As the foregoing illustrates, it is unthinkable to let the President of the United States exercise her veto power in exchange for a fine. Monetary sanctions fail in other contexts as well. In sports, for instance, it would appear inappropriate to request players to pay for violating the rules of the game. One needs to think, then, of alternative options to curb potentially harmful behavior.

Numerical caps constitute such an alternative. Indeed, in sports, the use of quotas is pervasive.¹² This is no coincidence, as more familiar regulatory tools, such as monetary fines, appear inappropriate. Tennis players, to give one example, are entitled to unsuccessfully challenge referee decisions only three times per set.¹³ Similar to other contexts, the “right” to challenge the referee should be limited—it serves a desirable goal (promoting accuracy) but also entails undesirable consequences.¹⁴ The three-challenge rule achieves a balance without imposing monetary sanctions—players will presumably invoke the right to challenge the referee only in the most crucial situations, as if “paying” from their pre-assigned quota. The use of quotas, of course, is not limited to sports. Caps exist in various other domains, including litigation,¹⁵ bankruptcy,¹⁶ and the provision of government services.¹⁷

This Article seeks to apply the insights from the use of real-world quotas to other legal settings, in which the common tools are not a viable alternative. It has several general goals. First, it attempts to analyze quotas as a mechanism to regulate potentially harmful behavior, comparing them to more familiar tools such as direct regulation, pricing, and case-by-case determinations. Second, it uses actual examples of quotas to expose the theoretical reasons that underlie their use. Third, it utilizes the theoretical insights to suggest extending the use of quotas to several concrete legal settings.

on the injurers, namely, strict liability).

¹¹ See *infra* notes 30–31 and accompanying text.

¹² See *infra* note 44.

¹³ See, e.g., *Hawk-Eye Challenge Rules Unified*, BBC (March 19, 2008), <http://news.bbc.co.uk/sport2/hi/tennis/7305404.stm>. For the implementation of the challenges system, see Ran Abramitzky et al., *On the Optimality of Line Call Challenges in Professional Tennis*, 53 INT’L ECON. REV. 939, 941–42 (2012).

¹⁴ For example, challenges delay the game. *Id.* at 941 (“When a challenge is initiated, it takes 20-30 seconds for the computerized path and final landing location of the ball to be calculated and shown to the umpire, players, and the crowd on a large screen.”).

¹⁵ For various litigation caps, see *infra* note 127 and accompanying text.

¹⁶ See *infra* note 19 and accompanying text.

¹⁷ For various public services caps, see *infra* notes 62 and 156 and accompanying text.

There are many alternatives to regulate potentially harmful behavior. In particular, economists have praised pricing as an efficient alternative. Appropriate pricing induces right-holders to take only the socially efficient acts, which are worth more than their price. What are, then, the justifications for using numerical caps in lieu of price mechanisms? In many areas, as the foregoing examples concerning vetoes and sports suggest, the more traditional mechanisms, such as pricing, are unavailable and/or are hard to implement. Along these lines, this Article lays out the theoretical framework for the use of quotas, outlining the reasons to prefer quotas to prices.

First, quotas simply obviate the use of money and hence can be superior when policymakers are reluctant to levy monetary charges. In an analogy to the classic arguments against alienability, forcing tennis players to pay for the right to challenge referee decisions seems inappropriate, and, in general, harmful to the very essence of sports. Quotas can alleviate these concerns. Similar considerations may drive the use of quotas in other contexts, such as imposing numerical caps on the right to disqualify judges and jurors.¹⁸ Second, price mechanisms allocate rights based on willingness-to-pay. However, in some contexts the willingness-to-pay criterion leads to an ineffective allocation. For instance, pricing prevents the poor from purchasing entitlements that they need. These settings invite the use of caps. To illustrate, individuals in the United States, among other restrictions, cannot file for bankruptcy more than once every eight years.¹⁹ The desire to curb over-use of the right to file for bankruptcy is evident. However, pricing is not a suitable option, as the *raison d'être* behind bankruptcy laws is to provide debtors, who lack financial means, a fresh start. As pricing the right to file for bankruptcy seems ineffective, one has to think of other mechanisms, such as quotas, to allocate that right. Third, quotas can also be superior to pricing when monetary values are difficult to calculate. Quotas save the need to translate quantifiable values to monetary terms, and it may be easier for policymakers to target a desirable quota rather than calculate a fine. Possible examples include the more familiar cap-and-trade regimes in environmental contexts.

Numerical caps are, then, a substitute for prices. When properly set, they discourage over-use and induce right-holders to prioritize within the limit and invoke their rights only in socially appropriate instances. Of course, compared to prices, numerical caps have notable disadvantages that

¹⁸ See *infra* notes 49–50 and accompanying text.

¹⁹ 11 U.S.C. § 727(a)(8) (West 2012).

should be taken into account. In particular, non-tradable caps, which cannot be exchanged for money, impose identical restrictions on all individuals without considering particular needs or capacities, and they require right-holders, in certain contexts, to utilize their assigned cap without information as to their future needs. While there exist avenues to mitigate the concerns, these limitations should be considered by the designers of actual numerical caps.

Against this theoretical backdrop, I argue that American legal procedure can benefit from employing numerical caps on the use of courts. While there are some sporadic quotas in legal procedure, this Article seeks to provide a general argument of litigation caps and concrete proposals thereof. On the one hand, litigation is beneficial, and it is considered an essential public service that the government has to provide to all individuals without an effective price tag. On the other hand, there seems to be over-use, or even abuse, of the federal courts. Numerical caps can present an alternative, striking a balance between conflicting considerations: avoiding over-use of the judicial system and simultaneously preserving equal access to it. Examples of concrete suggestions to implement quotas on litigation rights include quantity limits on interlocutory appeals and filings. More generally, one can think of a broader, pre-determined amount of litigation “vouchers,” to be used in lieu of money throughout all the steps of the legal process, from filing to appealing.

In addition to litigation, which I discuss in greater detail below, the logic of quota mechanisms can fit other, more general domains. One example is the provision of public services. There is a strong notion that access to essential services should not be price-dependent, since this would exclude the poor. Pre-defined numerical ceilings offer an effective way to limit the use of essential entitlements without employing money. Another area that can benefit from the use of numerical caps is regulation of government bodies. Consider a government organ that excessively uses its power such that restrictions are needed to control it. While the immediate regulatory alternative is to fine that organ, fining a government organ appears inappropriate and ineffective. As money is not an option, quotas should come to mind. These arguments can be demonstrated by the example that commenced this Introduction—numerical caps to restrict over-use of veto powers.

Numerical caps that regulate behavior—quotas—exist in everyday life. However, I am aware of no attempt to conceptualize quotas as a legal tool.

Voluminous literature, which the Article discusses, has compared the use of quantity and price as regulatory tools, but the discussion typically concentrates on environmental issues and tradable caps. The more general phenomenon of numerical caps, especially non-tradable ones, has been left without comprehensive analysis. By underscoring quotas as a unique policy mechanism, this Article aspires to provoke discussion on the appropriate ways to regulate behavior and to enrich the available tools to strike a balance between conflicting considerations. While I focus on extending the use of quotas to litigation settings, the potential of using quotas is vast, and it crosses diverse legal fields.

The Article proceeds as follows. Part I provides background for the use of quotas as regulatory instruments. It draws on insights from law and economics to posit quotas as a regulatory tool within a general array of policy alternatives to curb behavior. Part II outlines the justifications for using numerical caps in lieu of price mechanisms, relying on real-world examples. Part III discusses limitation of quotas. Part IV builds on these theoretical insights to suggest extending the use of quotas to litigation. The last part concludes and briefly extends the logic of quota mechanisms to other, more general domains.

I. REGULATORY ROADMAP

This part posits caps as a regulatory tool within a general array of policy alternatives to curb externalities. It sketches the relevant approaches to regulate harmful activity—direct, substantive restrictions on the scope of the relevant right; case-by-case determinations; and incentive-based mechanisms (price and quantity regulation).

A few preliminary clarifications are worthwhile. First, for the purposes of this Article, “quotas” or “caps” are quantity regulation mechanisms. They can be defined as restrictions on the number of times right-holders can invoke their rights, which are designed to restrain inefficient externalities and induce right-holders to prioritize and invoke their rights only in the most appropriate instances. Discussions on other numerical limitations are beyond the scope of this Article. Numerical ceilings that restrict a single dimension of an activity without directly encouraging participants to choose their best acts are not quotas for the purpose of this discussion. To illustrate, I do not treat “driving below thirty miles per hour” as a quota. A quota that directly forces drivers to take only the most important rides could be, for example, “each driver is assigned x miles per

day.”²⁰ This definition echoes the well-known distinction in torts between regulating wrongdoers’ level of care and level of activity.²¹

Second, this Article focuses on quantity limits that are designed to restrict harms to third parties, namely, negative externalities. There may be other quotas. To demonstrate, quantity limits can be used by policymakers to stimulate positive externalities—for instance, “each lawyer is obliged to commit three hundred hours of pro bono work, per her discretion.” While some of the insights that this Article presents do pertain to contexts other than negative externalities,²² such quotas merit separate discussion.

Finally, I discuss in this Article situations in which external regulation is required. It is well-known that in some settings externalities could be avoided by the possibility of bargaining between the victim and the right-holder.²³ However, in many typical real-world settings, such Coasean bargaining is impractical. A salient example is the case of numerous and dispersed victims.²⁴ Where bargaining is irrelevant, one should think how to intervene to regulate the right-holder’s harmful externalities. Real-world examples of quotas demonstrate such situations.

A. Curbing Externalities

Common human behaviors—driving, litigating, polluting, manufacturing drugs, etc.—often entail benefits to those who engage in the activity as well as harms to others. Driving and polluting endanger other participants and the environment, litigating burdens courts and rival litigants, and drugs have various side effects. The following paragraphs discuss three regulatory alternatives to handle these types of issues. This mapping might seem crude in some cases, and regulating harmful behavior can of course take many other forms;²⁵ nonetheless, this typology helps to

²⁰ To give another example, statutes of limitations restrict a single, substantive dimension of the right (“victims cannot bring an action after x years”) rather than the number of times it can be invoked (“ x lawsuits per year”), and hence I do not treat them as caps.

²¹ See, e.g., STEVEN SHAVELL, *FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW* 193 (2004) (“An injurer’s level of activity is to be distinguished from his level of care . . . [which measures things] such as slowing for curves, as opposed to the number of miles he drives.”). This distinction is perhaps crude, where some examples fall on the borderline. Nonetheless, I believe that this definition for quotas is sufficiently workable, helping to capture their unique advantages as a regulatory tool.

²² Another example is numerical ceilings that are used in contractual relations to restrict behavior that harms the parties to the contract, but not third parties. An example might be “the supplier can deliver later than agreed upon three times per year.” These contractual quotas are beyond the scope of this Article.

²³ See Ronald H. Coase, *The Problem of Social Cost*, 3 *J.L. & ECON.* 1 (1960).

²⁴ See, e.g., Jonathan S. Masur & Eric A. Posner, *Toward a Pigouvian State*, 164 *U. PA. L. REV.* 93, 103–05 (2015).

²⁵ Regulation can take place, for example, through courts after the harm materializes (for instance,

explain the use of real-world quotas and suggests other areas in which they can be implemented.

The first way to cope with harmful externalities is to directly limit the substance of the relevant right.²⁶ At one extreme, we can ban the relevant behavior altogether—e.g., forbid driving. By doing so, however, we lose socially desirable acts. More common are intermediate restrictions. One option is defining domains in which invoking the right is forbidden and areas in which there are no restrictions. In order to diminish road congestion and air pollution, for instance, the regulator could allow cars with odd-numbered license plates to drive on odd days only (and vice versa).²⁷ Likewise, the regulator could restrict a certain dimension of the relevant activity, e.g., limiting driving beyond thirty miles per hour. As these examples illustrate, attempts to directly define and control the relevant activity often result in arbitrary distinctions, facilitating socially inefficient acts and/or restricting beneficial ones. While direct regulation could be fine-tuned, it would require better information and increased efforts from policymakers. Policymakers instead may turn to alternative regulatory approaches.

A second way to avoid externalities is case-by-case determination: scrutinizing each and every single act and banning only the specific acts that are determined to be socially detrimental. One example is antecedent licensing of certain acts: drug manufacturers, for instance, cannot sell their products in the United States without approval by the Food and Drug Administration.²⁸ However, case-by-case determination is not always feasible. It could be complex and ineffective, as it requires policymakers to amass relevant data and make a decision in each instance.²⁹

tort litigation); or by agencies, based on the committed act and regardless of the harm. *See, e.g.*, Steven Shavell, *The Optimal Structure of Law Enforcement*, 36 J.L. & ECON. 255 (1993) (discussing the advantages and disadvantages of both approaches).

²⁶ This approach is often referred to as “direct,” or “command-and-control,” regulation. *See, e.g.*, Robert N. Stavins, *Experience with Market-Based Environmental Policy Instruments*, in 1 HANDBOOK OF ENVIRONMENTAL ECONOMICS 355, 358–59 (Karl-Göran Mäler & Jeffrey R. Vincent eds., 2003).

²⁷ *See, e.g.*, Cameron Hepburn, *Regulation by Prices, Quantities, or Both: A Review of Instrument Choice*, 22 OXFORD REV. ECON. POL’Y 226, 240, n.62 (2006) (noting that this unusual rule exists in Athens).

²⁸ *See, e.g.*, Yehonatan Givati, *Game Theory and the Structure of Administrative Law*, 81 U. CHI. L. REV. 481, 483–84 (2014) (discussing individual licensing as a regulatory alternative). Another example of case-by-case determination, in the post-harm context, is negligence litigation, which penalizes only faulty acts upon determination by a court.

²⁹ *See, e.g.*, Ian Ayres & Gideon Parchomovsky, *Tradable Patent Rights*, 60 STAN. L. REV. 863, 877 (2007) (criticizing the current, case-by-case policy with regard to patents, which “rel[ies] on the judgments of patent examiners ex ante or judges ex post . . .”).

A third way, incentive-based mechanisms, does not directly intervene in the relevant right; rather, these mechanisms impose a general constraint on an activity, allowing participants discretion to operate within the confines of the restriction. To illustrate, corrective taxes, often referred to as Pigouvian taxes, essentially put a price on the relevant activity. If properly set, they optimally encourage right-holders to embark on socially valuable acts and avoid disadvantageous ones. In principle, pricing does not require case-by-case determination or direct guidance. Examples can be found in the environmental-protection framework. Environmental taxes per pollution unit allow factories to “manage” their pollution emissions according to commercial justifications.³⁰ Charging appropriate congestion fees, to use another example, should similarly discourage drivers from worthless drives.³¹

Quantity regulation, or quotas, expands the foregoing alternatives. Numerical caps are also an incentive-based mechanism and, in principle, can achieve the same goal as pricing. Take the polluting factory example. Instead of charging a price per unit of pollution—which leads to an optimal level of production—the regulator can set the desired level of pollution, i.e., the maximum units of pollution allowed. Presumably, the proper pollution cap allows manufacturers to use the given limit to produce the most beneficial products, and only up to the socially worthy level.³² Other fields can adopt this concept. Drivers, for example, can be allocated a certain number of miles, forcing them to prioritize their driving behavior to the most important rides. In this sense, prices and quantities are interchangeable. They do not require policymakers to decide whether and how to restrict, sanction, or allow each specific act. Rather, they delegate this decision to those who are subject to the regulatory scheme. To the extent that policymakers can optimally set the relevant price/quantity, individuals can then internalize the harm they create, adopting socially optimal decisions.

³⁰ See, e.g., Stavins, *supra* note 26, at 363–73 (surveying similar regulatory programs).

³¹ See, e.g., Hepburn, *supra* note 27, at 240 (noting that such congestion charges exist in London). In addition to environment charges, other areas that are typically regulated through Pigouvian taxes are cigarettes and alcohol consumption. See, e.g., David S. Gamage, Note, *Taxing Political Donations: The Case for Corrective Taxes in Campaign Finance*, 113 YALE L.J. 1283, 1283–84 (2004). More generally, strict liability, which is triggered after the harm materializes, plays a role similar to corrective taxes, i.e., forcing defendants to pay the costs of their activity whether it is justified or not. See *supra* note 10 and accompanying text.

³² Or, in different words, prices and quantities are substitutes—“there is a formal identity between the use of prices and quantities as planning instruments.” Martin L. Weitzman, *Prices vs. Quantities*, 41 REV. ECON. STUD. 477, 480 (1974). The social planner can set “prices while the firms respond with quantities,” or it can “assign quantities while the firm reveals costs.” *Id.*, at 478.

Based on the foregoing typology, which direction is preferable? The answer is, of course, context-specific. Along these lines, Part IV elaborates on specific litigation contexts, explaining the reasons numerical caps can be superior to other alternatives in these domains. With that, in general incentive-based mechanisms—caps or prices—seem to be the most valuable tools.³³ While individual determinations heed the nuances of each case, they require costly inquiries into the merits of each act and typically delegate discretion to on-site decision-makers. Likewise, effective direct regulation of activity requires information regarding the various dimensions that need to be restricted; at least in some cases, it would be difficult to arrive at plausible substantive restrictions. Incentive-based mechanisms can overcome these problems.

The difficulty in using incentive-based mechanisms, which is discussed in greater detail below, is setting the appropriate price/activity level. Yet once policymakers set the price/activity level, price and quantity regulation obviate costly case-by-case determination and the necessity to control the precise manner in which the activity is performed. Rather, incentive-based mechanisms rely on participants' information, eliciting their preferences and inducing them to prioritize their behavior. Presumably, these mechanisms can more easily discourage socially inefficient over-use, achieving optimal outcomes. Against this backdrop, the next section discusses the choice between price and quantity regulation.

B. Price and Quantity

The preceding section articulates the rationale behind regulating harmful activities through incentive-based mechanisms. Policymakers opting for incentive-based mechanisms can use pricing or caps; caps can be tradable (“cap-and-trade” regimes) or inalienable. Economists, who typically advance the use of incentive-based mechanisms,³⁴ view pricing as the most straightforward policy alternative, and tradable caps as similar in spirit but somewhat inferior to prices.³⁵ Regardless, inalienable caps would be conceded inferior.

³³ For an argument concerning the general superiority of incentive-based mechanisms see, e.g., Masur & Posner, *supra* note 24.

³⁴ See, e.g., *id.*; N. Gregory Mankiw, *Smart Taxes: An Open Invitation to Join the Pigou Club*, 35 E. ECON. J. 14 (2009).

³⁵ See, e.g., Mankiw, *supra* note 34, at 18 (“cap-and-trade systems are better than heavy-handed regulatory systems. But they are not as desirable . . . as Pigouvian taxes . . .”); Masur & Posner, *supra* note 24, at 102 (“A cap-and-trade scheme is similar to a Pigouvian tax . . . [but for several subtle reasons] most economists prefer Pigouvian taxes . . .”). While the literature in economics has extensively discussed the choice between price and quantity regulation, especially in the context of

To illustrate, consider a situation involving multiple actors with different preferences and needs, but all subject to the same regulatory scheme—e.g., polluting factories. An appropriate price mechanism forces each participant to internalize the social costs of its acts, and those who highly value their activity would keep polluting after paying a price for it.³⁶ A cap-and-trade regime can achieve a similar result by enabling participants to buy and sell their allocations. These entitlements would be traded, for payment, from those who least value the activity to those who benefit the most from it. Hence, the argument goes, in these situations, tradable caps and prices can achieve an optimal result; however, non-tradable caps do not allow for optimal allocation of the relevant right.³⁷

This preference to pricing/cap-and-trade regimes notwithstanding, non-tradable numerical caps exist in various real-life situations. Moreover, Pigouvian pricing schedules, the most straightforward policy choice, do not seem to be popular, even in areas such as environmental regulation in which they were constantly advanced.³⁸ Why should policymakers choose non-tradable caps, or even tradable caps, over prices?

Previous literature that proposed to extend the use of cap-and-trade regimes beyond environmental law did not attempt to generalize the

environmental harms, this Article focuses on more basic considerations and abstracts away from many subtle issues that are less relevant to the quotas that I analyze in the text. See, e.g., Hepburn, *supra* note 27, at 240 (an overview of the merits of price and quantity instruments, in the context of road congestion). Under uncertainty, for example, if we care more about having the right quantity, it makes sense to fix quantities rather than prices. For an illustration see, e.g., *id.* at 231 (“suppose the relevant good is the provision of prompt medical treatment . . . If the marginal benefit of rapid treatment falls very quickly (perhaps because after a threshold delay, d , the patient will die), then the hospital should face a quantity instrument of the form ‘no patient shall face a delay of more than d days’”). Likewise, depending on the context and the available technology, charging a price may be more or less costly than allocating (and enforcing) a limited quantity. For an illustration, see Edward L. Glaeser & Andrei Shleifer, *A Reason for Quantity Regulation*, 91 AM. ECON. REV. 431 (2001). From a political economy perspective, to take into account another consideration, prices, essentially taxes, may on the one hand be preferable as they create public revenue; however, on the other hand, taxes in general are strongly resisted.

³⁶ See, e.g., Louis Kaplow & Steven Shavell, *On the Superiority of Corrective Taxes to Quantity Regulation*, 4 AM. L. ECON. REV. 1, 3–4 (2002).

³⁷ See, e.g., JOSEPH E. STIGLITZ, *ECONOMICS OF THE PUBLIC SECTOR* 229 (3d ed. 1999). See also Kaplow & Shavell, *supra* note 36, at 12–14. I elaborate on the issue of heterogeneity among participants and tradability of quotas in *infra* Part III.A.

³⁸ Stavins, *supra* note 26, at 420 (stating that “virtually no [complex pricing] systems have been adopted”); Ryan Bubb & Richard H. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. 1593, 1665 (2014) (maintaining that even simple pricing regimes are not common, “to the consternation of most economists”); Masur & Posner, *supra* note 24, at 94–100 (demonstrating that Pigouvian taxes are rarely used and encouraging policymakers to adopt them more commonly); Mankiw, *supra* note 34, at 14 (describing how, with respect to Pigouvian taxes “there is a large gap between the [supportive] beliefs of economists and those of the general public” and concluding that “economists are right and the general public is just ill informed”).

reasons for so doing.³⁹ More broadly, to my knowledge, there has been no attempt to conceptualize caps, especially non-tradable caps, as a general legal policy tool. The remainder of this Article seeks to integrate the foregoing discussion in economics with actual examples of quotas in order to delineate a more comprehensive framework for quotas as an effective regulatory device. It suggests that in certain circumstances, (inalienable) quotas do have simple and notable advantages over prices, and that these observations can explain real-world instances of quotas and be used to extend them to other domains.

II. THE CASE FOR A QUOTA REGIME

Synthesizing real-world examples, this Part discusses the considerations that are relevant to choosing quotas rather than prices. Specifically, policymakers turn to quotas when they are reluctant to use money due to commodification considerations, when allocation based on willingness-to-pay is not desirable, or when setting a price is difficult.

A. Commodification

A primary advantage of quotas over prices is that they simply obviate the use of money. Instead of incurring sanctions or fees, right-holders “pay” with their pre-allocated, limited quota. Though money is not involved, the fact that the right is limited in quantity forces right-holders to

³⁹ Examples of cap-and-trade proposals in more exotic settings include KENNETH E. BOULDING, *THE MEANING OF THE TWENTIETH CENTURY: THE GREAT TRANSITION* 135 (1965) (“[A] system of marketable licenses to have children is the only [solution to the overpopulation problem]”); Ian Ayres & Joshua Mitts, *Three Proposals for Regulating the Distribution of Home Equity*, 31 *YALE J. ON REG.* 77, 122 (2014) (proposing, regarding the problem of home equity mortgages, “a system of tradable leverage licenses [that] would [numerically] cap the number of high leverage loans [that lenders can offer]”); Ayres & Parchomovsky, *supra* note 29 (suggesting a regime of tradable patent rights); Christian Iaione, *The Tragedy of Urban Roads: Saving Cities from Choking, Calling on Citizens to Combat Climate Change*, 37 *FORDHAM URB. L.J.* 889 (2010) (extending cap-and-trade permits to road congestion); Peter H. Schuck, *Refugee Burden-Sharing: A Modest Proposal*, 22 *YALE J. INT’L L.* 243 (1997) (suggesting an allocation of countries’ responsibilities to host refugees through tradable quotas); Richard D. Smith & Joanna Coast, *Controlling Antimicrobial Resistance: A Proposed Transferable Permit Market*, 43 *HEALTH POL’Y* 219 (1998) (discussing a system of tradable permits to prevent the spread of resistance to the use of antimicrobials, which results from over-use of antimicrobial drugs); Richard Steinberg, *Economic Perspectives on Regulation of Charitable Solicitation*, 39 *CASE W. RES. L. REV.* 775 (1989) (raising the option of tradable permits for charity fundraising); AUSTL. BROAD. AUTH., *TRADING THE REGULATORY OBLIGATIONS OF BROADCASTERS* (2003) [hereinafter *AUSTRALIAN BROADCASTING REPORT*], http://pandora.nla.gov.au/pan/40791/20040311-0000/www.aba.gov.au/tv/research/projects/pdftrf/trading_oblig.pdf (proposing to permit networks to trade their mandatory programming obligations).

prioritize and use the right in the most important circumstances. In principle, if the quota is computed optimally, right-holders will behave optimally. These characteristics make quotas an attractive choice when the use of money directly violates the essence of the relevant right. Instead of money, quotas are employed.

Indeed, as demonstrated in the Introduction, the use of numerical caps is pervasive in sports, where money seems to have little meaning. Tennis provides one example: as the Introduction demonstrated, each player is allowed three chances per set to challenge referee decisions. (The quota only pertains to unsuccessful challenges and successful challenges are not deducted from the quota.) This quota serves a regulatory function, similar to the textbook examples of polluting factories, as it discourages over-use of the “right” to challenge the referee.⁴⁰ Moreover, the quota option appears as the best regulatory tool under the circumstances. One alternative is direct restrictions on the right to challenge the referee, but such restrictions are often arbitrary, and in the tennis context, reasonable substantive restrictions on challenging rights appear to be difficult to achieve. Another option, individualized decision-making, i.e., whether to allow each challenge, seems convoluted to implement. Finally, a pricing, incentive-based alternative, which induces players to invoke (and pay for) their best challenges, may be a viable option. However, forcing tennis players to pay for the entitlement to challenge referee decisions seems inappropriate. In analogy to the classic arguments against alienability,⁴¹ placing a price tag on the moves tennis players can choose would presumably transform the meaning of these moves in ways that are undesirable to the essence of the game.⁴² Imposing a quota, rather than a price, may corrupt the nature of the game to a lesser extent.⁴³ Along these

⁴⁰ See *supra* note 14 and accompanying text.

⁴¹ See, e.g., Margaret Jane Radin, *Market Inalienability*, 100 HARV. L. REV. 1849 (1987).

⁴² See generally John J. Sewart, *The Commodification of Sport*, 22 INT’L REV. SOC. SPORT 171 (1987) (arguing that the introduction of marketplace logic to modern sports degrades the nature of athletic activity). See also *id.* at 178 (“Sport has long been singled out as one of the few spheres of social life where rational meritocratic values are truly operational . . . one can hit or catch a ball or not. Commercialization and commodification have steadily eroded the ethic of skill democracy.”). Even if these arguments are unconvincing, and the foregoing commodification concerns are weak, the common use of quotas in sports can be explained by willingness-to-pay concerns. See *infra* note 56 and accompanying text.

⁴³ Conceivably, there may be situations in which caps also have an undesirable commodifying effect similar to prices. Indeed, previous literature has asserted that commodification arguments do not solely relate to the use of money. Tsilly Dagan & Talia Fisher, *The State and the Market—A Parable: On the State’s Commodifying Effects*, 3 PUB. REASON 44 (2011) (discussing non-market commodification, through state ordering—“in light of [the] inherent itemizing, categorizing and prioritizing nature” of regulatory interventions. *Id.*, at 44). This Article does not purport to identify the

lines, and given the reluctance to price, we observe in sports numerous (inalienable) caps,⁴⁴ as well as non-monetary sanctions.⁴⁵

These points about sports can be generalized to other contexts in which quotas can substitute for prices.⁴⁶ Consider the well-known numerical restrictions on the right to have children in China.⁴⁷ Presumably, policymakers want a qualified right.⁴⁸ Nonetheless, pricing the right to have children seems deeply inadequate, and regulatory alternatives such as case-by-case determinations and direct restrictions entail practical and conceptual difficulties. Other real-world caps seem to fit a similar rationale. The right of American litigants to disqualify potential jurors without stating a reason—peremptory challenges—is famously capped by a numerical ceiling.⁴⁹ A similar rule applies to the disqualification of judges in some jurisdictions.⁵⁰ While the benefits of peremptory challenges are clear—e.g., reducing the incidence of juror partiality which is harder to detect through for-cause challenges⁵¹—these quotas reflect an attempt to

exact boundaries of the commodification argument. Instead, it suffices to assume that commodification concerns are typically stronger when prices, rather than quotas, are used.

⁴⁴ In football, like tennis, each team can challenge the referee two times per game, and two successful challenges authorize a third. *See, e.g.*, John Clayton, *NFL Still Tinkering Under Hood*, ESPN (Mar. 30, 2014), http://espn.go.com/nfl/story/_id/10698781/mailbag-nfl-tinkering-replay-rules. In basketball, the number of timeouts each team can call is capped, and players are similarly assigned a “quota” of fouls. FÉDÉRATION INTERNATIONALE DE BASKETBALL, OFFICIAL BASKETBALL RULES Arts 18.2.5, 40.1 (Feb. 2, 2014), http://www.fiba.com/downloads/Rules/2014/Official_Basketball_Rules_2014_Y.pdf. In soccer, each team has a limited number of substitutions. FÉDÉRATION INTERNATIONALE DE FOOTBALL ASS'N, LAWS OF THE GAME 2015/2016 18 (2015), https://www.fifa.com/mm/Document/FootballDevelopment/Refereeing/02/36/01/11/LawsOfTheGameWEBEN_Neutral.pdf.

⁴⁵ For example, Rule 16.1 of professional ice hockey reads: “For a minor penalty, any player . . . shall be ruled off the ice for two . . . minutes.” NAT’L HOCKEY LEAGUE, OFFICIAL RULES 2016-2017 R16.1 (2016), http://1.cdn.nhle.com/downloads/2016-17_RuleBook.pdf.

⁴⁶ Note that in real-world settings non-monetary sanctions, which exist in sports, are less attractive. Pricing schemes transfer money from participants to the government, where non-monetary sanctions, such as incarceration, reduce social welfare. For a discussion in the context of litigation, see *infra* notes 113–14.

⁴⁷ For a description of the one-child policy in China *see, e.g.*, John Bongaarts & Susan Greenhalgh, *An Alternative to the One-Child Policy in China*, 11 POPULATION DEV. REV. 585, 586–89 (1985). This policy has recently been amended to allow Chinese couples to have two children. Barbara Demick, *Judging China’s One-Child Policy*, NEW YORKER (Oct. 30, 2015), <http://www.newyorker.com/news/news-desk/chinas-new-two-child-policy>.

⁴⁸ *Cf. infra* note 68 and accompanying text (discussing the target population that the one-child policy in China intended to achieve).

⁴⁹ The number varies according to the subject matter. *See, e.g.*, 28 U.S.C. § 1870 (West 2012) (three in misdemeanor cases and civil cases); FED. R. CRIM. P. 24(b) (twenty in capital cases).

⁵⁰ Litigants in California can remove the judge assigned to hear the case without proving bias—but only once. CAL. CIV. PROC. CODE, § 170.6(a)(4) (West 2012).

⁵¹ *See, e.g.*, Susan L. McCain, *Sex Discrimination in the Voir Dire Process: The Rights of Prospective Female Jurors*, 58 S. CAL. L. REV. 1225, 1250 (1985) (“a primary rationale for allowing

curb the right to disqualify judges and jurors without stating a reason. In theory, the numerical caps could have been replaced by pricing. However, “buying-out” judges and jurors seems contradictory to the very essence of judging. Alternatives to quotas and pricing, such as pre-defined rules to remove jurors or reliance on judges’ discretion, seem unsuitable to achieving the objective that peremptory challenges seek to achieve.⁵²

B. Willingness-to-Pay

Pricing allocates rights based on a willingness-to-pay. However, such an allocation can have drawbacks: income gaps, for example, prohibit the poor from purchasing the relevant right even when they highly value it, while allowing the rich to over-use their rights.⁵³ Policymakers may thus think that a pricing mechanism creates an ineffective allocation, especially when important rights are implicated. In that case, inalienable quotas become more attractive than prices. Specifically, choosing inalienable quotas can be justified based on paternalistic motivations, i.e., policymakers believe in the importance of an entitlement even for those who cannot afford it. Alternatively, a wide, equal allocation of the relevant right, among all strata of society, entails positive societal benefits.⁵⁴ While scaled pricing—charging a lower price from the poor and a higher one from the rich—is optional, it creates serious difficulties.⁵⁵

peremptory challenges in addition to challenges for cause” is that “most biases can only be intuited, not proven.”).

