JOURNAL OF LAW, ECONOMICS & POLICY

VOLUME 2 SPRING 2006 NUMBER 1

# CONTENTS

# THE INFLUENCE OF STATE LEGAL ENVIRONMENTS ON FIRM INCORPORATION DECISIONS AND VALUES

*Stephen P. Ferris,*[*] *Robert M. Lawless,*[**] *& Gregory Noronha*[***]

ABSTRACT

This study examines whether the state legal environment influences where IPOs elect to incorporate and their subsequent market value. To examine these questions, we develop a new measure of the state legal environment that incorporates both the presence of critical statutes and the willingness of a state to innovate. We conclude that this new measure offers legal academics, practitioners, and others interested in corporate finance a highly convenient and quantitative evaluation of a state's corporate legal climate. Our empirical use of this measure yields important cross-sectional variations in state legal environments, with the result that the most pro-management state is Pennsylvania. We also find that firms exhibit a willingness to separate their operational headquarters from the state of incorporation in a manner consistent with the pro-management orientation of the state legal code. Finally, we find that the state legal environment does affect firm value, but in a way that is consistent with a "race to the bottom" view of corporate law.

---

[*] Stephen P. Ferris is the J. H. Rogers Chair of Money, Credit and Banking and Professor of Finance at the University of Missouri-Columbia. He is also the director of the Financial Research Institute, an academic research center located on the MU campus and housed in the College of Business. Prior to his appointment to the Rogers chair, Dr. Ferris served as the chair of MU's department of finance for ten years. He received a B.A. from Duquesne University, a M.S.S. from the U.S. Army War College, and a M.B.A. and Ph.D. from the University of Pittsburgh. Dr. Ferris' research is in the areas of corporate finance, international finance, governance and the legal/finance interface. He has published over 100 referred academic research articles.

[**] Professor Robert Lawless is a nationally acclaimed expert in business organizations, corporate law, and bankruptcy law and joined the University of Illinois College of Law faculty after serving as the Gordon & Silver, Ltd. Professor of Law at the University of Nevada-Las Vegas William S. Boyd School of Law. Professor Lawless received both his B.S. (Accounting) with highest honors and J.D. summa cum laude from the University of Illinois, and served as Editor-in-Chief of the University of Illinois Law Review. Professor Lawless is intensely interested in empirical legal studies and interdisciplinary studies in business law.

[***] Greg Noronha is Professor of Finance at the Milgard School of Business at the University of Washington Tacoma. He has published articles in the Journal of Finance, the Journal of Banking and Finance, and the Journal of Financial Research, among others. He earned his Ph.D. in Finance from Virginia Tech. He is a CFA Charterholder.

## 1. INTRODUCTION

In the United States, laws in a firm's state of incorporation govern its internal affairs, regardless of where that firm's primary place of business is located. For example, the law of Delaware governs the internal affairs of Anheuser-Busch, Inc., although the headquarters and most of that brewer's operations are located in Missouri. Because Anheuser-Busch is incorporated in Delaware, that state's law governs whether the shareholders are entitled to vote on a particular matter or whether a plaintiff must first make a demand on the board of directors before bringing a lawsuit alleging that directors did not act in the best of interest of shareholders. Even a court from another state would apply Delaware law to answer these questions.

Because the state of incorporation need not bear a relationship to the firm's operations, the choice to incorporate in a particular state can be conceptualized as a "purchase" of that state's entire body of corporate law. Each state provides a set of rules (the law) and procedures (the judicial system) to resolve disputes that might arise within the corporation. If the decision to incorporate in a particular jurisdiction can be characterized as a "purchase," then each state can be characterized as competing for incorporations. Delaware is the clear winner of the competition, as most large corporations have chosen to incorporate in Delaware. The interesting question is "why Delaware?"

Cary (1974) posits a "race to the bottom," where Delaware wins by pandering to managers who essentially control the incorporation decision. Winter (1977) and others argue that there is a "race to the top." Specifically, Winter observes that pro-management rules penalize shareholders, and that, with a competitive market for capital, shareholders will demand a higher return from firms incorporated in pro-management states. In the long-run, this is not a sustainable equilibrium. Hence, Winter contends that Delaware must win what is ultimately a "race to the top" by having a body of corporate law that is most appealing to shareholders.

Bebchuk and Cohen (2003) examine the determinants of the firm's choice of where to incorporate and conclude that there is a significant home-state advantage in the market for corporate law. Specifically, they conclude that firms tend to incorporate in the state where their corporate headquarters are located. They attribute the impact of additional incorporation fees, the ability to influence local politics, and a perception of only marginal differences in state corporate law as factors that operate to keep firms in-state.

These arguments imply testable relationships between state corporate law and measures of firm activity. The "race to the top" and "race to the bottom" debate implies that both corporate managers and shareholders can identify critical differences in state corporate law, thus suggesting that a state's corporate legal environment will affect its rate of incorporation.

Bebchuk and Cohen (2003) argue, however, that firms will incorporate in those states where they are headquartered, and that managers do not shop for the most favorable state corporate law. Hence, the first research question examined in this study is the extent to which differences in state corporate law effects the incorporation decision of managers. Our empirical results suggest that businesses are sensitive to the opportunities provided by state corporate law. We find that firms separate their decisions of where to headquarter and to incorporate based on the favorableness of the legal environment to management.

Daines (2001) reports controversial findings that firms incorporated in Delaware are worth more than similar non-Delaware firms after controlling for size, industry, growth opportunities, and financial performance.[1] We examine this issue by constructing a new ranking measure for the corporate legal environment of each of the fifty states. Our ranking measure extends research by Romano (1985) who similarly constructs a variable examining state adoption of four pro-management statutes. More specifically, Romano examines the extent to which "director and officer" (D&O) indemnification, merger vote exemption, appraisal rights exemption, and antitakeover statutes influence a firm's choice of state for incorporation.

The state legal environment measure (LEM) we create simultaneously includes the presence or absence of key statutes relating to managerial authority and the ability of the state to innovate in corporate law. We construct our LEM such that higher values indicate the presence and more rapid adoption of pro-management laws. This new measure allows a more comprehensive assessment of the state's legal environment and permits greater sophistication in the empirical analysis than that provided by the coarse binary classification of Delaware versus all other states.

A pro-shareholder corporate legal environment should be associated with higher firm values while the opposite should hold for states with a pro-management tilt. Thus, the second research question of this paper examines the effect that a state's legal environment, measured with our more comprehensive metric, exerts on corporate values. We find in our empirical results that higher LEM values are positively associated with incorporation rates across the various states. We also determine that LEM values are inversely related to firm value.

Beyond introducing a new measure of individual state legal environments, this research is conceptually related to an emerging literature on the role of national legal regimes on firm behavior and value. In a series of studies, La Porta et al. (1997, 1998, 1999, 2000) find that international corporate law differs along a number of critical dimensions that ultimately impact how firms operate and are valued in the market. This study also

[1] More recent research by Bebchuk, Cohen, and Ferrell (2002), Subramanian (2002), and Bebchuk and Cohen (2003) contends that the findings of higher values for Delaware firms are spurious and do not hold after 1996.

examines the impact of legal institutions and other practices on corporate value, but over a much smaller physical territory. Thus, as firms become increasingly global, the "race to the top vs. the race to the bottom" debate will expand in geographical reach as international legal regimes compete with each other for business activity and incorporations.

## 2. THE STATE LEGAL ENVIRONMENT

### 2.1. *The nature of the legal environment*

To measure a state's legal environment, we examine the content of each state's corporate statutes. A state's corporate code is a particularly apt measure of the legal environment. The statutes form the basis for a state's corporate laws, and courts must conform their decisions to the statutes' content. If the legislature disapproves of a court's decision, the legislature can override the decision. Moreover, a legislative body sets its own agenda. It can enact legislation on any topic it desires as opposed to a judicial body which can only decide the issues placed before it. In considering a state's reputation for corporate innovation, firms should look primarily to the statutes enacted by its legislature.

To create the LEM, our empirical proxy for the state legal environment, we first consider whether a state has adopted any of the following eight statutes. As explained more fully below, each of these statutes generally has a pro-management orientation. Thus, higher LEM scores indicate a more pro-management state. The eight statutes are as follows:

(1) *Statutory Indemnification Standard* (first enacted by New York in 1961): These statutes eliminate the court-created doubt that a director or officer would be entitled to indemnification for litigation expenses and settlements. These statutes also expand the situations under which indemnification would be available and articulate clearer standards upon which a director or officer might rely.

(2) *Publicly Traded Company Appraisal Exception* (first enacted by Delaware in 1967): In a merger, shareholders of a target company often have a statutory right to bring a judicial action to ascertain the true value of their shares. These statutes eliminate this "appraisal right" if the acquiring company's stock is used as consideration in the merger and the target company's stock is publicly traded. Through these statutes, management has greater ability to override dissident shareholders in a merger.

