Bounded Institutions

By Yair Listokin

Abstract

This Article identifies and examines two alternative institutional structures for hierarchical institutions—"bounded" vs. "unbounded" institutional structures. In a bounded structure, a principal decides on a bounded numerical allocation and then an agent allocates to subjects while complying with the bound. In an unbounded structure, the principal provides no numerical cap or floor on agents, but instead provides some guidance to the agents regarding allocation. An example of a bounded institution is grading to a pre-arranged curve, while an example of an unbounded institution is granting a particular grade to every student who meets a particular threshold.

Bounded and unbounded institutions differ in their strengths and weaknesses. Principals should choose bounded institutions when there are consistent and large populations, when agents are likely to make systematic errors but otherwise share rank order preferences with the principal, and when it is difficult to devise rules along other dimensions of the agent's decision. If agents are biased but share a ranking order with the principal and principals know population traits but not individual traits, bounded institutions can produce a perfect allocation even though neither the principal nor the agent is fully informed or free of error. In other words, with many students and shared quality standards, grading to a curve can produce ideal grading, even if some professors are inclined to be generous with their grading while others are stingy.

I. Introduction

To bound, or not to bound? While the question is not framed in this way, policymakers face this choice throughout law. Should Congress give government programs a bounded budget (discretionary spending), or allow the program to be unbounded and spend as needed (entitlement spending)? Should Congress bound the EPA by subjecting its regulations to a "regulatory budget",[1] or should it allow the EPA to "unboundedly" enact any regulations that meet some standard? Should judges be bound in the length of the sentences they can give, or should the sentences be left to some discretion or rule? Should governments be required to award a bounded percentage of contracts to minority owned businesses, or should the government consider minority ownership as a factor without being bound to any target for contracts? Should the National Science Foundation (NSF) pick a bounded budget and then award grants to the best projects that conform to the budget, or should the NSF choose a quality threshold and award grants to all projects that meet the threshold, even if there are more or less than expected? Should emissions be governed by cap and trade, or should they be subject to a pigouvian tax or other mandates that do not bound the quantity of emissions? And, closer to home, should a law school instruct its professors to grade on a (bounded) curve, or should grading be left to the instructor's discretion or even a grading rule that does not limit the quantity of certain grades?

This Article examines the relative merits of these two alternative institutional structures for hierarchical institutions—"bounded" structures vs. "unbounded" structures-- that recur throughout lawmaking and policymaking. In a bounded structure, a policymaking body (the "principal") decides on a bounded numerical allocation or instruction, and then a subordinate body (the "agent") allocates to subjects while complying with the bound. In an unbounded structure, the principal provides no numerical cap or floor on subordinates, but instead provides some more or less specific guidance to the agent regarding allocation.

Bounded and unbounded institutional structures add an additional dimension to the long-standing rules vs. standards debate. Bounded institutional structures impose rule-like restrictions on one dimension of agent's behavior—the question of "how much in total"[2]—but can co-exist with either rules or standards on another dimension—such as the question of "to whom". For example, a grading curve specifies the number of each type of grade that the professor can award, but often tells the professor nothing about who should receive what grade. Similarly, an appropriation to an agency via the annual budgeting process specifies how much money can be spent, but does not specify how or where

[1] A recent Organization for Economic Cooperation and Development (OECD) report defined a "regulatory budget" as follows

> The regulatory budget operates by close analogy to the traditional fiscal process. For example, each year (or at some longer interval), the government would establish an upper limit on the costs of its regulatory activities to the economy and would apportion this sum among the individual regulatory agencies. This would presumably involve a budget proposal developed by a regulatory oversight body in negotiation with regulatory agencies, approved by the executive branch of government, and submitted for legislative review, revision and passage. Once final budget appropriations were in force, each agency would be obliged to live within its regulatory budget for the time period in question.

Nick Malyshev, *A Primer on Regulatory Budgets*, 2010 OECD J. ON BUDGETING, no. 3, at 2.

[2] Bounded structures, such as grading, may also specify a distribution of "how much", e.g. a certain number of As, Bs and Cs rather than just a total amount (e.g. an average grade of B+).

that money should be spent. Alternatively, a bounded institution can specify "how much" and allow the "to whom" problem to be solved by rule. For example, there are a bounded 435 seats in the House of Representatives.[3] And the way these seats are allocated to different states is also fixed by a formulaic rule.[4] Bounded institutional structures can thus coexist with standards or rules for resolving the "to whom" question.

Unbounded institutional structures entail no limits regarding "how much", but allow for anything from strict rules to unspecified discretion regarding the "who". For example, entitlement spending programs, such as Medicare, guarantee that all necessary spending is paid for, but leave many individual spending decisions to the discretion of doctors and patients.[5] By contrast, other entitlement programs, such as Social Security, use rules to determine which beneficiaries receive funding, and how much each beneficiary receives. Focusing on bounded and unbounded institutional structures thus allows us to see that, contrary to conventional analysis, the question of rules vs. standards is not a one-dimensional problem. Instead, there are two (and even more) dimensions to the problem—bounded and unbounded institutional structures with respect to "how much" can coexist with rules, standards, or discretion in determining "to whom"

And combining bounds with standards or discretion on other dimensions can, under certain assumptions, offer most of the benefits of rules without some of their costs.

When information acquisition is costly and agents are imperfect—conditions that are likely to characterize all of the contexts described in the initial paragraph, bounded and unbounded institutions differ in their strengths and weaknesses. Bounded institutions work better than unbounded institutions when the following considerations loom large. Bounded institutional structures work well when there are consistent subject populations. For example, demanding that a unit of government adhere to a budget works best when the unit's spending needs are reasonably consistent from year to year. Another factor weighing in favor of bounded institutions is the likelihood of systematic agent error or bias. If the EPA's regulators value a clean environment more than Congress or the American people (the principals), then a regulatory budget can effectively constrain this bias by limiting the number and value of the regulations that the EPA can issue.[6] Bounded institutional structures also become desirable when there is a lack of rules available for constraining agents. If costs and benefits were readily quantifiable and verifiable, for example, then Congress could constrain a biased EPA from issuing too many or too few regulations by commanding the EPA to only issue regulations that produce positive Costs Benefit Analyses. But benefits are difficult to reliably estimate, and so Congress may prefer to pick a total costs imposed, and let the EPA figure out where to impose the costs so as to maximize the social benefits.

---

[3] *See* Reapportionment Act of 1929, ch. 28, § 22, 46 Stat. 21..

[4] *See* 2 U.S.C. § 2a(a) (2012) (prescribing the formula call "the equal proportions" rule).

[55] See 42 U.S.C.§. 1395 (2012); 42 U.S.C. § 1320b-15(a) (2012); Patricia A. Davis et al., *Medicare Primer,* CONG. RES. SERV. 24 (2013), http://www.fas.org/sgp/crs/misc/R40425.pdf.

[6] If the EPA places a lower value on a clean environment than Congress, then a regulatory budget reduces the cost of the EPA's bias by requiring the agency to issue a certain number of regulations.

Unbounded institutions, by contrast, outperform bounded institutions when subject populations are small and inconsistent. Tight budgets for government departments subject to idiosyncratic needs, such as the Federal Emergency Management Administration (FEMA) are a bad idea. Instead, FEMA and similar departments are better off with a more flexible "unbounded" budget that responds to idiosyncratic needs, such as a catastrophic hurricane. In addition, unbounded institutions perform better when agents are relatively effective. If the NSF, for example, demonstrates expertise in evaluating grant applications and shares Congress's preferences over how many projects to fund, then it should be trusted with an unbounded budget. The unbounded budget would allow the NSF to respond when there an unusually high or low number of worthwhile applications.

In some circumstances, bounded institutions can produce ideal outcomes even though neither the principal nor the agent is fully informed or free of bias. Suppose, for example, that Congress is incapable of evaluating specific scientific proposals but has a general sense of how many scientific research projects are worth pursuing each year. The National Science Foundation (NSF) loves science more than Congress, and, left to its own devices via an unbounded budget, would fund more projects than Congress would like. But the NSF is good at judging proposals in relative terms, and the NSF's ranking of proposals would be the same as Congress's. Under these assumption, Congress should pick a bounded budget for the NSF that reflects Congress's general understanding of the public benefits of science. It should then order the NSF to allocate that budget according to the NSF's discretion. This bounded institution allows the right projects to receive funding—because the NSF is a good evaluator—but prevents the NSF's pro-science bias from producing too many projects through the use of a bound.

The Article is organized as follows. Section 2 provides more detail on the distinction between bounded and unbounded institutions. Section 3 examines the strengths and weaknesses of each type of institution in the face of common institutional stumbling blocks. Section 4 applies the theoretical insights to the institutional contexts listed in table 1. Section 5 concludes by arguing that the virtues of bounded institutions may be underappreciated and underutilized.

II. Bounded and Unbounded Institutions

Before defining the parameters of my argument, it is helpful to give a brief word regarding how the argument will proceed. In general, I will begin with an abstract description of the "problem", and then provide more concrete examples to illustrate my descriptions.

A. Defining Bounded and Unbounded Institutions

1. Principals, Agents, and Bounded Institutions

This Article examines hierarchical institutions in which a superior body (the principal) sets policy and subordinate bodies (agents) implement or allocate that policy.[7] The policy choices are made with reference to subjects, who are affected by both the principal's and the agent's decisions. The subjects differ in their quality. The principal wants the policy to be sensitive to the subjects' quality.

[7] While this paper focuses primarily on public institutions, bounded and unbounded institutional structures exist in both public and private settings. Most principal-agent relationships, for example, can be characterized by bounds.

The principal "knows" the distribution of subject quality within the population, [8] but cannot determine any individual subject's quality. Instead, the principal relies on one or many agents to determine subjects' quality and allocate benefits accordingly. The agent (or agents) observes each subject's quality.[9] However, in observing subject quality, the agent may have some "bias" relative to the principal. Bias means that the agent systematically perceives subjects as having higher or lower quality than the principal would if the principal were able to observe subjects' quality.

After observing subject quality, the agent determines the policy allocation. The principal may choose to impose or refrain from imposing a bound on the agent's allocation decisions. A bound is a numerical limitation to the agent's total allocation to subjects. When the principal imposes a bound, the agent allocates to the subject based on the agent's observation of the subjects' quality and the relation of the agent's total observations to the bound imposed by the principal. If there is no bound, then the agent allocates to the subjects based on the agent's observations.

The principal dislikes errors. An error occurs where the subject is allocated more or less from the program than the principal would think the subject "should" as determined by the subject's quality.

The principal decides to use a bound if the bound reduces the cost of allocational errors relative to going without a bound. When the principal imposes a bound on agents, the resulting allocational structure is called a "bounded institutional structure". When there is no bound, the allocating mechanism is "unbounded".

2. Examples

This abstract framework describes many important institutional contexts. Consider the process of government funding for scientific research. (Table 1 presents sketches of four other applications of this framework). Congress is the principal. [10] The agent is the NSF. The subjects are the grant proposals that are submitted to the NSF. The grant proposals differ in their quality—the prospective gain in knowledge per dollar. Congress wants to spend money on scientific research so long as the value derived from the resulting gains in knowledge is greater than the costs of the funding.[11]

Congress has a sense, from prior years of funding scientific research, of how much new knowledge is likely to be obtained from funding at different levels. Congress is unable, however, to evaluate the quality of individual scientific proposals. Instead, Congress must rely on the NSF, an agency with expertise in judging the merits of scientific projects.

---

[8] In reality, the principal does not "know" the probability distribution of quality. Some uncertainty regarding the distribution of quality is inevitable. Instead, the principal has a better sense of the distribution of quality in the population of subjects than the agent does. (This excludes the possibility that the principal may be "wrong" about subject quality in the sense that the principal cannot misunderstand the principal's own preferences.) .

[9] Agents may or may not know the overall distribution of subject quality.

[10] The ultimate principal is the "American people." To maintain a one level hierarchy, however, I make the simplifying assumption that Congress, representing the people, is the principal.

[11] The costs of funding include the private benefits of consumption foregone when the government raises revenue, as well as the deadweight losses associated with raising revenue via taxation. *See* JONATHAN GRUGER, PUBLIC FINANCE AND PUBLIC POLICY (4th ed. 2007).

So long as the NSF shares Congress's preferences for funding scientific research, then the delegation of spending allocations enhances the efficiency of Congress's spending. But if the NSF does not share Congress's preferences, then the delegation of decisions to the NSF may cause improper allocations. For example, if the NSF thinks the benefits of science are greater than Congress believes, (or if the NSF thinks the costs of funds are lower than Congress believes) then the NSF may fund more projects than Congress would deem appropriate. From Congress's perspective, these errors are costly—they direct funding to wasteful uses.

Congress chooses among several means of restricting the spending decisions of the NSF. In a bounded institutional structure, Congress gives the NSF a fixed budget for funding scientific research. This fixed budget can be combined with other sources of guidance, such as rules for the geographic distribution of funded projects, or the funding decision can left entirely to the NSF's discretion. Alternatively, Congress might have the NSF provide funding for all scientific projects that meet a certain standard or rule. For example, Congress might tell the NSF that all projects that receive a certain score on an NSF review process should be funded.

This Article examines when Congress should or should not bound the NSF by providing a fixed budget for scientific research funding.