⁵² See, e.g., Charles J. Ogletree, *Just Say No!: A Proposal to Eliminate Racially Discriminatory Uses of Peremptory Challenges*, 31 AM. CRIM. L. REV. 1099, 1141–43 (1994) (“In the realities of the courtroom, peremptory challenges may be necessary as a check on the occasional . . . unconscious racism, of a trial judge The elimination of peremptory challenges would give more power to trial judges . . . where trial judges already enjoy immense discretion and little potential appellate review.”). It should be noted that Ogletree finds the practice of peremptory challenges problematic, in particular, because “the harm caused today by the racial use of the peremptory.” *Id.*, at 1150. He concludes, then, that the option of peremptory challenges should be limited only to criminal defendants (“Legislatures or courts should . . . expand the for-cause challenge; and, where possible, abolish or drastically reduce peremptories for all but criminal defendants.” *Id.*, at 1151).

⁵³ See, e.g., Cass R. Sunstein, *Cognition and Cost-Benefit Analysis*, 29 J. LEGAL STUD. 1059, 1090 (2000) (“Poor people are willing to pay less than wealthy people In the face of disparities in wealth, willingness-to-pay should not be identified with expected utility or with the value actually placed on the good in question.”).

⁵⁴ The discussion here does not purport to exhaust the range of considerations for and against using the willingness-to-pay criterion. The literature on this topic is voluminous. Rather, the purpose is to highlight the main willingness-to-pay reasons to use non-tradable caps instead of prices. Cf. STIGLITZ, *supra* note 37, at 86–88, 362 (discussing “merit goods,” commodities that are not allocated based on willingness-to-pay as they are deemed essential to all individuals, due to paternalistic reasons or positive externalities).

⁵⁵ A differential price should ideally depend on the personal characteristics of each right-holder; hence, it reiterates the problems associated with individual determinations. Moreover, setting a differential price requires cumbersome calculations in order to tie the fee to each individual’s

The aforementioned numerical restrictions, which are presumably related to commodification issues, could also stem from willingness-to-pay concerns. In sports, for instance, even in the absence of commodification concerns, allocating rights based on willingness-to-pay would risk the value of competition.⁵⁶ Allocating the rights to have children based on willingness-to-pay would endanger diversity in the population.⁵⁷ For similar reasons, allowing a trade in these (inalienable) caps would defeat their purposes, as the entitlements would plausibly flow from the poor to the rich.

Willingness-to-pay concerns could also be demonstrated through other real-world quotas. Consider the rule that allows filing for bankruptcy once in an eight-year period.⁵⁸ Like the textbook examples of externalities, allowing debtors a fresh start is a valuable right, yet it simultaneously invites over-use.⁵⁹ The one-in-eight-years cap aims, therefore, to achieve a balanced outcome by eliminating at least some meritless filings.⁶⁰ Individual determination of bankruptcy rights may be difficult and time-consuming. More importantly, pricing the right is not a suitable option. The *raison d'être* behind bankruptcy laws, i.e., receiving a financial fresh start, makes requiring debtors pay for bankruptcy rights self-defeating. For similar reasons, it makes sense to ban the trade of these caps: we may think that fresh-start rights entail positive externalities for the community that we want to foster, and/or we suspect that at least some potential debtors will

willingness-to-pay. Furthermore, a differential price means that the rich pay more than the actual social costs they inflict—which can drive them to look for other options (for a discussion of this consideration in the context of litigation, see *infra* note 112 and accompanying text). For these reasons, perhaps, scaled prices do not seem to abound in real-life situations. See, e.g., Kevin Outterson, *The Vanishing Public Domain: Antibiotic Resistance, Pharmaceutical Innovation and Intellectual Property Law*, 67 U. PITT. L. REV. 67, 106–07 (2005) (raising similar considerations against a hypothetical differential tax on antibiotics).

⁵⁶ See, e.g., Natalie L. St. Cyr Clarke, *The Beauty and the Beast: Taming the Ugly Side of the People's Game*, 17 COLUM. J. EUR. L. 601 (2011) (discussing measures taken by sports associations to increase competition, such as salary caps).

⁵⁷ See, e.g., Demick, *supra* note 47, at 3 (describing the change in the one-child policy in China, where “the rich have been able to buy their way out of” the one-child quota, inciting concerns regarding “uneven[] and unfair[]” enforcement and pushing the government to allow couples to have two children).

⁵⁸ 11 U.S.C. § 727(a)(8) (West 2012).

⁵⁹ See, e.g., Katherine Porter, *Bankrupt Profits: The Credit Industry's Business Model for Postbankruptcy Lending*, 93 IOWA L. REV. 1369, 1371 (2008) (noting how recent reforms aimed to reduce incentives for debtors to “‘overborrow’ and then seek relief from the bankruptcy system”).

⁶⁰ Illustrative is President Bush's explanation for enacting a more limiting quota: “The new law will . . . make it more difficult for serial filers to abuse . . . bankruptcy protections. Debtors seeking to erase all debts will now have to wait eight years from their last bankruptcy before they can file again.” Press Release, White House, President Signs Bankruptcy Abuse Prevention, Consumer Protection Law (Apr. 20, 2005), <http://georgewbush-whitehouse.archives.gov/news/releases/2005/04/20050420-5.html>.

sell their bankruptcy rights to their detriment. One can find other examples of inalienable quotas that are presumably designed to restrict rights without using money, allowing the less well-off to exercise these rights.⁶¹ More generally, along these lines quotas can be employed to restrict the provision of important public entitlements, enabling the poor as well as the rich to enjoy those rights, albeit to a limit.⁶²

C. Difficulties in Monetizing

In some situations, policymakers are willing to charge a fee, but setting the optimal price may be a daunting task. This consideration invites the use of quotas because quantifying can be easier than monetizing.

Policymakers who desire to impose an accurate price should at the least compute the social harm from each relevant act. This is, however, an extremely difficult duty in complicated real-world situations. Setting a price tag can be particularly problematic when the behavior in question, e.g., litigating, implicates both negative and positive externalities that need to be accounted for, as well as when some of the relevant costs or benefits refer to abstract values, with no market price (the right to have children, for instance, presumably enhances human dignity). Thus, policymakers often have to use cruder mechanisms.

In those instances where policymakers engage in rough cost-benefit calculations, setting quantities rather than prices may be easier. Although quotas require quantification, they eliminate the need to monetize, or put a price tag on, the act. An identical distinction was endorsed by the U.S. Office of Management and Budget (OMB) as part of its guidance on cost-benefit analysis. The OMB recognized the difficulty of cost-benefit analysis for goods that “are not traded in markets.”⁶³ Accordingly, when

⁶¹ One example is numerical restrictions on the right to use small claims court. As with bankruptcy laws, the main idea behind small claims courts is to enable access to justice for those who cannot afford it. Charging fees to deter excessive use of small claims courts is thus contradictory to their justification. Instead, quota restrictions better achieve this task. Broader access to justice for the poor may constitute an independent societal value; hence, trade in these caps is unwarranted. Individual “permits” to sue in small claims courts and substantive limitations seem cumbersome, particularly in light of the small monetary stakes. *Cf.* Courts Law (Consolidated Version), 5744-1984 § 60(b), 38 (1983–1984) (Isr.) (establishing a quota of five lawsuits per year in small claims court).

⁶² I further discuss this point in the Conclusion. *See, e.g., Australian Passports Act 2005* § 15(b) (Austl.) (the Australian (discretionary) limitation on issuing a new passport to those who lost two passports in the previous five years); Names Law, 5716–1956, § 20, 10, (1955–1956) (Isr.) (the Israeli rule that allows citizens to change their name, in principle, once every seven years).

⁶³ OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, CIRCULAR A-4, REGULATORY IMPACT ANALYSIS: A PRIMER 9 (Aug. 15, 2011), https://obamawhitehouse.archives.gov/omb/inforeg_regpol_agency_review.

“monetization is not possible,” agencies should instead quantify.⁶⁴ To illustrate, “an agency may be able to quantify, but not to monetize, increases in water quality . . . resulting from water quality regulation. If so, the agency should attempt to describe benefits in terms of (for example) stream miles of improved water quality”⁶⁵

This lesson can be extended to other settings. While it may be difficult to assign a price to dignity or privacy, it is easier to quantify the number of beneficiaries.⁶⁶ Lives may be difficult to value, but it is easier to target and compare, say, a five percent reduction in fatalities. The link to caps is straightforward. Quotas save the need to translate quantifiable values to monetary terms.

More generally, we are sometimes more confident agreeing on and fixing in advance a quantitative target—a limited number of entitlements that we are willing to allocate—than we are setting a price and hoping to achieve the optimal quantity.⁶⁷ The one-child policy in China, for example, targeted a specific number: “hold[ing] the population at 1.2 billion by the end of the [twentieth] century.”⁶⁸ Policymakers (and human beings) are simply “more comfortable” with comparing quotas rather than prices.⁶⁹ While such an “attenuated way” to balance costs and benefits is perhaps less accurate than comparing money, this “second best” approach can be more effective, especially when uncertainty and measurement problems plague attempts to determine a price tag.⁷⁰ Quotas can be particularly attractive in value-laden areas, in which calculating an accurate price is

⁶⁴ *Id.* at 12.

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ See, e.g., Sandra Rousseau & Kjetil Telle, *On the Existence of the Optimal Fine for Environmental Crime*, 30 INT’L REV. L. & ECON. 329, 334 (2010) (“[W]hen fines cannot be optimally designed, the most ambitious goal a regulator might have is to avoid really harmful situations [through quotas].”).

⁶⁸ Bongaarts & Greenhalgh, *supra* note 47, at 587–88. By 2000, China’s population totaled 1.266 billion. Bingham Kennedy, Jr., *Dissecting China’s 2000 Census*, POPULATION REFERENCE BUREAU (2001), <http://www.prb.org/Publications/Articles/2001/DissectingChinas2000Census.aspx>.

⁶⁹ Nathaniel O. Keohane et al., *The Choice of Regulatory Instruments in Environmental Policy*, 22 HARV. ENVTL. L. REV. 313, 364 (1998). Indeed, other rough predetermined quantitative limitations play a similar role in various daily settings. Consider, for instance, a typical law review submission process. Law reviews often have a certain, but fixed, number of articles per volume, e.g., fifteen. This means that they are willing to publish the best fifteen articles they can get each year, even if the “objective” quality of these articles varies.

⁷⁰ Richard L. Revesz, Book Review, 11 ECOLOGY L. QUART. 451, 460–61 (1984). See also Iaione, *supra* note 39, at 910 (given the problems with pricing, “quantity instruments seem to be the most cost-effective tools [when] the socially acceptable ‘how much’ has been selected.”). The advantage of quotas over prices in this respect dissipates when costs and benefits cannot be quantified either. For a discussion on quantification problems, see Cass R. Sunstein, *The Limits of Quantification*, 102 CAL. L. REV. 1369, 1382–83 (2014).

nearly impossible and a politically-determined quantity can be easier to set.⁷¹

By the same token, quotas can be more effective in guiding the behavior of right-holders. Due to various cognitive limitations, prices sometimes fail to move individuals to fully appreciate the externalities they create.⁷² Quotas may serve as an “anchor” that is more conspicuous than a price, thereby better influencing individuals’ perceptions regarding the appropriate level of activity.⁷³

Translating human behavior into prices—monetizing human behavior—entails difficulties. These difficulties pertain both to those who set the price and those who respond to the price. Quotas offer an alternative. They can be easier to implement when monetizing is complicated and more powerful in shaping the behavior of right-holders. However, these differences between prices and quotas do not necessitate inalienable quotas, and caps that are motivated by monetizing obstacles can in principle be tradable. The more familiar cap-and-trade policies in environmental contexts can be explained by monetizing challenges rather than commodification or willingness-to-pay concerns.⁷⁴ Other cap-and-trade programs in the fields of broadcasting and housing obligations serve as additional examples of such a use of quotas.⁷⁵

III. LIMITATIONS

The use of quotas is not free of difficulties. This part discusses possible limitations on the use of quotas: the one-size-fits-all nature of quotas and information problems.

⁷¹ Moreover, as these contexts typically require intricate determinations, delegating discretion to lower-echelon, on-site decision-makers becomes more problematic. Hence, allocating the rights on a case-by-case basis may not be a favorable option.

⁷² See, e.g., Bubb & Pildes, *supra* note 38, at 1673–77.

⁷³ For a discussion of a similar phenomenon in the context of deadlines, see Eyal Zamir et al., *It’s Now or Never! Using Deadlines as Nudges*, 42 L. & SOC. INQUIRY 769, 771–74 (2017).

⁷⁴ See Keohane et al., *supra* note 69, at 364 (summarizing the reasons that pricing schemes are not employed in environmental contexts). See also Mankiw, *supra* note 34, at 16–17 (demonstrating the difficulty in pricing environmental harms through the need to presume a discount rate); Thomas Merrill & David M. Schizer, *Energy Policy for an Economic Downturn: A Proposed Petroleum Fuel Price Stabilization Plan*, 27 YALE J. ON REG. 1, 15–16 (2010) (“The degree to which prices should be raised to constrain [social] costs [in the context of fuel prices] is a matter of judgment [that] must ultimately be determined politically,” as the relevant social costs entail insurmountable measurement difficulties).

⁷⁵ See AUSTRALIAN BROADCASTING REPORT, *supra* note 39; ORG. FOR ECON. CO-OPERATION & DEV., PUTTING MARKETS TO WORK: THE DESIGN AND USE OF MARKETABLE PERMITS AND OBLIGATIONS, 36–39 (1997) [hereinafter OECD REPORT], <http://www.oecd.org/gov/regulatory-policy/1910849.pdf>.

A. One-Size-Fits-All

Quotas employ a one-size-fits-all approach, allocating identical rights to each and every person. As I mention above, diversity in preferences and needs makes uniform allocations inefficient. The more diverse the right-holders and their preferences, the greater the inefficiencies.

The solution to this problem is straightforward. Trade in quotas encourages an efficient internal allocation among diverse right-holders, ensuring that entitlements flow from those who have little need for them to those who value them the most, for each side's mutual welfare. However, the problems with this solution are also straightforward. The same considerations that hinder policymakers from charging a price indicate that trade is not desirable. In particular, tradability does not accommodate quotas that are driven by the desire to avoid using money, due to commodification or willingness-to-pay concerns. Actual quotas that seem to stem from reluctance to price—for example, bankruptcy rights—are indeed inalienable.

However, quotas that are not based on the desire to avoid using money need not be inalienable. Furthermore, even if they exist, in some contexts arguments against tradability are weaker; for example, minor income gaps or when the relevant right does not seem to be highly important. These contexts can benefit from a partial tradability of quotas.⁷⁶ Trade can be allowed with some restrictions, such as creating a limit on the rights that each right-holder can buy or sell such that each person is left with a minimal “floor” or “ceiling” of rights.⁷⁷ At least in some settings, these restrictions address the core problems that justified the use of quotas, for instance, diversity in the use of the right.⁷⁸

In other settings, the same reasons that led to quotas, namely, reluctance to price, would mandate inalienable quotas. In this case, policymakers could use means other than tradability to mitigate the one-size-fits-all problem and promote a more tailored assignment of inalienable caps. Quotas can be combined with individual decision-making to allow for deviations from the initial allocation. The injection of particular discretion

⁷⁶ For a discussion on the range of legal techniques that enable partial tradability, depending on the underlying considerations, see Tsilly Dagan & Talia Fisher, *Rights for Sale*, 96 MINN. L. REV. 90, 105–06 (2011).

⁷⁷ For real-world examples of limitations on the proportion of the quota that can be tradable, and the class of entities that can engage in a trade, see OECD REPORT, *supra* note 75, at 37, 43. Furthermore, governments could mediate the trade in caps—for example, through buying quotas from some right-holders and selling them to others under a price system.

⁷⁸ See *supra* note 57 and accompanying text.

narrows some of the appeal of quotas, but presumably preserves some of their benefits. Along these lines, policymakers can allocate the relevant rights based on finer-grained caps, and different individuals can receive different caps. A real-world example is subsidizing the poor through larger quotas.⁷⁹

To the extent right-holders are different from one another, an inalienable quota results in inefficient internal allocation. While this problem could be mitigated by some measures, such as finer-grained quotas, the one-size-fits-all nature of inalienable quotas diminishes their appeal relative to prices. In that case, quotas become closer to the alternative of direct regulation, though they regulate the entire activity rather than a single dimension thereof.⁸⁰

B. Information Problems

Information problems are relevant both to policymakers who set the quota and to the users of the quota, who have to manage their allocation over time.

1. Policymakers

Quantitative limits on entitlements may appear arbitrary, whereas pricing seemingly reflects a thoughtful balance of costs and benefits. Can policymakers set numerical limitations in a reasoned manner? I argue that policymakers' ability to set the relevant figure is not an insurmountable obstacle to the use of quotas.

First, the limit (a number) should be based on serious quantitative or qualitative research—akin to regular cost-benefit analysis. Consider the rule that allows litigants to depose ten witnesses without leave of court.⁸¹ The desired number of depositions in each case could be based on surveys of judges and practitioners. Somewhat along these lines, the ten-deposition limit is based on the discussions of a specialized committee, and it seems

⁷⁹ See, e.g., OECD REPORT, *supra* note 75, at 40 (discussing providing more landing slots to struggling airlines); Jonathan B. Wiener, *Global Environmental Regulation: Instrument Choice in Legal Context*, 108 YALE L.J. 677, 765–66 (1999) (discussing more generous allowances to pollute for poorer countries).

⁸⁰ Inalienable quotas may still be preferable to direct regulation. It might be easier for policymakers to target a desired level of activity than control a specific dimension thereof. See *supra* notes 67–69 and accompanying text. Relatedly, sometimes it would be difficult to find a single dimension of the activity that could be regulated in a plausible manner. Finally, unlike direct regulation, as demonstrated in the text, policymakers can allow a restricted trade in quotas, mitigating the one-size-fits-all concern.

⁸¹ FED. R. CIV. P. 30(a).

to be a reasonable limitation.⁸² Second, quotas, like prices, do not need to be static: over time they can be modified and adjusted. The process is straightforward when caps are tradable and their market price becomes clear.⁸³ Adjustments are also available regardless of the alienability of the entitlements, through a simple trial-and-error process. In the depositions example, ongoing surveys of judges and practitioners, after the implementation of the new policy, allow auditing the process and updating the numerical limit when needed. In this context, the Advisory Committee on Civil Rules has recently considered lowering the deposition cap from ten to five.⁸⁴ Third, the problem of inaccuracies in the relevant quota can also be mitigated through the combined use of caps and individual discretion—indeed, the current policy allows the first ten depositions without leave and depositions beyond that with judicial approval.⁸⁵ Finally, quotas are rough approximations, and part of their appeal stems from the inability to easily provide price tags. In this sense, quotas are second best, but may nonetheless outperform any feasible alternative.⁸⁶

2. Right-holders

In several settings, quotas provide a limited entitlement that stretches over time. As noted, individuals can file for bankruptcy only once in an eight-year period; likewise, tennis players can unsuccessfully challenge the referee three times within a set.⁸⁷ In these types of quotas, right-holders must temporarily ration their allocation. However, oftentimes right-holders face uncertainty with regard to the future, e.g., in the future, they may or may not need to file for bankruptcy. Uncertainty reduces right-holders' ability to prioritize their acts. They may employ their rights when the societal harm from doing so does not justify the benefit (over-use) or miss the opportunity to invoke their entitlement when it is needed the most (under-use).

⁸² See Thomas E. Willging et al., *An Empirical Study of Discovery and Disclosure Practice Under the 1993 Federal Rule Amendments*, 39 B.C. L. REV. 525, 538 (1998) (“75% of [surveyed] attorneys . . . said seven or fewer individuals were deposed . . .”).

⁸³ See Kaplow & Shavell, *supra* note 36, at 13. Too high a price suggests that the quota should be more generous (and vice versa).

⁸⁴ This proposal was later withdrawn. Adam N. Steinman, *The End of an Era? Federal Civil Procedure After the 2015 Amendments*, 66 EMORY L.J. 1, 19, 26 (2016).

⁸⁵ FED. R. CIV. P. 30(a).

⁸⁶ However, where the quota (or price) is arbitrary, other regulatory alternatives, such as direct regulation, become more attractive. See *supra* note 80 and accompanying text (discussing inalienable quotas and direct regulation).

⁸⁷ For other real-world quotas that assign rights for a limited period of time, see *supra* notes 61–62.

The degree to which information problems harm the ability to optimally employ assigned rights over time is a context-specific, empirical question. While right-holders' informational difficulties may render quotas less effective in some cases, they should not preclude the use of quotas altogether. First, in many settings, the problem is obviated because right-holders are not required to plan ahead. In the context of depositions, for example, litigants can typically submit a list of all witnesses to depose at the start of the proceedings. In many other instances, right-holders are at least somewhat knowledgeable about their future state, and it seems relatively easy for them to effectively manage their numerical allocations over time.⁸⁸ Second, the ability to trade or partially trade quotas alleviates these concerns. Individuals needing an entitlement at a later date can simply buy it from those who do not need it.⁸⁹ Third, to ease under-use concerns, quotas can be supplemented with individual decision-making in order to allow those who have exhausted their allocations to re-invoke their rights, or "borrow" against future rights,⁹⁰ at least in unique circumstances. Consider bankruptcy rights: we may think that debtors who recently filed for bankruptcy but soon thereafter file again, due to reasons beyond their control, should be entitled to receive this exceptional right at the discretion of a court. Along the same lines, policymakers can provide a more generous quota to allow right-holders a wider margin of error when invoking their rights over time and to prevent under-use of rights.

Finally, the pricing alternative may raise similar problems, at least in some contexts. Consider a right that is regulated through a meaningful price. When deciding whether to buy the entitlement, participants consider future fluctuations (i.e., uncertainty) in their wealth, and future uncertainty can distort their decision with regard to the correct timing to purchase the right.⁹¹ In this respect, the differences between quotas and prices are matters of degree. Accordingly, by providing right-holders with a more general quota that "bundles" several types of entitlements and allows right-holders greater control over their rights, policymakers can alleviate information problems and make quotas more like money. Along these lines, later I briefly discuss allocating litigants a broader cap of litigation

⁸⁸ Individuals often know, for example, when they would like to have children. *See also* Abramitzky et al., *supra* note 13 (presenting empirical findings that tennis players almost perfectly optimize their referee-challenge quotas).

⁸⁹ Of course, there are good reasons to restrict trade in certain quotas. *See supra* Part III.A.

⁹⁰ For a real-world example in the context of fuel economy standards, see Stavins, *supra* note 26, at 407.

⁹¹ An example, in a different context, is the purchase of a house.

“coupons” that can be used to “pay” for various procedures such as depositions, appeals, and amendments to pleadings.⁹²

* * *

The foregoing presented the case for a quota regime, highlighting through real-world examples the advantages of quotas over prices. In sum, when policymakers desire to avoid using money, due to commodification or willingness-to-pay concerns, or because prices are difficult to calculate, quotas offer an immediate policy alternative. The first two considerations also support non-tradable caps. The remainder of this Article moves from theory to practice, suggesting concrete implementations of quotas, particularly in legal procedure.

Before moving forward, it is important to note that once chosen, there are many variations to quotas, and this flexibility can assist in tailoring this regulatory tool to the relevant context and motivation behind its use. As mentioned, quotas can be tradable, partly tradable, or inalienable. To the extent that commodification and willingness-to-pay considerations are weak, policymakers can move from inalienability to partial tradability or tradability. Quotas can be intermingled with case-by-case determinations—one example is allowing for discretionary deviations from the quota when the quantitative limit has been exhausted.⁹³ Likewise, quotas and prices can be combined; for example, after reaching the numerical limit, an additional fee can be charged.⁹⁴ Policymakers can fine-tune caps, i.e., assign different quantitative limits for different individuals. Quotas can set a quantitative limit, e.g., once per year; but they can also set a desired ratio, e.g., allowing one frivolous suit for every three meritorious ones. More generally, quotas can be adjusted over time to fit the numerical limit to the exigencies of that time.⁹⁵

⁹² *Infra* Part IV.C.1.

⁹³ See *supra* notes 85 and 90 and accompanying text.

⁹⁴ The actual example of federal inmate litigation, *infra* notes 150-51 and accompanying text, demonstrates this mechanism.

⁹⁵ Constant adjustments and modifications also enable quotas to better mimic sophisticated pricing schedules. See Kaplow & Shavell, *supra* note 36, at 12–14. See also Brian Galle, *Tax, Command . . . or Nudge?: Evaluating the New Regulation*, 92 TEX. L. REV. 837, 860–64 (2014) (discussing ways to constantly improve information regarding the optimal price or quantity).

IV. APPLICATION: LEGAL PROCEDURE

Federal litigation, I argue, is particularly ripe for the use of quotas. Since litigation is beneficial, but also costly, we would like to allow only some litigation while discouraging over-use. Moreover, as I discuss below, alternative attempts to regulate litigation behavior have fallen short. It is extremely difficult to place a price tag on the use of the legal system; and even if it were possible, policymakers tend to view the legal system as an essential public service, unwilling to charge an actual price. Case-by-case determinations also have their own difficulties. They consume precious judicial time and can block litigants from using courts without a sufficient basis. Similarly, it can be challenging to construct plausible substantive restrictions on the right to use courts. Quotas, then, can enrich the available array of mechanisms that regulate litigation behavior.

A. To Price or not to Price

It is common to think that the federal courts are under immense pressure from a substantial workload, which may harm their capacity to administer justice.⁹⁶ While the reasons for this workload crisis and its implications are under a heated debate, and the American legal system is definitely plagued by various other afflictions,⁹⁷ abusive litigation behavior and meritless filings at the least seem to be a major concern.⁹⁸ To the extent abusive litigation is a problem, how should policymakers respond? The straightforward reaction is to charge litigants “user-fees,” i.e., force them to pay for the actual costs they inflict on the legal system (as well as on their rival litigants). User-fees can take the form of pre-filing tariffs on lawsuits and interim motions, and/or post-judgment sanctions on inappropriate litigation behavior. While litigating in the United States is far from free,⁹⁹ the current charges appear to fall short from reflecting the

⁹⁶ See, e.g., Bert I. Huang, *Lightened Scrutiny*, 124 HARV. L. REV. 1109 (2011) (discussing the notion of “crisis of volume” in the federal courts and empirically demonstrating its implications).

⁹⁷ Notably, evidence suggests that many plaintiffs with valid claims are not compensated. See David M. Studdert et al., *Claims, Errors, and Compensation Payments in Medical Malpractice Litigation*, 354 NEW ENG. J. MED. 2024 (2006) (discussing the exorbitant overhead costs and frequent denial of compensation in medical malpractice cases).

⁹⁸ See, e.g., Arthur R. Miller, *Simplified Pleading, Meaningful Days in Court, and Trials on the Merits: Reflections on the Deformation of Federal Procedure*, 88 N.Y.U. L. REV. 286, 357–71 (2013) (describing, from a skeptical perspective, the prevailing perceptions, and questioning their empirical foundations, *id.*, at 361–64). Discussions on the recent amendments to the Federal Rules of Civil Procedure are triggered by similar notions of abusive litigation. Steinman, *supra* note 84, at 14–18.

⁹⁹ The American legal system appears to be “neutral” to money. On the one hand, litigants are not charged meaningful fees and subsidies are generally not available. On the other hand, hiring lawyers is

actual social costs of litigation. Federal litigants are only charged fees upon filing, which seem modest at best; loser-pays rules are not the norm; and sanctions against abusive litigation are rarely imposed.¹⁰⁰ It would seem that, by and large, litigants are induced to over-use the legal system. Accordingly, numerous policymakers and commentators have proposed—to no avail—a radical reform of the pricing of American litigation.¹⁰¹

Of course, those who believe that abusive litigation is not a major problem oppose higher fees. Yet it seems that the opposition to pricing is more pervasive and also pertains to those who believe that the system is over-used. There seem to be two explanations for this wider opposition to a meaningful price on litigation. First, deciphering the correct price seems to be an insurmountable task. While the direct costs of the legal system can theoretically be measured—e.g., judges' time¹⁰²—litigation behavior

costly, and wealthy litigants presumably fare better. For a description of these conflicting views on pricing in the American legal system, see Judith Resnik, *Money Matters: Judicial Market Interventions Creating Subsidies and Awarding Fees and Costs in Individual and Aggregate Litigation*, 148 U. PA. L. REV. 2119, 2120–44 (2000).

¹⁰⁰ As Judge Frank Easterbrook observed, “[f]ederal courts are subsidized dispute-resolvers [as] filing fees defray only a small portion of the costs.” *Lewis v. Sullivan*, 279 F.3d 526, 528 (7th Cir. 2002). While filing fees in the federal courts are \$350, 28 U.S.C. § 1914 (West 2012), loser-pays rules are not the norm. A comparative look illustrates that the price for accessing federal courts is minimal. See Theodore Eisenberg et al., *When Courts Determine Fees in a System with a Loser Pays Norm: Fee Award Denials to Winning Plaintiffs and Defendants*, 60 UCLA L. REV. 1452, 1454 (2013) (noting that the prevailing norm in the world is the English, loser-pays rule); Elizabeth G. Thornburg, *Saving Civil Justice: Judging Civil Justice*, 85 TUL. L. REV. 247, 253, 259 (2010) (reviewing HAZEL GENN, *JUDGING CIVIL JUSTICE* (2010)) (noting that in England, “[f]iling fees alone can exceed £1000, and then each step in the process—such as . . . filing motions . . . also requires the payment of a fee”).

¹⁰¹ Prominent examples include JUDICIAL CONFERENCE OF THE U.S., LONG RANGE PLAN FOR THE FEDERAL COURTS 95–96 (1995) (discussing—and rejecting—a proposal to escalate court fees); RICHARD A. POSNER, *THE FEDERAL COURTS: CHALLENGE AND REFORM* 195–210 (1996) (discussing the overcrowding problem and suggesting an increased user fee with limited exceptions); 5A CHARLES ALAN WRIGHT ET AL., *FEDERAL PRACTICE & PROCEDURE* §§ 1331–32 (3d ed. 2004) (describing how amendments to the federal rules that enabled more meaningful sanctions on frivolous litigation were undone after a few years); Robert G. Bone, Twombly, *Pleading Rules, and the Regulation of Court Access*, 94 IOWA L. REV. 873, 933–34 (2009) (proposing a targeted fee-shifting approach coupled with limited pre-dismissal discovery); Bruce L. Hay et al., *Litigating BP’s Contribution Claims in Publicly Subsidized Courts: Should Contracting Parties Pay Their Own Way?*, 64 VAND. L. REV. 1919, 1921 (2011) (suggesting, with limited exceptions, mandatory user-fees in commercial contract disputes); Rex E. Lee, *The American Courts as Public Goods: Who Should Pay the Costs of Litigation?*, 34 CATH. U. L. REV. 267, 272 (1985) (the then-Solicitor General asserting that, at least in some cases, “the costs of courtroom services should be borne by those who use them”); Brendan S. Maher, *The Civil Judicial Subsidy*, 85 IND. L.J. 1527, 1528 (2010) (suggesting a system where “each litigant would bear responsibility for one half of court usage costs, collectible at the conclusion of the case . . .”); Gideon Parchomovsky & Alex Stein, *The Relational Contingency of Rights*, 98 VA. L. REV. 1313, 1362–66 (2012) (encouraging loser-pays rules).

¹⁰² Or, more generally, judicial overhead expenses. See Hay et al., *supra* note 101, at 1941 (noting that in contribution cases “the judicial overhead . . . is both substantial and reasonably calculable”). Cf. Maher, *supra* note 101, at 1543 (“Cost-minute tracking [can be] a powerful . . . tool that permits measurement of the cost [to the judicial system].”).

entails additional, broader social costs that are conceptually harder to gauge, such as the detrimental effect of the resulting delay on deterrence.¹⁰³ In addition to costs, litigation possibly begets benefits to others. These benefits are also difficult to calculate. Liberal litigation rules presumably enhance the accuracy of the legal system and provide better deterrence.¹⁰⁴ Similarly, broad access to courts is thought to promote their legitimacy and enhance democratic values.¹⁰⁵ Obviously, these direct and indirect negative and positive externalities must be calculated as well. Yet, at least in the current state of affairs, these calculations appear too complex to undertake.¹⁰⁶

The second, and seemingly more important, reason not to price litigation is the notion that the legal system is a public service that should remain available to all. User fees that reflect actual costs—at the very least, the substantial expenses associated with judges, clerks, and legal staff—would presumably be high. Therefore, placing a real price tag on litigation means excluding the poor from litigating their claims. The right to litigate, though, seems to be a fundamental entitlement, perhaps even akin to voting.¹⁰⁷ According to this notion, “[e]very person, regardless of means, is entitled to their day in court.”¹⁰⁸ Effective pricing—which would preclude

¹⁰³ See, e.g., Hay et al., *supra* note 101, at 1941–42.