(3) *Director & Officer Exculpation* (first enacted by Delaware in 1986): These statutes authorize companies to limit or eliminate the liability of directors for all but disloyalty and other intentional bad acts.

(4) *Short-form Merger Statutes* (first enacted by New York in 1949): A short-form merger statute authorizes abbreviated procedures for a merger

between a parent corporation and its controlled subsidiary, most notably excusing a shareholder vote and making it easier to override dissident shareholders.

(5) *First Generation Antitakeover Statute* (first enacted by Virginia in 1968): These statutes required registration of tender offers with state regulators and, in some states, authorized state regulators to review the merits of a proposed tender offer. Thus, these statutes made it much more difficult to eliminate management in a hostile tender offer. The United States Supreme Court declared these statutes unconstitutional in *Edgar v. MITE Corp.*, 457 U.S. 624 (1982).

(6) *Second Generation Antitakeover Statute* (first enacted by Ohio in 1982): Also called "control share statutes," these statutes specify certain thresholds of ownership (often zero percent, 33.3 percent, and 50 percent) that require shareholder approval for an acquirer to cross. Without shareholder approval, the "control shares" lose their voting authority. Management retains authority to call special shareholder meetings to remove the control share restrictions. Thus, these statutes again enhanced managerial power against hostile takeovers.

(7) *Third Generation Antitakeover Statute* (first enacted by New York in 1985): Under a third-generation statute, an acquirer crossing certain ownership thresholds, (e.g., 20 percent) needs shareholder approval. In the absence of shareholder approval, the acquirer is barred from continuing with a statutory merger. Management, acting through incumbent directors, retains the power to approve friendly transactions, meaning that the statutes effectively work against only transactions hostile to management interests.

(8) *Fourth Generation Antitakeover Statutes* (first enacted by Ohio in 1983): Also called "other constituency" statutes, these statutes empower the board of directors to reject value-increasing takeover offers on the grounds that the offer would harm the interests of other corporate constituencies, such as creditors, employees, or the community at large. These statutes effectively allow management to cloak self-aggrandizing decisions in the rhetoric of employee interests.

Within the universe of state corporate statutes, these eight statutes unambiguously enhance managerial discretion or rewards. Each of these statutes either financially benefits managers, protect managers from takeovers they disapprove, or increase managerial discretion. Our selection of statutes is consistent with the observation of Romano (1985) and Carney (1998) that the greatest variation in state corporate law will occur with rules that appeal most to managers. Our selection of statutes also compares favorably to that made by Gillan, Hartzell, and Starks (2002) in their construction of a corporate governance index. Our measure of the state legal environment captures the variability in state corporate statutes in those areas of greatest interest to managers as they contemplate the incorporation decision.

Not only the content of a state's corporate law, but also its reputation for innovation should matter in the incorporation decision. Firms will want to choose states that possess responsive legal structures as legal institutions and thought evolve. If a state lacks the reputation for responding to innovation, firms are likely to incorporate in other states to avoid the need for reincorporation at some future date. By including a state's reputation for innovation, the LEM score captures more variation between the states than a strict count of whether a state has adopted a particular provision. Similar to Romano (1985), our LEM is sensitive to the effect of innovation on incorporation rates in the fifty states. We improve upon Romano's measure, which had a 15-year horizon, by including more than 50 years worth of data, in an effort to capture the speed of state innovation in corporate law. We measure state innovation as the actual time it takes for a state to adopt a statute relative to its first adoption by any state. In addition, our measurement considers a greater number of statutes than those considered by Romano, and it includes the statutory exculpation statutes, which were not yet enacted at the time of her study.[2]

### 2.2. *Construction of the Legal Environment Measure (LEM)*

We begin our calculation of the LEM by constructing a vector of binary variables for each state that measures the presence or absence of each of the eight statutes described in section 2.1. Specifically, the binary variable, $\text{STATUTE}_{i,j}$ captures whether state $i$ has adopted statute $j$. The variable assumes a value of one if the state has adopted a particular statute and is coded zero otherwise. Each of these binary variables is then weighted by a factor that captures the relative speed with which the state adopted that statute.

To calculate the relative speed of adoption, we first determine $\text{YRADOPT}_{i,j}$, which is defined as the year in which state $i$ adopted statute $j$. The first adoption by any state is represented as $\text{YRADOPT}_{first,j}$. Thus, the difference as measured in years from state $i$'s adoption of a specific statute $j$ and when the amendment is first adopted can be expressed as:

$$\text{DIFF}_{i,j} = \text{YRADOPT}_{i,j} - \text{YRADOPT}_{first,j} \qquad (1)$$

---

[2] Our selection of statues partially overlaps with the four studied by Romano (1985) in the areas of statutory indemnification, appraisal exception, and first generation antitakeover. Areas of difference between our LEM and Romano occur with D&O exculpation, short form merger, and second, third, and fourth generation antitakeover statutes.

Because we need to initialize $DIFF_{i,j}$ for the first adopter of a statute with a value of one, we modify equation (1) as follows:[3]

$$DIFF_{i,j} = (YRADOPT_{i,j} + 1) - YRADOPT_{first,j} \quad (2)$$

We then invert $DIFF_{i,j}$ so that higher values imply an earlier adoption of management-friendly statutes by a specific state. The value of $DIFF_{i,j}$ for the first adopter remains at a value of one while those for subsequent adopters decline in appropriate relative progression.

To obtain the LEM value for a given state *i*, we begin by multiplying the vector of dummy variables indicating the presence or absence of a specific statute by the values of $DIFF_{i,j}$, which represent the speed of statute adoption as described in equation 2. The resulting values are then summed to yield the estimate of LEM for a specific state *i*:

$$LEM_i = \sum_{j=1}^{8} (STATUTE_{i,j}) \text{ X } (1/DIFF_{i,j}) \quad (3)$$

Higher LEM scores thus simultaneously imply the presence of more management-friendly statutes and the earlier adoption of these statutes by a particular state. To allow convenient scaling of LEM, we multiply the values obtained using equation (3) by 1000.

### 2.3. *Commentary on the LEM*

The LEM scores capture two attributes of the state's corporate legal environment. First, it includes the presence or absence of a specific statute in that state's corporate legal code. Second, it incorporates a measure of the speed (relative to the first adopter) with which a state adopts these amendments. Higher LEM scores imply not only a greater number of each of the eight statutes, but also greater speed in the adoption of these statutes.

---

[3] The favorableness of the state legal environment from a corporate perspective is influenced by both the adoption of these eight statutes as well as the speed by which a state adopts them. Hence, construction of the LEM requires that the influence of both of these factors is in the same direction. Consequently, first adopters of a statute can not have a value of zero for their speed of adoption, although that might appear intuitively appealing. Because the LEM requires that the statute adoption/non-adoption dummy variable is weighted by the relative speed of adoption, a weighting of zero for an initial adopter will cause that statute to be erroneously eliminated from the LEM calculation. Hence, we initialize first adopters with a $DIFF_{i,j}$ of 1. All other adopters have values of DIFF that decline from 1. This approach correctly allows the maximum weighting to the first adopters of a particular statute while maintaining the relative ranking of subsequent adopters.

The LEM reflects a state's long-term reputation for a timely response to new innovations in corporate law since it measures statutory adoptions since 1949. Because the eight statutes contained in the LEM enhance managerial perquisites, power or discretion, higher LEM scores imply a state that rapidly responds to pro-management innovations.

Table 1 reports the LEM scores for each of the 50 states along with select summary statistics. The scores range from a high of 27.77 for Pennsylvania to 5.78 for Vermont. Surprisingly, Delaware ranks only seventh in the LEM standings. In spite of its reputation for innovation, Delaware has enacted only one of the four antitakeover statutes we identify as essential for measuring a state's legal environment. The effect of the interaction between a willingness to innovate and statute adoption on LEM values is demonstrated with Idaho and Indiana. These states have adopted all four antitakeover statutes, but score lower than Delaware because they are slow to innovate and have an inconsistent record of adopting other statutes.

Because our construction of the LEM includes the responsiveness of the state legal environment in addition to the actual adoption, it is interesting to compare our rankings with Daines (2001). In his examination of the impact of Delaware incorporations on firm value, Daines identifies what he believes are the ten most responsive state legal environments using Romano's (1985) definition of that concept. The LEM rankings overlap with those of Daines for six of the states, suggesting a moderate consistency in the measurement of legal responsiveness.

The LEM values are not normally distributed and most states are clustered toward the lower end. Figure 1 presents a histogram of the LEM scores. Only a few states have high LEM scores. The mean (median) LEM score is 11.68 (10.20) with a standard deviation of 5.11. This distribution of LEM scores corresponds to the conventional legal perception of the competition for corporate charters. Generally, state competition for charters has been conceptualized as skewed, with most states ceding the market to a few states that vigorously compete such as Delaware, Nevada, and Pennsylvania.