Setting	Funding of Government Programs. E.g., Scientific Research	Environmental Regulation	Hiring of Racial Minorities	Law School Grading
Principal	Congress (the American People).	Congress	Congress or lower-level lawmaking body.	Dean/Law School Governing Entity
Agent	Administrative Agencies spending government funds.	EPA	Government agencies and employees.	Professors
Subjects	Set of expenditures that the government could undertake	Set of Possible environmental regulations.	Parties signing contracts with government agencies.	Students
Subject "Quality"	Benefit provided to the people per dollar of spending.	Amount of environmental benefit per unit cost.	Quality of Product/Service relative to cost.	Understanding of material as reflected in exam quality.
Goal of Principal	Spend on the right things—the things that provide more benefits than their costs. Don't spend too much or too little.	Improve the environment, so long as the cost of doing so does not exceed the benefit.	(1) Get the best combination of quality and price for government services. (2) Remediate Racial Inequality.	Give grades that accurately reflect quality of understanding.
Knowledge of	Congress has	Congress has	Congress/locality	Dean has sense of

Quality Distribution by Principal	sense of how much benefit is produced (in its opinion) from the set of expenditures that the government currently undertakes.	sense of the amount of amount it is willing to pay for certain improvements in environmental quality.	has sense of likely quality/price parameters for contractors of different races	quality of students in a typical law school class.
Policy Allocation	What does the government spend on.	Which environmental restrictions are enacted into regulation.	Which contractors win government contracts.	Which students receive certain grades.
Possible Source of Agent Bias and Flawed Outcomes.	Left to their own devices, agencies may overspend on their own salaries, and may also over- or under-value the benefits government spending produces relative to Congress.	The EPA may value improvements to the environment more than Congress. The EPA enacts too many environmental regulations.	Conscious or unconscious bias against racial minorities leads to inadequate number of minority contractors	Professor may be an easy or hard grader. Students of identical quality receive different grades.
Bounded Institutional Structure	Fixed Budget chosen annually. "Discretionary" Spending.	Regulatory Budgeting.	Quotas for value of government contracts signed with minority contractors.	Grading curve.
Unbounded Rules or Standards	"Entitlement spending" or "Mandatory Spending" such as Medicare or Social Security.	(1) level of quality that could be achieved through the use of the "best available technology economically achievable." (2)Cost Benefit Analysis: Approve all regulations with positive CBA.	Consider minority status of contractor as plus factor.	Professor's Discretion.
Do We See Bounds In Practice?	Yes.	No.	Yes, Until decision xxx (Croson?)	Yes.

3. Simplifying Assumptions

To make the analysis tractable, I make several unrealistic assumptions. These assumptions allow me to focus on the tradeoffs between bounded and unbounded institutions. I do not mean to imply that these other considerations are unimportant. Instead, I leave their analysis to future and past papers by me and others. I also discuss the consequence of relaxing these assumptions below.

a) Benevolent Principals

In many settings, the identity of the true principal is ambiguous. Congress may delegate to an agency, but the ultimate principal is the American people. A CEO delegates to a subordinate, but the CEO owes fiduciary duties to the shareholders of a company. For simplicity, this Article makes the unrealistic assumption that principals are benevolent social welfare maximizers. Congress's objective is thus the same as that of the American people when Congress delegates to an agent such as an administrative agency.

b) Two-Tiered Institutions

This Article focuses on two-tiered institutions with a policymaking principal and an implementing agent who makes decisions about subjects based upon the agent's perception of the subjects' quality. In reality, hierarchical institutions have multiple tiers. The choice between bounded and unbounded institutions explored here therefore exists at many levels. In a three tiered institution, the top level body may set an unbounded standard for the second level body, while the second level body imposes a bounded constraint on the third tier. The top level may also insist that a bounded constraint be applied to the third tier. In each case, the costs and benefits discussed here should govern the choice between unbounded and bounded instructions at any level of a hierarchy.

To illustrate, when I examine the relationship between Congress and the NSF, I assume that the NSF is a single agent or group of agents, ignoring the hierarchy within the NSF.

c) Sharing of Rank Order Quality Perceptions

While the agent may have bias in the sense that it has different perceptions of subject quality than the principal, I assume that the principal and agent's quality determinations are related in an important way. If the agent were asked to rank the subject's quality, the agent would give the same ranking as the principal. The agents may be biased in the sense of having a higher or lower observation of the subjects' absolute quality, but, if asked to make relative determinations, the agent and the principal would produce the same ordering.

To illustrate the rank order assumption in the context of grading, a professor may give a student a "B" while the dean would assign an "A," but if the professor were to rank the students in a class, the professor's ranking would be the same as the dean's. Professor's grading tendencies can be "generous" (grades are systematically too high) or "stingy" (grades are systematically too low), but the rank-order assumption means that they are not "weird" in the sense that the dean would disagree with how the professor ranked the students. In the context of funding for scientific research, the NSF may have a consistently higher or lower opinion of the benefits of scientific research than Congress, but agrees with Congress on which projects are more meritorious than others.

By assuming that principals and agents share rank-order assessments of quality, I assume that the agent's observations provide valuable information to the principal about the subject, even if the principal does not simply agree with the agent's judgments. By contrast, if there is no relationship between the principal's and agents rank order assessments, then the principal may not want to rely on the agent to perform any allocation to subjects, as the agent's allocation will share nothing with the principal's allocation. In this case, the principal may as well allocate to subjects at random and save the cost of paying for the agent.

d) No Strategic Behavior

I assume that the introduction of a bounded vs. unbounded restriction on agent behavior does not change the behavior of the subjects. Thus, a curve does not cause students to work harder because they are competing against their classmates. While such strategic behavior is undoubtedly present, it is the subject of a long literature.[12] In order to focus on the role of principal agent problems in the choice between bounded and unbounded institutions, I therefore assume that these undoubtedly important interactions do not play a role.

e) Random Assignment of Subjects to Agents

In cases where the principal relies one many agents to evaluate subjects, I assume that each agent gets a random sample of the subject population to evaluate. This assumption excludes, for example, the possibility that students of low quality choose professors who are known to be easy graders. This assumption means that the analysis developed below is most applicable to situations such as assigned classes or agents rather than other contexts.

f) Optimal Agent Response to Bounds

I assume that when the bound constrains the agent, the agent responds in a way that is best for the agent. For example, suppose that the NSF, left to its own devices, would like to award $2 billion in funding for research, but that Congress has only appropriated $ 1 billion. In response, the NSF will choose to fund the $1 billion in projects that it thinks will produce the most valuable scientific knowledge.

4. Rules vs. Standards and Bounded and Unbounded Institutions

While a bound resembles a rule and an unbounded structure reflects more discretion for the agent, bounds occupy a different dimension from the typical rules vs. standards analysis. A bound can co-exist with a conventionally defined standard or discretion and an unbounded structure can co-exist with a rule. To illustrate the relationship between bounded and unbounded institutional structures and conventional rules and standards, Table 2 presents a two by two matrix.[13]

Table 2 demonstrates how principals can bind agents numerically while allowing agents considerable discretion along other dimensions. Grading according to a curve provides a familiar

[12] *See, e,g.*, Edward P. Lazear & Sherwin Rosen, *Rank-Order Tournaments as Optimum Labor Contracts*, 89 J. POL. ECON. 841 (1981).

[13] Both rules/standards and bounded/unbounded institutions vary continuously rather than dichotomously. For example, a bound can be one dimensional (e.g. a floor or cap). Table 1's 2x2 matrix is therefore for illustrative purposes only. A more accurate table would have a continuum along both dimensions.

example. While many faculties or deans specify grading curves, they seldom specify how the professor is to grade an exam. Instead, the professor is free to grade as desired, so long as the distribution of grades adheres to the curve. The principal dictates the "how many", but gives little or no guidance regarding the "who".

	Rule	Standard/Discretion
Bounded	Grading: Give 10% As, Describe how to grade by rule. Regulations: Regulatory budget, approve all regs that comply with some rule. (e.g., strict CBA) Environmental: Cap and Trade with auction mechanism. Cap at fixed percentage of base year emissions. Criminal: Give n tickets. Give tickets to anyone going over 80 mph. Representation: 435 seats. Proportional to population. Spending: Budget of $X. Spend on all those over age Y.	Grading: Give 10% As. Choose A's based on excellence. Regulations: Regulatory budget. Pick the regulations based on appropriateness. Environmental: Cap, let subjects and agents determine how to allocate cap based on desert. Criminal: Give n tickets to those driving recklessly. Representation: 435 seats. Represent diverse interests/the entire population. Spending: Budget of $X. Spend on all those who need it.
Unbounded	Grading: Give As to all who get 90 on the final exam. Regulations: Strict CBA. Environmental: Pigouvian tax. Criminal: Give tickets to anyone going over 80 mph. Representation: One rep for every x population. Spending: Spend any amount of $ on all those over age Y.	Give As to whoever demonstrates excellence. Regulations: CBA plus. Pure discretionary regulation. Externalities: Give Agents discretion over issuing externality permits to subjects. Criminal: Give tickets to those driving recklessly. Representation: Represent diverse interests/the entire population.

		Spending: Spend on all those who need it.

Numerical bounds also can be combined with rules. The House of Representatives has 435 places, and these places are apportioned to states based on population according to a mathematical formula.[14] In this case, the principal provides rules to answer both "how many?" (435) and "who" (states with higher populations get more representatives according to a mathematical function of population).[15] When superior bodies combine rules with bounds and the rules provide an answer to "how many" as well as "who", the superior body must provide a conflict resolution mechanism. For example, if a principal issues rules for determining who should receive a speeding ticket, e.g. anyone going over 80 mph, and also binds the subordinate to issue a precise number of tickets (n), then the principal must specify what happens if the number of drivers going over 80 mph is different than n. If the rule yields to the bound and n tickets are issued even if there aren't n drivers going over 80, then the rule regarding "who" is less rule-like than it might initially appear.

Unbounded institutions can also be combined with rules or discretion. In the speeding example, an unbounded rule takes the form of "give tickets to any driver going over 80 mph." The number of tickets issued by the subordinate body ("how many") is determined by the number of drivers violating the 80 mph rule. An unbounded discretionary standard, by contrast, occurs when the subordinate body faces no clear rule for determining either how many or who. If the principal tells the subordinate, "give tickets to those driving recklessly," then this is an unbounded standard.

The default institutional structure is a combination of bounded and unbounded. If the principal does not impose a numerical limit on the subordinate's behavior, then the subordinate is formally unbounded. Nonetheless, any resource constraint on a subordinate body functions as a bound. If the subordinate body cannot simply increase its resources unilaterally, then the subordinate body's allocative capacity is limited by the resource limitation. For example, an administrative agency cannot make an infinite number of rules if it has limited resources. Nevertheless, the implicit bounded resource constraint presents a very different constraint on subordinate body behavior than a more specific bound. An agency with few resources and unbounded rulemaking authority may make very different rules than an agency with unlimited resources and strict bounds (such as a regulatory budget) on its rulemaking authority.

As the preceding paragraph indicates, hybrid unbounded/bounded institutions are possible. Lower level bodies can face bounded constraints along some dimensions and unbounded constraints in others. In addition, lower level bodies can face directions that combine elements of bounded and unbounded instructions along the same dimension. Again, grading proves illustrative. Grading curves are often not precisely bounded. Instead of prescribing an irreversible number or percentage for each grade, curves often prescribe a range. At the top of the curve, for example, professors may be allowed

[14] *See* Royce Crocker, *The U.S. House of Representatives Apportionment Formula in Theory and Practice,* CONG. RES. SERV. (2010), http://llsdc.org/attachments/files/233/CRS-R41357.pdf.
[15] *See id.*

to give 0%-5% of the class a grade of A+. Within this narrow range, the professor's ability to award an A+ is unbounded. Hybrid institutions such as the grading system described here will have some of the costs and benefits of both bounded and unbounded institutions. For simplicity, however, this Article focuses on the extremes—bounded and unbounded institutions. The Article addresses hybrid institutions when their strengths and weaknesses are not simply intermediate points between bounded and unbounded institutions

5. Bounded and Unbounded Institutions and Prices versus Quantities

The economics of regulation contains an extensive literature comparing the virtues of prices (Pigouvian taxes) versus quantities (e.g. emissions caps) for the control of externalities.[16] The choice of bounded vs. unbounded institutions is related to this debate, but also distinct from it.

Consider the debate between a Carbon tax (price) and cap and trade (quantity) for the control of carbon dioxide emissions associated with global warming. With a Carbon tax, the principal imposes a tax on agents (emitters) when deciding how much emissions (subjects) the agent should choose. The Carbon tax estimates the cost of global warming associated with a unit of emissions and imposes a "Pigouvian" tax equal to that cost on each unit of emissions. Once the tax is imposed, emissions will be at their efficient level, since emitters will only emit if the benefits of doing so are greater than the direct costs and the environmental costs (which are reflected by the tax). The price regulation corresponds roughly to an unbounded institution. So long as emitters pay the price of emitting, there is no limit to the amount of emissions.

Under a cap and trade system, Congress imposes a bound on total emissions for all agents (a cap). The agents then choose who will emit how much, but the cap ensures that total emissions will not exceed a certain amount.

Both pigouvian taxes and cap and trade have their benefits. Pigouvian taxes are desirable when the harm associated with emissions can be simply measured and taxed. Caps work better when there are sharply discontinuous harms associated with emissions and the Pigouvian tax cannot be adjusted to account for this discontinuity.

The prices vs. quantities debate concerns a much simpler problem than the generic bounded vs. unbounded quantities decision. In prices vs. quantities, one dimension of emissions quality—environmental harm—is easily measured or estimated. The other dimension of quality—the benefit associated with the emission, is known by the agent making the emission decision. Given these informational parameters, an unbounded Pigouvian tax is desirable. It will generate a first best outcome where agents emit until the costs exceed the benefits. Now, however, suppose that the regulatory problem is made slightly more complex. One dimension of quality—harm-- is no longer directly proportional to emissions. Instead, the relationship between harm is non-linear and rapidly increasing in some amount of total emissions. In addition, a non-linear Pigouvian tax is unfeasible. Under these conditions, a planner may prefer unbounded quantity regulation. A "cap" guarantees that the emissions

[16] *See, e.g.*, Louis Kaplow & Steven Shavell, *On the Superiority of Corrective Taxes to Quantity Regulation*, 4 AM. LAW & ECON REV. 1 (2002); Martin L. Weitzman, *Prices vs. Quantities*, 41 REV. OF ECON. STUD. 477 (1974).

above the cap—which may cause a much greater amount of harm per unit than lower emissions levels—never occur. The cost of the cap is that the principal must estimate the benefits of the emission. If these benefits are wrongly estimated, then the amount of emissions may be wrong. But this may be a price worth paying for the guarantee that emissions will not reach harmful levels.