¹⁰⁴ For a discussion see, e.g., *id.* at 1942–48; Maher, *supra* note 101, at 1536–39. See generally Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

¹⁰⁵ See generally Frank I. Michelman, *The Supreme Court and Litigation Access Fees: The Right to Protect One’s Rights—Part I*, 1973 DUKE L.J. 1153, 1172–77 (1973).

¹⁰⁶ See, e.g., Stephen B. Burbank et al., *Private Enforcement*, 17 LEWIS & CLARK L. REV. 637, 649 (2013) (asserting that “[d]etermining whether public funding of courts is adequate for their needs is an extremely challenging enterprise.”); Judith Resnik, *Constitutional Entitlements to and in Courts: Remedial Rights in an Age of Egalitarianism: The Childress Lecture*, 56 ST. LOUIS U. L.J. 917, 990 (2012) (arguing that, “[j]ust as tracking how much is spent in and around courts is difficult, so too is deciding whether to commodify and how to identify and to measure the outputs of court,” and referring to relevant econometric studies).

¹⁰⁷ “Perhaps no characteristic of an organized and cohesive society is more fundamental than its erection and enforcement of a system of rules” *Boddie v. Connecticut*, 401 U.S. 371, 374 (1971) (prohibiting states from charging access fees to indigents who seek good faith judicial dissolution of their marriages). Frank Michelman is known for drawing the parallels between effective access to courts and voting, stating that “[a]ccess to courts and access to legislatures are claims that merge into one another You cannot . . . call a person a citizen and at the same time sanction the exclusion of that person from that process.” Frank I. Michelman, *The Supreme Court and Litigation Access Fees: The Right to Protect One’s Rights—Part II*, 1974 DUKE L.J. 527, 539–40 (1974).

¹⁰⁸ Resnik, *supra* note 106, at 975 (quoting Jonathan Lippman, Speech at the Midyear Meeting of the National Association of Women Judges at Harvard Law School, Courts in Times of Fiscal Crisis—Who Needs Courts? 10–11 (Mar. 9, 2012) (on file with the author)). See also Maher, *supra* note 101, at 1534 (“[P]ublic adjudication is part and parcel of the healthy operation of pluralistic, constitutional democracies Permitting all citizens to participate . . . in public legal proceedings enhances the dignity of the individual and strengthens the communal bounds of the body politic.”).

the provision of this fundamental right to the poor—directly conflicts with these widely shared perceptions. Accordingly, “there is likely a deep-seated, intuitive conviction among Americans that to charge user fees of any type for court access is ‘unjust.’”¹⁰⁹

These notions reflect serious concerns regarding access to justice for the less well-off. One can argue that market mechanisms, such as contingent fees and non-recourse loans, which enable aggrieved parties to use their claims as collateral to finance lawsuits, eliminate opposition to pricing. However, these mechanisms do not remove the problem. Not all claims have a monetary value; for instance, prisoners attempting to improve their conditions cannot utilize the market to bring meritorious lawsuits. In addition, even with respect to those claims with a monetary value, the market does not currently seem to fully facilitate justified lawsuits of litigants with little means.¹¹⁰

Can the opposition to pricing be resolved through a more nuanced price? For instance, one could impose high user fees and simultaneously subsidize the poor. This solution, however, is limited. First, subsidies do not eliminate abusive litigation by the subsidized, since subsidized litigants do not pay for the services they consume and are free to externalize costs on others. The more one subsidizes litigants, the greater the problem becomes. Second, any subsidy for the poor should be coupled with a concomitant increase in the price that remaining litigants pay—resulting in sizeable user fees for non-subsidized litigants, contrary to the current practice. Third, and relatedly, in a high-fee, large-subsidy regime, those who are not entitled to fee-waivers would only use the legal system when their cases were sufficiently large to justify the high fee. Thus, as many average-size cases would be pushed out of the legal system, such a regime would lose the advantages of diversity.¹¹¹ For similar reasons, a more

¹⁰⁹ Maher, *supra* note 101, at 1545 (quoting Edward Brunet, *Measuring the Costs of Civil Justice*, 83 MICH. L. REV. 916, 930–31 (1985)). See also Bone, *supra* note 101, at 925.

¹¹⁰ See, e.g., Ronen Avraham & Abraham Wickelgren, *Third Party Litigation Funding—A Signaling Model*, 63 DEPAUL L. REV. 233 (2014) (discussing the flaws of the market and observing the benefits of admitting third-party funding agreements).

¹¹¹ See, e.g., Martin D. Beier, Comment, *Economics Awry: Using Access Fees for Caseload Diversion*, 138 U. PA. L. REV. 1175, 1195 (1990) (“[A] fee system that skews the system toward greater homogeneity of cases will delay, if not completely prevent, the creation of allocative rules”); Myriam Gilles, *Class Warfare: The Disappearance of Low-Income Litigants from the Civil Docket*, 65 EMORY L.J. 1531, 1562–67 (2016) (discussing the implications of excluding certain types of litigants—such as minorities and immigrants—from judicial decision-making); Judith Resnik, *The Privatization of Process: Requiem for and Celebration of the Federal Rules of Civil Procedure at 75*, 162 U. PA. L. REV. 1793, 1831 (2014) (discussing the concerns that “the federal courts would become places for poor people and criminal defendants, rather than attract . . . a diverse set of litigants”).

sophisticated pricing scheme, which accounts for the exact financial situation of each individual litigant, would be problematic. A differential price system would again drive the rich—who are now charged beyond the actual costs of adjudication—out of the public system, re-introducing the problem of non-diverse dockets.¹¹²

Are these arguments against pricing compelling? Should policymakers attempt to charge litigants the real price for accessing courts? How can one reconcile the widely shared notion against pricing with the strong perception of workload crisis? These questions exceed the scope of this Article, which takes the conflicting notions as a given. On the one hand, there is over-use, even abuse, of the legal process—an evident outcome in a regime that charges litigants less than the actual costs they inflict. On the other hand, meaningful pricing is not an appropriate option to regulate litigation behavior. While these two themes may stem from different worldviews, their mutual existence emphasizes the importance of finding alternative avenues to regulate litigation. The federal system has turned to routes other than pricing, which I sketch below, but these options seem unsatisfactory, making quotas—the immediate alternative to pricing—stand out as a new and potentially useful mechanism.

B. Regulating Litigation Behavior

This section briefly illustrates the differing reactions of federal courts and judges to the need to regulate litigation behavior without pricing. As a preliminary note, when pricing is irrelevant, an implicit price—a non-monetary sanction of the type we observe in sports—offers a possible alternative.¹¹³ With this in mind, delay may constitute such a “price.” By taking no action to regulate access to courts, policymakers generate delay in vindicating claims, which, like pricing, diminishes the value of legal rights. However, as is the case with other non-monetary sanctions, delay has social costs with no concomitant benefit. To illustrate, evidence tends to decay over time, reducing the accuracy of the legal process and increasing uncertainty.¹¹⁴ Therefore, to the extent policymakers regulate litigation behavior by delaying claims, or by other socially costly implicit prices, this is a suboptimal response.

¹¹² For similar claims, see POSNER, *supra* note 101, at 200. Moreover, as previously discussed, individually tailored pricing schemes of this sort face considerable practical difficulties. *Supra* note 55 and accompanying text.

¹¹³ See *supra* notes 45–46 and accompanying text.

¹¹⁴ POSNER, *supra* note 101, at 209.

Using the theoretical framework in Part I, other alternatives to regulate litigation can be classified as direct restrictions on the relevant right and case-by-case determinations. Direct restrictions, which eliminate the relevant right or curtail its substantive scope, can curb over-use of the legal system. Indeed, there seems to be a recent trend that cuts substantive access-to-justice rights in the federal courts.¹¹⁵ One example is the Supreme Court decisions concerning mandatory arbitration provisions in standard-form contracts.¹¹⁶ This line of cases essentially removes classes of plaintiffs from federal dockets.¹¹⁷ The 2015 amendments to the Federal Rules of Civil Procedure, which narrow the right to discovery, can also be interpreted as direct restrictions on the right to litigate.¹¹⁸

Substantive limitations on the right to access courts are of course prevalent. However, unlike quotas and pricing, substantive limitations on litigation rights do not rely on the information the parties have and do not induce parties to prioritize and undertake only their very best moves. Moreover, while fine-grained substantive restrictions are not necessarily undesirable, direct limitations tend to be crude. Indeed, various stakeholders have criticized the recent trend to curb litigation rights.¹¹⁹

A more nuanced approach attempts to balance advantages and disadvantages of a relevant litigation behavior on a case-by-case basis. Naturally, a judge, and in particular the judge who is already assigned to the case, is well-positioned to “license” beneficial litigation moves and disallow adverse ones. Of course, judges often regulate litigation behavior by various means. However, this approach requires particular determinations, which can be costly and/or time-consuming. Moreover, extensive reliance on judges can lead to controversial results. The recent Supreme Court precedents, which raised pleading standards, illustrate this point. Courts are now directed to dismiss, at the outset, those cases that do

¹¹⁵ See generally Miller, *supra* note 98.

¹¹⁶ *AT&T Mobility, LLC v. Concepcion*, 563 U.S. 333 (2011) (holding that the Federal Arbitration Act preempts state prohibitions on mandatory individual arbitration provisions and accordingly upholding these provisions); *Am. Express Co. v. Italian Colors Rest.*, 570 U.S. 228 (2013) (holding that mandatory arbitration provisions are valid even where the underlying right of action is based on federal, antitrust claims).

¹¹⁷ See, e.g., Resnik, *supra* note 106, at 932. These cases allow prospective defendants to eliminate class litigation through standard-form contracts.

¹¹⁸ For a discussion and criticism of the 2015 “anti-plaintiff” amendments, see Patricia W. Hatamyar Moore, *The Anti-Plaintiff Pending Amendments to the Federal Rules of Civil Procedure and the Pro-Defendant Composition of the Federal Rulemaking Committees*, 83 U. CIN. L. REV. 1083 (2015).

¹¹⁹ See Miller, *supra* note 98 (noting the Supreme Court’s preoccupation with early termination of lawsuits); Moore, *supra* note 118 (criticizing the new discovery rules); Resnik, *supra* note 106, at 995, 997 (criticizing the arbitration provisions decisions).

not initially present “enough facts to state a claim to relief that is plausible on its face.”¹²⁰ This doctrine motivates judges to screen out, on a case-by-case basis, unmeritorious claims, and it prevents plaintiffs with weak cases from proceeding to costly discovery and unnecessarily consuming precious judicial resources. While it seems straightforward to encourage judges to screen undesired cases, this move has evident difficulties. Indeed, this doctrinal shift has generated vigorous discussions and fierce criticism.¹²¹ For the current purposes, it suffices to briefly highlight two related lines of opposition—dismissing claims before the merits are known and granting wide discretion to trial court judges.

The more demanding pleading standards screen out cases without probing into their merits. Pre-merits screening can lead to unfortunate results, particularly when the plaintiff does not know the merits of her case and the relevant information resides with the defendant. Medical malpractice and civil rights cases serve as typical examples. In such cases, uninformed plaintiffs with good claims cannot present sufficient information to proceed to discovery.¹²² While pre-merits disposition is not necessarily undesirable, in these contexts it is coupled with relatively unfettered judicial discretion. Heightened pleading standards require judges to decide merits questions early on, with little evidentiary background, and invite them to dismiss cases “on instinct,” according to their subjective beliefs.¹²³ Relatedly, dismissing cases at the outset leaves no substantive record, allowing trial judges free range without meaningful supervision by higher courts.¹²⁴

This Article does not intend to convince the reader that existing approaches to regulating litigation behavior are necessarily wrong. Rather, its goal is to point to the existing tradeoff, the choice between imperfect alternatives. The less we trust judges’ individual, unfettered discretion—

¹²⁰ *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007). See also *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009).

¹²¹ For a description of the criticism see, e.g., William H.J. Hubbard, *A Fresh Look at Plausibility Pleading*, 83 U. CHI. L. REV. 693, 694–96 (2016).

¹²² See, e.g., Bone, *supra* note 101, at 925–26.

¹²³ Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 COLUM. L. REV. 433, 482 (1986). See also Bone, *supra* note 101, at 889 (“[C]ritics fear that [the new rule] gives too much latitude to district judges, who are eager to screen cases This fear is not unfounded”); Jeffrey J. Rachlinski, *Processing Pleadings and the Psychology of Prejudgment*, 60 DEPAUL L. REV. 413, 429 (2011) (arguing that the new regime may “feed[] the overconfidence and [cognitive] vulnerabilities that judges have when making intuitive misjudgments”).

¹²⁴ See, e.g., Marcus, *supra* note 123, at 444–47; Thornburg, *supra* note 100, at 267; Karen Petroski, *Iqbal and Interpretation*, 39 FLA. ST. U. L. REV. 417, 427 (2012) (surveying scholarly views). Relying on wide discretion of trial court judges also begets uncertainty. Bone, *supra* note 101, at 928.

particularly when judges have limited information on which to base their discretion—the more we should seek alternatives to case-by-case determinations. Likewise, the less we believe in pricing litigation, the more we need to restrict access to litigation through other means.¹²⁵

Quotas broaden the range of regulatory alternatives. They do not require direct intervention in the scope of the relevant litigation right, nor do they rely on judges to screen undesired litigation activities. Thus, they present a fresh approach to alleviate the pressure on the judiciary and decrease the number of unmeritorious issues that reach courts by allowing litigants to choose the most important instances for judicial treatment.¹²⁶ In particular, quotas respond to the problems associated with pricing. Non-tradable quotas do not deny the poor access to justice. Rather, they enable (limited) access to justice for all right-holders. Likewise, when prices are difficult to set, numerical caps reflect a judgment-call regarding the amount of litigation we are willing to allow.

It is true that non-tradable quotas are imperfect. But as previously discussed, quotas can be modified to minimize their weaknesses. Moreover, allocating rights through quotas can extract the social benefits that litigation presumably entails and, in particular, the benefits that are plausibly gained from a judicial system with diverse sets of right-holders. Finally, in the absence of caps, and given the reluctance to price, the range of regulatory responses to the over-use problem is smaller, and drastic restrictions seem almost inevitable. With these general principles in mind, the following section discusses concrete suggestions for numerical ceilings on litigation rights.

C. Implementation

Parts IV.A. and B. demonstrate that quotas can be a valuable tool to regulate litigation behavior. While there are sporadic quotas in different

¹²⁵ Indeed, those who oppose restrictions on access to justice sometimes explicitly invoke the idea of pricing as an alternative regulatory choice. In *Twombly*, the minority asserted (among other things) that instead of applying more demanding pleading standards “the district court has at its call . . . a wide array of Rule 11 [monetary] sanctions” to curb abusive litigation. *Twombly*, 550 U.S. at 593 n.13 (2007) (Stevens, J., dissenting).

¹²⁶ As litigants base their litigation decisions on their private benefit, they may invoke their rights where doing so is not socially valuable. However, similar problems exist in any regime that entrusts litigants with the power to trigger and manage litigation. Moreover, it seems plausible to think that there is some correlation between the private and social motivations to litigate. Cf. Steven Shavell, *On the Design of the Appeals Process: The Optimal Use of Discretionary Review Versus Direct Appeal*, 39 J. LEGAL STUD. 63, 77–79 (2010) (discussing similar issues in the context of a proposed regime that relies on litigants’ information regarding the strength of their appeals).

procedural contexts,¹²⁷ this Article attempts to systematically extend their use to regulate litigation. Specifically, this section discusses two potential access-to-justice uses for quotas: quotas on adjudication behavior and quotas on filing behavior.

1. Adjudication behavior

Quotas can be used to improve litigants' choices when litigation is underway—incentivizing them to prioritize and undertake only their very best moves. To demonstrate, one domain in which caps can balance the conflicting considerations is interlocutory appeals. The federal system is notorious for its strict adherence to the “final judgment rule,” as appeals are generally only allowed following the final decisions of district courts.¹²⁸ This policy has obvious drawbacks. Particularly, it prevents appellate courts from effectively reviewing and guiding lower courts, especially with regard to decisions that are not likely to be reviewed within final appeals (e.g., discovery orders). However, a liberal right to interlocutory appeals entails other difficulties—it invites tactical delays through frequent petitions for review and unnecessarily wastes the appellate court's resources.¹²⁹ Every legal system strikes a balance between these competing considerations. While the federal system strictly constrains interim appeals, other jurisdictions, such as New York, take a liberal stance toward interlocutory review.¹³⁰ The common goal of all these approaches “is to . . . permit desirable appeals to be taken, without encouraging large numbers of ill-founded appeals.”¹³¹

Quotas offer a new and perhaps better balance. Presumably, the majority of interim decisions do not justify interlocutory review; however, some interim orders do require immediate review. In principle, then, each litigant could have a right to, for instance, a single interlocutory appeal in

¹²⁷ As mentioned before, under the current rules, litigants can depose ten witnesses without leave of the court and attorneys can disqualify a certain number of jurors without stating a reason. *See supra* notes 49 and 81 and accompanying text. Parties can amend their pleading once as a matter of right within 21 days after serving it. FED. R. CIV. P. 15(a)(1). American inmates can bring three frivolous suits in their lifetime without incurring filing fees. 28 U.S.C. § 1915(g) (West 2012). Local rules in some parts of the United States limit the number of claim terms that parties can dispute in patent litigation. J.J. Prescott & Kathryn E. Spier, *A Comprehensive Theory of Civil Settlement*, 91 N.Y.U. L. REV. 59, 107 (2016).

¹²⁸ *See, e.g.*, JACK H. FRIEDENTHAL ET AL., CIVIL PROCEDURE 622 (4th ed. 2005). However, the final judgment rule has many exceptions that permit immediate appeals. 16 CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 3920 (3d ed. 2012).

¹²⁹ For a summary of these and other conflicting considerations *see, e.g.*, Shay Lavie, *Are Judges Tied to the Past? Evidence from Jurisdiction Cases*, 43 HOFSTRA L. REV. 337, 357–59 (2014).

¹³⁰ *See* N.Y. C.P.L.R. § 5701(a)(2) (McKinney 2017).

¹³¹ WRIGHT ET AL., *supra* note 128, at § 3920.

the life of a case.¹³² This proposal does not overly burden appellate courts, and it simultaneously guarantees that litigants will carefully ration their interlocutory review rights and use them only in their very best instances, enhancing the goals of effective review and law development. By setting a numerical cap, policymakers can also expect and manage the excess burden that results from interlocutory orders. While non-tradable caps present some difficulties, such as heterogeneity among litigants, some modifications to the proposed quota can alleviate the concerns.¹³³

Before turning to possible variations, it is important to note the range of relevant policy alternatives: on one extreme, interlocutory appeals are banned, and on the other, they are freely allowed. Both of these options have obvious flaws. The substantive scope of the right to interlocutory appeals can be curtailed, e.g., by allowing immediate review of certain types of cases or interim orders.¹³⁴ As the foregoing discusses, such substantive limitations present a viable regulatory option, but they also require thoughtful consideration and tend to be crude. Moreover, as these examples demonstrate, direct limitations do not elicit information from litigants on their very best appeals, eliminating the benefits of quotas and prices.

Another mechanism in this context is individual judicial decision-making, e.g., petitioning the appellate court for the right to appeal.¹³⁵ Such a regime also has apparent inefficiencies. If it aims to provide an effective opportunity to challenge the trial court's orders, it requires the appellate court to make a preliminary decision—to take the case or not—in each petition for interlocutory review. Clearly, such determinations are time-

¹³² For a recent suggestion along these lines see Kenneth K. Kilbert, *Instant Replay and Interlocutory Appeals*, 69 BAYLOR L. REV. 267 (2017) (proposing that “plaintiff and defendant each [would] ha[ve] the right to appeal one interlocutory order in the case immediately to the court of appeals, without the need for any permission by a judge” and arguing that such a proposal “strikes a better balance between the conflicting goals of appellate review, error correction and efficiency.” *Id.*, at 269).

¹³³ See generally *supra* Part III.A.

¹³⁴ 28 U.S.C. § 1292(a) (West 2012), for example, creates several exceptions to the final judgment rule such as receivership, admiralty cases, and interim injunctions. Along the same lines, appellate courts are generally guided to closely inspect the trial court's legal determinations, but not its factual findings. See, e.g., FED. R. CIV. P. 52(a)(6) (stating that the reviewing court “must not . . . set aside [findings of fact] unless clearly erroneous . . .”).

¹³⁵ The Supreme Court selects cases for review through a similar, discretionary process. SUP. CT. R. 10. To a limited extent this is also the current regime in the federal courts of appeals. For example, interlocutory appeals are permitted with the concurrent permission of both the district and the appellate court. 28 U.S.C. § 1292(b) (West 2012). Such parallel permissions, however, are uncommon. WRIGHT ET AL., *supra* note 128, at § 3929.

consuming, and allow judges wide discretion.¹³⁶ It is easy to see how quotas can fare better. They balance the conflicting considerations, and instead of relying on individual, open-ended, and complex judicial determinations, they exploit the information litigants already possess to entertain only the most important interlocutory appeals. Pricing is another tool to regulate interlocutory appeals—either as a sufficiently high fee or a substantial monetary sanction on the losing appellant/appellee. The difficulties with pricing were discussed above.

Interlocutory appeals quotas can be combined with other mechanisms to mitigate the one-size-fits-all and information concerns. One variation is to provide discretionary interlocutory appeals to litigants who have already used their quota. While this modification reduces some of the benefits of caps, it is more forgiving of those who failed to plan ahead.¹³⁷ Another variation is to combine the quota with a pricing regime, such that those who have exhausted their cap would have to pay a hefty fee for filing an interlocutory appeal.¹³⁸

Many other modifications are possible. The quota should be determined based on the amount of interlocutory appeals considered tolerable. This number can be adjusted over time based on continuous feedback from relevant stakeholders. Note, in this context, that the quota need not reflect integers. It can express any desired number, including fractions—policymakers can randomly assign, for example, one interlocutory appeal to ten litigants. Setting the interlocutory appeals quota more precisely through this process ensures a steady, albeit thin, stream of quality interlocutory appeals. In a similar vein, the proposed interlocutory appeals ceiling can be tradable, with restrictions. There can be other tweaks: the initial allocation can be more fine-grained—for example, policymakers can allocate additional interlocutory appeals in those areas they deem worthy of close appellate review. To the extent policymakers believe that it is important to subsidize the poor, additional appeals can be made available to those who lack financial means.

This demonstrates how interlocutory appeals quotas can be used to regulate the behavior of rival parties during litigation. The use of quotas

¹³⁶ For the problems that the wide discretion to allow interlocutory appeals creates *see* WRIGHT ET AL., *supra* note 128, at § 3929. For similar reasons, other legal systems have shifted from unrestricted appellate discretion regarding interlocutory appeals to the final judgment rule. Eisenberg et al., *supra* note 100, at 1466–67, n.78 (describing such legal changes in Israel).

¹³⁷ Quotas can also be supplemented with individual decision-making by allowing appellants who won their interlocutory appeals to still be able to use their initial allocation.

¹³⁸ In any case, as litigants can generally predict their needs in interlocutory appeals, the information problem seems manageable.

can be extended to similar domains. One example is a numerical ceiling on amendments to pleadings. Presently, the Federal Rules of Civil Procedure endorse a liberal policy—parties can automatically amend their pleadings once within twenty-one days of serving it, and additional amendments are “freely” given.¹³⁹ This permissive approach has merits, as it assigns greater weight to accurate decision-making. However, it imposes unnecessary costs on rival litigants.¹⁴⁰ Attempts to restrict the right to amend to specific categories or instances seem futile, and allowing the trial judge broad discretion might be a problematic move. The way out may be a qualified right in the form of a limited entitlement, such as the ability, once or twice per case, to automatically amend a pleading beyond the twenty-one-day window. Along the same lines, numerical ceilings can be implemented in other litigation contexts, e.g., curbing the capacity of litigants to postpone hearings through a quota.

Since procedural quotas are useful in controlling several forms of abusive adjudication behavior, a more general proposal may be possible: allowing litigation “coupons” for each filed lawsuit, for example, which could be spent at each stage of the proceedings—interlocutory appeals, amending pleadings, discovery requests, postponing hearings, etc. Such coupons would allow litigants broad autonomy to manage their cases, without actually charging money, and would rely on litigants’ information as the parties “pay” with their coupons and hence prioritize their moves. For similar reasons, such a system would minimize information concerns—it provides litigants considerable “resources” to handle, and it brings quotas closer to prices. Of course, this suggestion is not bulletproof either. Policymakers would have to set the “price” for different adjudication moves. More importantly, heterogeneity concerns persist under this approach, as some litigants deem their case worthy of more litigation coupons than others do. These concerns could be alleviated through the same mechanisms that were suggested in Part III.A. above.

2. *Filing behavior*

This section discusses more radical proposals to restrain the filing of meritless suits. To demonstrate, procedural quotas can be used in the context of pleading standards. As discussed above, the Supreme Court now

¹³⁹ FED. R. CIV. P. 15(a)(1)–(2). This rule has also been broadly interpreted. *See, e.g., Foman v. Davis*, 371 U.S. 178, 182 (1962).

¹⁴⁰ Moreover, the permissive approach appears to conflict with the recent policy that requires more demanding pleading standards. *See, e.g., Marcus, supra* note 123, at 440 (“The liberality of the pleading requirements is reflected [by the flexible approach to] . . . amendment of pleadings. . .”).

requires plaintiffs to meet a heightened pleading threshold in order to survive early dismissals.¹⁴¹ This doctrinal move seemingly harms misinformed plaintiffs.¹⁴² In these situations—typical examples are civil rights and medical malpractice cases—the defendant, but not the plaintiff, has access to the evidence and knows whether a good cause for action exists. As a result, the heightened standards may screen out, before discovery, those plaintiffs who have good cases but lack evidence. Heightened pleading standards, then, may be too drastic a tool, as they eliminate from courts good claims in important areas. Moreover, as discussed above, this tool is associated with additional difficulties, essentially granting judges increased discretion to screen on a case-by-case basis at the outset without proper evidence.¹⁴³ However, the alternative, permissive standards¹⁴⁴ allegedly trigger frivolous suits and pressure defendants with good defenses to settle, hence the shift in the Supreme Court’s jurisprudence.¹⁴⁵ Pricing also appears problematic for the usual reasons. Policymakers could perhaps do better by imposing direct restrictions, e.g., defining areas, such as medical malpractice, in which heightened pleading standards are not required and areas in which they are. While this is a sound proposal, the problem of misinformed plaintiffs presumably exists in other areas, albeit to a lesser extent.

An alternative to heightened pleading standards is using caps. Allocating potential plaintiffs, essentially any citizen, a limited right to bring a case without the need to provide more information up front ensures that at least some of these important cases will reach courts. As before, this proposal can be modified. The relevant figure—i.e., the number of times a victim can bring a case under the lax standards—should be determined. There are some parameters to consider to assist with this task: are there many asymmetric-information instances in which the heightened pleading standards present a major difficulty? Can the injured party reasonably present evidence? Depending on the answers to these questions and others, several opportunities in a lifetime to bring a case under the permissive standards may suffice. And, of course, adjustments can be made over time. Other variations relate to the integration of case-by-case judicial decision-

¹⁴¹ See *supra* notes 120–124 and accompanying text.

¹⁴² See *supra* note 122 and accompanying text.

¹⁴³ See *supra* notes 123–124 and accompanying text.

¹⁴⁴ See *Conley v. Gibson*, 355 U.S. 41, 45–46 (1957), *abrogated by Bell Atl. Corp. v. Twombly*, 550 U.S. 544 (2007) (representing the previous, permissive pleading policy).

¹⁴⁵ See *Twombly*, 550 U.S. at 559 (2007) (“[T]he threat of discovery expense will push cost-conscious defendants to settle even anemic cases . . .”).

making. Similar to the interlocutory appeals example, plaintiffs could be allowed to deviate from their quota—subject to the court’s discretion. Likewise, the quota could be based on unsuccessful invocations of the right, i.e., cases in which plaintiffs eventually lose. In addition, the poor can be subsidized through larger quotas, and different quotas can be assigned for different types of claims.¹⁴⁶

Admittedly, this suggestion substantially departs from existing practices. Yet it does attempt to directly tackle the core problems of access to justice. The gist of the pleading standards problem is its inability to distinguish between “purely” frivolous claims and meritorious claims that lack sufficient evidence at the filing stage. Caps exploit the “hidden” information that at least some plaintiffs possess. In many instances, litigants know whether their claim is likely frivolous or not; quotas elicit this information, as they urge litigants to prioritize and use their quota only when they believe they have good cause but lack sufficient evidence. Alternative avenues exist to conduct this screening, but they are costly and may be problematic.

Along these lines, quantitative ceilings on filing behavior can be useful in other contexts. The problem of forum-shopping, for instance, stems from legal authorization to file in several forums.¹⁴⁷ Plaintiffs, hence, may file in the forum that they believe has more favorable judges and juries.¹⁴⁸ Plaintiffs may also have legitimate reasons to file outside of their natural forum, such as a smaller caseload and shorter queues in the other forum. However, verifying the plaintiff’s true intentions in each and every case is a highly complicated task.¹⁴⁹ One option is to restrict apparent attempts of

¹⁴⁶ As this proposal essentially creates individual litigation rights, its extension to the collective litigation arena is possible, though by no means straightforward. A complete discussion of this point is beyond the scope of this Article.

¹⁴⁷ See, e.g., Ori Aronson, *Forum by Coin Flip: A Random Allocation Model for Jurisdictional Overlap*, 45 SETON HALL L. REV. 63, 73 (2015) (discussing “several familiar examples” such as “when personal jurisdiction laws permit a case to be litigated in more than one state; when venue laws allow for a case to be litigated in more than one federal district in a given state; and when subject-matter jurisdiction laws allow for a case to be litigated in either state or federal court . . .”).

¹⁴⁸ See, e.g., Julie Creswell, *So Small a Town, So Many Patent Suits*, N.Y. TIMES (Sept. 24, 2006), <http://www.nytimes.com/2006/09/24/business/24ward.html> (describing how a small town in Texas became attractive to patent suits, apparently because its courts lean toward plaintiffs in these issues). Cf., *TC Heartland, LLC v. Kraft Foods Grp. Brands, LLC*, 137 S. Ct. 1514, 1517 (2017) (holding that “a domestic corporation ‘resides’ only in its State of incorporation for purposes of the patent venue statute,” and effectively limiting forum-shopping in patent suits).

¹⁴⁹ In a very limited sense, this is the role of the forum non conveniens doctrine. See RESTATEMENT (SECOND) OF CONFLICT OF LAWS § 84 (AM. LAW INST. 1971) “A state will not exercise jurisdiction if it is a seriously inconvenient forum for the trial [and] a more appropriate forum is available.” The doctrine, however, “depends largely upon the facts of the particular case and is in the sound discretion of the trial judge.” *Id.*, cmt. b. Perhaps due to the complications that particular judicial decision-making

forum-shopping through quotas on the right to file outside of the plaintiff's natural forum. Other implementations of numerical caps in the context of filing behavior include restricting repeat defendants, such as insurance companies, from raising frivolous defenses and limiting recurrent litigation on different claims between the same parties. Following up on the litigation vouchers suggestion, filing caps could be bundled together, perhaps along with caps on adjudication rights, to move the non-monetary quota regime closer to pricing and mitigate information concerns in managing the quota.

A numerical cap on the ability of individuals to file lawsuits seems like a radical move, severely conflicting with access-to-justice notions. However, quotas might balance the conflicting considerations better than any other alternative. Similar numerical restrictions on filing are not unknown. American inmates can bring three frivolous suits in their lifetime without incurring filing fees.¹⁵⁰ While the merits of this cap could be questioned,¹⁵¹ this limitation was designed to restrict frivolous prisoner lawsuits. Similar quotas could be implemented in other areas.

CONCLUSION

Quotas can be beneficial in regulating litigation. Courts seem to be over-used, but pricing is currently not an available option. Procedural caps present an interim option, shedding the risks of abusive litigation without scuttling important values, such as access to justice for different and diverse classes. Procedural caps may constitute a second-best option, but they are a substitute for more drastic, substantive restrictions on litigation. Quotas offer an additional mechanism to balance the underlying, conflicting considerations. And as courts continue to suffer from drained

implicates, the Restatement does not attempt to effectively restrict forum-shopping. The plaintiff's "choice of a forum should not be disturbed except for weighty reasons." *Id.*, cmt. c.