The LEM presented in this study improves on previous measures of the states' hospitability for corporate law. First, it captures much more of the variation between state legal environments than the more commonly used dichotomous classification of Delaware versus all other states. Second, it accounts for a state's long-term reputation for adopting management-friendly innovations in corporate law. Still, it must be acknowledged that the LEM is not a perfect measure of the state legal environment since it does not account for case law. LEM considers the statutes that a state legislature adopts, but it does not capture rules developed in judge-made law. Although the state legislature has the primary responsibility for lawmaking, state judicial decisions can have important effects on the development of state corporate law. Delaware has

the most developed body of corporate case law, and consequently attracts a significant portion of the market for corporate charters.

Case law tends to be more ambiguous than statutory law, making empirical measurements difficult. For example, the Delaware Supreme Court arguably gave Delaware corporations the equivalence of an "other constituency statute" by declaring in 1985 that a board could consider the effect of a hostile bid on "constituencies" other than shareholders (i.e., creditors, employees and perhaps even the community generally).[4] A little over five months later, the same court refused a defense that a board of directors was acting to protect the interests of noteholders.[5] Assessing Delaware case law on this subject would require a qualitative judgment about the interplay of these two cases and perhaps other subsequent cases.[6]

Nevertheless, the LEM does capture some of the variability in case law between the states, as shown in Table 2. We examine each state's case law from 1945 through 2000, identifying the number of reported opinions in each state that the Westlaw database provides under the topical heading for "Corporations." We separately identify the number of corporate-law opinions from the lower state courts, the state supreme courts and the federal district courts in each state. The correlations between the LEM and these different measures of reported corporate law opinions are statistically significant. Also, the LEM score generally increases in value as we move across the quartiles based on the number of cases. For instance, the LEM increases from 9.98 for states in the bottom quartile of total state court cases and increases to 13.23 for the top quartile. The difference in the LEM score between these two quartiles is statistically significant as indicated by the t-statistics. Although the LEM expressly includes only state statutory law, it is generally responsive to variation in state case law.

### 3. DATA AND SAMPLE CONSTRUCTION

To test the usefulness of the LEM in capturing the state legal environment, we construct a sample of nearly 4,900 firms consisting of all Initial Public Offerings (IPOs) from 1994-2000 that have data contained on the Center for Research in Security Prices (CRSP) and Compustat databases. The Compact D/New Issues database provides additional data concerning the details of the issue, the state of incorporation, and the state location of corporate headquarters.

---

[4] Unocal Corp. v. Mesa Petroleum Co., 493 A.2d 946, 955 (Del. 1985).

[5] Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 182 (Del. 1985).

[6] *Compare* Credit Lyonnais Bank Nederland v. Pathe Communications Corp., 1991 WL 277613 (Del. Ch. Dec. 30, 1991) (directors may owe duties to creditors when corporation is "in the vicinity of insolvency") *with* Alderstein v. Wertheimer, 2002 WL 205684 (Del. Ch. Jan. 25, 2002) (directors owe duty to creditors in "vicinity of insolvency," but may not ignore interests of shareholders).

A firm's initial public offering is a propitious time to observe the choice of incorporation. A firm initially chooses its state of incorporation at the time of formation. It is easier and cheaper to change the state of incorporation prior to becoming public than after. A firm might subsequently reincorporate in another state, but reincorporation formally requires a vote by the shareholders. After the public issuance of securities, a shareholder vote requires the firm to satisfy significant regulatory hurdles, such as the preparation of a proxy statement and the sponsoring of a general shareholder vote. Compared with a post-IPO change of incorporation, a pre-IPO change is relatively costless. Therefore, we treat the date of the IPO as the firm's decision to incorporate in a particular state.

The sample is distributed randomly over our sample years, with no evidence of clustering. The sample firms are distributed across nine broad industrial groups. These industry groups and their accompanying number of IPOs are as follows: agriculture (15), mineral industries (107), construction industries (48), manufacturing (1,660), transportation, communications and utilities (321), wholesale (216), retail (370), finance, insurance and real estate (909), and service industries (1,240).

## 4. EMPIRICAL RESULTS

### 4.1. *State corporate law and business migration*

If state corporate law is an important consideration in the incorporation decision of firms, then we should expect to observe a pattern in such incorporations. To examine this issue, we introduce the concepts of emigrating and immigrating firms. When a firm lists its corporate headquarters in one state, but incorporates in a different state, we refer to that firm as an emigrating business. Specifically, from the viewpoint of the state in which it is headquartered, the firm is an émigré if it chooses to incorporate in another state. From the perspective of the state in which the firm is incorporating, the firm is an immigrant from another state. We use both terms in this examination of the firm's incorporation decision.

We begin our analysis by separately estimating for each state the percentage of its IPOs over our seven-year sample period that consists of emigrating firms. We repeat the calculation for the percentage of immigrating firms. Based on the medians for each of these variables, we construct four equally sized subsamples. Thus, we construct a subsample of high-emigration states and a subsample of low-emigration states. Then based on the median value of the percentage of immigrating firms in a state, we develop corresponding subsamples of high-immigration and low-immigration states.

Table 3 contains the results of our empirical analysis of patterns in corporate migration. In Panel A, we present the mean (median) LEM for

each of the sub samples described in the preceding paragraph. We also provide the mean LEM for the aggregate sample as a benchmark. Panels B and C contain the results of our comparison of the state legal environment across the subsamples. We first compare the LEM between states in the high-immigration subsample with those in the low-immigration subsample. We find that states in the high-immigration subsample have significantly higher LEM values than do those in the low-immigration subsample. This finding is consistent with the claim that firms will leave the state where they are headquartered and incorporate in a state where the legal environment is more attractive for managers. It is also interesting to note that the LEM for the high-immigration subsample is greater than the LEM for the overall sample. Further, the LEM for the low-immigration subsample is lower than the aggregate sample's LEM.

The converse of the above finding is that firms will leave states in greater numbers when that state has an unfavorable legal environment for managers. We test this proposition by comparing the average LEM of states with a high rate of corporate emigration with that for states having a low rate of corporate emigration. We find that low-emigration states have significantly higher mean (median) LEM values than do states with more emigrating firms. The comparisons with the aggregate sample provide further evidence that high-emigration states have an unfavorable state legal environment, while the low-emigration states enjoy a state legal code that is above average in its attractiveness to corporate management.

As our final comparison, we estimate the difference in average LEMs between states with a high percentage of immigrating firms and those with a high percentage of emigrating firms. If the state legal environment is important, then the average LEM should be larger for states receiving a high percentage of immigrating firms and correspondingly lower for states with a high percentage of emigrating firms. Our findings support this hypothesis. The mean LEM for the high-immigration states is 13.08 while that for the high-emigration states is 10.29. The difference between these values is statistically significant. The comparisons with the aggregate sample are likewise consistent with this hypothesis.

We conclude from Table 3 that the state legal environment does exert an influence on the decision by firms of where to incorporate. Firms demonstrate a willingness to separate their operational headquarters from the state of incorporation in a manner consistent with a rational assessment of the overall attractiveness of the state legal environment. Because the LEM captures the "management friendliness" of a state, our evidence is most consistent with a "race to the bottom" view of the state competition for corporate charters.

### 4.2. *Multivariate analysis of the incorporation decision*

#### 4.2.1. Model specification and design

In Table 4, we examine the explanatory power of the LEM in the presence of other variables that are theoretically related to the incorporation decision of firms. In the regressions that we report, we use four different dependent variables to capture the incorporation rates of the various states. We use: (1) the absolute number of firms incorporated in a state, (2) the absolute number of immigrant firms in a state, (3) the percentage of incorporations from immigrant firms within a state, and (4) the percentage of incorporations from non-immigrant firms within a state.

Our regression models presented in Table 4 contain a number of independent variables. Our variable of interest, the state legal environment, is represented by the LEM. Because the LEM values are highly skewed, as shown in Figure 1, we use its natural logarithm transformation to make it more consistent with the linear model employed. Our inclusion of the next two independent variables reflect the arguments of Alva (1990), Klausner (1995), and Fisch (2000), who contend that the depth of a state's case law, the expertise of a state's corporate bar, or the character of a state's political structure are important to a state's success in the market for corporate charters. Thus, we construct a variable that measures the total number of state and federal cases contained in the Westlaw database as catalogued under the subject of corporations. This variable is standardized by the state population. The variable labeled "political structure" attempts to capture a state's tendency to have institutions that are resistant to change (i.e., conservative)[7] and therefore promote stability and predictability for corporations. "Political structure" is a constructed variable that ranges from 1 to 5, with each state receiving one point for each of the following characteristics: (a) appointed state appellate courts, (b) appointed state trial courts, (c) no state intermediate appeals court (i.e., direct appeals to the state supreme court), (d) a legislature that meets biennially, and (e) a legislature that is constitutionally limited in the time it may stay in session. To better capture the depth and availability of counsel likely to be skilled in a state's corporate law, we also include as an independent variable the number of attorneys per capita engaged in private practice for each state.