In this Article, I consider a still more complicated regulatory environment. Neither the benefit nor the harm associated with any particular unit of emissions is known to the principal. Instead, benefits and harms vary depending on the subject. Imagine, for example, that the harm and benefit associated with an emission depends upon the time, place, and manner of the emission.[17] In order to estimate the harms and benefits associated with a particular emissions, the principal must rely on a different type of agent—a government official. In the typical prices vs. quantities question, this type of agent is unnecessary because the principal possesses perfect knowledge of harms, benefits, or both. When the principal does not possess this knowledge, the agent is needed to garner information to make feasible policy allocations. But this agent may be biased, and so this introduces a new set of concerns to the principal.

The bounded vs. unbounded institutions environment thus constitutes a generalization of the prices and quantities framework to the case where both benefits and harms ("quality") vary from unit to unit, and these benefits and harms must be estimated by an agent of the principal. This generalization allows the question of bounded vs. unbounded institutions to be applied to a much wider array of questions than prices vs. quantities. While the prices vs. quantities debate is exclusively applied to the (undeniably) pollutions regulation, the bounded vs. unbounded question applies to almost every domain in which the principal relies upon an agent to determine quality.

III. Bounded vs. Unbounded Institutional Structures

With the principal's choice between bounded and unbounded institutional structures now defined, this section analyzes circumstances favoring the use of one structure versus the other. Under some assumptions, bounds increase error costs, making them undesirable for the principal. Under other assumptions, bounds decrease the cost of errors.

A. Perfect Information, Unbiased Agents

First, assume that the principal has access to perfect information about the distribution of quality within a population. The principal thus knows the average quality in the population and how this

[17] While such variation in "quality" does not generally characterize carbon dioxide emissions that produce a global harm, it may characterize other emissions. Even carbon dioxide emissions do not have uniform harms. For example, carbon dioxide emissions from airplanes may be more harmful then other emissions because the airplane emits the carbon dioxide while in the sky.

quality varies from subject to subject.[18] This does not mean that the principal observes each subjects' quality. If the principal were able to do this, there would be no need to rely on agents. Instead, the principal has perfect knowledge about the distribution of quality in the population, but does not observe individual quality. If the principal is going to impose a bound, the principal chooses the bound best suited to minimize errors. This bound would reflect the principal's knowledge of the distribution of quality within the population.

Second, assume that the agent receives an informative signal about the subjects' quality. In addition, the agent is unbiased. That is, the agent shares the principal's preferences regarding allocation. As a result, the agent gives each subject the allocation that would be desired by the principal if the principal could observe each subjects' quality directly and without cost.

Under these circumstances, unbounded institutional structures dominate bounded structures. Unbounded institutional structures always produce the principal's preferred outcome because the agent makes no error and shares the principal's preferences. (If you have such an agent, there is no reason to constrain them with a bound.)

Bounded institutional structures, by contrast, make errors inevitable. Assuming the subjects are a random sample of the population,[19] the distribution of quality within the subject population is likely to randomly deviate from the distribution of quality in the overall population. Any such deviation leads to errors when the principal imposes a bound. The size of these errors depends upon the deviation of the agent's sample population from the total population. Accordingly, the superiority of an unbounded institution diminishes when the size of the sample population rises: when the sample population grows, the size of the errors associated with a bound goes down because the natural variation of quality within the sample population has more opportunity to "even out" and become more like the distribution of the overall population. Similarly, the size of the errors also goes down, and the superiority of unbounded institutions diminishes, when the distribution of quality within the population is less variable. If everyone in the population were identical (no variation), for example, then a bounded structure would produce the same error costs (zero) as the unbounded structure because the sample population would have the same traits as the subject population.

Overall, assuming perfect information and an unbiased agent, unbounded distributions are superior. This can be illustrated in the context of environmental protection regulations. Suppose the EPA chooses regulations exactly as Congress would want them to. In this case, a regulatory budget is harmful even if Congress knows the typical distribution of opportunities for environmental regulations. Without a regulatory budget, the EPA chooses Congress's preferred regulations. With a budget, by contrast, some regulations may be precluded. In any year, there may be more or less efficient and new environmental regulations than there would be in a typical year. However, the problems of a bounded curve are reduced when the population of possible regulations is stable, because the chance that a year will have an atypically strong or weak set of regulations is lower.

[18] In addition, the principal knows the "higher moments" of the population distribution, e.g. kurtosis, skewness, etc.

[19] *See supra*, Part II.I.e..

B. Perfect Population Information, Biased Agents

Now assume that the agent makes biased determinations of subject quality. In other words, the agent systematically over- or under-estimates the desirability of a particular allocation relative to the principal.

The principal chooses the limit for a bounded structure based on the true population mean or some other population parameter. Because the bound is based on the true population mean, it adjusts for bias: an agent who is downwardly biased in determining quality relative to the principal will be forced to raise their estimate, while an agent who is upwardly biased will be compelled to lower their average quality estimate. Introducing the bound, however, introduces some of the costs described in the previous section. The sample of subjects that come before the agent may differ from the population distribution. In this case, the bound may require over- or under-allocation, regardless of the agent's bias.

The principal should choose a bound when its bias correction benefits exceed the costs of the bound's rigidity and misallocation. The bound is most attractive to the principal when the agent's bias is high, the variability of quality within the population is low, and the agent faces a large sample of subjects. These factors either increase the benefits of bias reduction or reduce the costs of the rigidity imposed by a bound.

As the number of subjects gets very large, a bounded institutional structure produces a perfect allocation in spite of the agent's bias. With many subjects, the distribution of quality within the agent's sample population becomes indistinguishable from the total population distribution.[20] With the bound accurately capturing the distribution of the subjects (as we are assuming that the principal has perfect knowledge of the distribution of quality in the population), the rigidity costs of the bound become negligible. The bound's bias correction benefits remain robust as the number of subjects gets larger—the agent's bias does not go down with the number of subjects. Because of the bound, the agent must adjust allocations in proportion to the agent's degree of bias. Indeed, the bound eliminates the misallocation associated with bias. And because bias is the only source of agent error (the agent would rank the subjects the same way the principal would), the bound produces the best allocation from the principal's perspective.

The bound produces an optimal allocation even if the principal does not know the size or direction of the agent's bias. So long as the principal and agent share a rank ordering, the bound compels the agent to adjust the allocation in proportion to the size of the agent's bias. If the agent is too "optimistic", the bound imposes some "pessimism" by requiring the agent to allocate less than the agent's desired amount. But if the agent is biased in a pessimistic direction, then the bound imposes optimism by requiring the agent to allocate more than the agent would wish.[21]

---

[20] *See* discussion *supra* Section III.A.

[21] For an analogous result in the context of using property rules or liability rules when there is imperfect information, see Richard R.W. Brooks, *The Relative Burden of Determining Property Rules and Liability Rules: Broken Elevators in the Cathedral*, 97 NW. U. L. REV. 267, 269-71 (2002).

By contrast, an unbounded institutional structure produces the wrong allocation. The agent's systematic biases are relatively unconstrained in an unbounded institutional structure. As a result, these biases will be reflected in the ultimate allocation. A systematic positive bias in the subordinate body results in oversupply within an unbounded institutional structure. A systematic negative bias results in undersupply.

To illustrate, consider the NSF grantmaking process (as a proxy for all government spending programs). Suppose that, from past years, Congress has a sense of what type of projects will be funded if it provides different levels of bounded appropriations to the NSF and has a sense of how much these projects benefit society. If Congress chooses a bounded institutional structure for the NSF, it budgets a certain amount for funding research and then charges the NSF with allocating those funds. If Congress chooses an unbounded structure for the NSF, then it allows the NSF to provide grants to any scientific projects the NSF deems worthy.

Suppose the NSF is biased—it places a higher value on scientific research than Congress or the American people do. Now suppose Congress gives the NSF a budget that would fund the "right amount" of science in a typical year. So long as scientific quality does not vary greatly from year to year, the budget enables Congress to fund the "right" projects, even though Congress is incapable of judging scientific merit.

Why do we get the right amount of science funding even though the NSF is biased in favor of science and Congress is incapable of judging scientific proposals? Because Congress has a good sense of how much science it wants to fund and the NSF is a good judge of relative scientific merit. The budget constrains the NSF—instead of funding all the projects that it wants to, it has to confine funding to the amount that Congress determines. As a result, the NSF's bias cannot be expressed in policy. But because the NSF is still a good judge of relative merit, it will allocate its budget to the scientific projects that Congress would have picked had Congress been able to judge merit.

The budget produces a rigidity that the NSF cannot undo. This rigidity is good for reducing the impact of NSF bias or systematic error (as shown above), but the rigidity is bad for responding to unexpected shifts in the quality of science from year to year. So the less likely Congress is to predict the "right" amount of science, the more costly the rigidity of a budget and the more appealing an unbounded structure, such as funding for all projects that receive a certain "score" in the NSF's review process.

Congress is less likely to know the right amount of science if science is unpredictable from year to year. If, for example, the value of scientific projects unexpectedly goes up dramatically, then the NSF budget for that year is going to produce too little funding. The NSF is bound by the budget, even though the budget is too low. Relatedly, a bounded budget works better if there are many small projects to choose from rather than a few big ones. When there are only a few projects from year to year, then it is much more likely that any given year will produce too many or too few projects (relative to the bound) than Congress would wish to fund. With many projects, by contrast, the central limit theorem applies so that funding needs are less likely to change dramatically from year to year.

C. Perfect Population Information, Error Prone Agents without Bias

Instead of being systematically biased, an agent may make errors of the following kind: the agent wants to implement the principal's wishes—without bias-- but makes mistakes. In other words, the agent is a noisy but unbiased representative of the principal's wishes. The agent's errors are correlated; an agent's error in observing one subject's quality is likely to be repeated on other subjects.

Correlated errors occur when an agent's mistake in judging one subject is likely to be repeated with other subjects. If agents make correlated errors rather than just uncorrelated errors, bounded institutional structures become more attractive relative to unbounded structures.

Consider the extreme case where errors are perfectly correlated. If the agent makes a mistake with one subject, the agent will make the same mistake with all other subjects. This pattern of mistakes resembles bias in some respects but not others. Unlike biased agents, agents who make perfectly correlated errors are, from an ex- ante perspective, just as likely to over-allocate to agents as to under-allocate. But even without this ex-ante bias, an agent making the same error over and over systematically over-allocates or under-allocates to subjects, just as a biased agent systematically over-allocates or under-allocates to subjects. The systematic nature of mistakes is shared by both biased agents and agents who make perfectly correlated errors.

Bounded institutional structures outperform unbounded institutional structures when agents make errors that are both large and perfectly correlated, and when the number of subjects is high. The reasoning follows the reasoning for biased agents.

When an agent makes perfectly correlated errors, the agent's allocation is likely to deviate considerably from the optimal allocation because of the repetitive nature of the agent's errors. As a result, the constraints placed upon the agent via a bounded allocation prove attractive. The principal chooses the bound based on the principal's knowledge of the distribution of the population and forces the agent to modify the allocation to comport with the bound. If the typical size of perfectly correlated errors is small, however, then the cost of imposing the bounded constraint may not be worth the limitations on the agent errors. In addition, bounded constraints prove more costly when the number of subjects evaluated by an agent is small. Because the distribution of the sample evaluated by the agent becomes increasingly similar to the population distribution as the number of subjects grows large, the costs of the bound in terms of possible misallocation goes down. With smaller numbers of subjects per agent, by contrast, the bound may constrain the agent to give inappropriate allocations based on the subjects' quality, and this cost may be greater than the cost of misallocation due to correlated agent errors.

The bound imposed by the principal forces the agent to ignore its unbounded determination -- which has been systematically thrown off by correlated errors-- and instead choose its comparative ranking. This ranking is likely to closely resemble the true subject suitability for the program. The bound thus eliminates the systematic error caused by correlation and uses the agent for a job the agent is well suited for in spite of correlated errors—the task of making relative determinations of subject suitability for a program.

For example, a professor who makes perfectly correlated grading errors is not a systematically generous or ungenerous grader. Instead, the professor has an "angle" on each exam that is different from the one preferred by the dean or the faculty. Sometimes the professor's angle is more generous to students, sometimes it is less generous, and sometimes it is more or less the same. A dean choosing a curve when professors make perfectly correlated errors faces a problem analogous to the problem the dean faced when professors are systematically biased. A curve has benefits—it limits the scope for a professor's perfectly correlated errors to cause widespread errors—and costs—it restricts the professor's ability to adjust grades to reflect variation in class quality. With a small class or a student population with more variation in student quality, the costs of the curve -- precluding variation in grades for variation in class quality-- loom large. A small class drawn from a student population with wide variation in student quality is more likely to have random variation in student quality from the distribution of quality within the student population as a whole. As a result, a curve is less attractive for a small class because it produces incorrect grades. With larger classes and student populations with less variation in quality, the curve works better. The curve reduces the grading errors associated with the professor's "angle," but it is less likely to impose an incorrect distribution on the class because the larger class size and smaller variation in student quality makes an unusual distribution of quality unlikely.

When agent errors are correlated but not perfectly correlated, the best institutional structure depends on several different considerations. As ever, bounded structures become more attractive as the number of subjects grows large and as the ratio of population variance to variance caused by error gets smaller. If a bounded structure is preferable, it should constrain the absolute numerical target to comport with the distribution of quality within the population. This bounded structure closely resembles the constraint for biased agents. The bounded structure should also constrain the number of extreme allocations for extreme values of quality to a greater degree than allocations for more normal values. The extreme values are more likely to reflect idiosyncratic error in addition to systematic error and should therefore be curtailed to a greater degree.

D. Imperfect Population Information

The previous sections assumed that the principal possessed perfect information about the distribution of quality within the population. There was risk—the principal did not know the quality of any subject or group of subjects, but there was no uncertainty—the principal knew the distribution of quality. This is an unrealistic assumption. Not only is the principal unable to observe any individual subject's trait, but the principal is also unlikely to know precisely what the average value of quality is, or how quality varies from subject to subject within the population.

When the principal has more uncertainty regarding the distribution of quality within the population, bounded institutions become less attractive relative to unbounded institutions. Consider an extreme case in which the principal has no information regarding quality's distribution but the principal still desires to allocate to subjects according to quality. In this case, a bounded institution can only do harm. The principal has no information upon which to base the bound; any bound restricts the agent,

who has relevant information about quality, without reducing the agent's error. As a result, the principal should choose an unbounded institutional structure.