¹⁵⁰ 28 U.S.C.A. § 1915(g) (West 2012).

¹⁵¹ The quota is part of a wider reform that took place in 1996. The reform succeeded in significantly reducing the volume of inmate litigation, Alexander Volokh, *The Modest Effect of Minneci v. Pollard on Inmate Litigants*, 46 AKRON L. REV. 287, 312–13 (2013), but it also received criticism. See, e.g., Margo Schlanger, *Inmate Litigation*, 116 HARV. L. REV. 1555 (2003). In particular, this quota provision seems to punish inmates who have more suits, regardless of the "average" merits of those claims. Cf., *id.* at 1648–49 ("It may well be that the most frequent filers file not only a very large number of cases, but an especially high proportion of meritless cases . . . [but other] frequent filers are actually skilled litigators whose filings are particularly likely to have merit."). Plausibly, then, this quota harms inmates receiving the worst treatment, as they are also more likely to turn to courts more often, and presumably file more frivolous suits as well. A more careful design of the quota is possible—e.g., setting a ratio-quota that allows inmates a certain portion, say, 25%, of frivolous claims in their "portfolio" of suits.

resources, it is all the more important and timely to experiment with new approaches.¹⁵²

Beyond the context of litigation, this Article has attempted to demonstrate that quotas—in general—are a valuable regulatory tool, one which appears to be both under-theorized and under-used. Caps elicit information from their beneficiaries, induce them to prioritize, and discourage over-use. Quotas achieve these goals without charging fees, employing costly case-by-case determinations, or using direct, substantive restrictions on the relevant right. This Article advances a broader use of quotas, hoping to enrich the array of possible regulatory alternatives. I conclude in the following paragraphs that the discussion throughout the Article highlights other, more general domains that can benefit from a structured use of quotas. Providing a comprehensive list of such domains exceeds the scope of the current discussion, but I briefly present two more areas—the provision of public services and the regulation of government bodies—where quotas could effectively be used as regulatory alternatives.

A. Public Services

The context of public services embodies a particularly strong case for using numerical ceilings, as such (non-tradable) quotas obviate the need to charge money. Essential services should presumably be available to the poor as well as the rich, and allocation based on an ability to pay violates this notion. In a sense, the case for quotas in litigation is one manifestation of the more general argument for the use of quotas to regulate essential public services; just as access to litigation is perceived of as a fundamental right that should be available to all individuals regardless of wealth, essential public services are conceived of as fundamental rights that should be available to all individuals. Both litigation and essential public services could thus benefit from implementation of numerical ceilings.

With respect to many essential entitlements, there seems to be independent value in the exercise of the right by all individuals. To illustrate, consider the use of medical services. The sick can infect others; hence, it makes sense to provide all individuals with at least a basic level of non-transferable health services. As noted above, inalienable quotas that regulate the number of children in families could similarly achieve a broader purpose—diversity in the general population—which is

¹⁵² See, e.g., Resnik, *supra* note 106, at 969–70, 973–77 (surveying budget cuts in state courts and the measures taken in response).

unattainable under a pricing regime.¹⁵³ Charging a differential price to address these situations is complicated.¹⁵⁴ Policymakers could freely distribute the relevant right—but free allocation invites over-use. Other alternatives, such as directly restricting the substance of the relevant right and issuing case-by-case licenses, entail their own difficulties.¹⁵⁵

Pre-defined numerical ceilings offer an effective way to limit the use of essential entitlements without employing money. Several of the previous real-world examples of quotas fit this context—numerical limitations on bankruptcy rights and having offspring can be viewed as essential rights that are capped by a quota.¹⁵⁶ This logic can be extended to other essential public services. Simply put, where over-use is a problem and charging fees is not an option, quotas are almost inevitable.

Take, for instance, emergency telephone calls. We presumably want every citizen to have the ability to call 9-1-1. Accordingly, having insurance is not a precursor to being transported to the hospital by an ambulance, even though patients often avoid payment after the fact.¹⁵⁷ For similar reasons, wide discretion for dispatchers—whether to treat the call as an emergency or not—seems problematic.¹⁵⁸ Similar to screening lawsuits at the outset, any procedure to screen 9-1-1 calls by dispatchers would lack sufficient information and direct evidence—which would only be gathered after arriving at the scene. On the other hand, an unlimited 9-1-1 “right” invites abuse—i.e., calling in non-emergency situations. Indeed, 9-1-1 telephone calls seem to be over-used.¹⁵⁹ Quotas can achieve both

¹⁵³ See *supra* note 57 and accompanying text (discussing the advantages of quotas over pricing in the context of the right to have children).

¹⁵⁴ See *supra* note 55.

¹⁵⁵ See *supra* notes 26–29 and accompanying text.

¹⁵⁶ For other real-world examples of quotas on public services, such as issuing passports and changing names, see *supra* note 62. Similarly, in some American states, voters can change their early vote, but this right is limited through a quota. Daniel Victor, *On Election Day, Little Chance of Changing That Early Vote*, N.Y. TIMES (Nov. 8, 2016), <https://www.nytimes.com/2016/11/09/us/politics/change-early-vote.html> (discussing such a quota in Wisconsin).

¹⁵⁷ See, e.g., Parija B. Kavilanz, *911 Abuse: Calling with the Sniffles*, CNN MONEY (Aug. 24, 2009), http://money.cnn.com/2009/08/24/news/economy/healthcare_911_abuse/index.htm.

¹⁵⁸ See Karen Augé, *911 Non-Emergencies a Growing Problem Nationwide*, DENVER POST (Dec. 28, 2009), http://www.denverpost.com/ci_14084125 (“[E]mergency systems have a duty to respond ‘If you’re a system that responds to 911 calls, you must respond to every call.’”) (quoting a former president of the National Association of Emergency Medical Technicians).

¹⁵⁹ See, e.g., Augé, *supra* note 158 (stating that non-emergency 911 ambulance calls allegedly create an “enormous cost to health systems, taxpayers, and everybody with health insurance”); Gary Emerling, *Medics to Treat Overuse of 911*, WASH. TIMES (Mar. 27, 2008), <http://www.washingtontimes.com/news/2008/mar/27/medics-to-treat-overuse-of-911/?page=all> (“The D.C. fire department . . . estimates that 49,000 of the calls it receives each year [out of 127,000 annual calls] are for non-emergency situations.”).

ends—providing an essential service for free and restricting its use. While the use of a simple numerical cap on 9-1-1 calls appears extreme, several American communities have opted for a quota-style solution, combined with case-by-case determinations. Under these programs, “frequent users” of the right, those who have exceeded a certain number of calls, are identified and individually addressed.¹⁶⁰

The 9-1-1 example illustrates the idea of limiting over-use of public services through pre-defined quantity allocations. This example also demonstrates the drawbacks of such an idea, which relate to the one-size-fits-all and information difficulties. A quota on public services would presumably be inalienable, to avoid the pitfalls that the use of money creates. A regime of inalienable quotas on public services means that those who need the service but have exhausted their quota would not be able to access it.¹⁶¹ This is often a harsh result in the context of public services. The importance of the relevant right—emergency treatment, in the 9-1-1 example—may trump the desire to restrict over-use.

To implement a quota on essential public services, then, one needs to mitigate these concerns. The cap can be sufficiently generous to accommodate the particular needs of different groups in the population. Likewise, discretion can be integrated into such a scheme—deviations from the quota can be allowed in exceptional cases. The quota can be combined with a pricing scheme—such that those who have exhausted their allocation would be able to purchase the right at its appropriate price. The quota can also refer to different grades of public services, such that those who consumed their initial allocation would still receive the service, albeit of a lower quality.¹⁶² Finally, to remedy the right-holders’ information problems, governments can allocate a broader set of essential services through public service “credit points,” akin to food stamps. In the spirit of the litigation vouchers proposal, this public service credit can be valid for various essential government services—e.g., health services, litigation behavior, bankruptcy rights, 9-1-1 calls, etc.—to be used by its beneficiaries as they see fit. Such a system allows claimants the autonomy

¹⁶⁰ These excessive users typically suffer from minor, chronic—but non-emergency—health problems. The idea, in a nutshell, is to funnel them to a different, non-emergency channel, without burdening the emergency system. For attempts to provide a comprehensive solution to these frequent users in Denver and D.C., see Augé, *supra* note 158; Emerling, *supra* note 159.

¹⁶¹ Another problem is the over-use of services up to the quota. This problem seems less pressing in the implementation of quotas.

¹⁶² In a sense, the 9-1-1 “frequent-users” policy employs a similar tool—funneling heavy users to the non-emergency track. See *supra* note 160.

to manage a larger quota for various purposes over a long time, bringing caps closer to prices. These are, of course, preliminary directions, intended to provoke more systematic thought on the regulation of essential public services.

B. Government Bodies

Another domain that can benefit from (non-tradable) quotas is the regulation of government bodies. The idea is straightforward—charging a price from government agencies, at least in certain contexts, is not a viable option. As the foregoing suggests, commodification concerns (compelling agencies to purchase a certain right transforms the meaning of that right), or willingness-to-pay considerations (profitable and non-profitable government bodies should be able to have similar rights) may be the reason for the reluctance to price. Moreover, pricing may well be ineffective in regulating the behavior of government officials.¹⁶³ Be that as it may, when money is not an option, and alternative approaches, such as relying on the discretion of agencies, are unsatisfactory, quotas should come to mind.

I demonstrate this point through the example that started this Article—veto rights. Consider the veto right of the permanent members of the United Nations Security Council. Fifteen countries sit on the Security Council; five are permanent members—the United States, the United Kingdom, France, Russia, and China—who have the right to veto Council resolutions.¹⁶⁴ Presumably, providing veto power to five countries does serve some purposes.¹⁶⁵ However, this veto power seems to be too broad, allegedly leading to a continuous gridlock.¹⁶⁶ Indeed, several proposals to regulate the substance of these veto rights have been raised.¹⁶⁷

Caps offer an alternative approach to prevent over-use of veto rights—say, one per permanent member per year, or a certain fraction of the

¹⁶³ See Daryl J. Levinson, *Making Government Pay: Markets, Politics, and the Allocation of Constitutional Costs*, 67 U. CHI. L. REV. 345, 345 (2000) (“Government actors respond to political incentives . . . [and they] cannot be expected to respond to forced financial outflows like a private firm. If the goal of making government pay compensation is to achieve optimal deterrence with respect to constitutionally problematic conduct, the results are likely to be disappointing and perhaps even perverse.”).

¹⁶⁴ Posner & Sykes, *supra* note 9, at 204.

¹⁶⁵ See, e.g., *id.* (“At the time . . . it was believed that the five permanent members would be the world’s policemen These countries were too powerful to be compelled to use force by others.”).

¹⁶⁶ *Id.* at 204–05.

¹⁶⁷ See, e.g., Fredric L. Kirgis, Note, 93 AM. J. INT’L L. 975, 976 (1999) (reviewing BARDO FASSBENDER, U.N. SECURITY COUNCIL REFORM AND THE RIGHT OF VETO: A CONSTITUTIONAL PERSPECTIVE (1998)) (discussing a proposal to restrict the ability to invoke veto rights to certain matters). See also Posner & Sykes, *supra* note 9, at 206–07 (discussing proposed reforms).

Council's resolutions per year. More generally, veto power invites over-use by the holder of the right to veto; the proposal to limit veto power through quotas can fit other areas of law in which policymakers choose to create veto rights, e.g., presidential vetoes.¹⁶⁸

Of course, veto-quotas have familiar drawbacks. As they would presumably be inalienable, veto-holders may be limited in their ability to utilize their veto allocation over time; similarly, quotas are rigid and may be too inflexible in relevant circumstances (e.g., there may be a need for a larger quota in a certain year). These are complex and contentious issues, which merit a separate, comprehensive analysis. It suffices for the purposes of this Article to note that the unique context of veto rights mitigates at least some of these problems. These domains are highly politicized, and veto-holders typically carry substantial weight regardless of their formal veto power.¹⁶⁹ Hence, restricting veto-holders' capacity to invoke their veto does not seem to be a major problem. Furthermore, a quota on vetoes would force the veto-holder to prioritize and reveal her preferences, providing the public sphere with more information. Finally, the case-by-case approach seems irrelevant.¹⁷⁰ Veto rights are designed to provide veto-holders with ultimate decision-making power, and subjecting the ability to use the veto to a third party is contradictory to this purpose.

This brief example, then, illustrates how, in the absence of pricing, quotas provide a viable substitute to curb government bodies. Other examples, in more mundane contexts, also come to mind. For instance, quotas can be used to restrict, quantitatively, the number (or proportion) of plea bargains prosecutors can strike. Together with the foregoing suggestions to limit, through quotas, litigation and the provision of public services, this Article attempts to broaden the existing alternatives policymakers have at their disposal to regulate behavior, hoping to provoke further thought regarding the appropriate tools to do so.

¹⁶⁸ It should be noted that presidential vetoes are already restricted, at least to some extent, as a two-thirds vote in each chamber of Congress can override the veto. U.S. CONST. art. I, § 7, cl. 2. Hence, relative to other vetoes, a quota on presidential vetoes seems less urgent. On the other hand, presidential vetoes are rarely overridden, and the threat of a veto is "often sufficient to change the shape of a bill." Peter Raven-Hansen & William C. Banks, *From Vietnam to Desert Shield: The Commander in Chief's Spending Power*, 81 IOWA L. REV. 79, 117 (1995).

¹⁶⁹ See, e.g., Posner & Sykes, *supra* note 9, at 205 (discussing the influence of the permanent members of the Security Council).

¹⁷⁰ Practically, a case-by-case option could be, for instance, subjecting the veto to the approval of a court.

UNDERSTANDING THE RISE OF SUPER PREEMPTION IN STATE LEGISLATURES

Bradley Pough ♦

In 1957, the City of Tallahassee, Florida, enacted a local gun control ordinance. Now known as § 12-61(a) of the Tallahassee Code (hereinafter the “1957 ordinance”), the ordinance stipulates, “No person shall discharge any firearms except in areas five acres or larger zoned for agricultural uses.”¹ Although the law was never amended, it was restated in the Code’s 2003 codification and has remained in its current form ever since.² In 1984, Tallahassee passed another gun control law. Today referred to as § 13-34(b)(5) of the Tallahassee Code (hereinafter the “1984 ordinance”), it prohibits any person from discharging a firearm in a park or recreational facility owned by the City.³ The law was amended in 1988 and restated in the Code’s 2003 codification.⁴

In 1987, the Florida legislature passed § 790.33 of the Florida Statutes, establishing that the State “occup[ies] the whole field of regulation of firearms and ammunition.”⁵ Known as a “preemption” statute, this type of state legislation renders “null and void” all past and future local enactments that conflict with its terms.⁶ After the law’s passage, Tallahassee’s gun control ordinances amounted to little more than words on a page. In the past 10 years, there is no record of local police attempting to enforce either ordinance, and, in 2011, the Tallahassee Police chief officially advised all personnel that, due to the state legislation, the 1957 ordinance and the 1984 ordinance were unenforceable.⁷

If this is where the story ended, it would likely not warrant scholarly attention. Although state preemption of local ordinances causes much

♦ Law clerk for the United States Court of Appeals for the First Circuit. B.A., Yale University; J.D., Harvard University. Please note that this article was accepted for publication prior to the commencement of my clerkship. The opinions expressed in this article are my own. I would like to extend a special thanks to Professor Yishai Blank for his thoughtful comments on previous drafts. This piece would not be what it is without his guidance, and for that I am extremely grateful.

¹ TALLAHASSEE, FLA., CODE OF GEN. ORDINANCES § 12-61(a) (2009); *see also Fla. Carry, Inc. v. City of Tallahassee*, 212 So. 3d 452, 455–56 (Fla. 1st Dist. Ct. App. 2017).

² *Fla. Carry*, 212 So. 3d at 456.

³ TALLAHASSEE, FLA., CODE OF GEN. ORDINANCES § 13-34(b)(5) (2015); *see also Fla. Carry*, 212 So. 3d at 456.

⁴ *Fla. Carry*, 212 So. 3d at 456.

⁵ FLA. STAT. § 790.33(1) (2017); *see also Fla. Carry*, 212 So. 3d at 455.

⁶ FLA. STAT. § 790.33(1) (2017); *see also Fla. Carry*, 212 So. 3d at 455.

⁷ *See Fla. Carry*, 212 So. 3d at 456.

political consternation,⁸ there is little legal debate that state legislatures, generally speaking, possess broad authority to strike down local legislation.⁹ Indeed, in recent years state preemptive activity has become nearly ubiquitous, with states from Florida to Arizona overturning local enactments that curtail plastic bag use, ban hydraulic fracking, create municipal broadband networks, regulate the sharing economy, and establish municipal living wages.¹⁰ While many commentators are troubled by this rise in preemptive activity,¹¹ few see it as a material departure from the long-standing battle for power between states and their localities.¹²

But Tallahassee's story takes a surprising turn. No longer satisfied with simply "occupy[ing] the . . . field" of gun control", the Florida legislature amended its 1987 statute in 2011 to add an unusual provision: penalties against local officials who pass gun control ordinances in violation of the state's preemptive mandate.¹³ The amendment created a private right of action for declaratory and injunctive relief against any local ordinance that ran afoul of the state's preemption statute and allowed courts to impose civil damages of up to \$5,000 against local legislators charged with supporting the preempted law.¹⁴ Additionally, the law barred local officials named in these lawsuits from using public funds for their legal defense,¹⁵ and provided that violation of the statute could warrant removal from office or termination of employment at the governor's direction.¹⁶

In May 2014, two "gun rights" organizations brought suit against the Tallahassee Mayor and various city commissioners for noncompliance

⁸ See, e.g., Madeleine Davies, *The Republicans are Coming for Your Liberal Bubble*, THE SLOT (Jan. 6, 2017), <https://theslot.jezebel.com/the-republicans-are-coming-for-your-liberal-bubble-1790873529> (lamenting that Republicans, "[n]ot content to leave us with anything nice on this melting planet of ours," are marshalling a concerted effort to prevent liberal cities from passing progressive policies).

⁹ See generally Paul Diller, *Intrastate Preemption*, 87 B.U. L. Rev. 1113 (2007).

¹⁰ For a compilation of recent state preemptive activity, see Lori Riverstone-Newell, *The Rise of State Preemption Laws in Response to Local Policy Innovation*, 47 PUBLIUS 403, 408–17 (2017).

¹¹ See, e.g., Chris Conry, *Statewide Preemption: The Most Dangerous Bill You've Never Heard of*, MINNPOST (Feb. 8, 2017), <https://www.minnpost.com/community-voices/2017/02/statewide-preemption-most-dangerous-bill-you-ve-never-heard> ("Statewide preemption is a shameful attempt by powerful corporate interests to stop regular people from enacting popular laws using legitimate democratic processes. It is a reactionary, oppressive measure: end of story.").

¹² Cf. Kenneth A. Stahl, *Preemption, Federalism, and Local Democracy*, 44 FORDHAM URB. L.J. 133, 134 (2017) (arguing that while state preemption of local laws is "hardly unprecedented", the marked uptick in recent preemption has "rarely been seen in American history").

¹³ See FLA. STAT. § 790.33(3) (2017); *Fla. Carry*, 212 So. 3d at 456.

¹⁴ See FLA. STAT. § 790.33(3)(b)–(c) (2017).

¹⁵ See *id.* at § 790.33(3)(d).

¹⁶ See *id.* at § 790.33(3)(e). This portion of the law has been held unconstitutional solely as it pertains to county commissioners. See *Marcus v. Scott*, No. 2012-CA-001260, 2014 WL 3797314, at **3–4 (Fla. 2d Cir. Ct. Jun. 2, 2014).

with § 790.33.¹⁷ And, while the City ultimately prevailed in this legal battle, Florida's First District Court of Appeal notably declined to hold the state's new punitive measures unconstitutional.¹⁸ The City touted this as a moral victory for local decisionmaking,¹⁹ but in reality this holding did little to limit the State's preemptive power, or, perhaps more accurately, the State's "super" preemptive power.

The term "super preemption," coined in response to the proliferation of preemptive penalties like those in Florida's 2011 amendment, describes the diverse category of state preemption statutes aimed at holding local actors *personally* accountable for ordinances that impermissibly expand local power.²⁰ No longer is it enough for a state legislature to overturn local legislation. Instead, state legislatures have enacted super preemptive punitive measures that place local officials at risk of losing their jobs, paying civil damages, or even facing criminal charges for passing laws that conflict with state statutes.²¹ Moreover, many of these provisions appear to confer liability even in cases where the local law is no longer enforced.²² So long as the preempted law remains on the books, local officials may be held liable for their roles in its passage. Versions of these laws have passed

¹⁷ *Fla. Carry*, 212 So. 3d at 456; see also Kriston Capps, *A Florida Mayor Fights the Gun Lobby*, CITYLAB (Jan. 6, 2017), <https://www.citylab.com/equity/2017/01/a-florida-mayor-fights-the-gun-lobby/512345/>.

¹⁸ *Fla. Carry*, 212 So. 3d at 463–66. Specifically, the court declined to address the City's argument that the state's super preemption law violated its officials' rights to legislative immunity and free speech under both the Florida and United States Constitutions.

¹⁹ See Andrew Gillum, *How to Fight the NRA*, MEDIUM (Jan. 4, 2017), https://medium.com/@a_gillum/how-to-fight-the-nra-1a63f47d4a0c (touting the city's victory over "special interests and corporations . . . trying to intimidate and bully local communities").

²⁰ Given the relative newness of this legislative phenomenon, finding one agreed-upon definition has proven challenging. Indeed, some recent definitions of super preemption have been more encompassing than others. For example, in its September issue brief, the American Constitution Society defined "punitive" preemption (a synonym for super preemption) as any preemptive statute that "seeks to punish local governments and local officials for disagreeing with their states." RICHARD BRIFFAULT ET AL., AM. CONSTITUTION SOC'Y, ISSUE BRIEF, THE TROUBLING TURN IN STATE PREEMPTION: THE ASSAULT ON PROGRESSIVE CITIES AND HOW CITIES CAN RESPOND 9 (2017). Other commentators have defined super preemption so broadly as to include the often-related practice of "blanket" preemption, which bar cities from enacting *any* piece of local legislation that does not perfectly conform to state law. See Richard Florida, *City vs. State: The Story So Far*, CITYLAB (June 13, 2017) <https://www.citylab.com/equity/2017/06/city-vs-state-the-story-so-far/530049/>. This Article will deviate slightly from these broader definitions, defining super preemption as any preemptive legislation that attaches punitive measures directed at local officials in their individual capacities. This definition excludes both punitive measures that target the locality as an institution and blanket preemption measures that lack punitive provisions.

²¹ BRIFFAULT ET AL., *supra* note 20.

²² See, e.g., ARIZ. REV. STAT. ANN. § 41-194.01 (2017).

in states across the country, including Arizona, Mississippi, Texas, Oklahoma, and Kentucky.²³

The rise of statutes like Florida's presents two important questions about the modern relationships between states and localities that this Article seeks to address. First, why now? That is, if super preemption has always been legally permissible, what is it about the current relationship between states and their localities that has prompted so many state legislatures to adopt these punitive measures in recent years? Second, why super preemption? That is, if traditional preemption has historically been such an effective tool for overturning local legislation, what further purpose do these punitive add-ons serve their states? To be sure, if the Florida legislature's aim was to prevent the application of local laws conflicting with the state's gun policies, they achieved that goal with regard to Tallahassee well before passing their punitive amendment. In that way, the State's 2011 measure was, ostensibly, gratuitous—it added nothing beyond the original legislation's preemptive mandate. In order to understand super preemption provisions as something more than empty exercises in vindictiveness, this Article will need to develop a more nuanced picture of why states seek to suppress local legislative activity.²⁴

This Article puts forth two potential answers to the questions posed above. First, this Article argues that super preemption provisions are a symptom of a larger societal trend whereby the fortunes and demographics of our cities and rural communities have sharply diverged. Geography increasingly predicts both political affiliation and economic opportunity.²⁵ Many of America's urban centers are becoming increasingly liberal, affluent islands in seas of rural red. This hardening of political and economic identity along geographic lines helps explain why conservative state legislative leaders are striking an increasingly anti-urban posture.

However, this first theory only tells part of the story. While political geography may explain the timing of these policies, it does little to explain

²³ See ARIZ. REV. STAT. ANN. § 41-194.01 (2017); KY. REV. STAT. ANN. § 65.870(4) (West 2017); MISS. CODE ANN. §§ 45-9-53(5)(a), (c) (2017); OKLA. STAT. tit. 21, § 1289.24(D) (2017); Tex. S.B. 4, 85th Leg., R.S. (2017) (§ 5.02, adding a new § 39.07 to the Texas Penal Code).

²⁴ While this Article devotes much of its energy to developing a descriptive framework for understanding super preemption, many important questions still remain. Notably, this Article devotes little attention to super preemption's normative merits. Is super preemption a desirable practice? Is super preemption legal? If not, what strategies can localities take to prevent its harms? While these questions are important and will receive some cursory attention in the Article's conclusion, fuller explorations of their answers are both necessary, and, unfortunately, outside this project's scope.

²⁵ See Stahl, *supra* note 12, at 146 (noting that "rural residents are now solidly aligned with Republicans and urban dwellers with Democrats," and that cities and rural areas' economic interests have also diverged).

their strategic purpose. This Article's second theory serves that latter goal, arguing that super preemption targets aspects of local lawmaking that traditional preemption cannot reach. Using Professor Heather Gerken's three-part framework for understanding local decisionmaking,²⁶ this Article contends that, under a traditional preemption regime, state legislatures can only suppress one facet of local lawmaking: the act of self-governance. By striking down a local ordinance, state lawmakers have prevented local officials from changing the rules that govern their locality. But local lawmaking is more than simply enacting policy. According to Professor Gerken, local lawmaking serves the additional goals of adding to the marketplace of ideas and providing minorities with an opportunity to craft political identities.²⁷ These auxiliary goals occur irrespective of whether an actual policy ever goes into effect, and, for that reason, they are outside the reach of traditional preemptive measures. Super preemption, however, can reach these auxiliary goals. By preempting both the policy and *politics* of local lawmaking, super preemption has the ability to deflate local progressive action before it has a chance to take flight.

This Article proceeds in three parts. Part I provides a background for understanding local lawmaking power and the State's preemptive ability. Part II attempts to describe the relatively new landscape of super preemption laws. Using various state laws as examples, this Part seeks to develop a basic taxonomy of the super preemption provisions currently in existence. Part III aims to provide some answers to the two descriptive questions posed above. After illustrating that states cannot fully justify super preemption on traditional grounds, this Part will argue that their rise is both a product of America's changing political geography and state legislators' desires to curb both the policies and politics of local lawmaking.

PART I – CITY POWER AND STATES' PREEMPTIVE AUTHORITY

To better understand both super preemption and states' preemptive powers more generally, one must first understand the legal regime that permits such actions. Although both federal and state preemption are commonplace in the United States, it is important to recognize that the

²⁶ See Heather K. Gerken, *Dissenting by Deciding*, 57 STAN. L. REV. 1745, 1749 (2015) (organizing her analysis around three primary justifications for dissent: it contributes "to the marketplace of ideas, engages electoral minorities in the project of *self-governance*, and facilitates *self-expression*") (emphasis added).

²⁷ *Id.*

existence of preemption is not required by America's federalist structure of government. In fact, for much of American history, state preemption was either rare or non-existent.²⁸ Its prevalence today has as much to do with the modern *legal* rules governing the relationships between our cities and states as it does with the fractious political environment surrounding those rules. Indeed, it is quite easy to imagine a legal system where localities are afforded real autonomy over a particular area of policy – a similar arrangement governs the relationship between our federal government and our states. Although the federal government can certainly preempt states on some matters,²⁹ much of state action exists outside the reach of federal meddling. In the same way that the powers afforded to states are *legal* in nature—enshrined in the United States Constitution, statutes, and common law—the current regime of local disempowerment is also a product of well-established legal rules.

This Part describes the evolution of the legal rules that have given rise to state preemption and states' often unchecked authority over local matters. Starting with the theory of limited local authority known as Dillion's Rule, this Part charts the gradual expansion of city power through the Home Rule era into modern times. It then turns to the practice of state preemption, describing its evolution as part of a movement to cabin local autonomy in places where city power was at its height. This Part closes with a recitation of some of the common justifications for state preemption. Using various court opinions as examples, this Part illustrates that preemptive activity has historically been rationalized in three ways: as a mechanism for preserving uniformity, as a protection against extraterritoriality, and as a tool for limiting the subject matter of local action.

²⁸ See Diller, *supra* note 9, at 1123–25 (noting that under earlier legal constructions of city power, preemption was either unnecessary or difficult to achieve).

²⁹ In addition to preempting state legislative activity, the federal government has a long history of preempting local action. Although the federal-local preemptive relationship is not the topic of this paper, many of the same themes addressed in this Article apply to that relationship. See Paul S. Weiland, *Preemption of Local Efforts to Protect the Environment: Implications for Local Government Officials*, 18 VA. ENVTL. L.J. 467, 473–482 (1999) (highlighting several examples of federal preemption of local laws in the environmental context); Annie Decker, *Preemption Conflation: Dividing the Local from the State in Congressional Decision Making*, 30 YALE L. & POL'Y REV. 321, 35–368 (2012) (providing a framework for assessing when it is appropriate for the federal government to preempt local action).

A. The Evolution of City Power

Despite the persistent desire to characterize early American cities as bastions of democratic activity,³⁰ for much of America's history, localities possessed no inherent lawmaking authority. For most of the nineteenth century, cities were understood as little more than creatures of the state which only possessed powers expressly delegated to them from their state governments.³¹ This philosophy was grounded in the legal theories of jurist John F. Dillon, who described cities as state administrative agents only imbued with such powers as granted by the state.³² According to the eponymously-named "Dillon's Rule," if a city wanted to build a road, that city first needed to receive road-making authority through an express delegation from its state legislature. The Dillon's Rule conception of city power dominated city-state relations in the United States until the late 1800's,³³ eventually receiving the Supreme Court's endorsement in the landmark decision, *Hunter v. City of Pittsburgh*.³⁴

Although Dillon's Rule espoused a decidedly limited view of city power, it, perhaps surprisingly, left almost no room for the kind of state preemptive activity seen today.³⁵ Because city action required an express delegation of authority from the state, there were few opportunities for

³⁰ Even as far back as the early nineteenth century, political theorists like Alexis De Tocqueville extolled the virtues of local political activity in the United States. See ALEXIS DE TOCQUEVILLE, *DEMOCRACY IN AMERICA* 59–83 (Phillips Bradley ed., Henry Reeve trans., Vintage Books 1990) (1835).

³¹ See Diller, *supra* note 9, at 1122 (describing the Dillon's rule regime as one that "held that local units of government were mere administrative conveniences of the state with no inherent lawmaking authority"). However, despite Dillon's rather limited appraisal of local power, he did recognize that some localities possessed "inherent" powers that extended beyond explicit statutory grants coming from their states. According to Dillon, this was due to their many business-like characteristics and structures. JOHN F. DILLON, *COMMENTARIES ON THE LAW OF MUNICIPAL CORPORATIONS* § 15, at 34 (4th ed. 1890). See also David J. Barron, *Reclaiming Home Rule*, 116 HARV. L. REV. 2257, 2285–86 (2005) (describing the contours of Dillon's rule).

³² See Diller, *supra* note 9, at 1122. Dillon's narrow conception of city power was not merely anti-local bias. It instead stemmed from a gradual, national evolution in thought regarding the nature of the city. Prior to the 1800s, cities in the United States and England were understood as "municipal corporations," legally indistinct from the business corporations of the day. See Gerald E. Frug, *The City as a Legal Concept*, 93 HARV. L. REV. 1057 (1980) ("It must be understood that before the nineteenth century, there was no distinction in England or in America between public and private corporations, between businesses and cities."). Over time, this conception began to change in the United States. Corporations came to be seen as something private in nature that, if anything, needed protection from the state. Cities, by contrast, were increasingly public entities that needed few, if any, of those same protections. See David J. Barron, *Promise of Cooley's City: Traces of Local Constitutionalism*, 147 U. PA. L. REV. 487, 506 (1999).

³³ See Diller, *supra* note 9, at 1123 (describing Dillon's Rule's dominance through the mid- to late-nineteenth century).

³⁴ *Hunter v. City of Pittsburgh*, 207 U.S. 161 (1907).