We also include two independent variables suggested by Bebchuk and Cohen (2003). First, they find that a state's adoption of the Revised Model Business Corporation Act ("RMBCA") has a negative effect on incorporations. Because it is a model statute, a state's adoption of the

[7] We use the word "conservative" in its traditional sense, to imply a general tendency toward the status quo, rather than using the word in its more popular sense to denote a certain set of political beliefs.

RMBCA would make its corporate laws more uniform with the corporate laws of other states. Hence, adoption of the RMBCA will tend to reinforce the home-state bias as firms have less incentive to incorporate in another state. Thus, we include a dummy variable that assumes a value of one if a state adopts the RMBCA and is zero otherwise.

Bebchuk and Cohen (2002) also propose that a "liberal political culture" in a state might discourage firms from incorporating in the state. They hypothesize that, for example, a "liberal political culture" that encourages "judicial activism" might discourage incorporations.[8] To capture "liberal political culture," Bebchuk and Cohen use the percentage of voters who chose the Democratic candidate in the 2000 election. To capture a state's long-term reputation for political culture, we average each state's percentage vote for the Democratic candidate in the presidential elections of 1952 through 2000. As Table 4 reports, this variable is uniformly insignificant.[9]

Finally, we include a measure for the economic infrastructure of the state. A state with a robust economy might be more willing to offer the tax incentives to attract incorporations or may provide a stronger demand base for the firm's product. We use as our measure of an individual state's economic infrastructure, the average annual state per capita personal income over the seven years of our sample period. In untabulated results, we also use average annual aggregate state gross product to proxy the state economic infrastructure and obtain qualitatively identical results.

#### 4.2.2. Multivariate empirical findings

The regression results contained in Table 4 offer several interesting findings. First, we observe that the LEM is significantly positive across all

---

[8] It is not clear why Bebchuk and Cohen would hypothesize that the "liberal political culture" would have a negative effect on incorporations. On top of "liberal political culture," they heap on the label of "judicial activism." Regardless of their merits elsewhere, either label is unsatisfactory for the incorporation debate. The intriguing question is "liberal" and "judicially activist" to whom? Is a "liberal" or "judicially activist" political culture hostile to managers or shareholders? In the recent corporate accounting scandals, Democrats—who often are associated with the labels "liberal" or "judicial activist"—have tried to gain political capital by portraying the Republicans as pro-manager at the expense of investors. *See, e.g.*, James Gerstenzang, *Democrats Heap Criticism on Bush Politics*, L.A. TIMES, July 31, 2002, at A13; Dan Balz, *Democrats Assail Bush on Economy and Foreign Policy*, WASH. POST, July 30, 2002, at A5; Jill Zuckman, *Democrats Try to Score on Stock Slide*, CHI. TRIB., July 13, 2002, at 11. In a "race to the top" competition for corporate charters, a pro-investor political culture should attract firms.

[9] In regressions that we do not report, we replicate Bebchuk and Cohen's result for the year 2000 election in a separate analysis, but we do not find any significant relationships for the 1988, 1992, or 1996 elections individually or for averages from 1988 to 1996 or 1988 to 2000. Bebchuk and Cohen's result appears to be limited to only the election of 2000 and might be a coincidental relationship rather than a sign of underlying political culture.

regression specifications. This result indicates that the state's LEM is an important factor in the incorporation decisions of businesses. Because it captures a state's tendency to adopt pro-management statutes quickly, the LEM's significantly positive coefficient is consistent with the "race to the bottom" view of the corporate charter debate. The importance of the LEM is also shown by a comparison of the $R^2$s between model 1 and 2 estimated for each of the four dependent variables. We observe that the explanatory power of the model, as measured by $R^2$, increases when model 1's set of independent variables is expanded to include the LEM. Model 1 augmented by the LEM is the specification for model 2. Our estimation of two different models allows us to demonstrate further the usefulness of the LEM in explaining managerial decisions of where to incorporate.

Other independent variables also demonstrate statistical significance. Total state and federal corporate case law generally demonstrates a significant relation with the different measures of incorporation rates. The variable for political structure exhibits statistical significance with three of the four estimates of the dependent variable. As constructed, higher scores for political structure imply conservative legislative and judicial institutions, capturing the propensity for appointed judiciaries and time-limited legislatures. The significance of political structure supports those who hypothesize that institutional commitment to incremental change is significant in the market for corporate charters. Also, it is consistent with the reputational effects that the LEM suggests. The characteristics that comprise the political structure are deeply rooted institutions, often enshrined in state constitutions. The composition of a state's political structure will change infrequently. Just as the LEM implies, the importance of a state's long-term reputation for adopting statutes enhancing managerial discretion, the political structure variable implies the importance of institutions that are unlikely to change corporate laws quickly. The number of private legal practitioners has some association with attracting immigrant incorporations, although its statistical significance is sporadic across the various model specifications. Neither the binary variable for the adoption of the Revised Model Business Corporation Act nor the average state percentage Democratic vote is statistically insignificant, suggesting the unimportance of these factors for business incorporation decisions.

Finally, we observe that state economic infrastructure exerts a positive, but statistically insignificant impact on the firm's incorporation decision. This result might be due to firms emphasizing legal rather than local economic considerations in making their incorporation decisions. This could be especially true for firms with international operations or business holdings widely distributed throughout the United States.

### 4.3. *The LEM and firm value*

Because state corporate law determines an investor's rights, managerial duties and responsibilities, and resource allocations in various corporate transactions, it affects corporate value. When state corporate law reduces informational asymmetry between managers and investors, reduces the private benefits of insiders, or limits other kinds of agency costs, the value of firms operating under these rules should increase. If, however, the state legal environment facilitates the entrenchment of incumbent management or otherwise encourages managerial slack, then the worth of firms incorporated in such states should fall.

In this section we present our results from a multivariate analysis of the impact of state corporate law on firm value. Consistent with Daines (2001), we use an estimate of Tobin's Q to proxy for firm value and to serve as our dependent variable. Tobin's Q is defined as the ratio of the market value of a firm to the replacement value of a firm's assets. It is widely used as a measure of firm value in empirical corporate finance (e.g., Morck, Shleifer, and Vishny 1988; Lang, Stulz, and Walkling 1989; McConnell and Servaes 1990; Smith and Watts 1992; Berger and Ofek 1995; Cho 1998). Firms that are favorably regarded by the market and viewed as having profitable growth opportunities will have market valuations in excess of the simple replacement value of their assets. This will produce a value of Tobin's Q in excess of one. Thus, more-highly-valued firms will have higher Qs than will lower-valued firms.

Methodologically, it is intractable to estimate replacement value with any reliability. But, consistent with the empirical corporate finance literature (e.g., Berger and Ofek 1995), we can estimate Q as the ratio of the market value of the firm's equity and the book value of its debt to the book value of its assets.

The state legal environment is measured in two different ways. First, we follow Daines's (2001) methodology and estimate the state legal environment with a simple binary variable that assumes a value of 1 if the firm is incorporated in Delaware and 0 otherwise. We then measure the legal environment on a more continuous basis by using the LEM. Daines (2001) finds that the coefficient for the Delaware dummy variable is both positive and significant. We hypothesize, however, that the coefficient on the LEM will be negative because it is a measure of the pro-management orientation in state corporate law and is indicative of the potential agency costs to shareholders.

The remainder of our independent variables is largely consistent with the model specified by Daines (2001) to facilitate comparison. We recognize, however, that our sample, which consists of IPOs during the 1994 to 2000 period, differs from Daines's sample of industrial firms contained on the Compustat database from 1976 through 1996. Thus, any comparison of results can only be suggestive rather than definitive.

Because firm value will be directly influenced by corporate profitability, we include return on assets (ROA) as an independent variable in the model. Both a contemporaneous and a one-year lagged estimate of ROA are included to permit smoothing in reported profitability. Firm size is included in the model to recognize that poorly-designed corporate law is likely to be more disadvantageous and ultimately more costly to the shareholders of larger than smaller firms. To control for well-documented non-linearities in the influence of size on firm value, we estimate firm size as the log of total sales. We also estimate it as the log of total assets with no substantive change in our findings. Myers (1977) and Smith and Watts (1992) note that firm value is partially determined by future investment opportunities. Hence, consistent with Daines (2001), we use the firm's research and development expenses standardized by total assets as a proxy for future corporate investment opportunities. We also control for firm profitability and investment quality in an attempt to control for possible selection bias in our sample construction. We do so by using Altman's (1968) z score as a proxy for the quality of the firm's earnings. Finally, we control for possible industry effects by including a series of one digit SIC code dummies.