More realistically, the principal possesses imperfect information about the distribution of quality in the subject population. All other things equal, the principal should prefer unbounded institutions relative to bounded institutional structures when there is greater uncertainty about the population distribution.

When the population is stable, the principal is more likely to obtain accurate knowledge regarding the distribution of quality. Suppose, for example, that the principal learns something about the population distribution each year. When the population is stable, then, over time, the principal acquires highly accurate information. When the population is in flux, (or, more precisely, the population is in flux in an unpredictable way), however, then the value of the principal's previously acquired knowledge is reduced. With poorer quality information about the population distribution, the principal should be less inclined to impose the rigidities of a bound upon agents.

However, when the principal wants to allocate according to relative values of quality rather than absolute values, the case for bounded institutions increases. Even without knowing population parameters, the principal knows that large pools of subjects chosen at random should be relatively similar. A bounded institution insures that differences between agents observing large pools do not skew the relative rankings of subjects.

In the context of NSF funding, suppose that Congress does not know the distribution of quality for scientific grant proposals in any particular year. The more uncertain Congress is about the distribution, the more Congress should favor an unbounded funding budget. For example, if science is subject to sudden unpredictable lurches, in which the quality of proposals changes radically from one year to the next, then Congress should favor unbounded structures. The lurches introduce unpredictability in the value of scientific funding that fixed budgets cannot respond to, but unbounded funding arrangements can. In the grading context, suppose the dean knows nothing about the quality of the students in a class or class year.

.

E. Quantifiability

As defined here, bounded institutions work best with numerical bounds. Grades, regulations, budgets, among others, are all quantifiable in a way that makes a sharp and verifiable bound feasible. Bounds reduce the costs of agent bias and error by constraining the agent's freedom of action via quantifiable restrictions. If the agent's decision cannot be quantified accurately, then the advantages of bounded institutional structures diminish. Unquantifiable bounds offer much weaker constraints on agent behavior.

When an agent's choices cannot be constrained directly because there is no quantifiable metric on which to bound, the agent's decisions may still be bounded along some other dimension. For example, resource constraints do not impose hard bounds on agents' activities, but rather offer a fuzzy

bound on the harm that can be caused by bias or other error. A biased agent can make fewer errors when limited by a resource constraint than when given unlimited resources. The resource constraint is an indirect bound, however, because it does not explicitly address the action for which the agent is needed. Moreover, a resource constraint does not impose a constant bound. If the agent stretches resources, then the resource constraint is less restrictive.

IV. Rules, Standards, Non Linear Error Costs, and Bounded Institutions

The previous part examined conditions that made bounded institutional structures more or less attractive relative to unbounded structures. A brief summary of these conditions is useful.

Other things equal, bounded institutional structures are more attractive to the principal when

1. Agents exhibit greater bias.
2. Agents share rank order preferences with principals.
3. There is less variation of a trait within the sample population assessed by an agent.
4. The sample population assessed by an agent is larger.
5. Agents are more prone to correlated but unbiased error.
6. Principals have less uncertainty about the distribution of quality within the subject population.
7. Agent behavior can be bounded with a quantifiable metric.

This Part extends this analysis to two recurring debates in law: 1. The debate regarding rules vs. standards and 2. The debate about regulating prices versus quantities. This Part also considers the possibility that the costs of errors are asymmetric (an increment of over-allocation does not necessarily impose the same cost as an increment of under-allocation).

A. Rules vs. Standards and the Desirability of Bounded Institutional Structures

A vast literature examines and debates the desirability of rules or standards. Rules are laws given context ex-ante, before an event has occurred.[22] Standards are laws given content ex-post, after the event has occurred.[23] The advantages of rules are predictability and ease of application. The advantages of standards are accuracy and context dependence. Bounded institutional structures constitute a rule in one dimension that can offer the prospect of obtaining many (but not all) of the benefits of rules on alternative dimensions of interest without incurring the costs of rules.

[22] *See* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992).
[23] *Id.*

Framing a clear rule proves difficult when the evaluation of quality requires complicated and multi-dimensional analysis. As a result, multi-dimensional problems are often handled by standards. Yet standards introduce costs of their own. Because of their muddiness, standards provide relatively few constraints on agents. This allows possibly biased and error prone agents greater scope for decisions that deviate from those of the principal.

Bounded institutional structures can reduce the agency costs of standards without requiring a rule. Just as violations of a rule are relatively easy to detect, so too are violations of a bound. Moreover, the bound allows for flexibility on the "to whom" question that often vexes rules. Suppose the principal prescribes a multi-dimensional standard as well as a bound on the agent. The agent then applies the standard, while adhering to the bound. Because the agent applies the standard rather than an inapt rule, the agent does a better job of using ex-post information and context to identify the subjects most suited for a program. At the same time, the bound imposed on the agent by the principal limits the agent's ability to exploit the fuzzy standard to indulge the agent's bias. If an agent is too generous, the bound prevents this generosity from greatly affecting allocation. And, as established in the previous section, under certain conditions (large numbers of subjects, small population), a bound is not too costly. As a result, a bounded institutional structure proves desirable when the formulation of rules is infeasible, agents are prone to bias or error, and the subject population exhibits traits that limit the costs of the rigidity imposed by bounds.

When a rule can be formulated with reasonable cost and accuracy, by contrast, the case for bounded institutional structures diminishes. Because the rule limits the scope for the agent's biases, correlated errors, and correlated errors to adversely affect outcomes, the bound's benefits are less salient. The availability of a rule means there is less bias and error to bound. A bound's costs in terms of rigidity, however, remain relatively constant.

In total, a bound functions as a partial substitute for a rule in cases where a rule is not viable. A bound offers some of the benefits of a rule-- reduction in bias and easy measurability-- as well as some of the costs—rigidity and possible over or under-inclusiveness. The bounded institution proves attractive because it can impose some constraints on agents in cases where the direct method of constraining by rule is impractical because it will lead to inappropriate allocations. As a result, we would expect to see bounded structures coexisting with standards and rules co-existing with unbounded structures more often than combined bound/rules or unbounded/standard combinations.[24]

Rules, bounds, and standards all impose some kind of cost. A rule is too rigid in determining who to allocate to, a bound is too rigid in determining the amount of an allocation, and a standard allows too much scope for bias. The presence of bounded structures, however, introduces another

[24] Bounds can co-exist with rules, as highlighted in Table 1. In these circumstances, there must be a method for determining what to do when the bound and the rule conflict. One possibility that does not imply conflict is that the bound dictates "how much" and the rule dictates "to whom." For example, a bounded appropriations program may have a fixed budget, with a rule for allocating the budget to certain parties but not others. Consider the National Science Foundation grant-making process. The NSF may be given a fixed budget (a bounded allocation of "how much") and it may be told to allocate all of that budget to grant applications that receive a certain score ("to whom").

dimension of institutional design that allows for the costs of standards to be diminished in cases where ordinary rules are infeasible.

Bounded institutions may also encourage clarity of thought when agents are faced with standards. At present, an agent confronted with a standard announced by the principal must consider a multidimensional problem without concrete guidelines on how to distill these dimensions into a coherent policy. While the lack of rigidity allows agents to consider the entire context of the decision, it can also lead to muddled thinking—how should an agent weigh and balance the myriad of factors relevant to deciding whether a subject meets the standard? A bound creates some pressure for the agent to implicitly decide on a ranking without confining the agent to consider some relevant facts but not others. Because only subjects that comply with the bound can be allocated the program, the bound forces the agent to implicitly rank the subjects. While the agent's ranking function may not be explicit, the bound requires the agent to confront the fact that allocating a benefit is not free, and therefore requires hard choices. A standard without a bound, by contrast, may allow the agent to avoid hard but necessary choices by allocating too much in order to avoid the unpleasant feeling of allocating to some subjects but not to others who are relatively similar.

To be concrete, consider the decisions of an administrative judge working on immigration asylum cases. These cases could be decided by rule, such as deny (or accept) all asylum cases. Each of these rules, however, will produce many errors from the principal's perspective. Congress wants some foreigners facing danger at home to enjoy asylum, but not all of them. And Congress's perception of the desirability of asylum likely depends on many hard to define factors, such as reasonable fear of harm or denial of important rights. Any rule will miss important dimensions of these factors.

Instead of imposing a rule regarding asylum, Congress relies on a set of agents. It vests the power to grant asylum with both the U.S. Citizenship and Immigration Services (USCIS) in the Department of Homeland Security (DHS), and the Executive Office for Immigration Review (EOIR) of the Department of Justice.[25] These offices further specify standards and makes decisions, and asylum officers and Immigration Judges apply these standards.[26] This policy introduces its own problems. With such ambiguous standards, agents have the ability to indulge biases for or against asylum,[27] to opine on

[25] *See* 8 U.S.C. § 1158 (2012). The USCIS initially handles "affirmative claims" for asylum. *See* Exec. Office for Immigration Review, *Asylum Protection in the United States*, U.S. DEP'T OF JUSTICE 3 (Apr. 25, 2005), http://www.justice.gov/eoir/press/05/AsylumProtectionFactSheetQAApr05.pdf. Agents can grant asylum after an interview, or, if asylum is not granted, the individual is referred to an Immigration Judge of EOIR for a formal proceeding. *Id.* EOIR Immigration Judges also handle "defensive claims" of asylum during removal proceedings. *Id.*

[26] *See* 8 U.S.C. § 1158(b)(A) (2012); 8 C.F.R. § 208.13.

[27] *See* Jaya Ramji-Nogales et. al., Refugee Roulette: Disparities in Asylum Adjudication, 60 STAN. L. REV. 295, 302, 372 (2007)("[I]n the world of asylum adjudication, there is remarkable variation in decision making from one official to the next, from one office to the next, from one region to the next, from one Court of Appeals to the next, and from one year to the next, even during periods when there has been no intervening change in the law.").

issues on which the agents have little experience, or to indulge whimsical notions of how to apply the standards.[28]

A bound combined with a standard may provide an alternative that enjoys some of the benefits of rules without all of the costs. For example, Congress could specify that a bounded number of people should get asylum each year, but maintain a standard for determining who those people should be. The bound on asylum numbers should mitigate agent biases and restrict their ability to apply whimsical criteria. Biases will be directly reduced by the bound. If EOIR's appellate component, the Board of Immigration Appeas (BIA), is systematically biased against asylum seekers relative to Congress, then Congress's bound will force the BIA to grant more asylums than they otherwise would. If the BIA is biased in favor of asylum, then the specified number will restrict the number of asylums the BIA can grant. The bound may also discourage the BIA, or specific immigration judges, from indulging whimsical notions of eligibility. With a bound, indulging in whimsical notions has a cost. If the whimsical notion keeps a deserving applicant out, it also means that a less deserving applicant gets in. And so a biased administrative judge will be less likely to indulge their biases in an idiosyncratic way because the idiosyncracy does not facilitate the bias—e.g. fewer asylum recipients—but rather allocates the benefit in a way the judge does not want—e.g. grants asylum to less deserving applicants rather than more.

This is not to say that a bound is perfect. The number of people who Congress would deem worthy of asylum may fluctuate unpredictably from year to year, and setting a precise number of asylum recipients does not allow the agents to adjust to the fluctuation. If Congress could express its asylum preferences perfectly via an enforceable rule, then the rule would be preferable, because the rule would insure that the right applicants receive asylum without imposing a constraint on the number. But a rule is almost certainly infeasible in this case. And so Congress should weigh the bias reduction and thought clarification benefits of a bound combined with a standard against the cost of the rigidity the bound imposes. Congress may well decide that the benefits of the bound combined with the standard exceed the costs.

B. Prices and Quantities as Bounded and Unbounded Institutional Structures

Unbounded Pigouvian taxes perform well when it is easy to measure and calculate the harm associated with externalities such as carbon emissions. A Pigouvian tax is a rule that says "charge x dollars for each unit of emissions". Because harm is easy to quantify and measure, the Pigouvian tax perfectly expresses the principal's preferences.

By contrast, quantity restrictions fare poorly if the costs associated with mitigating the externality are hard to predict. A bound on the amount of negative externalities emitted proves too strict if mitigating the externality is costlier than expected. The bound is too lenient if the costs of

[28] *See, e.g.*, Benslimane v. Gonzales, 430 F.3d 828, 833 (7th Cir. 2005)(calling the rationale of the Board of Immigration Review's decision "completely arbitrary.").

mitigating the externality are unexpectedly cheap. With unpredictable costs of mitigation, the bound on quantity introduces rigidity. This rigidity is unnecessary with access to a perfect rule—the Pigouvian tax--and therefore the bound should be avoided.

Quantity regulation becomes more attractive when the costs of externalities above a certain quantity rapidly increase and there is no way for the Pigouvian tax to account for this nonlinearity in costs. For example, global warming might increase rapidly above a certain atmospheric carbon dioxide concentration. Under these circumstances, the rigidity imposed by the bounded quantity regulation may be worth bearing. An unbounded Pigouvian tax that does not account for the non-linearity runs the risk of triggering considerable harm if the costs of mitigation prove greater than expected, as the excess externalities associated with this eventuality are extremely costly. In total, the rigidity of the bounded quantity regulation may be more appealing to the principal than the possibility of costly overproduction of externalities associated with Pigouvian taxes.

C. Non-Linear Error Costs

In the standard prices versus quantities framework, non-linearities made bounded quantities more attractive to the principal. Therefore, one might expect the possibility of non-linear costs to generally favor bounded institutional structures relative to unbounded institutional structures. Not so. While the prices versus quantities framework allows for non-linearities in the harm associated with emissions, it does not consider the possibility of non-linearities in other important dimensions of the problem. If the framework is generalized to account for non-linearities in benefits and harms, then the existence of non-linearities does not push in favor of bounded institutions relative to unbounded ones. Instead, the desirability of bounded or unbounded institutions depends critically on the type of non-linearity that is considered relevant.