³⁵ See Diller, *supra* note 9, at 1123 (describing preemption as "a remote possibility" under Dillon's Rule regimes).

cities to promulgate policies in conflict with their states' wishes. If a state did not want a city to take a particular action, then, presumably, it would not have given the city the ability to take that action in the first instance. Instead, conflicts surrounding local action usually came through claims that the city had behaved *ultra vires*—that is, outside the bounds of the narrow delegations of powers that it had received from the state.³⁶ Although Dillon's Rule regimes have been eclipsed by more "robust" conceptions of city power in most places, the few cities still operating under Dillon's Rule continue to face accusations of *ultra vires* behavior from their states even today.³⁷

Drawing inspiration from the system of dual sovereignty enshrined in the United States Constitution, nineteenth century urban reformers began pushing for a protected sphere of local authority to fight a growing set of urban ills.³⁸ According to these advocates, state-level corruption and financial profligacy contributed to the era's high municipal tax rates, massive urban debt loads, poor housing conditions, and deplorable levels of urban sanitation.³⁹ Under the Dillon's Rule regime, cities interested in addressing these poor living conditions first required express policymaking authority from their state legislatures—the same state legislatures profiting off of urban disarray and under-regulation.⁴⁰ In an effort to protect their desired urban reforms from state legislative meddling, local leaders pushed for—and ultimately received—constitutional carve outs for protected, local lawmaking power.⁴¹ These early "home rule" provisions granted their cities the legislative autonomy to initiate, enact, and implement policies of "local" concern without state permission or oversight.⁴² Still in effect for many cities around the country, these early protections effectively created an "*imperium in imperio*," or "a state within a state," which ultimately contributed to their modern nickname: *imperio* provisions.⁴³

³⁶ *Id.*

³⁷ See *Arlington Cty. v. White*, 528 S.E.2d 706, 709 (Va. 2000) (striking down a domestic partnership ordinance in Virginia on the grounds that it was *ultra vires* the county's local power).

³⁸ See Barron, *supra* note 31, at 2289.

³⁹ *Id.*

⁴⁰ See *id.* at 2288 (describing late nineteenth century cities as being exposed to "state politicians in search of 'spoils.'" Barron argues that state politicians would often craft urban policy so as to place themselves in advantageous positions to obtain city "contracts and franchises," with little regard for how those policies impacted the cities and their residents. *Id.* at 2886–88).

⁴¹ See Diller, *supra* note 9, at 1124–25 (describing the rise of early home rule provisions).

⁴² See Barron, *supra* note 31, at 2290 (describing the package of early home rule powers as "charter power, some initiatory authority, and limited immunity rights").

⁴³ Diller, *supra* note 9, at 1125.

Although *imperio* provisions varied from state to state, these laws typically possessed two important features that protected cities against the kind of express preemptive interference seen today. First, these provisions were typically enshrined in their states' constitutions as opposed to simple statutory enactments.⁴⁴ This meant that state legislatures often had to clear a higher legislative bar if they wanted to overturn or amend these provisions at a later date. Second, and perhaps more importantly, these provisions were only understood to protect matters of *local* concern.⁴⁵ Embedded in this construction is the assumption that there are a set of matters that are distinctly local in nature and, therefore, exist outside the policymaking ambit of the state or federal government. In this way, *imperio* provisions created two nonconcentric legislative spheres—a truly local policy could not be enacted by the state, and a state policy could not be enacted by a locality.⁴⁶ By contrast, preemption requires overlapping spheres of legislative authority; both the state and the locality need to possess the authority to speak on a particular matter before one can make the determination that the state's voice supersedes that of the locality.⁴⁷ This constitutional restriction of *imperio* home rule to matters of local concern has been interpreted by many state courts as affording a degree of immunity from state interference in truly local matters.⁴⁸

However, the existence of an *imperio* provision did not mean that cities instantly had unfettered authority to legislate on matters of local concern. It instead meant that whichever entity was authorized to determine what constituted a “local matter” was also able to establish the metes and bounds of local power. That entity was, almost always, the judiciary. In the wake of the early home rule movement, courts occupied the important role of determining whether a newly-enacted local ordinance was sufficiently local in nature.⁴⁹ Given the term's vagueness, these early court opinions often turned on rather capricious notions of cities' traditional legislative qualifications. For example, land use decisions were typically considered the types of policies that cities enact, so they had to be local in nature.⁵⁰

⁴⁴ See Barron, *supra* note 31, at 2290.

⁴⁵ See Diller, *supra* note 9, at 1124–25.

⁴⁶ See *id.*

⁴⁷ See *id.* at 1125.

⁴⁸ *City of New Orleans v. Bd. of Comm'rs of Orleans Levee Dist.*, 640 So.2d 237, 242 (La. 1994) (noting that under early home rule provisions the city could act “without fear of the supervisory authority of the state government” when its activity was “local” in nature).

⁴⁹ Diller, *supra* note 9, at 1125; Barron, *supra* note 31, at 2325–26.

⁵⁰ See, e.g., *Town of Telluride v. Lot Thirty-Four Venture, LLC*, 3 P.3d 30, 43–44 (Colo. 2000) (Mullarkey, J., dissenting) (describing the land use power as one historically reserved for local actors).

Tax policy, on the other hand, traditionally fell to the state or federal government and therefore could be the type of policy envisioned by the term local.⁵¹ Suffice to say that, although *imperio* provisions greatly expanded local lawmaking authority on paper, in practice, they have been interpreted narrowly so as to provide very limited policymaking space for cities.⁵²

The vagueness of *imperio* provisions coupled with the significant way in which they empowered the courts prompted a second wave of reformers to push for a revised conception of home rule. Beginning in earnest around the 1950's, organizations of municipal leaders such as the American Municipal Association and the National Municipal League pushed for home rule provisions that mirrored the lawmaking authority of the state.⁵³ These "legislative" home rule provisions, which have become the most common approach to home rule, rejected the notion that there was some clearly identifiable set of local matters.⁵⁴ Instead, cities could ostensibly craft policy on any matter on which their states had the authority to legislate.⁵⁵ This broad grant of power was almost universally subject to one, important restriction: a local policy could not conflict with state law.⁵⁶

By greatly expanding the cities' policymaking authority, the legislative home rule provisions brought the separate spheres of state policymaking and local policymaking under one roof. What was once a state concern was now also local, and what was once purely local was now also a matter of state concern. Additionally, by stipulating that local policies not conflict with state statutes, these provisions shifted an important power from the judiciary to the state legislature.⁵⁷ Whereas the judiciary was the primary arbiter of local lawmaking authority under *imperio* regimes, with the passage of legislative home rule provisions, state legislatures gave themselves the final say over whether a city could legislate in a particular area. If the state felt that a matter should be off limits for cities, the

⁵¹ C.f. GERALD E. FRUG & DAVID J. BARRON, CITY BOUND: HOW STATES STIFLE URBAN INNOVATION 147–48 (2013) (discussing the limitations placed on city taxing power).

⁵² C.f. Kenneth A. Stahl, *Local Home Rule in the Time of Globalization*, 2016 BYU L. Rev. 177, 204–05 (2016) (arguing that the state/local distinction that has become so relevant in *imperio* home rule jurisprudence has been used to restrict local policymaking to primarily "family" affairs, while states are afforded policymaking power over "market" concerns).

⁵³ See Diller, *supra* note 9, at 1125–26.

⁵⁴ See *id.*

⁵⁵ *Id.*

⁵⁶ *Id.*

⁵⁷ *Id.* at 1126.

legislature need only pass a law stating as much. And with that, preemption power was born.⁵⁸

Before addressing the practice of super preemption specifically, it is worth noting that the formal categories of Dillon's Rule, *imperio* home rule, and legislative home rule do not perfectly capture the messiness and complexity of local legislative power. As Professor David Barron illustrates, early home rule provisions may have appeared similar on paper, but in practice, they placed very different limitations on city power, depending on the ideological leanings of their proponents as well as of the judges interpreting these provisions.⁵⁹ Indeed, while many cities may operate under a constitutionally-enshrined *imperio* provision, courts sometimes interpret these provisions in unpredictable or capricious ways. When judicial opinions interpret an *imperio* provision narrowly, they may limit the city's sphere of legislative immunity by allowing state preemption in matters that may be traditionally understood as local concerns.⁶⁰ In brief, the categories outlined above are not meant to describe city-state relations perfectly; instead, they are meant to serve as generalized typologies for the ways in which states delegate powers to their local subordinates.

B. Understanding State Preemption Doctrine

As the previous section illustrates, preemption is neither a necessary nor an intuitive practice in a system where subsidiary governments possess lawmaking authority.⁶¹ Instead, state preemption is a recent phenomenon responding to modern changes to the laws governing city power. For this reason, scholarly analysis on this topic is relatively sparse. While several scholars have addressed state preemption as a subset of broader discussions on state power, very few have explored the practice in depth or analyzed

⁵⁸ See *id.* (discussing the relationship between preemption and legislative home rule).

⁵⁹ Barron categorizes these three competing ideologies by the types of cities they aimed to create. These include the "Old Conservative City," whose proponents aimed to carve out just enough local legislative authority to combat state-level largess; the "Administrative City," whose advocates pushed for state delegation to an apolitical class of local government professionals tasked with addressing the complexities of rapid urbanization; and the "Social City," whose reformers saw city power as a political tool for redistributive ends. Barron, *supra* note 31, at 2292–309.

⁶⁰ See, e.g., *Town of Telluride*, 3 P.3d at 37 (describing a complicated three-tiered test for the permissibility of state preemption, whereby matters of truly local concern are afforded immunity from state preemption, but matters of statewide or "mixed" concern are subject to state legislative interference.).

⁶¹ Stephen A. Gardbaum, *Nature of Preemption*, 79 CORNELL L. REV. 767, 768 (1994) ("The granting of a power of preemption to the central government is a common, but not a necessary, feature of a federal state.").

how preemption statutes have been interpreted by state courts.⁶² However, in order to better understand the recent rise of super preemption provisions, it is important to situate these laws in the general landscape of state preemption.

Although the proliferation of legislative home rule provisions ostensibly reaffirmed the legislature's role as the primary arbiter of local power, courts still play an important role in the preemption battles across the country. Indeed, while legislatures most always have the ability to decide *if* they will preempt a particular local action, *whether* they have preempted or *what* they have preempted are often open questions that courts are enlisted to answer.⁶³ How the courts answer those questions is typically a function of the kind of preemption at play in a particular dispute. Most state courts, in keeping with the framework established in federal preemption jurisprudence, divide preemptive actions into two categories: express or implied preemption.⁶⁴ Whether a preemptive action is express or implied can determine everything from the type of analysis the court applies to the dispute, to the complexity of the legal questions at play, to whether the court will hear the case in the first place. For these reasons, understanding the contours of these two categories is necessary for furthering one's understanding of both traditional preemption and the more recent super preemption provisions.

Express preemption is perhaps the clearest category of preemption, although not the most common.⁶⁵ It occurs when a state legislature enacts a law that explicitly prohibits localities from taking a particular legislative action, or mandates that localities overturn a law that is already on their books.⁶⁶ This type of preemption can take a variety of forms; including specific prohibitions against local policies like gun control, fracking

⁶² Two notable outliers in this regard are recent articles by Professors Paul Diller and Kenneth Stahl. See generally Diller, *supra* note 9, at 1114 (suggesting courts addressing state preemption questions aim to maximize "good-faith" experimentation while minimizing exclusionary or parochial policies); Stahl, *supra* note 12 (teasing out the relationship between geographic political polarization and an increase in state preemptive activity).

⁶³ See Diller, *supra* note 9, at 1126 ("Thus, despite the second-wave home-rule reformers' intent to remove the responsibility for deciding the scope of local authority from the judiciary, legislative home rule traded the much-criticized judicial role of determining whether a subject matter was properly 'local' for the equally controversial task of applying the doctrine of preemption.").

⁶⁴ *Id.* at 1141–42 (noting that while Utah is the only state to explicitly adopt the Supreme Court's taxonomy, all state courts but Illinois recognize both conflict and implied preemption).

⁶⁵ Cf. Mary J. Davis, *The New Presumption Against Preemption*, 61 HASTINGS L.J. 1217, 1228 (2010) (noting that, historically, express preemption analysis has been rarely applied at the federal level).

⁶⁶ See Diller, *supra* note 9, at 1115 (defining express preemption).

ordinances, and rent control laws;⁶⁷ or blanket prohibitions against local laws on topics as broad as public health, social justice, or environmental protection.⁶⁸ In fact, several states have begun enacting even broader express preemption provisions, outlawing *any* municipal actions that do not perfectly conform to state law.⁶⁹ Texas, for example, recently introduced a bill that would have prohibited any local legislation that did not first receive express state approval.⁷⁰ Although that law was ultimately rejected,⁷¹ similarly broad express preemption provisions have appeared in Arizona and Oklahoma.⁷² These bills illustrate the sheer diversity, breadth, and ambition of express preemption provisions. Ultimately, the most important identifying features are that these provisions clearly point to types of policies that localities have enacted or could enact, and unambiguously establish that localities can no longer legislate in these areas.

As previously mentioned, a byproduct of the rapid expansion of legislative home rule has been a reduction in the role of the judiciary in disputes about city power. This is particularly true with regard to express preemption provisions. By passing an express preemption provision, state officials leave little room for ambiguity as to whether a locality can continue to take a particular course of action.⁷³ If courts have determined that a state has preemptive power over its localities, few questions remain after a state has expressly preempted a category of local law. With that said, the court's role in express preemption disputes is not immaterial. As a preliminary matter, courts still have to determine if the state can preempt local action in the first place.⁷⁴ In *imperio* states where home rule provisions are enshrined in the state constitution, localities are afforded a sphere of constitutionally protected lawmaking authority that even express preemption cannot pierce. Courts must therefore determine to what degree their state protects that kind of local power, and if the preempted action falls within the class of "local" policies that are often afforded constitutional protection.

⁶⁷ See Riverstone-Newell, *supra* note 10, at 407.

⁶⁸ AMERICANS FOR NONSMOKERS' RIGHTS, THE EVOLVING FACE OF PREEMPTION: NEW TACTICS TO LIMIT LOCAL CONTROL 2 (2016), <https://perma.cc/3SQC-R9CP>.

⁶⁹ Riverstone-Newell, *supra* note 10, at 418.

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² *Id.*

⁷³ See Diller, *supra* note 9, at 1158 (noting that Illinois' refusal to acknowledge implied preemption "severely reduces the judicial role in deciding questions of preemption").

⁷⁴ See, e.g., *Town of Telluride*, 3 P.3d at 37 (exploring the reach of Colorado's home rule power to determine if the state's express preemptive activity actually applied to Telluride's case).

Additionally, courts must determine if the local activity at issue is the kind of action covered by the express preemption provision.⁷⁵ In some cases that inquiry is fairly simple. If a state prohibits localities from “banning or imposing a fee for the use of paper or plastic bags,”⁷⁶ there should be little dispute as to whether a city’s tax on plastic bags has been preempted. However, not all disputes are this easily resolved. For example, in the *Florida Carry* case profiled earlier, the state law expressly preempted the “promulgation” of local firearm ordinances.⁷⁷ While there was no dispute as to whether Tallahassee’s two laws were firearm ordinances, the court nevertheless determined that they were not covered by the state’s express preemption provision because they were no longer enforced.⁷⁸ According to the court, unenforced ordinances were not “promulgated” in the way that the state law envisioned.⁷⁹ For that reason, applying the state’s preemption statute to Tallahassee’s laws made little sense, despite the legislature’s expressed intent to cover all local firearm regulations. This example illustrates that even under an express preemption provision, the judiciary plays an important but circumscribed role in determining the bounds of local power.⁸⁰ With that said, the opportunities for judicial discretion are few and far between under express preemption provisions.⁸¹ Given that super preemption laws are, by their very nature,

⁷⁵ See Diller, *supra* note 9, at 1158 (noting that despite only recognizing express preemption, “Illinois courts still play a role in determining whether the legislature has expressly preempted a certain field, and, if so, the extent of such a preemption provision”).

⁷⁶ Diana Barr, *Missouri Legislators Block Cities from Raising Minimum Wage*, ST. LOUIS BUSINESS JOURNAL (Sept. 17, 2015), https://www.bizjournals.com/stlouis/morning_call/2015/09/missouri-legislators-block-cities-from-raising.html.

⁷⁷ *Fla. Carry*, 212 So. 3d at 457.

⁷⁸ *Id.* at 458–59.

⁷⁹ *Id.* at 459.

⁸⁰ For other examples of the court playing a critical role in a battle over express preemption, see *Town of Telluride*, 3 P.3d at 37; *Fondessy Enters. v. City of Oregon*, 492 N.E.2d 797, 799 (Ohio 1986); *Dallas Merch.'s & Concessionaire's Ass'n v. City of Dallas*, 852 S.W.2d 489, 490 (Tex. 1993).

⁸¹ In opining on the opportunities for judicial discretion in a regime where express preemption was the only mechanism by which states could preempt local action, Professor Paul Diller made the following observation:

Express-only preemption also aims to deprive judges of discretion and the capability of rendering anything resembling a normative judgment. In this vein, Professor Elhauge and other proponents of default-rule theory have described the role of a judge as merely that of an “agent” carrying out the legislature’s instructions. As applied to preemption, an “express-only” default rule reduces judges to “agents” merely searching for a specific instruction from the legislature rather than partners in the process of interpreting state laws and developing the vertical distribution of power in a home rule system.

always express provisions, the fact that courts have historically played a minor role in adjudicating this category of preemptive dispute may indicate that super preemption will receive similarly short shrift from the judiciary moving forward.

Implied preemption, the second preemption category, is slightly more complicated. Most courts subdivide implied preemption into two further analytical categories. The first, conflict preemption, occurs when a local ordinance frustrates or directly impedes a state law's aims.⁸² For example, in *Casuse v. City of Gallup*, the New Mexico legislature passed a law requiring cities with populations of 10,000 or more to elect their city council members from single-member districts.⁸³ Despite having a population of more than 10,000 people, the city of Gallup, elected its council members via at-large districts.⁸⁴ Recognizing that the Gallup ordinance directly conflicted with the state's single-member district statute, the New Mexico Supreme Court held that the local law was preempted and struck it down.⁸⁵ The court reached this holding despite the fact that the state's statute included no express language explicitly preempting the local ordinance.⁸⁶ The fact that the two laws were incompatible was enough to indicate that the state legislature had impliedly preempted the ordinance and all others like it.

The second category of implied preemption, field preemption, requires even less of an affirmative statement from a state legislature for a determination that local law has been preempted. With field preemption, it is enough that the state legislature has simply "occupied the field" in a particular area for a court to preclude local action on that matter.⁸⁷ The theory behind field preemption is that when a legislature develops a comprehensive regulatory scheme on an issue, the legislature impliedly indicates its intent for that set of policies to be the final word on the issue.⁸⁸ In these cases, it does not matter if a local ordinance directly conflicts with the state's law. As long as the state has sufficiently occupied this policy field, the local law cannot stand.

Diller, *supra* note 9, at 1159 (footnotes omitted).

⁸² See, e.g., *Bloom v. City of Worcester*, 293 N.E.2d 268, 283, n.16 (Mass. 1973) (describing the test for conflict preemption as "whether the local ordinance . . . frustrates the fulfilment [sic] of the legislative purpose of any arguably relevant general law").

⁸³ *Casuse v. City of Gallup*, 746 P.2d 1103, 1104 (N.M. 1987).

⁸⁴ *Id.*

⁸⁵ *Id.* at 1105.

⁸⁶ *Id.* at 1104-05.

⁸⁷ See Weiland, *supra* note 29, at 470 (1999).

⁸⁸ See *id.* ("Field preemption may be implied from a pervasive scheme of federal regulation.").

State courts have decided field preemption cases on numerous occasions. For example, in *O'Connell v. City of Stockton*, Stockton, California passed an ordinance providing for the “forfeiture of [a]ny vehicle used . . . to acquire or attempt to acquire any controlled substance.”⁸⁹ Plaintiffs argued, in part, that the law was preempted by the California Uniform Controlled Substances Act (UCSA), which, among other things, authorized vehicle forfeiture for particular drug crimes.⁹⁰ The California Supreme Court, after considering the UCSA “*as a whole*,” ultimately agreed.⁹¹ According to the Court, even though the UCSA did not contemplate forfeiture for simple drug possession crimes and therefore did not directly conflict with the more stringent Stockton ordinance, the legislature’s host of regulations on the matter indicated a “clear intent” to reserve forfeiture for more serious crimes.⁹² In other words, because the state legislature had developed a “comprehensive scheme”⁹³ addressing forfeiture in drug crimes, they had fully occupied the field in that policy area as to preclude any further regulation from subsidiary governments. Cases like *O'Connell* depict the court’s role at its apex for preemption cases. With field preemption cases, courts are tasked with not only determining what constitutional or statutory power a city has relative to the state, but also with determining if a legislature has spoken expansively enough on an issue to foreclose local regulation on that matter. This latitude grants judges a level of interpretive (and normative) discretion that is almost always lacking in express preemption cases.

Implied preemption provisions, however, bear little resemblance to the super preemption provisions addressed in this Article. As discussed previously, super preemption provisions are punitive measures attached to *express* prohibitions against a category of local action. In this way, they will likely come to resemble other forms of express preemption, leaving little room for judicial discretion while maximizing the legislature’s power in intrastate disputes. Nevertheless, implied preemption cases highlight something notable about the recent proliferation super preemption provisions. As Professor Paul Diller has recognized, much implied preemption litigation is initiated by local business interests—not the city or state governments whose laws are implicated in these cases.⁹⁴ This lies in

⁸⁹ *O'Connell v. City of Stockton*, 162 P.3d 583, 586 (Cal. 2007) (quoting Stockton Mun. Code, § 5-1000).

⁹⁰ *Id.* at 588.

⁹¹ *Id.*

⁹² *Id.* at 590.

⁹³ *Id.*

⁹⁴ See Diller, *supra* note 9, at 1140.

stark contrast to the few super preemption cases that courts have heard to date. In super preemption cases in Florida, Texas, and Arizona, both the plaintiffs and defendants have come exclusively from state government, municipal offices, or advocacy organizations with an interest in the policy matter at hand.⁹⁵ In each of these cases the state has played an extremely active role in the litigation, submitting briefs and filing motions in defense of their preemptive provisions.⁹⁶ These few examples illustrate that super preemption cases are not dealing with parochial matters of purely local concern.⁹⁷ These are politically charged disputes in which the states have very real interests in prevailing. As this Article will soon argue, the deeply political nature of these provisions is one of the features that separates super preemption from much of the traditional preemptive activity.

C. Common Justifications for State Preemption

Given the ease with which preemptive provisions are passed by legislatures and upheld by many courts, it stands to reason that states and judges must have *some* justification for why they believe particular laws are best implemented at the state level. After all, many state preemption cases turn on whether the ordinance in question is sufficiently “local” in nature. In order for a court to make that determination—or for a state to assert otherwise—it should have some methodology for deciding what constitutes a local matter as compared to something best dealt with by the state. As it turns out, both states and courts rely on three common justifications for preemptive action: a desire for uniformity, a concern about extraterritoriality, and a distrust of local government’s ability to adequately handle certain challenges.⁹⁸ Understanding these justifications

⁹⁵ See *Fla. Carry*, 212 So. 3d at 455 (identifying advocacy groups Florida Carry, Inc. and The Second Amendment Foundation, Inc. as appellants); *City of El Cenizo v. Texas*, No. 17-50762, 2017 WL 4250186, at *1 (5th Cir. Sept. 25, 2017) (identifying the state of Texas as the defendant); *State ex rel. Brnovich v. City of Tucson*, 399 P.3d 663, 668 (Ariz. 2017) (noting that the litigation was prompted by the state filing a special action against the city).

⁹⁶ See Attorney General’s Motion for Summary Judgment and Response to Defendants’ Motion for Summary Judgment, *Fla. Carry*, 212 So. 3d 452 (No. 2014CA001168); Brief for Appellants, *City of El Cenizo*, 2017 WL 4250186 (No. 17-50762); Petitioner State of Arizona Ex Rel. Brnovich’s Supplemental Brief, *Brnovich*, 399 P.3d 663 (No. CV-16-0301-SA).

⁹⁷ See Erin Adele Scharff, *Hyper Preemption: A Reordering of the State–Local Relationship?*, 5 GEO. L. J. 1469, 1519–20 (describing the Arizona Attorney General’s apparent unwillingness to enforce its super preemption law against localities for “run-of-the-mill issues, driven by local losers in zoning disputes and other principally local matters”).

⁹⁸ See, e.g., *City and County of Denver v. State*, 788 P.2d 764, 768 (Colo. 1990) (noting that the three factors the court considers when assessing whether a policy falls within the state’s ambit include “need for statewide uniformity,” “impact of the municipal regulation on persons living outside the municipal limits,” and “whether a particular matter is one traditionally governed by state or by local government.”).

for traditional preemption will prove helpful in eventually teasing out a more nuanced justification for the recent spate of super preemption provisions.

One of the most common justifications for traditional preemption is a desire for state uniformity.⁹⁹ Of particular relevance in issues pertaining to business and mobile capital,¹⁰⁰ the theory holds that if mobile businesses have to navigate a patchwork of regulations in expanding from one municipality to the next, they will eventually grow frustrated and leave for a state with a less cumbersome regulatory landscape or potentially pass their increases in production costs on to consumers in the form of higher prices.¹⁰¹ In *American Financial Services Association v. City of Oakland*, the California Supreme Court relied on this justification to strike down a predatory lending ordinance passed in the City of Oakland.¹⁰² In that case, Oakland's law limited the amount in fees mortgage lenders could charge on subprime loans and mandated that subprime mortgage lenders not engage in various predatory or deceptive financial practices with prospective clients.¹⁰³ Plaintiffs pointed to similar legislation passed by the California legislature to argue that Oakland's more stringent law had been preempted and was therefore unenforceable.¹⁰⁴ Despite evidence indicating that the legislature had not intended to preempt local law with their statute,¹⁰⁵ the court ultimately sided with the plaintiffs.¹⁰⁶ According to the court, the California legislature had presumably balanced the risks of subprime mortgage lending with the benefits of providing their citizens

⁹⁹ See Lynn A. Baker & Daniel B. Rodriguez, *Constitutional Home Rule and Judicial Scrutiny*, 86 DENV. U. L. REV. 1337, 1349 (“[T]he two factors that seem to loom largest” when determining what fall should fall within the state’s policymaking power are “the extraterritorial effects of the local regulation[] and the need for statewide uniformity in the relevant regulatory area”).

¹⁰⁰ Professor Richard Schragger uses the term mobile capital to describe individuals and firms that have the ability to move from one jurisdiction, often in response to some local policy. See generally Richard Schragger, *Mobile Capital, Local Economic Regulation, and the Democratic City*, 123 HARV. L. REV. 482 (2009). Schragger draws a distinction between this type of highly mobile capital and “place-dependent capital,” which includes fixed assets like office buildings, homes, and railroads. *Id.* at 493.

¹⁰¹ See, e.g., *N. Calif. Psychiatric Soc’y v. City of Berkeley*, 178 Cal. App. 3d 90, 101 (Cal. Ct. App. 1986) (“Certain areas of human behavior command statewide uniformity, especially the regulation of statewide commercial activities . . .”).

¹⁰² *Am. Fin. Serv. Ass’n v. City of Oakland*, 104 P.3d 813, 823 (Cal. 2005) (“Moreover, it is beyond peradventure that effective regulation of mortgage lending, and in particular here abusive practices in such lending, ‘requires uniform treatment throughout the state.’” (quoting *Chavez v. Sargent*, 339 P.2d 801, 810 (Cal. 1959))).

¹⁰³ See *id.* at 818–19.

¹⁰⁴ *Id.* at 815.

¹⁰⁵ See *id.* at 826 (describing evidence that the legislature considered adding express preemption language into the statute and opted against it).

¹⁰⁶ *Id.* at 829.

easy access to liquidity.¹⁰⁷ By appending further prohibitions onto this regulatory baseline, Oakland risked “divid[ing] the state's economy into tiny geographic markets” and ultimately pushing lenders out of the state entirely.¹⁰⁸ For that reason, the court determined that the legislature must have impliedly preempted local laws like the one at issue, and chose to strike it down.

A second, related justification for state preemption is the fear of extraterritoriality.¹⁰⁹ Despite the best intentions of lawmakers, laws do not always obey political boundaries. Instead, the effects of particular laws often creep across jurisdictions, sometimes adversely impacting neighboring polities that had no say in the offending action. This “negative externality” problem is particularly pronounced in the context of local governments. With small geographic boundaries and many neighboring jurisdictions in close proximity, a local government’s law could have far-reaching impacts for citizens across a metropolitan region.¹¹⁰ The Colorado Supreme Court addressed this exact issue in *Town of Telluride v. Lot Thirty-Four Venture*.¹¹¹ In that case, the Town of Telluride passed a rent control ordinance mandating that all new development include a certain percentage of affordable units.¹¹² In assessing whether the matter was best characterized as one of state or local concern, the court pointed, in part, to the law’s extraterritorial impact.¹¹³ By requiring the construction of affordable units, Telluride was, in effect, limiting the supply of market-rate units that could be developed in its borders.¹¹⁴ According to the court, this limitation could cause a “ripple effect” across the entire region’s housing market, foisting the unsatisfied demand for market-rate construction upon neighboring localities that had no say in Telluride’s policy decision.¹¹⁵ For this reason among others, the Colorado Supreme Court determined that rent control policies were better decided by the state, and held that Telluride’s policy had been preempted.¹¹⁶

¹⁰⁷ *Id.* at 824. It is worth noting that this case was decided in 2005, well before the subprime mortgage crisis that would ultimately sour even the most positive perspectives on subprime loans.

¹⁰⁸ *Id.* at 825.

¹⁰⁹ See Baker & Rodriguez, *supra* note 99.

¹¹⁰ Cf. Marygold Shire Melli & Robert S. Devoy, *Extraterritorial Planning and Urban Growth*, 1959 WIS. L. REV. 55, 55 (1959) (“However, in an urbanized area consisting of several governmental units, it is not enough that each unit individually prepare for the future. Political boundaries are arbitrary in the sense that they may have no relationship to the economic and social units.”).

¹¹¹ *Town of Telluride*, 3 P.3d at 38–39.

¹¹² *Id.* at 33.

¹¹³ *Id.* at 38–39.

¹¹⁴ *Id.* at 39.

¹¹⁵ *Id.*

¹¹⁶ *Id.* at 40.

A final, perhaps less common, justification for state preemption addresses the comparative competencies of state and local governments. Although courts and state leaders may be reluctant to uphold preemptive action on institutional competency, this issue often bubbles just beneath the surface of most preemption conversations. Indeed, although it was never stated explicitly, institutional competency seems to have influenced the court's decision in the aforementioned *American Financial Services Association* case. Despite formally justifying their holding on uniformity grounds, the majority frequently alluded to concerns about the complexity of the problem at hand. The court notes that, while Oakland may in fact bear a disproportionate burden from subprime lending tactics, those burdens "do not give the City a license to regulate a *highly complex* financial area comprehensively addressed by state law."¹¹⁷ The court goes on to extol the legislature's "reasoned assessment"¹¹⁸ of the complicated situation, and argues that the modern reality of mortgage-backed securities would "confound" a system of locally-based regulation.¹¹⁹ This language suggests that the court is simply more comfortable with the state legislature's ability to grasp and analyze the details of the financial system.

It is important to note that opinions regarding institutional competency are not necessarily grounded in unfounded prejudice. There are many reasons to believe that state governments have some technical superiority over their local counterparts. For one, state governments tend to be larger than local governments and can therefore probably provide more manpower to solving a problem than individual localities. Additionally, state governments likely draw from a wider pool of job applicants than local governments, increasing the likelihood that they will be able to hire someone with a niche but valuable skillset. Finally, state governments likely have better financing and therefore can pay their employees higher salaries and provide them with better resources. Assuming qualified applications are at least partially motivated by pay and institutional resources, this financial disparity may result in a noticeable skill gap between state and local governments.

These three justifications for preemption are neither collectively exhaustive nor mutually exclusive. Courts and state governments often rely on these three justifications in tandem, weaving arguments from one

¹¹⁷ *Am. Fin. Serv. Ass'n*, 104 P.3d at 825 (emphasis added).

¹¹⁸ *Id.*

¹¹⁹ *Id.* at 823.

justification to the next to support a preemptive decision.¹²⁰ Additionally, scholars and judges have put forth various other arguments to justify preemption at both the federal and state levels.¹²¹ These examples merely serve to illustrate the philosophical underpinnings beneath preemptive action and centralized governing, more generally. As this Article will soon argue, while state leaders lean on these same justifications in their support of super preemption, these traditional arguments for centralized decisionmaking fail to fully explain the purpose behind these punitive measures.