#### 4.3.1. Results from a dichotomous classification of the state legal environment

In panel A of Table 5, we capture the state legal environment through a binary classification variable. This binary variable assumes a value of one if the state of incorporation is Delaware and is zero otherwise. We find that the impact of a Delaware incorporation is generally insignificant. In the one model specification where the estimate is significant, it has a negative sign. This result is inconsistent with Daines (2001), who reports a significantly positive impact on firm value with Delaware incorporations. Because our sample varies from Daines (2001), our findings are consistent with other studies that are unable to replicate Daines's results with differing samples over different time periods (Bebchuk, Cohen, and Ferrell 2002; Subramanian 2002; Bebchuk and Cohen 2003).

We find that other variables are more significantly related to firm value than this simple dichotomous measurement of the state legal environment. We find that larger firms, as measured by total assets, have higher valuation ratios, as do firms with greater expenditures on research and development. The firm's ROA also impacts the valuation ratio, with the current ROA positively impacting valuation. The strong inverse relation between the firm's market-to-book ratio and the preceding year's ROA is more difficult to explain, but it might be driven by the divergence between realized and expected ROA for that year. Finally, we observe that the Altman's z score measure for firm quality has a positive influence on

the valuation ratio. This is consistent with the premise that higher quality firms will enjoy greater market valuations.

#### 4.3.2. Results from a continuous classification of the state legal environment

In panel B of Table 5, we present our findings regarding firm value when we measure the state legal environment using the LEM. The results for the coarser, binary classification measured that way did not produce statistically meaningful results, but we find that the LEM produces significantly negative coefficient estimates across the different model specifications. This result is consistent with the argument that the pro-management orientation of high LEM states generates agency costs that are borne by shareholders, and it is ultimately reflected in lower market valuation for the shares of these firms. These findings suggest that the impact of the state legal environment is both subtle and meaningful. It is sufficiently subtle that a coarse binary classification such as Delaware and non-Delaware incorporators is unlikely to capture its effect. The state legal environment is meaningful in that it exerts a statistically significant influence on firm value, even in the presence of other factors that are related to firm valuation.

We find that this result is robust across a variety of model specifications that control for other possible determinants of firm value. Further, our results are robust to the time period in which these variables are measured. More specifically, the variables presented in Table 5 are estimated in the year immediately following the IPO. Re-estimation of the model over the second year following the IPO produces qualitatively identical results and, hence, they are not separately reported.

## 5. CONCLUSION

This paper proposes that state reputation plays an important role in the market for corporate charters. To capture state reputation, we construct a new measure, the LEM, which simultaneously captures eight different statutes that unambiguously enhance managerial authority as well as the rate of legal innovation. Because the states adopt these statutes over a 50-year period, the LEM captures a state's long-term reputation in the market for corporate charters.

Our findings offer several useful conclusions. First, we observe significant cross-sectional variability in LEM scores. Pennsylvania possesses the highest LEM value. Its LEM value is nearly five times as great as that of Vermont, which has the lowest score. Other states long thought to be aggressive players in the market for corporate charters, such as Delaware and Nevada, also have high scores. Yet, surprisingly,

Delaware is behind such states as Tennessee, Wisconsin, New York, Nevada, and Virginia. In spite of Delaware's reputation for innovation and extensive case law, it has only one of the four different antitakeover statutes, thus leading to lower LEM values. LEM also demonstrates a correlation with state legal and political characteristics such as state lower court and state supreme court caseloads. Thus, LEM captures some fundamental state qualities important in the market for corporate charters. We believe that the LEM offers legal academics, practitioners, and others interested in corporate finance, a highly convenient and quantitative evaluation of a state's corporate legal climate.

We find that firms demonstrate a willingness to separate their operational headquarters from the state of incorporation consistent with a rational assessment of the favorableness of the state legal environment. Firms tend to leave the state where they are headquartered and incorporate in states where the legal environment is more favorable toward managers. Likewise, states with an unfavorable legal environment toward managers experience high rates of business emigration while those states with a favorable legal environment have low rates of emigration.

The LEM is negatively associated with firm value as measured by Tobin's Q. Again, this negative association is stronger than that obtained by using a simple binary variable to represent a Delaware incorporation. Because the LEM captures the states' long-term reputation for adopting statutes that enhance managerial authority and power, these results are most consistent with the "race to the bottom" theory of the market for corporate charters.

What is clear from our analysis is that state reputation, as captured by LEM, does have a significant relationship with incorporation rates. For states seeking to attract incorporations, our study suggests that it is partially out of the state's control in the short term. Reputation is a sunk investment, one that the states cannot easily change. In the long term, however, our study suggests that states win incorporation business by adopting statutes and cultivating a reputation for enhancing managerial discretion and prerogatives.

TABLE 1: INDIVIDUAL LEGAL ENVIRONMENT MEASURE (LEM) SCORES AND SUMMARY STATISTICS

The LEM scores simultaneously reflect the presence or absence of eight state corporate statutes that unambiguously enhance managerial discretion as well as the rate of legal innovation for a given state.

*Panel A: Decreasing LEM scores by individual state*

| State Rank | State Name | LEM Score | State Rank | State Name | LEM Score |
|---|---|---|---|---|---|
| 1 | Pennsylvania | 27.77 | 24 | Iowa | 10.20 |
| 2 | Tennessee | 27.02 | 24 | Nebraska | 10.20 |
| 3 | Wisconsin | 22.72 | 24 | South Carolina | 10.20 |
| 4 | New York | 20.00 | 29 | Idaho | 10.00 |
| 5 | Nevada | 19.60 | 30 | Kentucky | 9.61 |
| 6 | Virginia | 19.23 | 31 | Missouri | 9.52 |
| 7 | Delaware | 16.66 | 32 | Colorado | 8.92 |
| 8 | Louisiana | 16.39 | 33 | South Dakota | 8.77 |
| 8 | New Jersey | 16.39 | 34 | Mississippi | 8.69 |
| 10 | Maine | 14.70 | 35 | Illinois | 8.54 |
| 11 | Kansas | 14.49 | 36 | Arkansas | 8.26 |
| 12 | Georgia | 14.08 | 37 | Wyoming | 8.13 |
| 13 | Florida | 13.51 | 38 | California | 8.06 |
| 13 | Massachusetts | 13.51 | 39 | Oklahoma | 7.93 |
| 15 | Michigan | 13.33 | 40 | Washington | 7.87 |
| 16 | Indiana | 12.50 | 41 | North Carolina | 7.81 |
| 17 | Arizona | 12.04 | 42 | Alaska | 7.75 |
| 17 | Maryland | 12.04 | 43 | Texas | 7.09 |
| 19 | Utah | 11.76 | 44 | New Mexico | 6.99 |
| 20 | Minnesota | 11.11 | 45 | New Hampshire | 6.80 |
| 21 | Oregon | 10.87 | 46 | Montana | 6.49 |
| 22 | Ohio | 10.75 | 46 | North Dakota | 6.49 |
| 22 | Rhode Island | 10.75 | 48 | Alabama | 6.17 |
| 24 | Connecticut | 10.20 | 49 | West Virginia | 5.98 |
| 24 | Hawaii | 10.20 | 50 | Vermont | 5.78 |

*Panel B: Summary LEM statistics*

| Statistic | Value |
|---|---|
| Mean | 11.68 |
| Median | 10.20 |
| First quartile | 8.03 |
| Third quartile | 13.65 |
| Minimum | 5.78 |
| Maximum | 27.78 |
| Range | 22.00 |
| Standard deviation | 5.11 |

FIGURE 1: DISTRIBUTION OF LEGAL ENVIRONMENT MEASURE (LEM) SCORES

A histogram of Legal Environment Measure scores. All individual state scores are approximated to their nearest integer value.

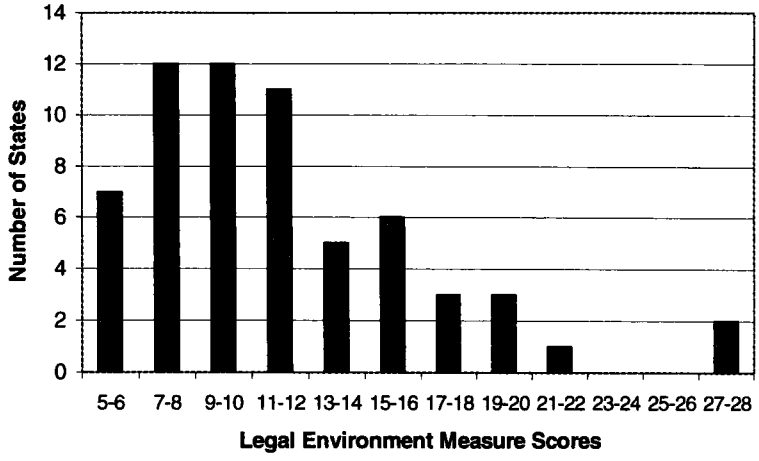


TABLE 2: ASSOCIATION BETWEEN THE LEGAL ENVIRONMENT MEASURE (LEM) AND CASE LAW

State lower court cases are the total number of cases in the Westlaw database for all courts other then the state supreme court for each state. State supreme court cases are the total number of cases in the Westlaw database from each state supreme court between 1945 and 2000 under the West topical heading for "Corporations." Total state cases represent the number of all cases in all state courts in the Westlaw database. Total state cases and federal cases represent the number of cases in state court and federal district court in the Westlaw database. State quartiles are constructed on the basis of the number of cases for each relevant court or courts. Correlations are estimated as the nonparmetric Spearson coefficients. Statistical significance at the one and five percent levels is indicated by ** and *, respectively.