Consider, for example, the possibility that the costs of mitigating carbon emissions follow a non-linear process. Mitigation costs will either go down dramatically, or they will not go down at all.[29] Under these circumstances, quantity regulation may prove extremely costly. If the quantity regulation assumes that carbon mitigation costs will go down by an average of the two extremes and chooses the quantity accordingly, the quantity chosen is guaranteed to be wrong. If carbon mitigation costs go down dramatically, then the quantity restriction is far too lenient. If mitigation costs stay relatively constant,

[29] This may characterize different aspects of carbon mitigation. For example, the cost of generating electricity via solar panels, one appealing form of mitigation, has been going down exponentially for many years. *See, e.g*., Ramez Naam, *Smaller, Cheaper, Faster: Does Moore's Law Apply to Solar Cells?,* SCI. AM. BLOGS, (Mar. 16, 2011), http://blogs.scientificamerican.com/guest-blog/2011/03/16/smaller-cheaper-faster-does-moores-law-apply-to-solar-cells ("Over the last 30 years, researchers have watched as the price of capturing solar energy has dropped exponentially. There's now frequent talk of a "Moore's law" in solar energy. In computing, Moore's law dictates that the number of components that can be placed on a chip doubles every 18 months.").
The costs of storing energy in batteries, by contrast, has not experienced the same type of decline. *See* Seth Fletcher, *40 Years Later: Electric Cars and the OPEC Oil Embargo*, SCI. AM. BLOGS, (Oct. 8, 2013), http://blogs.scientificamerican.com/observations/2013/10/08/40-years-later-electric-cars-and-the-opec-oil-embargo ("Why did it take four decades to get viable electric cars on the road? Part of the answer has to do with the intrinsic difficulty of battery chemistry. Batteries are messy, disobedient, devilish machines. They don't obey Moore's law. It has simply taken researchers a long time to work through all of the problems.").

then the quantity restriction is far too strict. As the quantity restriction moves farther from the point at which the costs of carbon mitigation equal the harm caused by another unit of carbon, the costs of the bound go up in a non-linear fashion. For example, if the quantity restriction is too strict, then it might cause a severe economic contraction and financial panic. If a global depression is a possibility, the possibility of triggering non-linear costs of excess carbon emissions may be worth the risk.

As with carbon emissions, so too with bounded and unbounded structures generally. Many factors relevant to the decision to use bounded versus unbounded structures may have non-linear components. These non-linearities can push principals towards either bounded or unbounded structures.

Consider government funding for scientific research, as an example of government funding more generally. Suppose that scientific knowledge increases spasmodically, and that there are complementarities for different projects—research produced by one scientific project increases the value of research produced by another project. In this case, a bounded budget for the NSF may prove foolish. If science has a particularly good year, and one project facilitates another project, then a budget limits funding at precisely the times that funding most desirable. Under these circumstances, allowing the NSF to fund all worthy projects may lead to less costly errors than limiting the NSF via budget.

Alternatively, suppose that very low levels of scientific funding have extreme costs, because they permanently shut down laboratories, prevent training of new scientists, and generally damage scientific institutions that cannot simply be reinstated. In this case, an unbounded NSF funding scheme may be undesirable. If the NSF happens to be run by agents who are biased against science, then, without a budget, the NSF may dramatically curtail scientific funding. This reduction will have the long-term effects on scientific research—such as shutting down labs and plugging the pipeline of new scientists. To avoid this occurrence, Congress may prefer to give the NSF a budget that is sufficiently large to prevent long term damage to scientific institutions, even if this brings the risk of overfunding science if the NSF is not biased.

The previous two paragraphs presented two types of non-linearities in the funding of scientific research. One non-linearity-- complementarities arising from additional funding in rapidly developing areas of research—strengthened the case for unbounded funding schemes. The other non-linearity--permanent harm to scientific institutions arising from inadequate funding—strengthened the case for a scientific budget exceeding the level that would cause long term harm. As a general matter, therefore, non-linearities do not favor the use of bounded or unbounded institutions. The finding that quantities are preferred to prices when there are non-linearities in the costs of harm from externalities is an artifact of the structure of the prices vs. quantities problem, rather than a general point about non-linearities and bounded vs. unbounded institutions.

V. Bounded vs. Unbounded Structures in Action

The previous sections developed the concepts of bounded and unbounded structures and identified situations in which bounded or unbounded institutions were more likely to be effective. This section applies the insights developed above to several examples.

A. Traditional Regulatory Oversight vs. Regulatory Budgeting

The analytical framework just developed applies to the problem of controlling administrative agencies. Administrative agencies are "agents" of a principal that could be Congress, the President, or the people. Agencies evaluate a "population" of potential regulations. The principal cannot evaluate the entire population, and therefore delegates the task to the agency. Each potential regulation has a "trait", which is the regulation's suitability for achieving the goal of the principal.[30]

Bias presents a recurring concern in the analysis of administrative agencies. Agencies are often presumed to have interests that diverge from those of the principal. The agency may differ from the principal in its evaluation of potential regulations. Indeed, mitigating such conflicts is one of the central questions of administrative law and administrative law scholarship. The administrative law literature debates the efficacy of different mechanisms, such as Cost Benefit Analysis, Judicial oversight, Executive Branch oversight, and public oversight (FOIA), for reducing the costs of errors.[31] Each method brings plusses and minuses, but all of the methods leave agencies unbounded.[32] No matter how tough the oversight, any method that satisfies the oversight process becomes regulation. So long as the cost-benefit analysis proves that the regulation has positive net benefits, the regulation follows the statute, or the regulation passes through appropriate executive or judicial oversight, the regulation may be issued. Because there is no hard numerical cap or floor on regulation, agency regulations are promulgated in an unbounded institutional environment.

The regulatory environment need not be unbounded. The number, or more plausibly the value,[33] of regulations could be constrained by statute. The much-discussed concept of a "regulatory budget" imposes limitations on the costs that may be imposed by agencies via regulation. A recent Organization for Economic Cooperation and Development (OECD) report described a regulatory budget as follows

[30] This goal could be maximization of social welfare or a different goal such as minimization of environmental harm.

[31] *See, e.g.,* Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999); Harold H. Bruff, *Presidential Management of Agency Rulemaking*, 57 GEO. WASH. L. REV. 533 (1989); Charles H. Koch, Jr., *Judicial Review of Administrative Discretion*, 54 GEO. WASH. L. REV. 469 (1986); Richard J. Pierce & Sidney A. Shapiro, *Political and Judicial Review of Agency Action*, 59 TEX. L. REV. 1175 (1981); Sidney A. Shapiro, *Political Oversight and the Deterioration of Regulatory Policy*, 46 ADMIN. L. REV. 1 (1994).

[32] More precisely, they leave the agency without a direct bound. The agency is constrained by its resource constraint.

[33] If the number of regulations is constrained but not their value, agencies could combine similar regulations into larger regulations. Such regulations would comply with the numerical limit on regulations, but have the same economic impact as an unbounded regulatory system.

> The regulatory budget operates by close analogy to the traditional fiscal process. For example, each year (or at some longer interval), the government would establish an upper limit on the costs of its regulatory activities to the economy and would apportion this sum among the individual regulatory agencies. This would presumably involve a budget proposal developed by a regulatory oversight body in negotiation with regulatory agencies, approved by the executive branch of government, and submitted for legislative review, revision and passage. Once final budget appropriations were in force, each agency would be obliged to live within its regulatory budget for the time period in question.[34]

A regulatory budget provides the bounded institutional counterpart to the conventional unbounded regulatory environment.[35] Although the concept of a regulatory budget is more than thirty years old, regulatory budgets have been implemented sparingly. This section analyzes when to choose regulatory budgeting vs. conventional regulatory oversight as methods of reducing the costs of agency errors and bias.

The analysis of the previous two Parts offers several reasons to believe that bounded institutional structures such as regulatory budgeting may prove superior to traditional unbounded oversight methods. Bounded structures are particularly attractive when agent bias and error are more pervasive, when there are no accurate rules to restrict agent discretion, when the cost of agent errors are non-linear, and when the sample population assessed by an agent is large. These features describe many regulatory environments. At the same time, other features of the regulatory environment, such as the principal's probable ignorance of the population distribution of regulations, and the difficulty of quantifying a regulatory budget, counsel against universal application of a bounded institutional structure such as a regulatory budget. Instead of an all or nothing approach to regulatory budgeting, wherein a regulatory budget is either applied to all agencies or none, the analysis provided here suggests that a regulatory budget may be appropriate for some agencies but not others. Alternatively, a regulatory budget may be appropriate for the head of an agency to delegate to a sub-agency, but inappropriate for Congress to demand of an entire agency.

1. EPA Regulation—The Case for Bounded Institutions

To be concrete, consider possible environmental regulation issues by the EPA. When promulgating environmental regulations, the EPA serves as an agent of Congress and the President under a variety of statutes, including the Clean Air Act, etc. Many have accused the EPA of having a pro-environmental, anti-business, bias. Oversight mechanisms such as cost-benefit analysis, executive oversight via OIRA, and judicial oversight, focus heavily on the EPA's regulations. The EPA has authority to consider a wide range of regulations, from clean water standards[36] to greenhouse gas emissions.[37] In

[34] Nick Malyshev, *A Primer on Regulatory Budgets*, 2010 OECD J. ON BUDGETING, no. 3, at 2.
[35] A deregulatory agenda could be accomplished by a negative regulatory budget, which would force regulators to lower the cost of existing regulations.
[36] *See, e.g.,* 33 U.S.C. § 1251 (2012).
[37] *See, e.g.,* 42 U.S.C. § 7521 (2012).

addition, crafting environmental regulations is hard to specify by rule. Indeed, many environmental statutes specify vague standards for the EPA to follow.[38]

Suppose that the EPA is considering a set of regulations to remove air pollutants subject to its authority under the Clean Air Act.[39] Further suppose that the EPA is biased: it places a value on clear air that is double the value that would be prescribed by Congress. In addition, Congress lacks the ability to specify by rule the regulations that it views as desirable. Finally, suppose that Congress has a sense of the amount of GDP it is willing to spend on clean air and specifies this number in a regulatory budget.[40]

Under these conditions, a regulatory budget outperforms conventional regulatory oversight mechanisms. The budget compels the EPA to regulate to Congress's preferred amount. But even if the EPA is constrained in the costs of regulations it can promulgate ("how much regulation"), how do we know that the agency will choose the right regulations ("which regulations")? The EPA chooses the right regulations because it values clean air twice as much as Congress does. It will try to maximize the value of clean air, and doing that requires that the EPA choose the regulations that provide the cleanest air possible subject to the budget constraint imposed by Congress. These are the same regulations that Congress would choose. The EPA values clean air twice as much as Congress, but its ranking of "clean air per dollar" is the same. As a result, the EPA chooses the "right" regulations. While the EPA would prefer more clean air regulations than Congress, the regulatory budget prevents the EPA from promulgating the extra regulations. Because Congress has specified its clean air "budget", the fact that the EPA values clean air more than Congress imposes no costs.[41]

Conventional oversight mechanisms, by contrast, fail to guarantee that the EPA's bias in favor of Clean Air produces the right regulations. Judicial oversight means that the EPA will choose the regulations that are easiest to justify under the relevant oversight standard. Choosing regulations in this manner leads to over-regulation because of the EPA's bias. In addition, the EPA's regulations may not produce the cleanest air per unit cost; instead of seeking the most efficient regulations, the EPA seeks the regulations that are most likely to pass judicial muster.

A similar story applies to executive oversight. The EPA will attempt to implement its clean-air bias by choosing regulations that appeal to its executive overseers at the Office of Information and Regulatory Affairs (OIRA) or other relevant centers of oversight within the White House. As with judicial oversight, these regulations will likely be greater in quantity and less efficient than the regulations that would be chosen with a regulatory budget.

Cost benefit analysis (CBA) provides the closest analogue to a regulatory budget among unbounded regulatory oversight mechanisms. CBA requires agencies to quantify the benefits and the

[38] *See, e.g.*, 42 U.S.C. § 7403 (2012); 33 U.S.C. § 1254 (2012); 42 U.S.C. § 6912 (2012); 42 U.S.C. § 11002 (2012).
[39] 42 U.S.C. §§ 7401-7671 (2012).
[40] The validity of this questionable assumption will be discussed below.
[41] This assumes that Congress, as the Principal, places the "right" value on environmental quality such as clean air. This value may be high or low. A high value would correspond to a quantitatively large regulatory budget. The key assumption is that, whatever Congress or another Principal desires, the EPA has alternative preferences (bias).

costs of each regulation. Regulations are warranted if benefits exceed costs. Cost estimation is therefore required of both regulatory budgeting and cost benefit analysis. CBA differs from regulatory budgeting in asking for quantification of regulatory benefits.

CBA's quantification of benefits yields advantages and disadvantages. On the plus side, accurate calculation of costs and benefits produces optimal regulation. Congress wants regulations for which benefits exceed costs and accurate CBA realizes this aim. The bounded structure of regulatory budgeting, by contrast, introduces rigidity that is costly if agents accurately estimate costs. An agency subject to a regulatory budget may forego positive value CBA projects, or undertake negative value CBA projects.

CBA's disadvantages arise when the agent has a biased view of the benefits of a regulation. In our hypothetical, cost benefit analysis produces too much regulation because the EPA values clear air (the benefits of regulation) more than Congress. Using CBA, the EPA will choose the right regulations—the ones that produce the most clear air per dollar -- but there will be too much regulation because of the EPA's over-valuation of clear air. If OIRA or some other oversight body can "correct" the EPA's overvaluation, then CBA can produce the right amount and types of regulations, but this requires the oversight body to effectively monitor the EPA's estimate of both costs and benefits.

By bounding the cost of regulations, by contrast, regulatory budgeting does not demand that Congress or any other oversight body obtain good information about the value of clean air. So long as the EPA shares Congress's rank ordering of benefits of clean air and the costs of the EPA's regulations are quantifiable, the bounded regulatory budget produces an efficient outcome.