PART II – THE RISE OF SUPER PREEMPTION

Given the frequency with which traditional preemption provisions are enacted and upheld by state courts, why should one think about super preemption any differently? On the one hand, these policies are simply additional manifestations of the states' supremacy over their local governments, grounded in the same, well-established legal tradition as any other preemption provision. On the other hand, super preemption is unique—and therefore noteworthy—for two reasons. First, prior to the birth of super preemption in 2003,¹²² legislators had never tied punitive provisions to preemptive legislation. Although these punitive provisions come in a variety of forms, as a whole, they signal a marked shift in the way in which states approach the practice of preemption. Second, most super preemption provisions aim to pierce the governmental veil of the localities that they target. These provisions are not simply concerned with attacking the policies passed by city officials, nor are they simply concerned with holding cities as institutions accountable for the policies' passage. Instead, many super preemption provisions aim to hold the *individual local officials* accountable for their legislative actions. This, of course, changes the power dynamics of state preemption. What was once a battle for authority between states and their cities has now become a battle over individual legal consequences between states and city officials. For these reasons, scholars should view super preemption as something more than a mere continuation of traditional preemption's reign. These

¹²⁰ See, e.g., *City and County of Denver*, 788 P.2d at 768 (noting that Colorado courts consider relevant all three justifications to make preemption determinations).

¹²¹ See generally, e.g., Mark D. Rosen, *Contextualizing Preemption*, 102 NW. U. L. REV. 781 (2009) (arguing that preemption exists as a mechanism for addressing the maladies of concurrent governmental powers).

¹²² See BRIFFAULT ET AL., *supra* note 20 (describing Oklahoma's 2003 super preemption law as the first of its kind).

provisions signal a paradigm shift in the relationship between states and localities; therefore, they deserve specific attention as a category unto themselves.

This Part aims to begin some of that work. First, this Part will address some of the historical precedent for super preemption provisions. While nothing quite like super preemption has ever occurred in the past, these laws do carry some thematic similarities to nineteenth century “ripper” legislation, as well as to the jurisprudential thread that attributes corporate-like fiduciary duties to city officials. These two legal practices foreshadowed super preemption in that they conceptualized the role of city officials differently than other government actors, and therefore afforded them fewer protections or required additional responsibilities of them. After addressing these historical trends, this Part will then turn to super preemption provisions in earnest by outlining some of the common features in these modern provisions and providing various examples from around the country.

A. Historical Precedence for Super Preemption

As previously mentioned, one of the most salient features of super preemption provisions is the way that they move past the city as an institution to attach damages to local elected officials or city administrators. Traditional preemption pits different levels of government against each other in battles where the victorious party is awarded the ability to enact and enforce a particular piece of legislation and the losing party (often the city officials) bears no residual damage beyond their inability to implement its desired policy. Super preemption changes this dynamic. With these provisions, the opposing parties are no longer state and city, but state and *city officials*. Moreover, the terms of the battle have also changed. Either victorious party is still awarded authority to enact their desired law; but if the city officials lose, they not only lose the ability to enact a particular law, but also may experience civil damages, criminal penalties, and/or loss of employment.

If these individual damages seem odd, they should. Legislators, even at the local level, have traditionally been afforded a wide degree of legislative immunity for work performed in their elected capacity.¹²³ This means that officials cannot be held personally liable for the government decisions they make while acting in their official capacity. Indeed, some local officials

¹²³ See Max Schanzenbach & Nadav Shoked, *Reclaiming Fiduciary Law for the City*, 70 STAN. L. REV. 565, 635 (2018).

operating under super preemption regimes have already raised the legislative immunity as a legal defense to some of these punitive provisions.¹²⁴ However, legislative immunity for local officials sits uneasily next to the myriad of ways in which the law has historically assigned vulnerabilities and responsibilities to local actors unexperienced at higher levels of government.

One example of how state laws historically disempowered local officials are “ripper” bills. Ripper bills were legislative acts that “ripped” authority from local officials and vested it at the state level.¹²⁵ These laws were commonplace throughout the nineteenth and early twentieth centuries,¹²⁶ despite frequently drawing the ire of local officials and city residents. For example, in 1871, the Michigan state legislature passed a bill that took the authority to appoint a board of public works away from the city of Detroit and placed it in the hands of a state body.¹²⁷ In 1857, the New York state legislature enacted a similar statute removing New York City’s ability to organize its local police and granting that power to the state’s governor.¹²⁸ Perhaps most shockingly, in 1870, state legislators in Harrisburg took over managing the construction of Philadelphia’s City Hall from local officials.¹²⁹

Each of these ripper bills was promoted by the state legislature as a legislative change aimed at empowering state governments. However, the unspoken corollary to state empowerment in these cases was the *disempowerment* of local actors. These bills not only ripped authority from the city as an institution; they ripped responsibilities away from local individuals who were tasked with carrying out these mandates. In this way ripper legislation bears a striking resemblance to modern super preemption. Both of these legislative tactics aim to empower state government at the expense of local office holders. Power and authority that was at one time unquestionably vested at the local level is in both cases taken by the state, enfeebling local actors by restricting their scope of power.

¹²⁴ See, e.g., *Amicus Curiae in Support of Cross-Appellants at 12, Fla. Carry*, 212 So. 3d 452 (No. 1D15-5520).

¹²⁵ See Lyle Kossis, *Examining the Conflict between Municipal Receivership and Local Autonomy*, 98 VA. L. REV. 1109, 1126 (2012).

¹²⁶ See Stahl, *supra* note 12, at 145.

¹²⁷ *People ex rel. Le Roy v. Hurlbut*, 24 Mich. 44, 53 (1871).

¹²⁸ *People ex rel. Wood v. Draper*, 15 N.Y. 532, 535 (1857).

¹²⁹ See Kossis, *supra* note 125.

Another example that illustrates the unique legal treatment of city officials is the attachment of fiduciary duties to local actors.¹³⁰ Fiduciary duties, typically applied to agents controlling trusts or corporations, are divided into two categories: duties of care and duties of loyalty.¹³¹ Under a duty of loyalty, a fiduciary agent is required to avoid conflicts of interest when managing whatever assets are under their control.¹³² Under a duty of care, a fiduciary agent must exercise sound management of those assets.¹³³ Although fiduciary duties are typically associated with *private* law, courts have historically held that, when city officials act in their role as marketplace participants (e.g. when cities behave like parties to private contracts), it is appropriate to attribute a form of fiduciary duties to local actors.¹³⁴ For example, in *Milhau v. Sharp*, the New York city council agreed to allow a private party the right to run a passenger railway down a public street.¹³⁵ There was little question that the city possessed the legal authority to make such a grant—after all, the street was public and the private party would be paying for access.¹³⁶ The plaintiffs, however, took issue with the *amount* of money that the city was willing to accept to allow the railway to operate.¹³⁷ According to them, by awarding the street access for a “trifling sum,” the city was paid less than fair value.¹³⁸ The court agreed.¹³⁹ It stated that, while it typically avoided passing judgement on the wisdom of political acts, when the city council acted “with reference to its private property,” it was no longer acting within its legislative capacity—instead it was “as if it were the representatives of a private individual, or of a private corporation.”¹⁴⁰ As the proprietors of public assets, council members were bound by a fiduciary duty that did not exist when exercising traditional legislative powers. Because the council ignored that duty by accepting considerably less than fair market value for sale, the court struck down the transaction.

¹³⁰ See Schanzenbach & Shoked, *supra* note 123, at 573 (describing city officials “long-dormant status” as fiduciaries when transacting in city assets).

¹³¹ *Id.* at 568.

¹³² *Id.*

¹³³ *Id.*

¹³⁴ *Id.* at 573 (“[A] long line of forgotten common law decisions from the nineteenth and early twentieth centuries held that city officials are fiduciaries when transacting in city assets and making contracts on the city’s behalf.”).

¹³⁵ *Milhau v. Sharp*, 15 Barb. 193, 206–07 (N.Y. Gen. Term 1853).

¹³⁶ *Id.* at 207 (citing *Drake v. Hudson River R.R. Co.*, 7 Barb. 528 (N.Y. Gen. Term 1849)).

¹³⁷ *Id.* at 214–15.

¹³⁸ *Id.* at 198.

¹³⁹ *Id.* at 194.

¹⁴⁰ *Id.* at 193–94.

It is important to highlight just how powerful this decision is. This, unlike previous cases in this Article, is not an instance in which the court struck down a local action because it was preempted by state law. The state was not a party to this matter, and the court did not doubt that the city had the *legal* authority to make this transaction. Instead, the court chose to strike down a lawful action performed by an elected legislative body because it decided the transaction was a bad *business* deal. Here, the Court treated the city council as an agent of city residents, held to a higher standard when entering business transactions regarding public property. Although the court was not disempowering the council as was the case with the ripper legislation, it was attaching additional responsibilities to the position that other legislators did not have.¹⁴¹ Similarly, this decision illustrates how courts may conceive local officials as distinct from other categories of elected governmental agents. In this way, decisions like *Milhau* and ripper legislation may have presaged the unorthodox treatment of local officials in super preemption provisions.

B. The Current Landscape of Super Preemption Provisions

Despite sharing some thematic similarities to the historic trends just outlined, modern super preemption provisions come in a variety of forms. Indeed, while all super preemption provisions include punitive measures leveled at localities and local actors, no two punitive measures are exactly alike. This Section describes several of the most common punitive features in super preemption provisions, including reductions in state funds, private rights of action, civil damages, criminal penalties, removal from office, and restrictions on the use of government funds in legal disputes. In describing these features, this Section will introduce various pieces of super preemption legislation as examples of how states implement these features in practice.

1. Private rights of action

One common feature in many super preemption provisions is the creation of private rights of action. Under these provisions, any private citizen or organization who believes they have been adversely impacted by the local ordinance has a statutory right to initiate litigation against a

¹⁴¹ *C.f.* Schanzenbach & Shoked, *supra* note 123, at 576–78 (arguing that the modern scholarly trend of trying to ascribe fiduciary duties to federal or state officials sits on uneasy ground and does not comport with the way judges and policymakers have historically understood these actors).

locality in violation of the states super preemption statute.¹⁴² This feature played a prominent role in the *Florida Carry* case.¹⁴³ In that dispute, the plaintiff notably was not the state government or some agent thereof. Instead, two advocacy organizations (Florida Carry, Inc. and The Second Amendment Foundation, Inc.) leveled complaints against Tallahassee for its gun control ordinances.¹⁴⁴ Although the state played an active role in the litigation as an *amicus*, the case was initiated by private actors exercising their rights under the legislation's private right of action.

Following Florida's lead, Mississippi enacted a firearm super preemption statute in 2016.¹⁴⁵ Under this law, "[N]o county or municipality may adopt any ordinance that restricts the possession, carrying, transportation, sale, transfer or ownership of firearms or ammunition or their components."¹⁴⁶ Similar to the statute in Florida, this law also created a private right of action, establishing that "a citizen of this state . . . who is adversely affected by an ordinance or posted written notice adopted by a county or municipality in violation of this section may file suit for declarative and injunctive relief against a county or municipality."¹⁴⁷ If the actions of local officials conflict with the statute, then the local officials may be civilly liable for up to \$1,000 as well as for the cost of the opposing party's attorney's fees.¹⁴⁸

By creating private rights of action, the Florida and Mississippi laws relieve their states of two responsibilities. First, under these statutes, the state does not have to bear the entire burden of identifying local violators. While some local violations are easily identifiable, many potential violations could go unnoticed by state officials.¹⁴⁹ Especially with regard to laws that are on the books but not currently enforced, private rights of action decrease the likelihood that violators will slip through the cracks. Second, under these statutes, the state does not have to bear the entire burden of litigation. Without a private right of action, state attorney general

¹⁴² See, e.g., Daniel P. Tokaji, *Public Rights and Private Rights of Action: The Enforcement of Federal Election Laws*, 44 IND. L. REV. 113, 160 n.1 (2010) (defining private right of action in the federal context as "a nongovernmental litigant's ability to bring suit to enforce a federal statute").

¹⁴³ *Fla. Carry*, 212 So. 3d at 455–56.

¹⁴⁴ *Id.*

¹⁴⁵ MISS. CODE ANN. § 45-9-51(1) (2017).

¹⁴⁶ *Id.*

¹⁴⁷ MISS. CODE ANN. § 45-9-53(5)(a) (2017).

¹⁴⁸ *Id.* at (5)(c).

¹⁴⁹ *C.f.* Trevor W. Morrison, *Private Attorneys General and The First Amendment*, 103 MICH. L. REV. 589, 608 (2005) (arguing that one of the benefits of private attorney general laws, which are laws that empower private actors to bring suits against those who violate public interests, is that they "valuably supplement the government's enforcement efforts without taxing state resources").

offices would have to litigate every case against violating localities. Often operating with limited resources, these offices likely would have to choose which cases to litigate and which to let go. With a private right of action, this decision becomes less daunting. Even if the state chooses to pass on a particular violation, there is still the possibility for a private actor, like a local advocacy organization, to play the role of attorney general and litigate the case.

2. Civil penalties and damages

Most super preemption statutes include some provisions for civil damages or penalties in the event that the locality is found to have violated the statute's terms. Some of these provisions take the form of civil *penalties* or *finer*, which suggests that the defendant would have to pay the fee whether or not the plaintiff proves monetary damages.¹⁵⁰ Other provisions are expressed as caps on civil *damages*, which suggests that payment would only occur after an assessment of the plaintiff's monetary damages due to the violation.¹⁵¹

One particularly noteworthy statute is Arizona's 2016 firearm provision.¹⁵² Similar to the laws in Florida and Mississippi, Arizona's law states, "(E)xcept for the legislature, this state and any agency or political subdivision of this state shall not enact or implement any law, rule or ordinance relating to the possession, transfer, or storage of firearms other than as provided in statute."¹⁵³ What is most curious about this provision is that it provides for both civil damages *and* penalties at varying amounts. First, the law establishes that the court may assess a civil *penalty* of up to \$50,000 when a political subdivision has knowingly and willfully violated this section.¹⁵⁴ Then, it states that if the plaintiff prevails under the private right of action, the court shall award "actual *damages* incurred not to exceed one hundred thousand dollars."¹⁵⁵ It is not clear how these two subsections are expected to operate or if the legislature's use of the words "penalty" and "damages" is purposeful or inartful. It is possible that the penalty provision is only meant to apply in cases where the state is the

¹⁵⁰ See, e.g., FLA. STAT. § 790.33(3)(c) (2017) ("[T]he court shall assess a civil fine of up to \$5,000 against the elected or appointed local government official or officials or administrative agency head under whose jurisdiction the violation occurred.").

¹⁵¹ See, e.g., NEV. REV. STAT. § 269.222(7)(c) (2015) (referring to "[l]iquidated damages in an amount equal to three times the actual damages").

¹⁵² ARIZ. REV. STAT. ANN. § 13-3108 (2017).

¹⁵³ ARIZ. REV. STAT. ANN. § 13-3118(A) (2017).

¹⁵⁴ ARIZ. REV. STAT. ANN. § 13-3108(I) (2017) (emphasis added).

¹⁵⁵ *Id.* at (K)(2) (emphasis added).

plaintiff. In those situations, it might be difficult to calculate how the state has been “damaged” by a local firearm ordinance. Therefore, the legislature may have decided it best to impose a penalty, which is easier to apply, because it does not require the court to determine the actual injury suffered by the plaintiff. On the other hand, this language may mean that, in cases where the plaintiff is a private party, both the penalty *and* the damages are applicable, which could potentially expose the city to \$150,000 of liability. What is clear is that both of these amounts are much larger than the civil fees in most other super preemption provisions.¹⁵⁶ This is likely due to the fact that the fees are attributable to the city itself and not an individual official.

Whether civil liability takes the form of damages or a penalty, the effect is generally the same. Local governments and local officials are rarely in a financial position where they can comfortably afford these awards. For that reason, individuals operating under these super preemption regimes will likely take extra care to ensure they do not run afoul of one of these provisions.¹⁵⁷

3. *Criminal liability*

At least two states have provided for criminal liability for officials who violate super preemption statutes. Kentucky passed a firearm statute similar in scope to many of the super preemption laws previously profiled in this Article.¹⁵⁸ However, in addition to creating the relatively common private right of action, this law took its punitive measures a step further, establishing that “a violation of the law’s provisions by a public servant constitutes a criminal infraction.”¹⁵⁹ These criminal provisions can result in up to a year of imprisonment for a local official found in violation of the preemption statute.¹⁶⁰

Following Kentucky’s lead, Texas recently passed anti-sanctuary city legislation that, in addition to including the traditional civil penalties common in super preemption provisions, also included criminal penalties

¹⁵⁶ Compare ARIZ. REV. STAT. ANN. § 13-3108(I) (2017) (establishing \$50,000 in civil penalties), *and id.* at (K)(2) (establishing \$100,000 in civil damages), with FLA. STAT. § 790.33(3)(c) (2017) (establishing only \$5,000 in civil fines), *and* MISS. CODE ANN. § 45-9-53(5)(c) (2017) (establishing only \$1,000 in civil damages).

¹⁵⁷ Local officials might be less risk averse if they knew that their city government would indemnify them for their damages or cover their legal fees. However, many of these super preemption laws prevent the use of public funds for this purpose. *See* Scharff, *supra* note 97, at 1501 (describing such a provision in the Florida firearm preemption statute).

¹⁵⁸ *See* KY. REV. STAT. ANN. § 65.870(6) (2017).

¹⁵⁹ KY. REV. STAT. ANN. § 65.870(6) (2017).

¹⁶⁰ KY. REV. STAT. ANN. § 532.090(1) (1975).

for local violators. The law forbids localities from adopting policies that would prevent law enforcement officers from complying with federal immigration detainer requests.¹⁶¹ This law was passed in response to many Texas cities (and cities across the country) refusing to comply with federal immigration detainer requests on the grounds that such federal mandates ran afoul of the Supreme Court's anti-commandeering doctrine.¹⁶² A critical feature of the anti-commandeering doctrine, however, is that it only protects cities from federal commandeering in their capacities as political subdivisions of *states*.¹⁶³ Through Texas's law, cities lose their ability to justify their sanctuary activities on anti-commandeering grounds because their actions now violate both federal *and state* policy. As a penalty for noncompliance, this law states that an official who "knowingly fails to comply with the detainer request" can be charged with a Class A misdemeanor resulting in up to \$4,000 in fines and one year in prison.¹⁶⁴

Criminal penalties are particularly powerful in that they carry a degree of moral opprobrium that civil damages lack. While the primary aim of civil proceedings is to make the wronged party whole again, the American criminal justice system has the added purpose of *punishing* the party that has violated some norm that our state holds dear. By attaching the "criminal" label to local officials in violation of these preemption statutes, the state is not only signaling that the official inflicted damage against the opposing party, but also that the official committed an offense that was morally reprehensible from the perspective of the polity.¹⁶⁵

¹⁶¹ Tex. S.B. 4, 85th Leg., R.S. (2017) (§ 1.01, adding §§ 752.051–752.057 to the Texas penal code).

¹⁶² See Ian Millhiser, *Breaking: Federal Judge Blocks Trump's Attack on 'Sanctuary Cities'*, THINKPROGRESS (April 25, 2017), <https://thinkprogress.org/jeff-sessions-amateurish-unconstitutional-assault-on-immigrants-dd6ab8a1671e/> ("Under the Supreme Court's 'anti-commandeering doctrine,' the feds cannot order a state or local government to participate in a federal program. Thus, while a state or municipality may voluntarily agree to have its police force participate in federal immigration enforcement, state and local governments also have an absolute right to refuse to do so.").

¹⁶³ See *Printz v. United States*, 521 U.S. 898, 935 (1997) ("The Federal Government may neither issue directives requiring the States to address particular problems, nor command the States' officers, or *those of their political subdivisions*, to administer or enforce a federal regulatory program.") (emphasis added); see also Defendants' Response to Applications for Preliminary Injunction at 16-17, *City of El Cenizo v. Texas*, 264 F. Supp. 3d 744 (W.D. Tex. 2017) (No. SA-17-CV-404-OLG) (arguing that while anti-commandeering doctrine restricts Congress's ability to direct state action, states do not have similar constraints on their ability to direct local action).

¹⁶⁴ Tex. S.B. 4, 85th Leg., R.S. (2017) (§ 5.02, adding a new § 39.07 to the Texas penal code).

¹⁶⁵ See generally, Paul D. Carrington, *The Moral Quality of the Criminal Law*, 54 NW. U. L. REV. 575 (1959) (discussing the role morality plays in our criminal justice system).

4. Payment of legal fees

An additional feature of super preemption provisions is that they often stipulate that local officials accused of violating the statutes cannot use city funds to pay their legal fees.¹⁶⁶ On its face, this policy might seem egalitarian. After all, if a local official did violate the statute, why should city taxpayers have to foot the bill for their legal expenses? However, this appraisal ignores how these provisions tilt lawsuits in the state's favor—irrespective of which party has the better legal argument. Many local officials lack the personal funds necessary to mount a successful defense against a deep-pocketed state. While a local official may believe that she committed no wrong, her personal financial situation might force her to settle with the state. Faced with the options of either settling the case and simply paying damages, or paying an expensive legal team to mount a defense that they still might lose, it is not hard to see why some local officials chose the former.

The *Florida Carry* case illustrates how these financial constraints can play out in practice. As previously discussed, two gun-rights organizations brought the lawsuit against multiple Tallahassee city commissioners.¹⁶⁷ The super preemption statute stipulated that local officials could not use public funds to pay for their legal defense.¹⁶⁸ Meanwhile, the defendants were able to secure the legal and financial support of over a dozen advocacy organizations,¹⁶⁹ which was fortunate given that the court ultimately decided that the city committed no wrongdoing.¹⁷⁰ Had the city defendants lost, they likely would have been liable for expensive legal fees. Without the help of pro-bono support, the Tallahassee defendants likely would not have been able to mount a legal defense and instead may have been compelled to settle with the plaintiffs. This shows just how powerful these legal fees provisions can be: without adequate representation for defendants, plaintiffs may be all but assured of receiving an outcome favorable to the state regardless of the case's strength.

5. Removal from office

A final common feature of many super preemption statutes are provisions that provide for the removal from office of local officials who

¹⁶⁶ See, e.g., MISS. CODE ANN. § 45-9-53(5)(c) (2017); FLA. STAT. § 790.33(3)(d) (2017).

¹⁶⁷ *Fla. Carry*, 212 So. 3d at 455–56.

¹⁶⁸ FLA. STAT. § 790.33(3)(d) (2017).

¹⁶⁹ Sean Rossman, *City Gets Support in Pro-Gun Lawsuit*, TALLAHASSEE DEMOCRAT (June 15, 2016), <https://www.tallahassee.com/story/news/2016/06/15/city-gets-support-pro-gun-lawsuit/85944658/>.

¹⁷⁰ *Fla. Carry*, 212 So. 3d at 465–66.

violate the law's terms. Both Florida and Arizona's firearm statutes include language that calls for the termination of employment or removal from office of any local official who passes a law in conflict with the state's gun policy.¹⁷¹ Similar to the civil penalties discussed previously, these termination provisions can have a chilling effect on local legislative activity if city officials are concerned that taking the wrong vote could cost them their jobs. These provisions also mirror some of the more retributive effects of the criminal penalties in the Kentucky and Texas laws in that they are solely concerned with punishing a recalcitrant local official. Finally, as this Article will argue, these removal provisions go beyond both civil and criminal penalties in one crucial way: they *permanently* end an individual official's ability to create policy change. While civil and criminal penalties may have a strong deterrent effect on the passage of future conflicting policies, the only way the state can ensure that a particular local official never again violates their preemption statute is to take away their lawmaking power entirely.

PART III – UNDERSTANDING SUPER PREEMPTION'S MODERN APPEAL

One clear takeaway after exploring the landscape of super preemption provisions is that states are taking unprecedented measures to thwart particular policies of their urban centers. Progressive local action on gun control and immigration has been met with strong pushback from conservative state legislatures, resulting in overturned local ordinances, contentious court battles, and the potential for damaging punitive measures leveled against local actors. What is less clear is why these provisions have proliferated so quickly, and what additional purpose they serve beyond traditional preemptive legislation. Assuming super preemption has always been a lawful mechanism for combating undesirable local policies, why have states only recently chosen to enact these types of policies? Similarly, if traditional preemption has historically been an effective mechanism for stopping local policies, are super preemption's punitive measures adding any value?

This Part offers two potential answers to these pressing questions. First, this Part argues that, in order to understand super preemption's recent rise, one must first recognize the way in which partisan differences have hardened along the urban-rural divide. Today, perhaps more than at any other time in America's history, political ideology correlates almost

¹⁷¹ ARIZ. REV. STAT. ANN. § 13-3108(J) (2017); FLA. STAT. § 790.33(3)(e) (2017).

perfectly with a person's proximity to the urban core.¹⁷² Whereas historically urban and rural residents may have found common cause over politics,¹⁷³ increasingly the policy preferences of urban residents are diametrically opposed to those of their rural neighbors. This fact, coupled with the dominance of rural legislators in state politics,¹⁷⁴ helps to explain why state legislatures are striking down politically charged local policies with unprecedented impunity. Second, this Part argues that, while traditional justifications for super preemption fail to explain the purpose behind these punitive policies, by taking a more expansive view of local politics, one can begin to see that these measures serve very real ends. Using Gerken's three-part explanation for the value of local, minority decisionmaking,¹⁷⁵ this Article contends that, although traditional preemption has been effective at stopping expressions of local self-governance, state legislatures use super preemption to combat the two other goals of local policymaking: contributing to the marketplace of ideas and allowing minority communities the opportunity to develop their political identities.

A. The Increasing Political Importance of the Urban-Rural Divide

In *Federalist Number 10*, James Madison warned against the dangers of factionalism in America's fledgling republic.¹⁷⁶ According to him, although factions—particularly local factions—were an unavoidable reality in democratic governance, factionalism's more corrosive effects could be dulled by extending the republic's geographic sphere.¹⁷⁷ With a large enough polity, no one faction could obtain and hold on to power. Instead, factions would rise and fall over time as the polity's size and diversity caused political coalitions to shift gradually.¹⁷⁸ While a particular group's interests might align on one issue, that faction would almost certainly break apart in future political battles when its members found cause to

¹⁷² Cf. Stahl, *supra* note 12, at 136–39 (describing this phenomenon as it has manifested in North Carolina's urban centers).

¹⁷³ See *id.* at 149 (“In the past, when both parties had rural and urban voters, partisanship eased the tension between them by uniting them against a common enemy—the other party.”).

¹⁷⁴ See *id.* at 136–143 (describing the way rural, legislators have used gerrymandering tactics to create Republican majorities in red states countrywide).

¹⁷⁵ Gerken, *supra* note 26.

¹⁷⁶ See Gerald E. Frug, *The City as a Legal Concept*, 93 HARV. L. REV. 1057, 1106, 1127 (1980) (discussing how Madison's fear of factionalism presaged modern American suspicion of institutions that exist between the state and individuals).

¹⁷⁷ See THE FEDERALIST NO. 10, at 64 (James Madison) (Jacob E. Cooke ed., 1961).

¹⁷⁸ See ROBERT A. DAHL, A PREFACE TO DEMOCRATIC THEORY 30 (1956) (describing how the instability of democratic majorities protects minority groups from political exploitation).

partner with other diverse interests on some other issue.¹⁷⁹ For Madison, this theory of factionalism helped justify the move toward a larger, more centralized government.¹⁸⁰ By expanding the geographic boundaries of the government's constituency, Madison hoped to thwart the entrenchment of local factions that he believed poisonous to a well-functioning republic.

For much of American history Madison's solution to factionalism has appeared effective.¹⁸¹ By funneling our political activity through two national parties, geographic difference could only gain so much political traction.¹⁸² For a party to find political success, it would have to strive to appeal to northern and southern, eastern and western, urban and rural constituencies. This political necessity for the most part ensured that no party could completely adopt one locality's provincialism. America's large national stage also helped ensure that geographic coalitions shifted from time to time. Citizens saw that, while they may be on the losing side during one political battle, their enemies could become their allies during the next fight, preventing the formation of sectional "cleavages" along consistent geographic, racial, or ideological lines.¹⁸³ With the notable exception of the violent battle between the north and south over slavery, American politics never truly metastasized along geographic lines.¹⁸⁴ For years urban Democrats in the north occupied the same party as rural Democrats from southern states.¹⁸⁵ That kind of geographic diversity in our political parties has become increasingly rare.¹⁸⁶

In the current political environment, politics and geography are becoming increasingly intertwined. These political cleavages have not formed along northern and southern, or eastern and western divides as they might have in the past, but along urban and rural lines. As evidence of this

¹⁷⁹ See ROBERT A. DAHL, *DEMOCRACY IN THE UNITED STATES* 279 (4th ed. 1981).

¹⁸⁰ See THE FEDERALIST NO. 10, at 64 (James Madison) (Jacob E. Cooke ed., 1961) ("Extend the sphere, and you take in a greater variety of parties and interests; you make it less probable that a majority of the whole will have a common motive to invade the rights of other citizens. . . .").

¹⁸¹ See Stahl, *supra* note 12, at 148 (noting that for much of American history this system worked "reasonably well," with the notable exception of the Civil War).

¹⁸² See *id.* ("Our modern two-party system, for example, has tended to give our political system a remarkable degree of stability by ensuring that political differences are channeled through the two major national parties."); see also Yascha Mounk, *The Rise of McPolitics*, THE NEW YORKER (July 2, 2018), <https://www.newyorker.com/magazine/2018/07/02/the-rise-of-mcpolitics> (describing how, historically, the two party system "yoked" "socially progressive Democrats in the North" to "segregationist Democrats in the South").

¹⁸³ Cf. DAHL, *supra* note 178 (warning that "[i]f all the cleavages occur along the same lines, if the same people hold opposing positions in one dispute after another, then the severity of conflicts is likely to increase").

¹⁸⁴ See Stahl, *supra* note 12, at 148.

¹⁸⁵ Mounk, *supra* note 182.

¹⁸⁶ See *id.*

fact one need not look any further than America's recent presidential elections. In 2012, President Obama won fewer counties nationwide than any Democratic candidate in recent memory.¹⁸⁷ And yet, Obama was reelected by a healthy margin over opponent Mitt Romney due in large part to his garnering of 69 percent of the votes in cities with over half a million people.¹⁸⁸ Hillary Clinton, the Democratic presidential candidate in 2016, was able to improve on that number, winning 71 percent of the vote in those metro areas.¹⁸⁹

Multiple explanations abound for the stark political cleavage along urban-rural lines. One explanation, attributed to journalist Bill Bishop, is that this geographic divide is the result of a decades-long geographic reorganization that he calls "the big sort."¹⁹⁰ According to Bishop, America's partisan differences have gradually bled outside the boundaries of the political arena, coming to characterize the near entirety of personal identities.¹⁹¹ Increasingly, how Americans see themselves politically has become synonymous with how they see themselves culturally, socially, racially, religiously, sexually, and, often, economically.¹⁹² Historically, Democrats and Republicans attracted supporters of different races, religions, and ideologies.¹⁹³ Today, however, both parties have become more homogenous in these regards, with Republicans increasingly becoming the party of white, evangelical, conservatives, and Democrats becoming the party of everyone else.¹⁹⁴ This alignment of the various facets of personal identities along political lines has become so strong that individuals no longer want to live next to neighbors of opposing political stripes.¹⁹⁵ If being a Republican suggests that a person is not simply opposed to a set of policies that Democrats support, but to the very identity of Democrats as individuals, why would that person want to live next to someone of the opposite party? Bishop argues that this trend has resulted in a "post-materialist" geographic reorganization whereby people no longer

¹⁸⁷ *Id.* at 142.

¹⁸⁸ *Id.*

¹⁸⁹ *Id.*

¹⁹⁰ BILL BISHOP, *THE BIG SORT* (2008).

¹⁹¹ *Id.* at 8–9.

¹⁹² For a detailed analysis of how our politics have gradually come to align across racial, religious, and economic lines, *see generally* LILLIANA MASON, *UNCIVIL AGREEMENT: HOW POLITICS BECAME OUR IDENTITY* (2018).

¹⁹³ *See id.*; *see also* Mounk, *supra* note 182.

¹⁹⁴ *Id.* ("Party affiliation is influenced more by factors like race and religion than by local interests or political traditions.")

¹⁹⁵ BISHOP, *supra* note 190.

choose where they live purely based on economic considerations, but instead on lifestyle choices that closely mirror political divides.¹⁹⁶

Unlike Bishop, who sees our current geopolitical divide as the result of *voluntary* sorting, other scholars point to the current “stickiness” of residential patterns as another explanation for why our politics have hardened along urban/rural lines.¹⁹⁷ These theorists posit that, due to various land use, housing, and occupational licensing policies, disadvantaged demographics have become increasingly unable to access high-opportunity locales.¹⁹⁸ Whether it be the rural high school student stuck in a struggling town because her family cannot afford the booming metro center’s artificially high housing prices,¹⁹⁹ or the low-income minority individual stuck in a disadvantaged urban neighborhood because of the surrounding suburb’s exclusionary zoning practices, for many Americans, where they live is a product of the legal forces that keep them stuck in a particular place. In this way, it is less that our politics determines our geography as Bishop would contend, but that our geography, and all of its attendant economic realities, determines our politics.