| Statistical Analysis | State Lower Court Cases | State Supreme Court Cases | Total State Court Cases | Total State Cases and Federal Cases |
|---|---|---|---|---|
| **Correlation with LEM** | 0.31** | 0.23* | 0.28** | 0.29** |
| | | | | |
| **Mean LEM for:** | | | | |
| **Quartile 1** | 10.09 | 10.42 | 9.08 | 9.66 |
| **Quartile 2** | 10.00 | 10.94 | 12.87 | 12.01 |
| **Quartile 3** | 12.41 | 11.22 | 11.61 | 11.23 |
| **Quartile 4** | 14.06 | 14.23 | 13.23 | 13.82 |
| | | | | |
| **t-statistic for LEM comparison: top quartile vs. bottom quartile** | -1.98* | -1.50 | -2.03* | -1.97* |

TABLE 3: THE INFLUENCE OF THE STATE LEGAL ENVIRONMENT (LEM) ON PATTERNS OF CORPORATE MIGRATION

An emigrating business is a firm that lists its corporate headquarters in one state, but incorporates in a different state. From the viewpoint of the state in which it is headquartered, the firm is an émigré. From the perspective of the state in which the firm is incorporating, the firm is an immigrant. Statistical significance at the one and five percent levels is indicated by ** and *, respectively.

*Panel A: Mean (median) LEMs for subsamples based on the incidence of corporate migration*

| **Overall Sample** | **Subsample 1: High Immigration States** | **Subsample 2: High Emigration States** | **Subsample 3: Low Immigration States** | **Subsample 4: Low Emigration States** |
|---|---|---|---|---|
| 11.68<br>(10.20) | 13.08<br>(12.05) | 10.29<br>(9.62) | 10.30<br>(8.48) | 13.07<br>(11.11) |

*Panel B: Tests of mean (median) differences in LEMs between corporate migration subsamples*

| **Subsamples compared** | **Description** | **t-statistic (Wilcoxon z) for difference** |
|---|---|---|
| 1 vs. 3 | High immigration states compared with low immigration states | -2.86**<br>(-2.75**) |
| 4 vs. 2 | Low emigration states compared with high emigration states | 2.90**<br>(2.42*) |
| 1 vs. 2 | High immigration states compared with high emigration states | 2.41*<br>(2.32*) |

*Panel C: Comparison of LEM for corporate migration subsamples with aggregate sample*

| Subsample | Subsample Mean (median) LEM | Mean (median) LEM for Aggregate Sample | t-statistic (Wilcoxon z) for difference |
|---|---|---|---|
| Subsample 1: High immigration | 13.08 (12.05) | 11.68 (10.20) | 2.68** (2.49*) |
| Subsample 2: High emigration | 10.29 (9.62) | 11.68 (10.20) | -2.10* (-2.06*) |
| Subsample 3: Low immigration | 10.30 (8.48) | 11.68 (10.20) | -2.69** (-2.51*) |
| Subsample 4: Low emigration | 13.07 (11.11) | 11.68 (10.20) | 2.10* (2.16*) |

TABLE 4: STATE LEGAL AND POLITICAL EFFECTS ON INCORPORATION RATES

This table examines four different dependent variables: "Number of Firms Incorporated in State" representing the absolute number of firms incorporated within the state; "Number of Immigrant Firms" representing the absolute number of firms incorporated within the state that have headquarters in other states; "Percentage of Incorporations from Immigrants" representing the percentage of incorporations in a state from firms that are headquartered in another state; "Percentage of In-state Firms Incorporated in State" representing the percentage of firms that are headquartered in the state and also are incorporated in that state. Independent variables are the natural logarithm of the LEM score; the total number of state and federal district court cases on corporate law in the Westlaw database from 1945-2000 standardized by state population; political structure, which captures the state's court and legislative structures; the number of attorneys per capita in 2000 practicing in a state; a binary variable representing whether the state has adopted the Revised Model Business Corporation Act; the average of each state's percentage vote for the Democratic candidate from 1952-2000; and the averge annual state per capita personal income over the sample period. Statistical significance at the one, five and ten percent level is indicated by ***, **, and *, respectively.

| | **Number of Firms in State** | | **Number of Immigrant Firms** | | **Fraction of Incorporations from Immigrants** | | **Fraction of In-state Firms Incorporated in State** | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **Model 1** | **Model 2** | **Model 1** | **Model 2** | **Model 1** | **Model 2** | **Model 1** | **Model 2** |
| Intercept | -7.331 (-0.551) | -19.770 (1.702)* | -3.513 (-0.772) | -22.571 (-1.433) | -21.477 (-1.222) | -33.782 (-1.703)* | 27.477 (1.512) | 9.77 (0.605) |
| Log(LEM) | | 9.66 (2.689)*** | | 8.992 (2.283)** | | 9.455 (2.179)** | | 10.371 (2.112)** |
| Total Cases, Rank | 0.827 (3.221)*** | 0.557 (4.227)*** | 0.751 2.333)** | 0.671 (2.512)** | 0.481 (1.225) | 0.442 (1.877)* | 0.873 (1.808)* | 0.762 (1.799)* |
| Political Structure | 1.993 (1.862)* | 2.893 (1.891)* | 4.249 (2.778)*** | 3.937 (2.111)** | 6.099 (3.004)*** | 5.896 (2.904)*** | -0.891 (-0.517) | -1.117 (-0.882) |
| 2000 Prac. Attys., Rank | 0.307 (1.557) | 0.177 (0.982) | 0.471 (2.121)** | 0.433 (1.401) | 0.772 (1.023) | 0.561 (1.003) | -0.899 (-1.047) | -0.337 (-2.233)** |
| RMBCA Jurisdiction | 1.332 (0.671) | 1.444 (0.882) | -0,099 (-0.442) | -0.397 (-0.103) | 1.699 (1.003) | 1.398 (0.447) | 1.229 (0.873) | 1.379 (0.437) |
| Dem. Vote 1952-2000 | 8.885 (0.752) | 2.065 0.332 | -0.117 (-0.083) | -5.338 (-0.887) | 38.037 (0.887) | 26.541 (1.099) | -9.661 (-0.247) | -17.448 (-0.663) |
| Economic Infrastructure | 1.022 (0.995) | 0.972 (1.122) | 1.329 (1.277) | 0.893 (1.055) | 1.115 (1.125) | 1.086 (1.003) | 1.212 (0.995) | 1.271 (1.023) |
| | | | | | | | | |
| Adjusted $R^2$ | 0.600 | 0.661 | 0.499 | 0.557 | 0.244 | 0.355 | 0.083 | 0.171 |
| F-Statistic | 14.09*** | 15.93*** | 7.81*** | 8.70*** | 2.89** | 3.88** | 0.77 | 1.66 |

TABLE 5: THE LEM'S EFFECT ON FIRM VALUE

The LEM scores simultaneously reflect the presence or absence of eight state corporate statutes that unambiguously enhance managerial discretion as well as the rate of legal innovation for a given state. Firm value is proxied with the firm's asset market-to-book ratio. In Panel A the state legal environment is proxied with a binary variable that assumes a value of one for Delaware incorporations and is zero otherwise. In Panel B the state legal environment is captured with the LEM. Total assets, R&D, and sales are year-end values as obtained from Compustat. Both the current return on assets (ROA) and the one-year lagged return on assets ($ROA_{-1}$) are calculated as net income divided by total assets. The z score is based on Altman's (1968) multivariate discriminate model of financial failure. SIC dummies are a set of binary variables to capture industry membership based on a one-digit SIC code. Statistical significance at the one, five and ten percent level is indicated by ***, **, and *, respectively.