2. EPA Regulations—The Case Against Bounded Institutions

The previous section told a rosy story about the benefits of a bounded institution-- regulatory budgeting— relative to the performance of more conventional unbounded institutions in the context of EPA regulation. The case for regulatory budgeting, however, rested on several assumptions that may be unrealistic. This section considers the efficacy of the bounded institution when these assumptions are relaxed.

a) Quantifiability

In order to impose a bound, Congress (the principal) must be able to quantify and measure the value of the bound. In the regulatory budgeting context, the bound is the cost imposed upon the public by regulations issued by the agency. Unlike the discretionary budgeting process that regulatory budgeting aims to emulate, the cost of regulations cannot be known precisely. Congress knows the size of the check it writes to the EPA to cover expenses. The costs the EPA imposes on others via regulation, by contrast, has no neat answer. Instead, regulatory costs must be estimated. Regulatory budgeting thus requires that agency discretion is curtailed by a bound premised on an estimate. Other oversight mechanisms, such as judicial oversight, may be flawed, but they don't rely on a false precision.

The critique is an important one. Regulatory budgeting is not like ordinary budgeting. If estimates of regulatory costs prove to be inaccurate and/or manipulable, then regulatory budgets do

not yield the intended outcomes. If the EPA manipulates costs to appear lower than they actually are, for example, then the EPA will produce too much clean air regulation. And if some types of regulations have costs that are easier to manipulate than others, then we may get the wrong regulations in addition to having too much regulation.

But we should also not overstate the quantifiability problem for regulatory budgeting. CBA, which assumes a large role in the EPA regulatory process, requires the EPA to estimate both costs and benefits. Regulatory budgeting requires only costs to be estimated.[42] And costs, which are often direct and pecuniary (such as the cost of pollution reducing equipment for a power plant), are likely easier to estimate than benefits.[43] If CBA can work, then so can regulatory budgeting.

b) Knowledge of the Distributional Parameters

A more problematic assumption that boosted the case for a regulatory budget for the EPA was the presumption that Congress knows the distribution of regulatory outcomes. This means that, although Congress cannot properly evaluate any particular regulation, it has a good sense of the "population parameters" of environmental regulations.[44] A Congress that knows the universe, if not the particulars, of environmental regulations, can reasonably choose the dollar value of regulation it wants, leaving the identity of the particular regulations to the EPA as agent. But if Congress doesn't have a sense of what is out there, than its bound will likely be flawed. And an EPA subject to a flawed bound may produce worse outcomes than an unbounded but biased EPA.

Legislators are not environmental rule-makers, and so Congress is unlikely to have a good sense of the distribution of the effects of environmental regulations. This provides a strong argument against the implementation of a regulatory budget by Congress upon the EPA.

But bounded institutions in the form of regulatory budgets may have benefits at other points in the regulatory oversight process. For example, OIRA, which is more sophisticated than Congress in evaluating environmental regulations, may be able to acquire a sense of the distribution of possible environmental regulations, even if does not have the resources to examine every possible regulation. In addition, OIRA may have less of a pro-environmental bias than the EPA.[45] As a result, a regulatory budget formed by OIRA may have the advantages of a bounded institution without some of the disadvantages that accompany a regulatory budget formed by Congress. Indeed, learning the

---

[42] This is not to say that regulatory budgeting eliminates the problem of estimating benefits. The next section considers the problematic assumption that Congress knows the distribution of outcomes from different regulations. This assumption essentially assumes away the problem of estimating benefits.

[43] *See, e.g.*, Thomas O. McGarity, *A Cost-Benefit State*, 50 ADMIN. L. REV. 7, 57 (noting "the cost side of the equation implicates fewer "soft" considerations than the benefits side."); Thomas O. McGarity & Ruth Ruttenberg, *Counting the Cost of Health, Safety, and Environmental Regulation*, 80 TEX. L. REV. 1997, 2000 (2002)("The benefits side of a typical cost-benefit analysis is quite controversial, laden with huge uncertainties, based largely upon numerous modeling exercises, and does not rely to any large degree upon empirical analysis. The cost side of the analysis is less controversial, but still fraught with uncertainty.").

[44] Because there are presumably an infinite number of very bad regulations, knowledge of the population parameters requires Congress to know the distribution of plausible environmental regulations.

[45] I take no position on the statutory framework necessary to allow OIRA to form a regulatory budget.

distribution of regulatory possibilities and then "getting out of the way" by imposing a regulatory budget may prove to be an easier task for OIRA than the task of overseeing all regulations and cost benefit analyses.

We also should not be too quick to dismiss the possibility of a Congressional regulatory budget for the EPA. Congress knows little about the distribution of environmental regulations, but it also knows little about budgeting requirements for different agencies and somehow passes an annual appropriations bill (or at least a continuing resolution). Regulatory budgeting may not be all that different from conventional bounded budgeting procedures.[46]

The EPA for example, might submit an annual regulatory budget request to Congress each year. These submissions might give Congress the opportunity to learn more about the distribution of regulatory effects for environmental regulations. While Congress is unlikely to attain a comprehensive understanding of the population of existing and possible environmental regulations, it, or at least the relevant committee staffs, may be able to achieve a rough sense of the possibilities. If regulatory bias is a significant problem, then Congress may be better off with a regulatory budget based on imperfect distributional information rather than the unbounded regulatory systems currently in operation.

B. Mandatory vs. Discretionary Spending

We have already examined the problem of government spending on scientific research through the NSF. But the bounded vs. unbounded divide pervades government appropriations. Government spending takes two primary forms—discretionary and mandatory. Discretionary spending means "the budget authority controlled by annual appropriations acts and the outlays that result from that budget authority."[47] Appropriations acts (or continuing resolutions extending previous appropriations acts), specify spending amounts for federal government activities. For example, the Consolidated Appropriations Act of 2012 specifies that, "For compensation of the President, including an expense allowance at the rate of $50,000 per annum as authorized by 3 U.S.C. § 102, $450,000."[48]

Mandatory spending means "budget authority and outlays provided by permanent laws."[49] Medicare, for example, is enacted by 42 U.S.C. § 1395 et seq and does not require annual renewal. Instead, Medicare and Social Security are funded by Federal trust funds that require mandatory transfers from designated revenue sources.[50] Many mandatory spending programs, such as Social Security and Medicare, are commonly known as "entitlement" programs.

---

[46] *See infra*, Section V.B.

[47] Office of Mgmt & Budget, *OMB Circular No. A-11* § 20.9, at 35 (2013), http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/s20.pdf.

[48] Consolidated Appropriations Act of 2012, Pub. L. No. 112-74, § 5, 125 Stat. 786, 892 (2011).

[49] *OMB Circular No. A-11 § 20.9, supra* note 47, at 35. .

[50] See 42 U.S.C. § 1320b-15(a)(2012), providing that "No officer or employee of the United States shall—
(1) delay the deposit of any amount into (or delay the credit of any amount to) any Federal fund or otherwise vary from the normal terms, procedures, or timing for making such deposits or credits." Federal funds are defined in 42 U.S.C. § 1320b-15(c) to mean "(1) the Federal Old-Age and Survivors Insurance Trust Fund [Social Security for retirees];
(2) the Federal Disability Insurance Trust Fund [Social Security for the Disabled];

Entitlement spending programs constitute unbounded institutional structures. Congress is the principal, and the agency charged with administering the entitlement program (such as the Social Security Administration, or the Centers for Medicare and Medicaid) is the agent. There is no restriction on how much or how little is spent on entitlement programs. Instead, the costs of entitlement programs are generally determined by eligibility requirements. For example, Medicare applies to all citizens and permanent residents aged 65 and older.[51] In entitlement programs, the "to whom" question is specified by law, but the question of "how much" is unbounded. Indeed, the "how much" question is generally determined on an as-needed basis. If doctors and patients demand more medical spending for those over age 65, then Medicare will cost more. The beginning of the Medicare Act, for example, states that

> Nothing in this subchapter shall be construed to authorize any Federal officer or employee to exercise any supervision or control over the practice of medicine or the manner in which medical services are provided, or over the selection, tenure, or compensation of any officer or employee of any institution, agency, or person providing health services.[52]

Discretionary spending programs, by contrast, are bounded institutional structures. Congress is the principal, and the department or agency receiving the appropriation is the agent. A department or program cannot spend more in a given year than its appropriation. Per the Consolidated Appropriations Act, the President's salary is $450,000, no more nor less.[53] Once spending has been appropriated, the executive branch even has limited ability to spend less than the appropriated amount.[54] The appropriation provides a bound to the cost of any programs. This bound contrasts with the explicitly unbounded nature of Medicare spending described above.

Most mandatory spending programs can be converted into discretionary spending programs. Medicare is an unbounded entitlement spending program, but it could be converted to a bounded discretionary spending program by appropriating a certain amount each year for each of its functions. This would require Medicare "rationing", but rationing is commonplace in government spending. The administrator of the Medicare program, the Centers for Medicare and Medicaid Services (CMMS), was appropriated a certain amount, $3,879,476,000 in fiscal year 2012,[55] and must "ration" this appropriation in order to carry out CMMS's responsibilities as well as possible. So Medicare's administrators are subject to rationing, even if the program they administer is not. Indeed, there are many proposals to convert unbounded entitlement spending programs into bounded programs. Proposals to convert Medicaid payments to states from sharing formulas to block grants of fixed dollar

---

(3) the Federal Hospital Insurance Trust Fund [Medicare Part A]; and
(4) the Federal Supplementary Medical Insurance Trust Fund [Medicare Part B]."

[51] *See* 42 U.S.C. § 1395c (2012).

[52] 42 U.S.C.§. 1395 (2012).

[53] *See* Pub. L. No. 112-74, § 5, 125 Stat. at 892.

[54] Under the Congressional Budget and Impoundment Control Act of 1974, Pub. L. No. 93-344, 88 Stat. 297 (codified as amended at 2 U.S.C. §§ 621 – 691f (2012)), the President can propose that certain amounts appropriated by Congress be rescinded. If both Houses of Congress do not approve this proposal, then the appropriation must be made available for obligation. *Id.*

[55] *See* Pub. L. No. 112-74, § 5, 125 Stat. at 1075.

amounts seek to transform Medicaid from an unbounded spending program to a bounded spending program.[56]

Conversely, most discretionary spending programs can be converted into mandatory spending programs. Instead of appropriating money to departments and agencies on an annual basis as it does with bounded discretionary spending, Congress could pass laws funding such departments in perpetuity, on an as-needed basis. The President's salary could be determined by an agency that set salaries to be comparable to similar positions in other sectors. And while as-needed funding may sound curious, recall that this is exactly the unbounded framework that characterizes entitlement programs such as Medicare and Social Security.[57] Medicare pays for all eligible expenses, with no explicit rationing.

Scholars have offered surprisingly little analysis of the efficacy of mandatory vs. discretionary spending programs. The bounded vs. unbounded institutional structure developed here provides a framework for understanding this question. Mandatory spending rules work best in contexts that favor unbounded rules. Discretionary spending formulas function better when a bound cabins agent error effectively.

From a positive perspective, the bounded vs. unbounded framework developed above offers some traction for examining when discretionary or mandatory spending programs should be applied. Both discretionary and mandatory spending programs appear to be deployed in contexts where they are more likely to be effective.

The largest entitlement spending programs often determine eligibility by rule according to a clearly definable metric. Medicare, for example, applies to permanent residents over age 65.[58] Social security benefits and eligibility are determined by a statutorily defined rule that is a function of contributions to the program and age.[59] Medicaid and S-Chip eligibility are primarily functions of income, age and family status—all quantifiable and verifiable metrics.[60] The analysis above suggested that when rules regarding verifiable metrics are available, unbounded institutional structures tend to outperform bounded structures. Bounded institutional structures impose some rigidity but reduce bias. Because rules reduce the scope for bias, imposing a bounded structure when a rule is available—as in the case of Medicare, Social Security, and Medicaid—introduces rigidity with relatively little scope for bias reduction.

---

[56] *See, e.g.*, Paul Ryan, *The Path to Prosperity: Restoring America's Promise* 38-40 (Apr. 5, 2011), http://budget.house.gov/UploadedFiles/PathToProsperityFY2012.pdf. While the Ryan proposal for providing block grants to states for Medicaid expenses combined the transformation of the program to a bounded structure with an attempt to cut costs, this need not be so. Block grants could be adjusted annually to account for actual medical price inflation, but retain a bounded structure in the sense that they are block grants.

[57] It is also not dissimilar from how salaries are set for executives of major companies. Compensation consultants typically choose a peer group of executives to the subject executive and propose a salary for the subject executive based on the salary of this peer group. *See* Lucian Arye Bebchuk et al., *Managerial Power and Rent Extraction in the Design of Executive Compensation*, 69 U. CHI. L. REV. 751, 790-91 (2002).

[58] *See* 42 U.S.C. § 1395c (2012).

[59] *See, e.g.*, 42 U.S.C. § 402 (2012).

[60] *See* 42 U.S.C. § 1396(a) (2012).

But the efficacy of rules is not constant across all entitlement programs. Rules function particularly well in the context of the Old-Age and Survivors (OAS) insurance component of the Social Security system.[61] Because the Social Security benefit rule for retirees answers both the questions of "to whom" and "how much", imposing a bounded dollar amount on the program would introduce rigidity without diminishing any obvious sources of bias.[62] For Medicare and Medicaid, rules provide a clear answer to the eligibility question ("to whom"), but no obvious answer to the question of "how much" the program should spend. At present, the amount is determined via the discretion of other agents, such as doctors and patients. These agents may well be biased, as a host of studies demonstrate. As a result, Medicare may be a candidate for a rule/bound combination, in which the rule determines eligibility and the bound determines how much can be spent. Indeed, the bound may be expressed in per capita terms, e.g. a budget of $X per eligible participant, to minimize the rigidity imposed by the bound.[63] With a hard budget in place for a fixed group of subjects, agents such as doctors and administrators would find biases constrained by the budget. Forced to keep spending within the bound, they may allocate more effectively in the partially bounded context than in the unbounded context that characterizes Medicare and Medicaid currently.