A third, albeit related, theory points to the way globalization has caused the economic fortunes of our cities and their surrounding rural areas to drastically diverge. According to Professor Kenneth Stahl, “globalization has created a huge geographic imbalance in economic fortunes as capital investment is increasingly directed towards urban centers and away from rural areas.”²⁰⁰ In the past, urban and rural areas had a symbiotic relationship inside small, self-contained regional economies. The city depended on the surrounding rural areas for agricultural production, while the rural areas relied on their cities as markets where rural residents could sell their goods.²⁰¹ In this way the economic fortunes of these two

¹⁹⁶ *Id.*

¹⁹⁷ See Ross Douhat, *We Should Treat Big Cities Like Big Corporations and Bust Them Up*, DALLAS NEWS (Mar. 28, 2016), <https://www.dallasnews.com/opinion/commentary/2017/03/28/treat-big-cities-like-big-corporations-bust> (arguing that while our urban centers may be economically successful, their economic benefits have not been equally accessible to disadvantaged demographics due to their high costs of living and highly segregated residential patterns).

¹⁹⁸ See David Schleicher, *Stuck! The Law and Economics of Residential Stagnation*, 127 YALE L.J. 78, 78 (2017).

¹⁹⁹ See Scott Beyer, *The Verdict Is In: Land Use Regulations Increase Housing Costs*, FORBES (Sept. 30, 2016), <https://www.forbes.com/sites/scottbeyer/2016/09/30/the-verdict-is-in-land-use-regulations-increase-housing-costs/#287858c4162a> (arguing that restrictive zoning regulations decrease the supply of available housing in high demand cities, thereby artificially raising the prices of the existing housing stock).

²⁰⁰ Stahl, *supra* note 12, at 150.

²⁰¹ Cf. Cecilia Tacoli, *Rural-Urban Interdependence*, in ACHIEVING URBAN FOOD AND NUTRITION SECURITY IN THE DEVELOPING WORLD (2000) (describing the importance of the surrounding rural areas to urban centers in the developing world).

geographies were linked: if the city failed, the surrounding rural areas failed, and if the city succeeded, the surrounding areas also, likely, succeeded. Today that economic link between cities and their rural neighbors has been severed. As America moves away from agriculture and manufacturing and toward a knowledge-based economy, cities become less reliant on the surrounding land for their economic success.²⁰² Moreover, as trade barriers fall, immigration policies become more liberal, and mechanized agriculture becomes the norm, America's rural citizens are seeing their economic fortunes decline as a result of policies often championed by urban residents.²⁰³

Stahl notes that this divergence of economic fortunes has turned urban-rural politics into a zero-sum game.²⁰⁴ Whereas in the past, it may have harmed rural residents to resist policies supported by their urban neighbors, today rural denizens likely will not experience serious repercussions for taking that political stance.²⁰⁵ For example, pro-immigrant policies like sanctuary city provisions directly benefit urban areas because they increase cities' abilities to access both the high-skill and low-skill workers that their economies require to operate.²⁰⁶ Conversely, those same policies have the potential to undercut the economic opportunities of residents living in surrounding rural areas who may have to compete with low-skill immigrant workers for the shrinking pool of agricultural jobs in their communities.²⁰⁷ For this reason rural residents may be more inclined to support anti-urban preemption measures than they would have in years past.

No single theory provides a full explanation for the convergence of political identification and geography. Instead, each of these three theories (voluntary sorting, geographic "stickiness," and globalization—along with numerous others) probably play some role in America's growing

²⁰² Stahl, *supra* note 12, at 152.

²⁰³ *Id.*

²⁰⁴ *Id.*

²⁰⁵ *Id.* at 154–55 (noting that Republican state legislators “have little disincentive to take actions that harm cities because, in today's global economy, cities are already so completely disconnected from rural areas that an urban economic downturn is unlikely to have ripple effects on the places Republicans care about”).

²⁰⁶ *Cf. id.* at 153 (describing the aim of sanctuary city policies as signaling “friendliness to immigrants”).

²⁰⁷ See Katherine Fennelly, *Why Immigration Worries Americans—Especially Rural Residents*, SCHOLAR STRATEGY NETWORK (Feb. 2012), <http://www.scholarsstrategynetwork.org/brief/why-immigration-worries-americans-%E2%80%93-especially-rural-residents> (“In rural focus groups, hostility toward immigrants and the belief that there are too many in the community was strongest among low-income, white residents who worry that they face competition for jobs and believe that foreign residents have access to undeserved benefits.”).

geopolitical divide. But the sway that geography holds over modern American politics can only partially explain the rise of super preemption. After all, the fact that urban and rural residents hold different political beliefs does not necessarily lead to a world where rural interests dominate state legislatures.²⁰⁸ Fortunately, political science may have an answer to the question of what fuels the success of rural conservatives in American state legislatures: gerrymandering.

Gerrymandering is the practice by which state legislative leaders draw legislative districts to advantage one political party over the other.²⁰⁹ In order to understand its political power, one need not look any further than the state that provided the backdrop to the *Florida Carry* saga. Florida's status as a perennial swing state needs little explanation. The state is almost evenly divided between Republicans and Democrats.²¹⁰ In 2008, Obama won the state by less than three percentage points.²¹¹ In 2012, he won it by less than one.²¹² And in 2016, Trump won it by less than two.²¹³ In the 2012 presidential race, no state in the nation produced a closer electoral result than Florida.²¹⁴ One would therefore be forgiven for believing that the state's political parity in presidential elections must carry over to state elections. But instead of a near-even split between Republicans and Democrats in Tallahassee, Florida's legislative chambers skew overwhelmingly Republican, with conservatives holding virtual supermajorities in both houses.²¹⁵

What causes this glaring partisan disparity? Over the past two decades, Republican leaders have engineered a legislative map that almost perfectly

²⁰⁸ See Stahl, note 12, at 136–43.

²⁰⁹ See Christopher Ingraham, *This is the Best Explanation of Gerrymandering You Will Ever See*, WASH. POST (Mar. 1, 2015), <https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see> (defining gerrymandering as “drawing political boundaries to give your party a numeric advantage over an opposing party”).

²¹⁰ Ledyard King, *Democrats Seize Advantage Registering Florida Voters*, USA TODAY (Oct. 10, 2016), <https://www.usatoday.com/story/news/politics/elections/2016/10/10/democrats-seize-advantage-florida-registering-new-voters/91865978/> (noting that, as of 2016, Florida had 4.6 million registered Republicans and 4.7 million registered Democrats).

²¹¹ *Election Results 2008: Florida*, N.Y. TIMES (Dec. 9, 2008), <https://www.nytimes.com/elections/2008/results/states/florida.html>.

²¹² *Election 2012: Florida*, N.Y. TIMES, <https://www.nytimes.com/elections/2012/results/states/florida.html> (last visited Dec. 12, 2017).

²¹³ NEW YORK TIMES, *Florida Presidential Race Results: Donald J. Trump Wins*, N.Y. TIMES (Aug 1, 2017), <https://www.nytimes.com/elections/results/florida-president-clinton-trump>.

²¹⁴ David Catanese, *The 10 Closest States in the 2016 Election*, U.S. NEWS (Nov. 14, 2016, 4:39 PM), <https://www.usnews.com/news/the-run-2016/articles/2016-11-14/the-10-closest-states-in-the-2016-election>.

²¹⁵ Dan Sweeney, *GOP Easily Keeps Control of Florida Legislature*, SUN SENTINEL (Nov. 8, 2016, 10:54 PM), <http://www.sun-sentinel.com/news/politics/fl-election-florida-legislature-20161108-story.html>.

maximizes their electoral advantage across the state.²¹⁶ That Republican leadership have been able to do this with such ease is due in part to the very political sorting that this Article describes. Democratic voters have gradually coalesced inside a handful of Florida's urban areas.²¹⁷ Although these urban communities do not perfectly align with the state's legislative districts, those tasked with drawing legislative maps have packed these urban voters inside a small number of left-leaning urban districts, while generously dispersing rural Republican voters across the majority of the remaining districts. This process, known as "vote wasting," creates a handful of solidly-blue urban districts that may vote 80 or 90 percent Democratic, as well as a sizeable number of rural districts that will reliably vote for Republicans, but only at a rate of 55 or 60 percent.²¹⁸ Under a fairer map, those excess Democratic urban voters would have been more evenly distributed throughout the surrounding rural and suburban districts. Because Republicans control the mapmaking process, they are able to draw districts that both advantage their party *and* almost perfectly mirror the rural-urban divide.

None of this is to dispute the notion that there may be benefits to the compact urban districts drawn in states like Florida. Indeed, compact and homogenous districts may actually lead to better political representation for multiple reasons. First, elected officials in these districts likely do not have to travel great distances to meet with their constituents, which means more time listening to constituent concerns and less time on the road. Additionally, elected officials in these districts are more likely to reflect the demographic make-up of their district. If a district is drawn to include mostly members of one race, one religion, or constituents from one city or one neighborhood, it is quite likely that district's representative will share those demographic traits and therefore be more acutely attuned to the needs of those groups. However, the fact that state legislatures' current gerrymandering practices may come with some benefits does not diminish the fact that these practices have helped harden the differences between

²¹⁶ Cf. Stephen Wolf, *One Key Reason Why Republicans Dominate in State Legislatures: They Drew Most of the Districts*, DAILY KOS (Nov. 26, 2016, 1:47 PM), <https://www.dailykos.com/stories/2016/11/26/1599841/-One-key-reason-why-Republicans-so-dominant-in-state-legislatures-They-drew-most-of-the-districts> (attributing the Florida Republican Party's legislative success to their ability to control the districting process).

²¹⁷ Matthew Isbell, *The Disappearing Swing Counties of Florida*, MCI MAPS (June 8, 2015), <http://mcimaps.com/the-disappearing-swing-counties-of-florida/> (describing the increasing urbanization of Florida's Democratic voters).

²¹⁸ See Stahl, *supra* note 12, at 167.

America's rural and urban communities along political lines and helped fuel the modern rise in preemptive activity.

This method of gerrymandering occurs in states across the country.²¹⁹ Facilitated by geopolitical distribution, Republicans from Arizona to North Carolina have been able to draw districts that reliably elect a majority of conservative legislators representing rural interests, and a minority of liberal legislators representing urban communities. This modern political trend helps explain super preemption's rise and the uptick in preemptive activity more generally. Unlike in decades past, where state legislative officials may have depended on both urban and rural voters for support, today, legislators rely on geographically- (and politically-) homogenous constituencies for electoral success. Given the already divergent economic fortunes of urban and rural communities, state legislatures' recent willingness to strike down policies that benefit urban constituencies should come as little surprise.

B. Minority Dissent and the Purpose Behind Super Preemption

While modern geopolitical trends help explain the timing of super preemption, they shed little light on its purpose. Indeed, the hardening of America's urban-rural divides should only suggest an uptick in preemptive activity generally, not the creation of a *new* mechanism for preempting local policies. And yet, a new mechanism has been created. This rapid proliferation of super preemption laws indicates that there must be something attractive about this tactic beyond what state legislatures have already achieved through traditional preemption legislation. But what is it? Why have state legislatures specifically chosen to enact these untested punitive measures instead of relying on the traditional instruments in their preemption toolkits?

Before offering an answer to that question, it is important to walk briefly through why it is that super preemption cannot stand solely on traditional justifications for preemptive activity. As stated previously, three of the most common justifications that courts and state officials offer for promoting preemptive action have been the desire for state uniformity, the state's interest in curbing extraterritoriality, and the benefits of preserving

²¹⁹ Christopher Ingraham, *This is Actually What America Would Look Like Without Gerrymandering*, WASH. POST (Jan. 13, 2016), <https://www.washingtonpost.com/news/wonk/wp/2016/01/13/this-is-actually-what-america-would-look-like-without-gerrymandering> (illustrating the ubiquity of gerrymandering in the United States).

state and local core competencies.²²⁰ Although each of these justifications is helpful for describing the purpose behind traditional preemption,²²¹ they fail to justify the punitive measures at play in super preemption. Take, for example, the arguments for preserving state uniformity. Courts traditionally argue that forcing businesses and individuals to navigate a patchwork of regulatory provisions as they move from one municipality to the next can be cumbersome.²²² Therefore, in order to promote economic efficiency and transparency in the law it is often beneficial to have a *single* set of laws on a topic as opposed to having regulations promulgated by every one of a state's subsidiary governments.²²³ Traditional preemption promotes this end by providing state legislatures with a mechanism to strike down laws that deviate from the state's overarching regulatory scheme. But once the law is no longer operative, no further uniformity goals are advanced by punishing the locality or local official who voted for the city's ordinance. Businesses do not have an easier time navigating the state's regulatory landscape because a city official paid them civil damages. The harm in that scenario—too many business regulations—has been rectified via the traditional preemptive measure. Therefore, adding super preemption's punitive measures seems gratuitous. If it is not serving the state's underlying preemptive goal, why do it?

One argument is that super preemption is necessary because traditional preemption is actually ineffective at achieving its stated goals. While the state may attempt to strike down extraterritorial municipal laws through traditional preemption bills, localities are not obeying the state's directives and instead continue to enforce their local regulations. Though plausible, this response is unsatisfying as a justification for super preemption. There is little evidence that localities openly flout preemptive measures in

²²⁰ See, e.g., *City and County of Denver*, 788 P.2d at 768 (noting that the three factors the court considers when assessing whether a policy falls within the state's ambit include "need for statewide uniformity," "impact of the municipal regulation on persons living outside the municipal limits," and "whether a particular matter is one traditionally governed by state or by local government").

²²¹ This Article, however, does not address the merits of these justifications from a normative perspective.

²²² See, e.g., *N. Calif. Psychiatric Soc'y*, 178 Cal. App. 3d at 101 ("Certain areas of human behavior command statewide uniformity, especially the regulation of statewide commercial activities.").

²²³ See, e.g., *Am. Fin. Serv. Ass'n*, 104 P.3d at 823 ("Moreover, it is beyond peradventure that effective regulation of mortgage lending, and in particular here abusive practices in such lending, 'requires uniform treatment throughout the state.'" (quoting *Chavez v. Sargent*, 339 P.2d 801, 810 (Cal. 1959))).

violation of their states' directives.²²⁴ In the *Florida Carry* case, Tallahassee ceased enforcing its gun control ordinances years before the lawsuit commenced.²²⁵ The city did not need a punitive measure to compel compliance; simply knowing that their law conflicted with the state's policy was incentive enough. Moreover, even in scenarios where cities continue enforcing preempted local laws, states have the ability to sue to compel compliance. Assuming a court finds that the locality's laws have, in fact, been preempted, a judge can strike down the ordinance and threaten to hold local officials in contempt of court should they continue their violation.²²⁶ In this way, super preemption's punitive measures at best serve as a legislative proxy for the judiciary's powers of contempt. While that may animate some of the attraction to these policies, it seems too weak a justification to warrant super preemption's increasing popularity.

If traditional justifications for preemption do not explain super preemption's role, what can? One example that may help illustrate super preemption's subtle power is the story of San Francisco's 2004 decision to issue marriage licenses for same-sex couples. Between February and March of 2004, San Francisco Mayor Gavin Newsom issued approximately 4,000 marriage licenses to same-sex couples.²²⁷ Within two weeks after Newsom issued his first license, California Governor Arnold Schwarzenegger and Attorney General Bill Lockyer filed petitions with the California Supreme Court, requesting a declaration that the Mayor's policy was unlawful.²²⁸ Six months later, the court did just that.²²⁹ Declaring that the Mayor's policy had been preempted by state law, the court ordered Newsom to end the unlawful practice and voided all licenses issued under the mayor's same-sex directive.²³⁰

²²⁴ See Scharff, *supra* note 97, at 1506 ("Lawmakers also have not put forward significant evidence of local governments undermining state laws in ways that traditional preemption doctrine cannot address.").

²²⁵ See *Fla. Carry*, 212 So. 3d at 456.

²²⁶ Scharff, *supra* note 97, at 1506 (noting that it is "possible to hold local officials in contempt for refusing to follow a court order"). Although this was not a preemption case, the Second Circuit took this very approach when the City of Yonkers, New York, refused to adopt the Court's required desegregation plan. See James Feron, *First Contempt Fine Is Paid by Yonkers over Housing Plan*, N.Y. TIMES (Aug. 4, 1988), <http://www.nytimes.com/1988/08/04/nyregion/first-contempt-fine-is-paid-by-yonkers-over-housing-plan.html>.

²²⁷ See Dean E. Murphy, *California Supreme Court Voids Gay Marriages in San Francisco*, N.Y. TIMES (Aug. 12, 2004), <http://www.nytimes.com/2004/08/12/national/california-supreme-court-voids-gay-marriages-in-san-francisco.html>.

²²⁸ See Rusty Dornin & David Mattingly, *Same-Sex Marriage Breaks for Weekend*, CNN (Feb. 21, 2004, 6:43 PM), <http://www.cnn.com/2004/LAW/02/21/same.sex/>.

²²⁹ See Murphy, *supra* note 227.

²³⁰ *Id.*

On paper, San Francisco's same-sex marriage story reads like a traditional case of successful state preemption. The city tried to pass a policy out of step with the state's laws, the state sued arguing that the law was preempted, the court agreed and overturned the local measure, and the city complied. Same-sex marriage licenses would not be issued again in San Francisco for another four years,²³¹ and the Supreme Court would not permanently legalize them for an additional five.²³² And yet, marriage equality advocates often tout San Francisco's month-long policy as a political success, citing the way same-sex marriage laws spread in the years after Newsom's directive.²³³ How can one reconcile these two competing depictions? On one hand, it was a local policy that was overturned and voided only six months after it went into effect. On the other, it was a political act that helped precipitate national change. Understanding how these two portrayals can describe the same policy will help elucidate the power of local policymaking as well as the purpose behind super preemptive measures.

In her 2005 article *Dissenting by Deciding*, Professor Heather Gerken argues that while traditional forms of dissent (e.g. civil disobedience, casting a dissenting vote, drafting a dissenting opinion, etc.) receive outsized attention, one often-overlooked strategy is for minority communities to express dissenting views through local policy enactments.²³⁴ According to Gerken, cities, states, juries, and courts give national minority groups the opportunity to "wield the authority of the state" by implementing their policy preferences through real laws with real implications.²³⁵ Gerken describes this strategy as "acting radically," and contrasts it with more traditional forms of dissent where minorities "speak radically" (e.g. protest), or "act moderately" (e.g. bargaining for concessions with a minority vote).²³⁶

Acting radically provides minorities multiple advantages that traditional forms of dissent lack. First, acting radically allows minorities the opportunity to inject their views into the national marketplace of policy

²³¹ See Jesse McKinley, *Same-Sex Marriages Begin in California*, N.Y. TIMES (June 17, 2008), <http://www.nytimes.com/2008/06/17/us/17weddings.html>.

²³² See Bill Mears, *Supreme Court Dismisses California's Proposition 8 Appeal*, CNN (June 27, 2013, 8:27 AM), <http://www.cnn.com/2013/06/26/politics/scotus-prop-8/index.html>.

²³³ See Carla Marinucci, *Will Gavin Newsom's Post-Pride Glow Get Him the Governorship?*, S.F. CHRON. (July 5, 2015, 9:12 PM), <http://www.sfchronicle.com/politics/article/Will-Gavin-Newsom-s-post-Pride-glow-get-him-the-6367587.php> (quoting activists who credit Newsom's policy with being "the catalyst" for the events leading up to the Supreme Court's decision in *Obergefell v. Hodges*).

²³⁴ See Gerken, *supra* note 26, at 1746–47.

²³⁵ *Id.* at 1747.

²³⁶ *Id.* at 1746–47.

ideas in a way that traditional dissent cannot.²³⁷ Second, acting radically gives local minority communities the chance to “express and define” their *community’s* identity as opposed to the identity of a single dissenter.²³⁸ Finally, acting radically grants minorities the opportunity to take part in the practice of self-governance, thereby forging valuable civic ties that will serve them well in future political endeavors.²³⁹ To better understand how these three advantages work in practice, it may be helpful to view them through the context of the San Francisco same-sex marriage license fight.

First, Mayor Newsom’s directive illustrates how “acting radically” allows minority groups to engage with the marketplace of ideas more effectively than they could through traditional forms of dissent. Under traditional dissent, an outlier view expressed to the public may never warrant a response or may be dismissed outright as unworkable.²⁴⁰ Dissenting by deciding, however, engages with the marketplace of ideas in a manner that is much harder to ignore.²⁴¹ Mayor Newsom’s policy did not just indicate to the country that cities could impact the theoretical debate over the definition of marriage, it illustrated that same-sex marriage was a viable option, in practice. In the wake of the Mayor’s decision, several other cities across the country followed suit, emboldened by the real example of an action that they may have never thought was possible.²⁴²

Moreover, Newsom’s decision forced the majority to respond in a way that traditional dissent often does not. If Governor Schwarzenegger, Attorney General Lockyer, and the members of the California Supreme Court disagreed with the Mayor’s decision, they could not simply let minority policy die through inattention. Majority leaders had to spend time and political capital to defeat the policy, making public their competing vision of marriage in the state and hoping it held up under public scrutiny. This response was particularly difficult for Democratic Attorney General Lockyer, who recognized that, by opposing the Mayor’s policy, he may alienate a sizeable portion of his political supporters. In siding with the

²³⁷ *See id.* at 1749.

²³⁸ *Id.* at 1750.

²³⁹ *See id.*

²⁴⁰ *See id.* at 1761–62.

²⁴¹ *See id.* at 1762.

²⁴² *See e.g., San Jose Recognizes Gay Marriage*, CHI. TRIB. (Mar. 10, 2004), http://articles.chicagotribune.com/2004-03-10/news/0403100284_1_gay-marriage-marriage-licenses-same-sex (noting that both San Jose, California, and Asbury Park, New Jersey, began issuing same-sex marriage licenses after Newsom’s directive).

Governor, Lockyer nevertheless made a point to declare his support for same-sex policies like domestic partnerships and civil unions.²⁴³

This response would not have been necessary had Newsom simply written an op-ed or worked with his local legislative delegation to file a bill in the state legislature. His decision to “act radically” injected a policy into the marketplace of ideas in a way that both demanded a reaction from political opponents and provided cover for other cities to follow suit. His decision even compelled the California Supreme Court to rethink the issue, eventually leading to the court’s 2008 decision in *In re Marriage Cases* holding restrictions on same-sex marriage unconstitutional.²⁴⁴ This illustrates the power of decisional dissent to affect real change outside the bounds of its immediate political jurisdiction.

Additionally, the Mayor’s decision provided gay rights activists with an unprecedented opportunity for community building and identity formation. After the Mayor’s decision, leaders on the left, prominent members of the gay community, and supporters of marriage equality engaged in a heated public debate about the appropriateness of the Mayor’s actions.²⁴⁵ Massachusetts Congressman Barney Frank, one of the most prominent gay elected officials at the time, famously criticized the Mayor’s decision as an “illegitimate act” that undermined the rule of law.²⁴⁶ Democrat and California Senator Diane Feinstein agreed, contending that the Mayor’s actions were “too much, too fast, too soon” when questioned regarding her opinion on the matter.²⁴⁷ Conversely, gay activist and San Francisco Assemblyman Mark Leno rallied to the Mayor’s defense.²⁴⁸ Previously content with simply getting a win on domestic partnerships, Leno stated that seeing couples become “spouses for life” changed his position on the matter.²⁴⁹ Similarly, Matt Foreman, the head of the National Gay and Lesbian Task Force, admitted that when he first heard about the Mayor’s decision he was “skeptical.” But he went on to say that “the minute those pictures came out, waiting in line, going in, and getting married, it put a human face on this issue.”²⁵⁰

²⁴³ See Dornin & Mattingly, *supra* note 228.

²⁴⁴ See *In re Marriage Cases*, 183 P.3d 384, 453 (Cal. 2008).

²⁴⁵ See Joe Dignan, *Way Out Front*, GAY CITY NEWS (Mar. 3 2004), https://gaycitynews.nyc/gcn_309/wayoutfront.html; Joe Dignan, *Too Much, Too Fast, Too Soon*, GAY CITY NEWS (Nov. 17 2004), https://gaycitynews.nyc/gcn_346/toomuchtoofast.html.

²⁴⁶ Dignan, *Too Much, Too Fast, Too Soon*, *supra* note 245.

²⁴⁷ *Id.*

²⁴⁸ Dignan, *Way Out Front*, *supra* note 245.

²⁴⁹ *Id.*

²⁵⁰ *Id.*

The Mayor's same-sex marriage decision forced a much-needed *internal* debate within the ranks of gay rights activists. What were the community's goals? How would they communicate those goals to the rest of the country? What methods would they use to achieve their ends? Who would lead the community and speak on its behalf? The way that the community answered those questions would have ramifications for the marriage equality movement moving forward. According to Professor Gerken, this internal debate is a common feature of local dissent, allowing "[o]pportunities for group members to hash out the connection between group and civic identity."²⁵¹ Had the Mayor simply stated his opinions in a speech, other members of his community could have dismissed the Mayor's words as representing his views and his alone. However, by expressing this minority opinion through the instruments of state authority, Newsom precipitated a conversation amongst his allies about how best to advance the goals of their movement.

The final advantage of dissent through local decisionmaking is that minority groups are able to engage in the act of self-government without having to compromise their views. For the month that Newsom's policy was in effect, San Francisco's gay community was able to do just that: govern themselves under the policies that they preferred. But that act of self-governance was cut short by the court's decision holding San Francisco's policy preempted by state law. In this way, traditional preemption was able to completely undermine one leg of Gerken's three-part framework for local decisionmaking: when the state successfully preempted the local action, the policy no longer carried the force of law and the act of self-governance ended.

Notably, while the court's preemption decision undermined San Francisco's act of self-governance, it had almost no impact on the other two advantages of local decisionmaking. The Mayor's decision still diversified the marketplace of ideas in a powerful way, prompting copycat legislation,²⁵² action from the Governor,²⁵³ and comments from the President.²⁵⁴ Additionally, the decision influenced the internal strategy of the gay rights community for years to come as more activists and

²⁵¹ Gerken, *supra* note 26, at 1796.

²⁵² See Anne Hull, *Just Married, After 51 Years Together*, WASHINGTON POST (Feb. 29, 2004) https://www.washingtonpost.com/archive/politics/2004/02/29/just-married-after-51-years-together/94c88fff-045d-46e3-b31c-04ab658608c6/?utm_term=.384dd748b131 (describing how city officials in New Mexico, New York, and Chicago followed Newsom's lead).

²⁵³ See Dornin & Mattingly, *supra* note 228.

²⁵⁴ See Hull, *supra* note 252 (noting that "President Bush called for a constitutional amendment banning gay marriage" in the wake of the San Francisco directive).

policymakers gradually came to follow Newsom's lead on the issue.²⁵⁵ Calls for domestic partnerships and civil unions waned as activists instead pushed for full marriage equality under the simple but powerful justification that "love is love."²⁵⁶

In effect, the Mayor's decision was able to help spark a societal movement without permanently changing local law. But would that have been the case if California's marriage provision had instead been a super preemption law? Engaging in this thought experiment helps illustrate the important ways super preemption differs from traditional preemption as a strategic tool. As a preliminary matter, had California's marriage law resembled modern super preemption provisions, it is highly unlikely that Newsom would have ever issued his marriage license directive in the first place. If conflicting with the state's law could have resulted in civil damages, criminal penalties, or termination from office, Newsom probably would have been deterred from taking such bold, dissenting action and instead opted for political self-preservation.

Moreover, even if Newsom had chosen to enact his policy, its lifespan would have likely been cut short if the Mayor had to pay for his legal costs out of pocket. The moment the Mayor was served process in a suit initiated by the state or a well-funded non-profit, he may have had to settle as opposed to engage in a potentially expensive legal battle where he stood a real risk of losing. This initial deterrence would have been particularly beneficial for the Governor and Attorney General. By forcing the Mayor to concede before his policy got off the ground, these statewide leaders would have been able to avoid spending valuable political capital in a fight as contentious as the one over same-sex marriage. Finally, if the Mayor had chosen to enact his policy and fight the legal battle to its end, he would have lost. This loss would not have simply ended the same-sex marriage policy in San Francisco, it likely would have resulted in the Mayor's removal from office as well as the removal of any local official who helped advance the policy. This would have had the effect of crippling the ranks of local leadership in San Francisco and stunting the City's emerging gay rights movement.

This hypothetical illustrates super preemption's strategic superiority over traditional preemptive measures. For a state legislator interested in suppressing local movements, traditional preemption is an unsatisfying

²⁵⁵ *See id.*

²⁵⁶ Ben Jacobs, 'Love is Love': Obama Lauds Gay Marriage Activists in Hailing 'a Victory for America,' *GUARDIAN* (June 26, 2015), <https://www.theguardian.com/us-news/2015/jun/26/obama-gay-marriage-speech-victory-for-america>.

tool in that it only prevents the act of local self-governance (i.e. enacting policy). It does little to stop a locally-initiated policy from entering the marketplace of ideas and from influencing the way we conceptualize legislative options. Additionally, traditional preemption cannot stop the process of political mobilization that precedes legislative enactment. That process is critical for shaping a minority group's political and civic identity. It forces that group to grapple with internal disagreements and allows them to forge a cogent, battle-tested voice that they can use in future contests. Super preemption has a more expansive reach than traditional preemption in that it is able to address all aspects of local decisionmaking. By creating a system whereby local leaders will have to bear personal liability for initiating counter-majoritarian legislation, super preemption stands to stop both the political movement and the policy itself.

CONCLUSION

Although the primary purpose of this Article is to provide a descriptive account of this recent legislative trend, it is important to address just how troubling super preemption is from a normative perspective. Cities provide an unrivaled forum for democratic empowerment and community building. Indeed, early political theorists like Alexis de Tocqueville and John Stuart Mill described localities as schools for democratic empowerment, teaching citizens the skills they need to become active and responsible stewards of their democratic polity.²⁵⁷ More recently, Professor Gerald Frug opined on the many civic advantages of decentralized government, including the ability to actively participate in the policy decisions that affect one's surroundings, the ability to experiment in solving local problems, and, perhaps most importantly, the "energy derived from democratic forms of organization."²⁵⁸ According to Frug, these advantages, when employed correctly, allow us to abandon the idea that government is centered on the individual, and instead embrace a decentralized conception of government that places the public as its primary subject.²⁵⁹

But the advantages of local policymaking do not stem simply from the mechanics of small government. Local government without the ability to

²⁵⁷ See DE TOCQUEVILLE, *supra* note 30, at 60 ("Municipal institutions are to liberty what primary schools are to science; they bring it within the people's reach, they teach men how to use and how to enjoy it."); JOHN STUART MILL, *CONSIDERATIONS ON REPRESENTATIVE GOVERNMENT* (1890), *reprinted in* ON LIBERTY AND OTHER ESSAYS 205, 413 (John Gray ed., 1991) (discussing the way local government can serve as political education for citizens in a democracy).

²⁵⁸ GERALD FRUG, *CITY MAKING: BUILDING CITIES WITHOUT WALLS* 10 (1999).

²⁵⁹ *Id.* at 9–10.

affect real change is not democracy. And yet, the proliferation of super preemption provisions inches America toward that world. By greatly disincentivizing local action on matters that cut against the states' wishes, super preemption chills the kind of democratic energy de Tocqueville, Mills, and Frug celebrate. If local officials are no longer willing to take actions that are out of step with the politics of the state as a whole, citizens—especially minority citizens—will gradually come to see local government as a forum unable to address their needs.

If super preemption stands to have such a damaging effect on local empowerment, what recourse do cities and their local officials have to push back? While the law on super preemption is still evolving, several promising legal tactics have emerged as potential defenses against these punitive measures. One such defense looks to the source of the locality's home rule power as a possible shield against state legislative meddling. As discussed previously, depending on their language, constitutional home rule provisions are sometimes interpreted as affording cities a protected sphere of legislative immunity over local affairs.²⁶⁰ Just how far that sphere reaches depends on a variety of factors, including the whims and caprices of whatever judge is assigned to decide the matter.

However, successful home rule defenses are possible. Although this was not a super preemption case, recently in *City of Cleveland v. State of Ohio*, an Ohio trial court struck down a state statute preempting Cleveland's residential employment requirement for city-funded projects.²⁶¹ According to the Court, the state could only exercise its preemptive powers through "general laws" that regulate statewide conduct—not simply through laws that limit local authority.²⁶² While this case will likely find new life on appeal, it is immediately important in that it interprets an Ohio home rule provision that does not, on its face, afford localities more protection than many of the states profiled in this Article.²⁶³ Indeed, a recent decision out of a trial court in Florida seemed to follow a similar line of thought, striking down a Florida statute that prevented the

²⁶⁰ See *supra* text accompanying notes 41–49.