*Panel A: Binary variable measurement of the state environment*

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Intercept | 2.351 (27.89)*** | 2.336 (2.69)*** | 2.291 (2.67)*** | 3.447 (1.68)* | 3.224 (1.91)* | 2.892 (1.67)* |
| DE binary variable | -0.027 (-0.25) | -0.108 (-1.01) | -0.136 (-1.29) | -0.495 (-2.15)** | -0.339 (-1.77)* | -0.335 (-1.67)* |
| Log of Total Assets | | | 0.000 (7.76)*** | 0.000 (5.26)*** | | |
| R&D / Sales | | | | 3.193 (6.90)*** | 2.149 (3.62)*** | 1.609 (2.01)** |
| Log of Sales | | | | | 0.051 (0.88) | 0.124 (1.74)* |
| ROA | | | | | 0.614 (1.70)* | 0.740 (1.67)* |
| $ROA_1$ | | | | | -1.326 (-4.50)*** | -1.822 (-4.82)*** |
| Z Score | | | | | | 0.030 (7.04)*** |
| SIC Dummies | | YES | YES | YES | YES | YES |
| | | | | | | |
| Adjusted $R^2$ | -0.000 | 0.052 | 0.074 | 0.077 | 0.066 | 0.125 |
| F Statistic | 0.06 | 14.92*** | 19.77*** | 7.83*** | 5.41*** | 6.64*** |

TABLE 5 CONTINUED: THE LEM'S EFFECT ON FIRM VALUE

*Panel B: LEM measurement of the state legal environment*

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Intercept | 2.836<br>(13.47)*** | 2.781<br>(3.14)*** | 2.720<br>(3.10)*** | 4.295<br>(2.05)** | 3.927<br>(2.38)** | 3.717<br>(2.09)** |
| LEM | -0.033<br>(-2.46)** | -0.108<br>(-2.53)** | -0.034<br>(-2.62)*** | -0.080<br>(-2.79)*** | -0.061<br>(-2.57)*** | -0.069<br>(-2.34)** |
| Log of Total Assets | | | 0.000<br>(7.75)*** | 0.000<br>(5.22)*** | | |
| R&D / Sales | | | | 3.112<br>(6.76)*** | 2.091<br>(3.54)*** | 1.541<br>(1.93)* |
| Log of Sales | | | | | 0.046<br>(0.80) | 0.121<br>(1.71)* |
| ROA | | | | | 0.630<br>(1.75)* | 0.760<br>(1.72)* |
| $ROA_1$ | | | | | -1.313<br>(-4.47)*** | -1.832<br>(-4.83)** |
| Z Score | | | | | | 0.030***<br>(7.16) |
| SIC Dummies | | YES | YES | YES | YES | YES |
| | | | | | | |
| Adjusted $R^2$ | 0.002 | 0.054 | 0.076 | 0.080 | 0.069 | 0.129 |
| F Statistic | 6.04** | 15.54*** | 20.33*** | 8.14*** | 5.69*** | 6.92*** |

REFERENCES

Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *Journal of Finance* 23 (1968): 589-609.

Alva, Curtis. "Delaware and the market for corporate charters: History and agency." *Delaware Journal of Corporate Law* 15 (1990): 885-920.

Bebchuk, Lucian A., and Alma Cohen. "Firms' decisions where to incorporate." *Journal of Law and Economics* 46 (2003): 383-425.

Bebchuk, Lucian A., Alma Cohen, and Allen Ferrell. "Does the evidence favor state competition in corporate law?" *California Law Review* 90 (2002): 1775-1821.

Berger, Phillip G., and Eli Ofek. "Diversification's effect on firm value." *Journal of Financial Economics 37 (1995): 39-65.*

Carney, William J. "The production of corporate law." *Southern California Law Review* 71 (1998): 715-77.

Cary, William L. "Federalism and corporate law: Reflections upon Delaware." *Yale Law Journal* 83 (1974): 663-705.

Cho, Myeong H. "Ownership structure, investment, and the corporate value: an empirical analysis." *Journal of Financial Economics* 47 (1998): 103-21.

Daines, Robert. "Does Delaware law improve firm value?" *Journal of Financial Economics* 62 (2001): 525-58.

Fisch, Jill E. "The peculiar role of Delaware courts in the competition for corporate charters." *University of Cincinnati Law Review* 68 (2000): 1061-1100.

Gillan, Stuart L., Jay Hartzell, and Laura Starks. "Industries, investment opportunities, and corporate governance structures." (2002) Working paper, Center for Corporate Governance, University of Delaware.

Klausner, Michael. "Corporations, corporate law, and networks of contracts." *Virginia Law Review* 81 (1995): 757-852.

Lang, Larry H., Rene Stulz, and Ralph Walkling. "Managerial performance, Tobin's Q, and the gains from successful tender offers." *Journal of Financial Economics* 24 (1989): 137-54.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. "The legal determinants of external finance." *Journal of Finance* 52 (1997): 1131-50.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. "Law and finance." *Journal of Political Economy* 106 (1998): 1113-55.

LaPorta, Rafael, Florencio Lopez de Silanes, and Andrei Shleifer. "Corporate ownership around the world." *Journal of Finance* 54 (1999): 471-517.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. "Investor protection and corporate governance." *Journal of Financial Economics* 58 (2000): 3-27.

McConnell, John J., and Henri Servaes. "Additional evidence on equity ownership and corporate value." *Journal of Financial Economics* 27 (1990): 595-612.

Myers, Stewart C. "Determinants of corporate borrowing." *Journal of Financial Economics* 5 (1977): 147-75.

Morck, Randall, Andrei Shleifer, and Robert Vishny. "Management ownership and market valuation: An empirical analysis." *Journal of Financial Economics* 20 (1988): 293-316.

Romano, Roberta. "Law as a product: Some pieces of the incorporation puzzle." *Journal of Law, Economics & Organization* 1 (1985): 225-83.

Smith, Clifford W., Jr., and Ross L. Watts. "The investment opportunity set and corporate financing, dividend, and compensation policies." *Journal of Financial Economics* 32 (1992): 263-92.

Subramanian, Guhan. "The influence of antitakeover statutes on incorporation choice: Evidence on the "race" debate and antitakeover overreaching." *University of Pennsylvania Law Review* 150 (2002): 1795-1893.

Winter, Ralph. "State law, shareholder protection, and the theory of the corporation." *Journal of Legal Studies* 6 (1977): 251-92.

# PRINCIPAL-AGENT OBSTACLES TO FOSTER CARE CONTRACTING

*Kelsi Brown Corkran*[*]



## INTRODUCTION

By all accounts, America's foster care system is a mess. Bad decision-making and carelessness abound: foster children are placed with known sex offenders, beaten to death by abusive foster parents, moved through a dozen foster homes in a year, forgotten by overworked caseworkers, and returned to drug-addicted parents.[1] With an annual turnover rate of up to seventy percent, social workers carrying exorbitant caseloads have stopped returning phone calls, visiting foster homes, or attending court hearings at

* Law Clerk to the Honorable David S. Tatel, U.S. Court of Appeals for the D.C. Circuit. J.D. and M.P.P., University of Chicago; B.A., University of Pennsylvania. I would like to thank Benjamin Glatstein, William J. Martin, David Weisbach, and three anonymous reviewers for helpful comments and suggestions.

[1] *See* Timothy Roche, *The Crisis of Foster Care*, TIME Nov. 13, 2000, at 74.

all.[2] In at least twenty states, lawyers have filed class action suits alleging severe maltreatment of foster children by the child welfare system and demanding massive reform.[3]

Over the past fifteen years, researchers and policymakers have put considerable effort into devising and implementing strategies to improve the quality of foster care services. Nevertheless, severe dysfunction continues to pervade the child welfare system and substantial improvement remains illusive. The primary purpose of this article is not to contribute to the growing body of recommendations for foster care reform, but to examine a more preliminary yet still unanswered question: Why has providing adequate foster care services been such a difficult task? I am specifically interested in the extent to which the failings of the foster care system can be explained by economic theory. Although the poor quality of foster care services is a classic example of what economists call a market failure, the lens of economic theory has been surprisingly underutilized by researchers and policymakers interested in reforming the foster care system.

Part I of this article provides a brief history of child welfare services in the United States and identifies several characteristics inherent to foster care and important in understanding economic barriers to the provision of high quality services. Most significantly, the individualized nature of service goals and the absence of meaningful performance measures make it extremely difficult to determine how well foster parents, social workers, and child welfare agencies are serving the children in their care. A central theme of this article is that the failures of the foster care system stem in large part from the inability of the purchaser (i.e., the government) to monitor the quality of services that the "sellers" of foster care provide. Part II discusses the information asymmetry inherent in foster care contracts in more detail, situating this asymmetry in principal-agent theory and explaining the obstacles faced by principals in ensuring that their own goals are met in their contractual relationships with agents. The difficulties created by information asymmetries in foster care contracting raise an important question: Why contract out these services at all? Part III explores the various rationales offered for privatization and discusses their application to contracting in the foster care system. The most important conclusion of this discussion is that although competition-induced efficiency is the rationale most frequently provided for privatization, the monitoring problems inherent in foster care make it inapplicable to foster care contracts. Instead, the primary reasons for privatization in foster care are the expertise, ethos, and community relationships of private agencies.

These first three Parts thus conclude that although contracting in foster care is problematic from a principal-agent perspective, private agencies are still at least arguably more qualified than the government to provide foster

---

[2] *See id.*

[3] *See id.*

care services. The challenge facing the government as principal is to design and implement foster care contracts that utilize the unique skill set of private agencies while also protecting against the pitfalls associated with information asymmetry. Part IV argues that the predominant approach to enforcing foster care contracts, which focuses on monitoring measurable but intermediate outcomes, is misguided and frequently undermines the quality of foster care by creating perverse incentives. Instead, this article proposes that an inputs-based approach to measuring service quality provides the government with the best opportunity to achieve its contractual goals in the foster care setting.