The disability insurance ("DI") component of Social Security[64] also creates tensions for the unbounded structure. For disability, the "how much" per person question is determined by rule. Once someone is determined to be disabled, they receive benefits according to a formula based on past earning and the type of disability.[65] Determining disability ("to whom"), however, is not amenable to rule delimited decisionmaking. At present, disability determinations are made through a complex administrative process that must consider a wide variety of illnesses and definitions of disability.[66] This process, which is controlled by administrators, may be prone to bias or systematic error. In addition, the population of individuals applying for DI is likely to be relatively constant or changing predictably from year to year. Bounds are well suited to cabin agent bias or error under these conditions.

The bound on DI could be introduced in one of two related ways. First, Congress could place a bound on the number of disability recipients per year. If each administrator of the system receives a random draw of a large number of DI applicants, then the bound could be subdivided such that each administrator can make a bounded number of disability determinations. A "curve" for DI determinations

---

[61] 42 U.S.C. § 402 (2012).

[62] One may think that the benefit rule is overly generous or overly stingy. This is not a problem of bias by the agent, but rather a choice of the principal that formulated the benefit rule. As a result, a bounded budget constraint as with discretionary spending would not solve the problem. The principal's generosity or stinginess would carry over to the bounded budget.

[63] If there are an uncertain number of Medicare participants but a fixed budget, then the uncertainty about the number of participants adds another layer of uncertainty to the costs of the program. If the bound is expressed in per capita terms, the uncertainty about the number of participants does not affect the total uncertainty about the (per capita) costs of the program.

[64] 42 U.S.C. § 423 (2012).

[65] *See* Ellen O'Brien, *Social Security Disability Insurance: A Primer*, AARP PUB. POL'Y INST. 11 (Apr. 2009), http://assets.aarp.org/rgcenter/econ/i28_ssdi.pdf.

[66] *See* SOC. SEC. ADMIN., *Disability Evaluation Under Social Security*, http://www.ssa.gov/disability/professionals/bluebook/index.htm (last modified June 26, 2013).

would fulfill many of the criterion for an optimal bounded structure. DI determinations are made by agents who may be biased and error prone, but likely share a general ranking of disability severity with Congress. In these circumstances, a fixed number of "yes" determinations per agent ameliorates bias and error and does not cause a great deal of incorrect DI decisions because each administrator sees a large number of randomly drawn DI applicants and the distribution of disability in the population is likely constant or predictably changing from year to year. Second, a bounded dollar budget could be placed on the disability insurance system. In this case, some primacy rule would have to decide what to do when the bounded budget conflicted with the benefits mandated by rule to the applicants granted disability insurance.

Discretionary spending programs often display the characteristics of programs that would benefit from bounded institutions. Consider the budgets of most cabinet departments and administrations. Unlike income support programs such as Social Security, it is difficult to specify budgets via rule—the decision regarding how much should be spent is too multidimensional. Delegating budgeting decisions to agents in an unbounded way may also lead to considerable bias. Thus, a mandatory program calling for agencies to receive whatever budget they request would likely lead to too much spending. The agency will often have a more generous view of its spending needs than Congress. Just as the EPA, for example, may issue regulations that exceed Congress's ideal level, so too might the EPA request a greater appropriation than deemed warranted by Congress because a higher appropriation enables the EPA to better protect the environment. In addition, the annual nature of the appropriations process gives Congress an opportunity to get a sense of the distribution of needs in different agencies and how likely these needs are to vary. This learning process allows Congress to set reasonable bounds. In total, many program subject to discretionary spending allocations have attributes that make them conducive to bounded institutional structures.

While the bounded institutional structure of many discretionary spending programs is appropriate, there are other programs that are currently bounded that the analysis developed above suggests should be unbounded. Consider FEMA spending. Disasters, and especially large disasters, are rare events. Their number and cost may change dramatically from year to year. With such a variable distribution, a bounded institutional structure can produce bad outcomes. If the budget is fixed but there is a much higher than expected number of disasters, then disaster victims may receive much less than Congress would ideally like. If there are fewer disasters and FEMA has a fixed budget, then disaster victims may receive too much. Both possibilities are costly and will occur frequently if FEMA has a fixed budget to provide for disaster relief. A mandatory spending program that dictates that disaster victims should receive what FEMA thinks appropriate, though prone to bias would avoid the bad outcomes of over-allocation in the event of too few disasters or under-allocation if there are too many disasters. Of course, Congress can augment a discretionary annual appropriation with an emergency appropriation in the event of higher than expected disaster costs. This effectively replaces a bounded discretionary spending allocation with an unbounded allocation. But passing supplementary bills imposes costs. Congress may not be able to pass bills efficaciously, as the recent brouhaha concerning a relief bill for

the costs of Hurricane Sandy demonstrates.[67] Congress is also unlikely to pass a bill reducing FEMA spending if there are fewer disasters than expected.[68]

Some types of defense spending also appear ill-suited for bounded spending restrictions. While some aspects of defense spending have predictable distributions from year to year (e.g., regular personnel costs, long term procurement programs), other parts of military spending are less predictable. Wars are expensive, and we cannot rely on the law of large numbers to guarantee that the expense average out. Exclusive reliance on bounded defense spending would produce too much spending in peaceful years and too little in times of war. As a practical matter, our budget process recognizes this mismatch. While much of defense spending is bounded and "discretionary", additional expenditures required for war are often the subject of supplementary budget proceedings.[69] But, as with FEMA spending, this is simply another way to make a bounded program less bounded. Passing additional war related spending also imposes transaction costs and it is unlikely that times with less than expected war will lead to a reduction in military expenditures. Both of these considerations weigh in favor of an unbounded spending program for war related costs.

The fact that wartime spending does not fit all the criteria for an ideal discretionary program does not mean that an unbounded program would be superior. War related decisions may be subject to considerable bias from agents, which a bound mitigates. In addition, war related decisions are difficult to specify via rule. As a result, the hybrid system of today, with bounded defense discretionary spending and periodic war related supplementary budgets, may do the best job of containing bias (by requiring debate) while enabling the flexibility needed to handle uncertain war expenses.

C. Minority Set Asides and *Croson*

The bounded vs. unbounded framework applies to the question of "minority set asides" discussed by the Supreme Court in *Fullilove*[70] and *Croson*.[71] Minority set asides sought to remedy past and current discrimination by requiring minimum percentages of the value of government contracts be used to hire minority business enterprises. In *Croson*, for example, the City of Richmond, with an African American population of 50%, "required prime contractors to whom the city awarded construction contracts to subcontract at least 30% of the dollar amount of the contract to one or more Minority Business Enterprises (MBE's)."[72]

The *Croson* policy and those of other states constitutes a bounded institutional structure. The principal-- the city of Richmond—sought to ameliorate past and present discrimination. It therefore bounded the choices of its agents—the prime contractors. Concerned about bias by the prime

[67] *See, e.g.*, Raymond Hernandez, *Stalling of Storm Aid Makes Northeast Republicans Furious*, N.Y. TIMES (Jan. 2, 2013), http://www.nytimes.com/2013/01/03/nyregion/congressional-members-blast-house-for-ignoring-storm-aid-bill.html.
[68] If FEMA can withhold its spending, then the over-allocation problem is reduced. Allowing FEMA to withhold spending, however, transforms MA's institutional structure from bounded to partially bounded.
[69] *See, e.g.*, Emergency Wartime Supplemental Appropriations Act of 2003, Pub. L. No. 108-11, 117 Stat. 559.
[70] *See* Fullilove v. Klutznick, 448 U.S. 448 (1980).
[71] *See* City of Richmond v. J. A. Croson Co., 488 U.S. 469 (1989).
[72] *Croson*, 488 U.S. at 477.

contractors towards some subjects (MBE subcontractors), Richmond required that 30% of subcontracts be granted to MBEs.[73]

The City of Richmond could have chosen an unbounded structure to compensate for possible agent bias. For example, Richmond could have insisted that prime contractors give MBEs a "close look". With this unbounded policy, Richmond would be stating its policy preferences but giving the agents discretion over fulfilling the city's preferences. Alternatively, Richmond could have said that any bid by MBEs that was within 5% of the lowest non-MBE bid must be accepted. This would impose a rule on subcontracting with minorities, but would not have bounded the dollar value of subcontracts with MBEs.

As detailed in the previous sections, Richmond's choice of minority set asides—a bounded structure-- offered advantages and disadvantages. If the City of Richmond cannot judge subcontractor bids but has a good sense of how many MBEs it is wants to hire in order to overcome discrimination, then the minority set aside can produce a first best outcome. The prime contractors would choose the best MBEs using their superior judgment of quality, but the prime contractor's racially biased preferences relative to the city would be offset by the set aside requiring 30% of value be granted to MBE subcontractors.

In other circumstances, the set aside policy would bring inefficiencies relative to unbounded alternatives. If prime contractors are not particularly biased against MBEs, then the set aside introduces rigidity without any benefit. If Richmond does not have a realistic sense of its preference for MBE subcontractors, then the set aside could produce very inefficient results, with too many expensive or ill-qualified MBEs. If the quality of MBEs varies widely, or if there are only one or two subcontracts per contractors, then there is a much greater chance that the set aside will be much greater than what the city would want if it had perfect knowledge of subcontractor quality and race.

This rigidity was criticized by the Supreme Court in finding Richmond's Policy unconstitutional under the Fourteenth Amendment:

> The Richmond Plan denies certain citizens the opportunity to compete for a fixed percentage of public contracts based solely upon their race. To whatever racial group these citizens belong, their "personal rights" to be treated with equal dignity and respect are implicated by a rigid rule erecting race as the sole criterion in an aspect of public decisionmaking.[74]

There is no doubt that, as described, the bounded set aside creates a "rigid rule". But the analysis of this paper demonstrates that, under the right conditions, many of the Supreme Court Majority's critiques of Richmond's policy are inapt. Indeed, a bounded set aside can produce an ideal outcome that would not be feasible with a seemingly less rigid rule.

---

[73] *Id.* The bound was one dimensional. Prime contractors could hire more than 30% MBEs. On the high side, MBE subcontracting was unbounded (or bounded at 100%).

[74] *Id.* at 493.

When there are many MBE and non-MBE contractors, a set aside becomes one factor among many that determine who receives subcontracts. For example, a bounded rule, such as a set aside, does not "erect[] race as the sole criterion of public decisionmaking."[75] The set aside insures that race is accounted for, but the use of bounded criterion allows for consideration of an infinite number of dimensions of "quality". Instead of requiring subcontractors to comply with a formula for how to weigh MBE status against other dimensions of quality such as price and expertise, the set aside lets each contractor makes decisions based on the criteria they think is most important, while insuring that the interest in ameliorating discrimination is realized. In addition, the bound may treat people with "equal dignity."[76] If the size of the set-aside quota accurately reflects Richmond's legitimate interests in ameliorating past discrimination, then the set aside allows contractors to make subtle and individualized decisions among MBEs and non-MBEs.

In *Croson*, there were not many (indeed only one) MBEs. In these circumstances, bounded institutions such as set asides perform poorly. Set asides introduce rigidity, such as (in *Croson*) requiring the only marginally qualified MBE to be chosen as a subcontractor. Such rigidity indeed precludes equal dignity and respect. But this does not mean that set asides are generally a bad idea. In cases where current or past racial bias is prevalent, there are many MBE and non-MBE subcontractors, there is no viable rule dictating the weight to be placed on race relative to other factors, and the principal cannot judge the quality of subcontractors, then, as shown in Part III, set asides can produce excellent outcomes.

To sum up, analyzing *Croson* in light of our study of bounded vs unbounded institutions suggests that the problem is not racial set-aside per se, but rather the use of set asides in appropriate circumstances.

D. Judicial and Quasi-Judicial Decision-Making

In many contexts, society charges an agent with making a series of decisions about subjects. The agents' decisionmaking process is often unbounded—even though the agent may well be biased in a socially undesirable manner. The determination of disability benefits described above is one such circumstance. Society has some sense of who should receive disability and who should not. If an administrative law judge systematically awards benefits to too many or too few applicants than society would prefer, then there is a cost. Bounded institutional structures can mitigate the costs of agent bias.

Bounded institutional structures work best when the distribution of the subject population coming before the agent is relatively constant and large, the agent's decisions are quantifiable, and the agent's shares a rank ordering of subjects with the principal.

These conditions may be met in many other judicial and quasi-judicial decisionmaking processes. Consider criminal sentencing. A judge makes sentencing decisions on a large number of offenders who are randomly assigned to the judge. Some judges impose harsher sentences, while others

[75] *Id.* at 493.
[76] *Id.*

are more lenient. Society demonstrates great concern about the possible biases of these judges.[77] Indeed, the Federal Sentencing Guidelines attempted to reduce the scope of judicial bias in sentencing[78] by imposing intricate sentencing rules upon judges. The Guidelines generated and continue to generate considerable controversy.[79]

A bounded sentencing structure offers an alternative means of constraining judicial bias without intruding into the judicial decision making process. Each judge in a district[80] could be given a "sentencing budget" that is determined by Congress.[81] The budget could be specified for each offense type[82] or for all offenses put together and could span several years if necessary to minimize the problem of random variations in subject population. With a sentencing budget and random assignment of offenders to judges, each judge should have a relatively similar population of offenders, particularly if the time frame is long enough to allow natural variation to balance out. Under these conditions, a bounded structure (a sentencing budget) reduces bias with little cost in rigidity so long as the judge shares a rank ordering of offense severity with society.[83] In effect, society says "we expect you to have to give out x years of

[77] *See* James M. Anderson, et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J. L. & ECON. 271, 275 (1999); Ryan W. Scott, *Inter-Judge Sentencing Disparity After Booker: A First Look*, 63 STAN. L. REV. 1, 6 (2010).

[78] *See* 28 U.S.C. § 991 (2012)(stating the purpose of the Sentencing Commission is "avoiding unwarranted sentencing disparities among defendants with similar records who have been found guilty of similar criminal conduct"); U.S. SENTENCING COMM, *Guidelines Manual* 13 (2012), http://www.ussc.gov/Guidelines/2012_Guidelines/Manual_PDF/Chapter_1.pdf; Anderson, *supra* note 57, at 273.