²⁶¹ *City of Cleveland v. State of Ohio*, No. CV-16-868008, 2016 WL 9414089 (Ohio Ct. Com. Pl. Aug. 30, 2016) (judgment entry, with opinion and order granting permanent injunction).

²⁶² *Id.* at 4–5.

²⁶³ Compare, e.g., OHIO CONST. art. XVIII, § 3 ("Municipalities shall have authority to exercise all powers of local self-government and to adopt and enforce within their limits such local police, sanitary and other similar regulations, as are not in conflict with general laws.") with FLA. CONST. art. VIII, § 2(b) ("Municipalities shall have governmental, corporate and proprietary powers to enable them to conduct municipal government, perform municipal functions and render municipal services, and may exercise . . . power for municipal purposes except as otherwise provided by law.").

City of Coral Gables from regulating Styrofoam containers on the grounds that it violated the state's constitutional home rule amendment for Miami-Dade County.²⁶⁴

A second potential strategy for local officials interested in pushing back against super preemption provisions is to appeal to legislative immunity. As stated previously, legislators, even at the local level, have traditionally been afforded a degree of legislative immunity for work performed in their elected capacity. This means that officials cannot be held personally liable for the governmental decisions they make while in office. In *Bogan v. Scott-Harris*, the Supreme Court made clear that common law principles of legislative immunity extended to local officials, noting that “voting for an ordinance” or “signing into law an ordinance” are legislative acts that are afforded legal protection.²⁶⁵ Although this case specifically dealt with legislative immunity as it related to federal statute 42 U.S.C. § 1983,²⁶⁶ a judge may be willing to import a similar standard to state super preemption cases.

Finally, local officials might consider arguing that super preemption provisions violate their constitutional rights to free speech under the First Amendment to the United States Constitution. Under this theory, local legislators would contend that taking a political vote as a government officer is no different than expressing a political opinion through some other forum. If the latter is protected by First Amendment doctrine, it makes little sense why the former would be exposed to reprisal through super preemption provisions. Unfortunately, the Supreme Court held in *Nevada Commission on Ethics v. Carrigan*, that restrictions on a local legislator's votes are not restrictions on their speech as it is understood by the First Amendment.²⁶⁷ While this case did not address the kinds of punitive provisions that have come to characterize super preemption laws,²⁶⁸ it does at least indicate an initial unwillingness to afford legislative votes the protections of political speech. Therefore, in order to succeed in a First Amendment defense, local officials will have to draw a distinction between some of the severely punitive measures included in super preemption provisions (i.e. heavy fines, criminal penalties, termination of employment) and the relatively mild punishment at issue in *Carrigan*

²⁶⁴ See Fla. Retail Fed'n, Inc. v. City of Coral Gables, No. 2016-018370-CA-01 (Fla. Cir. Ct. Mar. 9, 2017).

²⁶⁵ *Bogan v. Scott-Harris*, 523 U.S. 44, 55 (1998).

²⁶⁶ *Id.* at 47.

²⁶⁷ *Nev. Comm'n on Ethics v. Carrigan*, 564 U.S. 117, 125 (2011).

²⁶⁸ The case instead addressed a decision by the state ethics commission to censure a local commissioner. *Id.* at 117.

(legislative censure). Although the case law on this matter is still in its most nascent stages, a recent decision on Texas's sanctuary cities provision (SB 4) indicates that judges may be suspicious of the way these rather draconian measures curb political expression.²⁶⁹

Given super preemption's relative newness, it is unclear if any of these defenses would convince a judge to strike down these punitive measures. But it is imperative that local officials try. Super preemption has the ability to chill the kind of local political activity that has come to characterize our cities as "laboratories of democracy." Indeed, by targeting not only the policies enacted by our localities, but also the politics surrounding those enactments, the punitive provisions in super preemption laws can ground local political movements before they begin. Given that local government is one of the rare forums where political minorities can create real change, this chilling effect runs the added risk of further alienating an already-ostracized demographic. As America's residential patterns continue to break along partisan lines, the incentive for Republican state legislators to attack local policymaking will only increase. It is therefore critical that judges and local officials find ways to prevent some of the worst effects of this troubling practice.

²⁶⁹ See *City of El Cenizo v. Texas*, 264 F. Supp. 3d 744 (W.D. Tex. 2017) (order granting preliminary injunction), *aff'd in part, vacated in part sub nom.* *City of El Cenizo v. Texas*, 885 F.3d 332 (5th Cir. 2018).

IN WHAT SENSE A COUP? A REVIEW OF *THE FRAMERS' COUP: THE MAKING OF THE UNITED STATES CONSTITUTION*
BY MICHAEL J. KLARMAN
George Rutherglen ♦

In his magisterial history of the fraught and compromised origins of the Constitution, Professor Klarman explores in absorbing detail all the dimensions of the tumultuous events that led to the drafting, ratification, and amendment of the Constitution by the Bill of Rights. The book aspires, in the author's words, to put the entire history of these developments "between two covers."¹ This vast panorama of issues, actors, and historical context unfolds before the reader as each of the crises of the years from 1787 to 1791 comes into focus. To forestall any sense that the result was foreordained, Klarman regularly reports on the anxieties of the Framers, particularly James Madison, who worried that the whole effort would collapse, followed shortly thereafter by the descent of the country into anarchy and civil war, and then partition of the country among foreign powers and a return to monarchy. Madison was not alone in these apprehensions, as Federalists and Antifederalists alike recognized the weaknesses of the general government under the Articles of Confederation.²

Klarman invites the reader to relive the making of the Constitution in all its contingencies, some predictable and perennial, like the opposition between large and small states, some nearly forgotten, like disputes over navigation of the Mississippi River. The latter arose from southern suspicions that a commercial treaty with Spain would bargain away rights of settlers beyond the Appalachians to navigate on the Mississippi River.³ Klarman gives the reader the good, the bad, and the ugly: the good in the farsighted vision of the Framers; the bad in their tolerance of slavery and general distrust of democratic government; the ugly in the often cynical processes of ordinary politics that led to ratification of the Constitution. If

♦ George Rutherglen joined Virginia's law faculty in 1976. He teaches admiralty, civil procedure, employment discrimination and professional responsibility. He is grateful to have spent so many years with Prof. Klarman as a colleague and as a scholar, from whom he has learned so much, exactly in proportion as he has initially disagreed with him.

¹ MICHAEL J. KLARMAN, *THE FRAMERS' COUP: THE MAKING OF THE UNITED STATES CONSTITUTION* x (2016).

² *Id.* at 69–72.

³ *Id.* at 48–69.

you want a history unencumbered by hagiography of the Framers—and by the same token, of their Antifederalist opponents—this is the book for you. And it should be the book for everyone who wants a comprehensive and unvarnished look at the framing of the Constitution.

If Klarman does not come to praise the Framers, neither does he come to bury them. Like most historians, he accords Madison a preeminent place in the debates over the Constitution. Madison emerges from his account with his reputation intact as the genius behind the Constitution, although something of an evil genius in his adamant opposition to popular government. Klarman also acknowledges that Madison might have skewed the historical record in his favor by his prolific correspondence and his role as the principal chronicler of the otherwise secret deliberations of the Constitutional Convention.⁴ Still, Klarman observes that at the outset of the convention, its agenda “had pretty much existed only in Madison’s head,”⁵ and that throughout the process of framing and adopting the Constitution, Madison “played a critical role at almost every stage of this process.”⁶

Klarman’s most serious qualms about the Framers begin with the words that appear at the very beginning of the Preamble: “We the People.” He finds the Framers’ appeal to the people to be instrumental, if not entirely opportunistic. They had little hope of getting the state legislatures to ratify the Constitution since the increased powers of the national government came mainly at the expense of the states and state legislatures. That left ratifying conventions as the most likely means of securing approval of the Constitution, and even that course, as events bore out, proved to be a very close call. On a higher level of principle, republican political theory at the time held the people to be the ultimate source of political power and the only one sufficient to create national law that would be supreme over state law.⁷

But even though the Framers accepted these propositions in the abstract, their deliberations and proposals were rife with distrust of “the People.” Edmund Randolph of Virginia made a typical remark: “Our chief danger arises from the democratic parts of our constitution.”⁸ Klarman recounts in great detail the many provisions of the Constitution designed to insulate government from the people, foremost among them the allocation of seats

⁴ *Id.* at 135 n.*.

⁵ *Id.* at 536.

⁶ *Id.* at 596.

⁷ *Id.* at 312–13, 415–16, 603.

⁸ *Id.* at 209 (quoting James McHenry, Report (May 29, 1787), in 1 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, at 26 (Max Farrand ed., rev. ed. 1966)).

in the Senate, the indirect election of senators, and their comparative lengthy term of office. Randolph again observed that “[t]he democratic licentiousness of the state legislatures proved the necessity of a firm Senate.”⁹ The same checks against democracy were necessary to protect the interests of the property holding classes, from which virtually all the Framers were drawn. These elites had become alarmed by the issuance of paper money in the 1780’s and by Shays’s Rebellion.¹⁰ The prohibitions in Article I, Section 10, against impairing the obligation of contract, issuing bills of credit, and making paper money a form of legal tender were all directed at these populist measures by the states. This critical theme in Klarman’s book has led some reviewers to see it as an updated and sophisticated version of Charles Beard’s “An Economic Interpretation of the Constitution of the United States.”¹¹ He duly attends to this charge in a careful account of his differences with Beard.¹²

We can take him at his word and ask a different question. Just as Klarman has doubts about the Preamble, we can have doubts about his title. In particular, in what sense was the adoption of the Constitution a “coup”? Klarman apparently takes the primary sense of the term to be a “coup d’état”: a sudden forcible overthrow of the government and a seizure of power by a small group of conspirators.¹³ But the broader meaning of the term, taken from the original French, is a sudden, successful stroke.¹⁴ This sense does not exclude the other one, but it opens up the possibility of a more charitable interpretation of what the Framers accomplished. A coup in this sense could be a stroke of genius, in addition to simply a coup d’état. The Framers could both have displaced the system of government established by the Articles of Confederation and have done so through inspired political thought and action. Klarman might well resist this second sense of this term, since it could lead to the hagiography of the Framers that he sets himself against.

Yet neither sense of the term appears to be quite right. The entire series of events that led from the initial call for a constitutional convention to the

⁹ *Id.*

¹⁰ *Id.* at 88–101.

¹¹ Published in 1913. See Matthew C. Simpson, *The Founding Fathers’ Power Grab: Was the Constitution Designed to Make the United States Less Democratic?*, THE NEW REPUBLIC (Sept. 29, 2016), <https://newrepublic.com/article/137310/founding-fathers-power-grab>.

¹² KLARMAN, *supra* note 1, at 376–77.

¹³ “[A] sudden and decisive stroke of state policy; *spec.* a sudden and great change in the government carried out violently or illegally by the ruling power.” *Oxford English Dictionary Online Edition*. 2018. <http://www.oed.com>. (last visited July 12, 2018).

¹⁴ “[A] blow, stroke.” *Id.*

ratification of the Bill of Rights was drawn out over several years, with periodic crises that could have brought the process to an abrupt and untimely end. It was not as if the Framers enlisted military force to surround the Continental Congress and then seize power from it. If the Framers accomplished a coup in any sense, they did it in slow motion.

By referring to a “coup,” Klarman seems to be making a point different from the direct implications of the term: a point about the antidemocratic procedures that the Framers followed and about the antidemocratic proposals they adopted. The procedures they followed were disingenuous and hypocritical, with an appeal to “the People” at the same time as they were excluding ordinary people from the secret deliberations of the convention and manipulating them in the ratification process. The substance of the Constitution pushed this strategy forward to the operation of the federal government as a whole, in provisions like the composition of the Senate and the Electoral College. This review explores both of these features of “The Framers’ Coup” and then draws out the implications that a coup in the sense of an antidemocratic exercise of power has for current interpretation of the Constitution.

I. HOSTILITY TO DEMOCRACY

The Constitutional Convention originated in a call to revise, not replace, the Articles of Confederation. At the Convention, the Framers almost immediately abandoned this modest objective and transformed their commission into one of replacing the Articles in their entirety. The Framers, already drawn from the elite classes, worked to prevent anyone else from knowing of their deliberations. After the Convention proposed the Constitution, the Framers exploited the prevailing consensus that the Articles suffered from severe defects to force an all-or-nothing choice upon the ratifying conventions. Madison, ever present at critical moments, deftly removed from the Bill of Rights any structural amendments that would have weakened the powers of the federal government or restructure it in a more populist direction.¹⁵ In a telling detail, Klarman notes that Madison removed the word “expressly” from the Tenth Amendment, making clear that Congress enjoyed implied powers under the Constitution that it did not have under the Articles.¹⁶

¹⁵ KLARMAN, *supra* note 1, at 592.

¹⁶ *Id.* at 579.

The Framers plainly failed to accept, and in fact, rejected principles of democratic transparency. But we have to ask how, if at all, that compromised the legitimacy of their actions. Their procedures did, of course, depart from those for amending the Articles, which required unanimous approval of the states.¹⁷ Article VII of the Constitution required the ratification of only nine states for the Constitution to go into effect, although those nine could only bind themselves. That put great pressure on the few remaining holdouts to ratify or to be excluded from the union. Rhode Island was the most notorious holdout, refusing to ratify the Constitution until 1790. But the New York and North Carolina ratifying conventions were still in session after the Constitution had gone into effect, after ratification by New Hampshire and Virginia as the ninth and tenth states. Dispensing with the Articles' unanimity requirement gave extra force to the all-or-nothing choice that the Framers presented to the ratifying conventions. Those that ratified last, after the Constitution went into effect among the ratifying states, faced a hostile federal government with the resources to coerce ratification, for instance, by threatening a trade war, as the Antifederalists in Rhode Island soon discovered.¹⁸

The Framers' strategy dispensed with any pretense that the Constitution followed in a chain of authority from the Articles. Although Antifederalists soon abandoned arguments that the Constitution was an unauthorized replacement for the Articles, they objected strongly on this ground while the acceptance of the Constitution still remained in doubt.¹⁹ What weight did those objections have? Klarman seems to come around, as the Antifederalists did, to something like the modern view that replacement of one legal system by another does not require pre-existing authority. In the words of H.L.A. Hart, "all that succeeds is success."²⁰ In Hart's version of legal positivism, success in changing the foundations of a legal system requires acceptance by the officials of the system and obedience by the general population to the rules then promulgated by the officials.²¹ The elitism of the Framers, on this theory, did not detract from the authority of what they accomplished but assured it by obtaining the assent of a significant proportion of the legal elite. Objections to the legitimacy of the Constitution faded away as Antifederalists accepted it as the basis for their

¹⁷ ARTICLES OF CONFEDERATION OF 1781, art. XIII.

¹⁸ KLARMAN, *supra* note 1, at 516–30.

¹⁹ *Id.* at 619–22.

²⁰ H.L.A. HART, *THE CONCEPT OF LAW* 116–17, 153 (3d ed. 2012).

²¹ *Id.* at 116–17.

own exercise of federal power in the early decades of the nineteenth century.

No fundamental change in a legal system can be expected to take place strictly according to the terms for amendments under the status quo ante. The Framers' rejection of the Articles of Confederation and their terms for amendment cannot simply be attributed to their antidemocratic tendencies. The veto that any state could exercise over amendments under the Articles, and particularly the stubborn independence of Rhode Island, effectively precluded that avenue of constitutional change. To be sure, Rhode Island was also a hotbed of populist agitation for paper money and debtor relief, but that hardly made the procedure for amendment under the Articles a model of democratic government. Twice under the Articles, a single state—Rhode Island and then New York—blocked ratification of an amendment approved by all the other states.

If entrenchment is the measure of departures from democracy, the Articles were effectively more entrenched and less democratic than all but one of the provisions in the Constitution—equal representation of the states in the Senate.²² That provision can be abrogated only with the consent of the state “deprived of its equal Suffrage in the Senate.”²³ Klarman regards this provision as one of the great antidemocratic defects of the Constitution, along with the related provision for representation of states in the Electoral College.²⁴ Even so, the Constitution does not take entrenchment nearly as far as the Articles, which gave all states exactly one vote in the Continental Congress and entrenched that provision, and all the others, by requiring unanimous consent of the states to any amendment.²⁵ And, again, Klarman accepts the consensus of historians that the Constitution could not have been approved by the Convention, let alone ratified, without the Connecticut Compromise on representation of the states in the Senate and the House.²⁶ He finds the roots of this compromise in the equal representation that the small states received under the Articles, which then became the model for voting at the Convention and for the

²² This does not count the entrenchment of the provision on importing slaves, but it expired by its own terms in 1808. U.S. CONST. art. V.

²³ *Id.*

²⁴ *Id.* at art. II, § 1, ¶ 2. This provision is not explicitly protected from amendment without state consent, but a critical minority of 13 states that benefit from this provision could block any amendment to this effect. And they would, of course, have no incentive to support such an amendment.

²⁵ ARTICLES OF CONFEDERATION OF 1781, arts. V, XIII.

²⁶ KLARMAN, *supra* note 1, at 200–01 & n.*.

terms of ratifying the Constitution itself.²⁷ Once ingrained in the national government, equal representation of the states could not be displaced.

By the same token, the baseline for judging the Framers' procedures to be antidemocratic cannot be set by the Articles or by the existence of some feasible, more democratic, alternative available to the Framers at the time. Klarman ultimately appeals to our contemporary views, finding the principle of "one person, one vote" to be basic to our sense of democracy.²⁸ To his credit, he does not attempt to impose a "presentist" conception of political theory on the Framers, retrospectively holding them to our views centuries later. In his description of what the Framers did in their time, he remains agnostic about how it should influence what we should do in ours. Most readers would probably agree with equal representation in voting and apportionment as an axiom of acceptable democratic procedures today, assimilating "one person, one vote" to the rejection of the racist and sexist views of the Federalists—and no doubt Antifederalists as well. To this we might add the modern acceptance of paper money and the need for emergency debtor relief.²⁹

Klarman might well be right about this, but the minimalist strategy that he adopts of accepting only the least controversial principles of current democratic theory creates problems of its own. If, for example, a new constitutional convention were convened today, no one could safely predict exactly what it would propose and whether its proposals would be ratified by the necessary three quarters of the states.³⁰ That was, as Klarman astutely points out, exactly the objection of the Federalists to a second constitutional convention and to conditional ratification of the Constitution.³¹ They worried that, absent unconditional acceptance of the Constitution, all bets would be off. And today, just as much as in the Founding era, we could not expect the delegates to a constitutional convention to put themselves behind some kind of "veil of ignorance" that hid their own interests from themselves. They would not place themselves in an "original position" in which they adopted only the widely shared principles of an overlapping consensus.³² They would be subject to all the

²⁷ *Id.* at 202.

²⁸ *Id.* at 625–26.

²⁹ The first was rejected in the Legal Tender Cases, *Knox v. Lee*, 79 U.S. 457 (1871) and *Juilliard v. Greenman*, 110 U.S. 421 (1884), and the second during the Depression in *Home Bldg. & Loan Ass'n. v. Blaisdell*, 290 U.S. 398 (1934).

³⁰ U.S. CONST. art. V.

³¹ KLARMAN, *supra* note 1, at 536–39.

³² JOHN RAWLS, *A THEORY OF JUSTICE* 15–19 (rev. ed. 1990).

forms of ordinary politics, conflicting interests, and devious strategies that Klarman documents so well.

II. THE JEFFERSONIAN PRINCIPLE

Klarman does not so much deny this indeterminacy as embrace it, following a famous passage in a letter by Thomas Jefferson on revision of the Virginia Constitution: “Each generation is as independent as the one preceding it, as that was of all which had gone before. It has then, like them, a right to choose for itself the form of government it believes most promotive of its own happiness”³³ This principle makes constitutions good for this day and generation only. It is basically inconsistent with constitutionalism as an established structure for governance. But again, the problems run deeper than escaping the dead hand of the past. No one would deny, and probably could not consistently deny, the virtues of stability in the basic structure of government. Even Jefferson supposed that generational revision of a constitution would amount to “wisely yielding to the gradual change of circumstances, of favoring progressive accommodation to progressive improvement.”³⁴

Moreover, as Jefferson also recognized, determining what a majority of the current generation wants is itself a fundamental choice: “But how collect their voice? This is the real difficulty.”³⁵ He proposed a fixed hierarchy of wards, counties, and general government, presumably extending to the federal level.³⁶ To neglect this issue would be inconsistent with democratic government itself, which presupposes a structure of rules that determine who is eligible to vote, how voters and representatives are apportioned, and what the procedures are for legislation and governance. No system of representative democracy can get off the ground without some provisions like those to be found in Article I of the Constitution specifying how a bill becomes law.³⁷

Such provisions have their own entrenching effect by conferring advantages and disadvantages on different groups, and by inviting the groups initially with power to gain more. There is no getting away from the normative presuppositions of any form of government, democracy

³³ Letter to Samuel Kercheval (June 12, 1816), in THOMAS JEFFERSON, WRITINGS 1402 (Library of America 1984).

³⁴ *Id.* at 1401.

³⁵ *Id.* at 1402.

³⁶ *Id.* at 1399–1403.

³⁷ U.S. CONST. art I, §§ 1–9.

included, and the entrenching effects that selecting any one structure of government inevitably has. Klarman's vivid illustration of this point is in the far-reaching effects of equal state representation under the Articles of Confederation, which led directly to the Senate and the Electoral College we now have.

The Jeffersonian principle that each generation has the right to decide its own form of government does not dispense with these normative and consequentialist questions. At best, it evades them. It could be taken simply as a truism: that every generation could possibly exercise the power to overthrow the existing form of government. This formulation leans very heavily on the difference between a conceptual possibility—if the people decided to overthrow the government, it would be gone—and a live possibility—the people actually have a realistic opportunity to make this decision. In the latter form, it cannot generally be true. Only a few generations in our history have been faced with this fateful choice.

But in either form, the principle does not reach the further question whether the people should exercise the right to change their government. That is the real question that Jefferson evaded. The cost of constitutional change might not be worth the benefits, either because the process itself would lead to civil strife or because the resulting government would be worse than what it replaced. The proposition that this gamble should be taken every twenty years or so is not, to use Jefferson's own language in the Declaration of Independence, "self-evident."³⁸ It becomes so only on optimistic assumptions about the likely costs and benefits of the process. Imponderables abound at every turn.

Hard as it is to extract an "ought" from an "is,"—or for historians, what we should do now from what was done then—Klarman's account could yield either the Jeffersonian conclusion or exactly the opposite. The Federalists themselves thought they were fantastically lucky to have secured adoption of the Constitution, some of them attributing their success to divine intervention.³⁹ When the country revisited the constitutional principles adopted in the Founding Era, the result was the Civil War, which killed more Americans than all the other wars in our nation's history combined, and which could at several points have turned into a victory for the southern secession.⁴⁰ Could the enormous costs of that conflict, not to mention the continuation of slavery for the better part

³⁸ THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

³⁹ KLARMAN, *supra* note 1, at 539–40.

⁴⁰ JAMES M. MCPHERSON, *BATTLE CRY OF FREEDOM: THE CIVIL WAR ERA 854, 857–58* (1988).

of a century, have been forestalled through a more democratic process of regularly amending the Constitution? It is hard to know how even to begin to answer this question.

Contingency on a smaller scale and with less violent immediate consequences can be found, as Klarman emphasizes, throughout the process of framing and ratifying the Constitution. It all would have failed if any of several conditions had not been met: if George Washington had not attended the Constitutional Convention and lent his enormous prestige to it; if the Connecticut Compromise between large states and small states had not been reached; if the Framers had not accepted the continuation of slavery; if the ratifying conventions in Virginia and New York had come out differently, which they probably would have had they been held a few months earlier; and if James Madison had not changed his mind about the Bill of Rights, been elected to Congress, pushed it through to ratification, and in the process omitted most of the structural provisions in the amendments favored by the Antifederalists. Klarman's book has the great virtue of not making any of this look easy, even when we know what happened in the end.

His narrative raises the question—without offering an answer to it—of why we should go along with Jefferson in urging that our nation go through similar crises regularly every few decades. Jefferson's advice is all the more puzzling coming from someone who did not participate in the Constitutional Convention because he was the ambassador to France, where he witnessed the first stages of the French Revolution.⁴¹ Perhaps this is not a recommendation, but a factual claim: that fundamental constitutional change regularly recurs with all its attendant costs. Any such claim of cyclical recurrence would, of course, have to extend beyond the single example of the U.S. Constitution and would raise interpretive questions of its own in assessing what counts as fundamental change and what its costs are. Yet another alternative, one that aligns with Klarman's distaste for making the Framers the prophets of constitutional law as our secular religion, makes Jefferson's principle into an interpretive warning: Do not give too much weight to the Constitution out of reverence for the Framers. The next section takes up this perspective on the lessons of the Founding era.

⁴¹ That did not stop him, however, from recalling in his memoirs that it all would have come out differently if the French had followed his advice. Thomas Jefferson, *Autobiography*, in THOMAS JEFFERSON, *WRITINGS* 85 (Library of America 1984).

III. IMPLICATIONS FOR INTERPRETATION

Klarman's catalogue of the antidemocratic provisions in the Constitution poses especially severe problems for originalist interpretation of the Constitution. Do originalists have to accept the Federalists' antipathy to democracy, at least as it has survived in unamended provisions of the Constitution? These extend beyond the composition of the Senate and the Electoral College to the several restrictions on state power in Article I, Section 10. More telling is the disdain frequently expressed by the Framers for the participation of ordinary people in government and their attempt to set up the federal government as one composed mostly of the elite. Should these antidemocratic tendencies become a touchstone for resolving uncertain issues of constitutional interpretation? Ironically for originalists, who espouse a conventionally conservative agenda, such distrust of the people would reinvigorate the case for judicial review to enforce individual rights. The familiar and accepted provisions for tenure of federal judges during good behavior, the supremacy of federal law, and the individual rights enumerated in the first eight amendments to the Constitution would all be given added force as the expression of the Framers' overriding intent to constrain majoritarian government.

Klarman obviously is no originalist, since he concludes his book with the admonition that "those who wish to sanctify the Constitution are often using it to defend some particular interest that, in their own day, cannot in fact be adequately justified on its own merits."⁴² Few constitutional scholars, however, would willingly concede that they have sanctified the Constitution, but would instead disclaim, with Noah Webster, any intent to opine that the Framers could "judge for future generations better than they can judge for themselves."⁴³ Yet conceding the fallibility of the Framers leaves open the question about the force of the Constitution itself. Absent an unlikely amendment to Article VI, the Constitution remains "the supreme Law of the Land." It therefore continues to constrain the ordinary processes of lawmaking, both in the states and in the federal government. We might not like how the Framers reached their compromises over the Constitution, or the compromises themselves, but that does not settle the question whether we should continue to be bound by them.

⁴² KLARMAN, *supra* note 1, at 631.

⁴³ *Id.* at 628.

Klarman takes a surprisingly meliorist view of how to reconcile the continued force of the Constitution with democratic principles.⁴⁴ Amendments have taken the edge off many of the antidemocratic provisions of the Constitution, notably the Thirteenth Amendment, which abolished slavery and therefore rendered the Three-Fifths and Fugitive Slave Clauses inoperative; the Seventeenth Amendment, which required direct election of Senators; and the Fifteenth, Nineteenth, Twenty-Fourth, and Twenty-Sixth Amendments, which extended the franchise in a variety of ways. Changes in state law also extended the franchise so that it now approaches universal suffrage. Judicial decisions, beginning with the Warren Court, brought the principles of equality in the Fourteenth Amendment into effective operation, so that the Framers' views on race, sex, and apportionment of the right to vote have been discarded. Even the expansive interpretation of the Necessary and Proper Clause, which Klarman finds to have been a Federalist maneuver to enhance federal power, still had its good side in reconciling the country to the need for expanded regulation of an expanded economy. No doubt the same could be said of the New Deal decisions that further increased federal power over the economy. Only the composition of the Senate and the Electoral College remain immune to these developments.

Klarman endorses a strategy of minimal entrenchment, which finds the political commitments of the Framers to be more tolerable as they are more easily subject to revision.⁴⁵ This leads to a backhanded defense of progressive judicial activism as one way to counteract entrenchment. It might be an antidemocratic form of government but one better than being ruled by the dead hand of the past. This strategy, of course, applies to all decisions of the Supreme Court, regardless of how progressive they actually are. The Jeffersonian principle, by contrast, tries to preserve the power of each generation to choose by democratic means. In the process, it ends up diminishing both the disappointments and achievements of the Constitution, whose force would be diminished close to that of a statute.

The Constitution, on this view, would resemble another landmark of that era, the Judiciary Act of 1789,⁴⁶ which has been subject to far more amendments than the Constitution. It also has received far less of the adulation directed at the Constitution, although it is still celebrated as the

⁴⁴ *Id.* at 622–28.

⁴⁵ He has elsewhere defended this view in greater detail. Michael J. Klarman, *What's So Great About Constitutionalism?*, in *PRINCETON READINGS IN AMERICAN POLITICS* 81 (Richard M. Valelly, ed., 2009).

⁴⁶ Judiciary Act of 1789, 1 Stat. 73 (1789).

foundation of the federal judicial system.⁴⁷ Yet this example remains instructive because one of the central antipopulist features of the federal judicial system has remained a constant since 1789. It is diversity jurisdiction,⁴⁸ originally made available to the federal courts to protect creditors, particularly British creditors, from unfair treatment in state courts.⁴⁹ Diversity jurisdiction must be conferred by Congress and it can equally well be repealed by Congress. Yet it has become a fixture of the federal judicial system. Such is the force of tradition and established practice. If that holds for what some regard as an esoteric feature of federal jurisdiction, why shouldn't it hold more broadly? The cost of change might well outweigh any advantage gained from making the change.

Of course, that is an empirical question on which we have scant evidence, even with respect to particular issues such as the representation of the states in the Senate, and still less so with respect to the Constitution as a whole. Under a very simple model of assessing constitutional change by reference to the preference of the median voter, a number of variables complicate the analysis immediately: how preferences of the electorate are distributed around the median; whether the costs of transition are fixed or depend upon the magnitude of change from the status quo; and how much control key officials have in setting the agenda and how they exercise that power.⁵⁰ Further complexities ensue, of course, if constraints on democratic majoritarianism, such as protection of individual or minority rights, generate independent arguments for entrenchment. A constitution that entrenches least does not necessarily result in a regime that governs best.

In fact, on the plausible assumption that any change generates transition costs, the minimal level of acceptable entrenchment might be quite high. We can only tell as a matter of investigation, experiment, and historical experience. That lesson diminishes the significance of the Framers' antidemocratic tendencies. It could be that, despite their elitist motives, they hit upon about the right degree of entrenchment in our Constitution. Looking backwards to the origins of the Constitution does not tell as much

⁴⁷ An enduring pattern in the judicial history of the United States has been "the relative stability of the structure of courts established by the First Judiciary Act—at least as is based in the district courts and its apex in the Supreme Court." RICHARD H. FALLON, JR. ET AL., HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL JUDICIAL SYSTEM 26 (7th ed. 2015).

⁴⁸ *Compare* Judiciary Act of 1789 § 11 *with* 28 U.S.C. § 1332 (2012) (current codification).

⁴⁹ KLARMAN, *supra* note 1, at 166-67, 349-50.

⁵⁰ Michael D. Gilbert, Optimal Entrenchment of Legal Rules (May 17, 2017) (Virginia Law and Economics Research Paper No. 2017-10), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2970164###.

about this question, as opposed to looking forward to experience under the Constitution. Perhaps this conclusion generally accords with an implication to be found in Klarman's book: that we should give no special reverence to the founding moment or to the Founders themselves. But once we see that entrenchment can be justified on other grounds, the question of how much continued force to give the Constitution, and the amendments enacted under it, becomes largely independent of the motives of the Framers, good or bad.

CONCLUSION

It is the great virtue of Klarman's book that he forces the reader to soberly reflect on the Constitution and its origins. We can still celebrate it, and the shrewd political judgment of the Framers, but we have to realize that their views and the Constitution they created and forced through the process of ratification ill accord with many of our contemporary values. Klarman brings this realization home, as starkly as possible, by allowing the Framers and their opponents to speak in their own words. The reader will often find their words understandable but strange, coming from a generation we might have revered but never knew. If what we hear makes us deeply uncomfortable—as it should—then his book has succeeded in its ambitious aims. For he has shown that the contingency that marked the process of drafting and ratifying the Constitution extends to our own time and our obligation to be bound by it.