# 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

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/21e03f4267a84eeea4e6750ecf3128e9/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/pdfrtf/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 challenges[51]—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'LHOCKEY 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. McCoin, *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 tradablity 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' time[102]—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 standards[144] 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.