[79] *See, e.g.*, Albert W. Alschuler, *Disparity: The Normative and Empirical Failure of the Federal Guidelines*, 58 STAN. L. REV. 85 (2005)("The Federal Sentencing Guidelines have failed to reduce disparity and probably have increased it."); Frank O. Bowman, III, *The Failure of the Federal Sentencing Guidelines: A Structural Analysis*, 105 COLUM. L. REV. 1315 (2005); Owen S. Walker, *Litigation-Enmeshed Sentencing: How the Guidelines Have Changed the Practice of Federal Criminal Law*, 25 U.C. DAVIS L. REV. 639 (1992).

[80] The budgets should be decided at the district level because this is the level at which defendants are randomly assigned. Different districts have different populations of offenders and should therefore have different sentencing budgets that reflect the difference in defendants.

[81] If the current sentencing schedule is viewed as appropriate on average but beset by excess inter-judge disparities, then each judge in the district's budget could be set to the average of the current sentence.

[82] The advantage of specifying the budget per offense type is that it reduces the possibility that sentencing will be misallocated because offense types are randomly but unevenly allocated across judges. The disadvantage of specifying the budget per offense type is that the number of offenders in each offense type will likely be smaller, raising the possibility of error. One possible solution is to specify the judge's overall budget by offense type, but allow the judge to allocate sentences across offense types. To illustrate, suppose that offense A has a budget of 2 years and offense B has a budget of 3 years. The average judge sentences the same number of offense A violators as offense B, meaning that the average sentence per offender is 2.5 years. Over a time period, a particular judge sentences 10 "A" offenders and 6 "B" offenders. If the sentencing budget is set per judge with no revision for offense type, then our judge will have a total of 16*2.5=40 years of sentences to the 16 offenders that need to be sentenced by our judge. If the sentencing budget is per offense type, then the judge will have a budget of 20 years to sentence the 10 violators of A and 18 with which to sentence the 6 violators of B. If the sentencing budget is allocated per sentence type but the judge can average across sentence types, then the judge will have a budget of 38 years with which to sentence both the 10 type A offenders and 6 type B offenders.

[83] The assumption that judges share a rank ordering of offense severity with society is reasonable but far from inevitable. As a member of society, the judge may have internalized many norms that are difficult to specify by guidelines but are shared by most citizens. A judge who has internalized society's norms will share society's rank ordering. The judge, however, may not share society's rank orderings. A racist judge, for example, may deviate

sentences, you decide how to do it." The x years insures that neither harsh judges nor lenient judges are able to impose their biased views of sentencing upon the offenders who happen to come before them. If the judges share society's rank ordering of "to whom", then a sentencing budget mitigates or even eliminates bias without the intrusiveness of an intricate rule based system such as the sentencing guidelines.

As discussed above, bounded institutional structures also offer promise in the context of judging immigration asylum cases. Indeed, bounded institutional structures can ameliorate bias by agents in almost every context in which bias by agents is a problem. We can imagine "stop, question, and frisk" budgets assigned to the police to bound their ability to engage in a particularly aggressive form of policing; "summary judgment budgets" to standardize the difficult to police burden of "summary judgment"; "tort budgets" to reign in judges or districts that are perceived as too plaintiff friendly or defendant friendly; "discrimination budgets" to address disparities by judges in propensity to allow employment discrimination suits"; sentencing budgets for prosecuting attorney's offices to mitigate the problem of prosecutorial over-reach or under-reach; patent budgets to insure that the Patent and Trademark Office or individual patent examiners do not award patents to innovations that are insufficiently novel or fail to reward innovation by refusing to grant patents.[84]

In all of these cases, and many others that have not been identified, the bounded institutional structure—some form of hard budget constraint--- offers bias-reduction at the cost of rigidity. Whether this trade-off is justified depends upon the institutional setting. Settings that meet more of the bounded institution favoring factors described above are more conducive to the use of the bounded budget.

One particularly important factor in all of these contexts is the ability to quantify and not manipulate the item subject to the budget. For example, if a "stop, question and frisk" is ambiguously defined, then the ability to bound agent's behavior is limited.[85] Agents can manipulate the ambiguity in the definition to avoid the bias-ameliorating features of the bounded constraint. Thus, the ability to quantify and not manipulate the bounded constraint is worth re-emphasizing here.

E. Timing

When grading, a professor views the entire distribution of traits at one time.[86] In the examples mentioned above, however, the agent evaluates the subject's over time. This may lead to some

---

from society's ranking by imposing overly harsh sentences on the offenders of a certain race. In these conditions, the sentencing budget does not eliminate the problem of bias. So if the problem of sentencing disparities is due to over-harshness or leniency on the part of judges, then a bounded sentencing budget would work well. If the problem is racism, however, then a bounded sentencing budget would not solve the problem, unless the sentencing budget could be racially adjusted. In this case, the judge would receive a sentencing budget by race, limiting the judge's ability to indulge their anti-social racial preference orderings.

[84] See Jonathan Masur, *Patent Inflation*, 121 YALE L.J. 470, 479 (2011).

[85] While "stop question and frisk" seems like it would be subject to greater ambiguity than other quantities discussed here, it is sufficiently definable and quantifiable to collect data on the subject that has been used to critique the police departments who use the tactic. See Andrew Gelman, Jeffrey Fagan, & Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. OF AM. STAT. ASS'N 813 (2007).

[86] I thank Sara Light for pointing out this concern.

complications. For example, early within a sentencing budget period, a judge or prosecutor may be excessively cautious, lest there be more egregious criminals later in the period with a lack of sentencing resources to allocate.[87] Or a judge may find that she has allocated too much of her sentencing budget to her earlier offenders and finds that there is little left in the "budget" for a heinous offender.

Several considerations suggest that, though the timing problem is real, it should not overly restrict the scope for bounded institutional constraints. First, any budget has a timing problem. When Congress allocates money for a purpose, the recipients of the money must insure that the funds last the year. With practice and careful observation, agents can be taught to use the budget appropriately.[88] Second, there may be ways to adjust the treatment of the early arrivals ex-post to insure equal treatment with late arrivals. For example, if a judge finds that she has been too strict with her early offenders, she can be allowed to cut some of the sentences she has already meted out in order to insure an adequate budget for the remainder of the period. Third, variation of treatment by timing is undesirable, but it may be less troublesome than the existing inter-agent variability. The inter-agent variability is due to systematic factors that differ between agents while the variation due to timing is harder to attribute to pernicious forces.

Fourth, and most importantly, the potential costs of the timing problem for bounded constraints can be mitigated via the use of a bounded interval rather than an exact numerical bound. Judges, for example, can be given a sentencing budget interval rather than a precise sentencing number. In our example of the previous section, the judge may be granted a sentencing budget of between 35 and 45 months rather than a hard bound of 40 months. The interval allows for some slack in the budget, mitigating the harm caused by over or under-aggressive use of the budget on early arrivals relative to late arrivals. Of course, the interval also partially undermines the purpose of the bounded constraint because the interval allows some scope for biases to be realized. Nevertheless, a bounded interval combines some of the benefits of bounded constraints—preventing the realization of systematic bias that is greater than the size of the interval—while mitigating some of the bounded constraints most significant costs.

VI. Conclusion

This paper asked whether public policy should or should not use bounds. Under the right conditions—a principal who has a good sense of the overall distribution of subject quality, biased agents who share a rank ordering of subjects with the principal, and large numbers of subjects for each agent, bounded institutions can overcome the principal agent problem and produce excellent outcomes. As a result, bounded institutional structures should be seriously considered whenever traditional rules are unfeasible but unfettered standards give too much discretion to agents.

[87] For an analysis of the allocation of a given budget to subjects that arrive over time, see Yair Listokin & Kenneth Ayotte, *Protecting Future Claimants in Mass Tort Bankruptcies*, 98 NW. U. L. REV. 1435 (2004).

[88] This consideration favors a gradual rather than abrupt introduction of a bounded constraint where there has not been one previously. A gradual introduction allows learning that will prevent large-scale errors.

VII. Appendix

VIII. Appendix

A. Model Setup

There are *N i.i.d.* subjects. Quality of subject i is given by θ_i.These attributes are distributed according to (θ) . At some cost, k, the principal can learn $f(\theta)$. Suppose $\theta \sim U[0,1]$.

Agent a observes quality according to the following process :

$\delta_{i,a} = b_a\theta_i + \varepsilon_{i,a}$ (Equation 1)

Where b_a is the agent's bias expressed multiplicatively (relative to the principal) and $\varepsilon_{i,a}$ is an mean zero error term. Determining δ is free for the agent. For simplicity, assume that there is one agent (meaning we will drop the a subscript).

The agent makes an allocation to subject i, d_i based on the agent's observation of the subject's trait.

For simplicity, assume that the principal values error costs in a linear fashion and puts equal weight on over-allocations and under-allocations. For each subject, error costs are equal to

$E_i = |\theta_i - d_i|$ (Equation 2)

Assume that the agent attempts to make the best allocation (minimize errors) given the observations of subjects and the constraints imposed by the principal. When the agent is subject to an unbounded institutional structure, $d_i = \delta_i$ (UB).

With a bounded institutional structure, the agent is given a per subject bound equal to the population mean, . (With a uniform distribution on the unit interval, the bound is equal to $.5$.

Each agent will now award each subject $d_i = \delta_i * \frac{\mu}{\frac{1}{N}\sum_{i=1}^{N}\delta_i}$ (B).

For each subject, error costs are

$E = \left|\theta_i - \delta_i * \frac{\mu}{\frac{1}{N}\sum_{i=1}^{N}\delta_i}\right|$ (3)

The principal aims to reduce per subject (and therefore total) error costs. The principal should choose an unbounded structure if (2)<(3) and a bounded structure if (3)<(2).

B. Section III.A.

Assume that the principal knows μ with certainty and that the agent makes no errors and has no bias. ($b = 1$ and $\varepsilon = 0$ are zero.) Thus, $\theta_i = \delta_i = d_i$. In this case, the error costs with unbounded structures, equation 2, are zero. The agent observes the subject's trait without error or bias and allocates to the subject according to the signal.

With a bounded structure, by contrast, error costs (equation 3) are positive. $\theta_i = \delta_i$, and so error costs are zero only if the ratio of the population mean to the sample mean is always one. The ratio of the population mean to the sample mean is one in expectation, but the sample mean is a random variable, implying the existence of errors. The errors are proportional to the sample means's standard error, which is $\frac{\sigma}{\sqrt{N}}$, where σ is the standard deviation of the population. Thus, the added cost of choosing a bounded institutional structure when agents make no errors rises when the distribution of a trait in the population is more variable. The error cost of a bounded structure goes down when there are more subjects evaluated by the agent. With more subjects, the standard deviation of the sample mean goes down, making the bound more accurate.

C. Section III.B.

Now assume that $b = \bar{b} \neq 1$ and that $\varepsilon_i = 0$. With condition UB, combining equation 1 and equation 2 yields $E_i = \left|\theta_i - \left(\bar{b}\theta_i\right)\right| = \left|(1 - \bar{b})\theta_i\right|$.

With condition B, equation 3 becomes $E_i = \left|\theta_i - \left(\bar{b}\theta_i\right) * \frac{\mu}{\frac{1}{N}\sum_{i=1}^{N}(\bar{b}\theta_i)_i}\right| = \left|\theta_i - \frac{\mu(\bar{b}\theta_i)}{\bar{b}\frac{1}{N}\sum_{i=1}^{N}(\theta_i)}\right| = \left|\theta_i - \theta_i \frac{\mu}{\frac{1}{N}\sum_{i=1}^{N}(\theta_i)}\right|$.

The principal chooses a bound whenever $\left|\theta_i - \theta_i \frac{\mu}{\frac{1}{N}\sum_{i=1}^{N}(\theta_i)}\right| < \left|(1 - \bar{b})\theta_i\right|$.

As N goes up, $\frac{1}{N}\sum_{i=1}^{N}(\theta_i)$ converges to μ. Thus, as N gets very high, a bounded allocation produces almost no error. The bound corrects for the agent's bias, and the lack of agent error means that, correcting for bias, each subject gets the appropriate allocation. The unbounded allocation, by contrast, has a constant error per subject as a result of the agent's bias. For large N, therefore, the bounded structure proves superior to the unbounded structure.

D. Section III.C

Now assume that $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. With uncorrelated errors, $Corr(\varepsilon_i, \varepsilon_j) = 0$ .With correlated errors, $Corr(\varepsilon_i, \varepsilon_j) \neq 0$.

Assume that the agent does not know $f(\theta)$. The principal continues to know the distribution of θ. As a result, the agent cannot make inferences regarding error terms from the final distribution of observed traits.

1. Uncorrelated Errors

Assume that θ is normally distributed with mean μ and variance σ_1^2. In this case, the agent's signal, δ, is the sum of two normally distributed variables. The signal , δ , is normally distributed with mean μ and variance $\sigma_1^2 + \sigma_\varepsilon^2$. The variance of the signal, δ ,is therefore greater than the variance of quality, θ. For the agent to report a distribution of quality that reflects the true population distribution, the agent should allocate not solely according to the signal, but rather the Bayesian posterior probability of quality, given the signal.

$$f_{\theta|\delta=\bar{\delta}}(\theta) = \frac{f_\theta(\theta) L_{\theta|\delta=\bar{\delta}}(\theta)}{\int_{-\infty}^{\infty} f_\theta(\theta) L_{\theta|\delta=\bar{\delta}}(\theta) d\theta}$$

Because the agent not know $f(\theta)$, the agent cannot use this conditional distribution. If, as a result, the agent is inclined to allocate based on the signal δ, then the agent will find too many high and low values of δ relative to the true population, θ.

2. Correlated Errors

For simplicity, assume that the agent errors are perfectly correlated. $Corr(\varepsilon_i, \varepsilon_j) = 1$. In this case, equation 1 becomes $\delta_{i,a} = b_a \theta_i + \varepsilon$. When agents make correlated errors, then the case for bounds is stronger because the unbounded structure will produce an per-unit error of $E_i = |\theta_i - d_i| = E|\varepsilon|$. The unbounded structure produces consistently positive error. A bounded structure that adjusts for additive errors in a way analogous to the multiplicative error adjustment described in Appendix Part C will therefore have lower expected error costs.