## I. THE DEVELOPMENT AND DESIGN OF THE AMERICAN FOSTER CARE SYSTEM

The notion of providing substitute care for abused, neglected, or orphaned children did not originate as a public policy matter, but instead traces back to efforts by private philanthropic groups to protect vulnerable children in their own communities.[4] These private agencies tended to be religiously based, although some grew out of larger, secular charities such as the Society for the Prevention of Cruelty to Animals.[5] Their activities included running orphanages, finding adoptive homes, and even seeking court intervention on behalf of children severely abused by their parents.[6]

It was not until the 1960s that states and the federal government became involved in the provision of child welfare services. State legislatures enacted statutes requiring investigation of child abuse cases and providing public funding for the nonprofit agencies that were already caring for abused and neglected children. In 1961, Congress amended the Social Security Act to include reimbursement for state and local foster care expenditures, and in 1967 it again amended the Act to explicitly allow states to use federal funding to contract with private, nonprofit foster care providers.[7] Today, all fifty states have publicly administered foster care programs financed through a mix of federal and state funds.

Although some states have created public agencies that directly place children in foster homes and employ social workers to monitor their care,

---

[4] *See* Susan Vivian Mangold, *Protection, Privatization, and Profit in the Foster Care System*, 60 OHIO ST. L. J. 1295, 1301-02 (1999) (providing an overview of the historical involvement of private providers in the foster care system).

[5] *See id.*

[6] *See id.* at 1301-06.

[7] For a discussion of the political climate that led Congress to become involved in the provision of child welfare services, *see id.* at 1306-10 (describing Theodore Roosevelt's efforts to provide welfare benefits to needy families as an important predecessor of federal and state laws dealing with child abuse and neglect).

most continue to contract these services out to private nonprofit organizations.[8] Federal and state statutory mandates shape some of the content of these contracts, but most of the details are determined by the public agency responsible for administering child welfare services. The private provider agency in turn contracts with individual families to provide foster homes and employs social workers to supervise the placement of children in those homes.

Despite variation across states and localities regarding the structure and content of agency contracts, the primary goal of foster care is always the same: child well-being. Within this framework of child well-being, there are numerous secondary goals: reasonable efforts should be made to place children with relatives instead of foster parents; biological parents should be assisted in remedying the problems that led to foster care placement; children should be returned to their parents as soon as the parents can provide an adequate home; and adoptive parents should be found as soon as possible for children who cannot be returned to their parents or placed with other relatives.

The pursuit of child well-being, however involves much more than compliance with these secondary goals. Foster children often have mental and physical health problems that require medical treatment, and many need special education services. Extensive counseling may be necessary to help a child cope with the grief and trauma associated with foster care placement. Because adolescents in foster care are at high risk for delinquent behavior, drug and alcohol abuse, and early sexual activity,[9] their well-being might require efforts to protect them from these dangers. Adolescents who will "age out"[10] of the foster care system need help in preparing for the transition into adulthood and self-care. In other words, although the goal of well-being applies to all children in foster care, the optimal path for achieving that goal is different for every child and may be hard to identify.

Even more difficult is determining whether the goal of well-being has been met—monitoring each child's educational progress, emotional wellness, and mental and physical health takes a great deal of effort, and meaningful measures are not always available. How does one know if a child's difficulties in school result from poor caretaking by the foster

---

[8] It is important to note, however, that most states do *not* contract out their duty to *remove* children from abusive or neglectful homes. It is only after a child has been taken into state custody that a private nonprofit takes responsibility for his care and places him in a foster home. *See generally* Jan McCarthy et al., A FAMILY'S GUIDE TO THE CHILD WELFARE SYSTEM 13-23 (2003), *available at* http://www.cwla.org/childwelfare/ familyguide.htm (summarizing the responsibilities of the public child welfare system).

[9] *See* Joel F. Handler, *"Ending Welfare as We Know It": The Win/Win Spin or the Stench of Victory*, 5 J. GENDER RACE & JUST. 131, 164-65 (2001).

[10] "Age out" means that a foster child has reached the age of majority and been discharged from state custody.

parents and provider agency, or instead stem from abuse and neglect prior to placement? Mental health and emotional wellness might be measured by the absence of behavioral problems, but, again, the extent to which such problems are attributable to the quality of foster care rather than pre-placement trauma is unclear. Furthermore, behavioral indicators fail to adequately capture some of the most important aspects of emotional wellness, such as happiness, peace of mind, self-esteem, and sense of security. Physical health appears to be the most easily observable outcome, but evidence of child abuse, especially sexual abuse, can be very difficult to detect.[11] Ultimately, the most important indicator of child well-being—becoming a happy, productive, and ethical adult—cannot be observed until long after the child's foster care placement has ended.

As mentioned in the Introduction, these two characteristics of foster care—the individualized nature of service goals and the absence of meaningful performance measures—make it very difficult for public officials to determine how well private child welfare agencies are serving the children in their care. Part II situates this monitoring problem in principal-agent theory and explains how the information asymmetry between the government and foster care providers impedes the government in meeting its contractual goals.

## II. THE INFORMATION ASYMMETRY PROBLEM: PRINCIPAL-AGENCY THEORY AND FOSTER CARE CONTRACTS

Whenever one party contracts with another to provide a good or service, the possibility of a principal-agent problem arises. The primary concern of a principal is ensuring that the agent performs her contractual duties according to the principal's preferences.[12] When the agent's tasks are specifiable and her performance easy to monitor, the principal-agent relationship is not particularly problematic because the agent will have sufficient motivation to produce the outcomes desired by the principal. Similarly, if the incentives of the principal and agent are aligned—i.e., the principal and agent have the same goal—the agent's self-interest will ensure that the principal's interests are also served by their relationship.

Most principal-agent contracts are not inherently characterized by either of these conditions. Instead, an information asymmetry frequently exists between the principal and the agent which allows the agent to serve

---

[11] *See, e.g.*, Pennsylvania v. Ritchie, 480 U.S. 39, 60 (1987) (describing child abuse as "one of the most difficult crimes to detect").

[12] *See* David E. M. Sappington, *Incentives in Principal-Agent Relationships*, 5 J. ECON. PERS. 45, 45 (1991) (discussing the key incentive problems that arise in principal-agent relationships and arguing that simple principal-agent models by themselves do not provide a complete understanding of the structure and operation of complex operations).

her own interests at the expense of the principal.[13] In situations like this, principals seek ways to choose agents, structure contract terms, and monitor performance so as to maximize the extent to which they meet their own goals through their contractual relationship with the agent. This Part discusses the application of principal-agent theory to foster care contracts. Section A describes the information asymmetry between the government and private service providers in more detail, focusing specifically on the ways in which this asymmetry is exacerbated by the severance of purchaser and consumer in the foster care system. Section B examines the possibility of incentive alignment in foster care contracts, discussing both the problem of adverse selection and the extent to which the nonprofit nature of private agencies mitigates the principal-agent problem in the foster care context.

### A. *Information Asymmetry in Foster Care Contracts*

As discussed above, the individualized nature of service goals and the absence of meaningful performance measures make it very difficult for the government as principal to determine the quality of foster care services provided by its agents—i.e., the nonprofit agencies with which it contracts. This monitoring problem creates an information asymmetry between the government and contracting agencies—the agencies know the quality of the services they provide, but the government does not. This kind of asymmetry is common in situations where, as in foster care, the purchaser of the service is different than the consumer. Under normal market conditions, individualized service goals pose less of a problem because each consumer can monitor the extent to which her contractual relationship with the service provider meets her personal needs, and can switch providers when those services are unsatisfactory. Similarly, even when performance measures are complex and difficult to identify, individual consumers usually have some sense of the quality of the services they receive. When the buyer and the consumer are two different people, however, "the market institutions that address quality uncertainty work far less effectively."[14]

For some publicly-funded social services, policymakers have attempted to mimic this market-disciplining mechanism through the use of

---

[13] *See* ELLIOTT D. SCLAR, YOU DON'T ALWAYS GET WHAT YOU PAY FOR: THE ECONOMICS OF PRIVATIZATION 103 (Cornell Univ. Press 2000) (discussing information asymmetry in principal-agent relationships).

[14] Michael Krashinsky, *Transaction Costs and a Theory of the Nonprofit Organization*, *in* THE ECONOMICS OF NONPROFIT INSTITUTIONS: STUDIES IN STRUCTURE AND POLICY 114, 117 (Susan Rose-Ackerman ed., 1986). *See also* Lester M. Salamon, *The Marketization of Welfare: Changing Nonprofit and For-Profit Roles in the American Welfare State*, 67 SOC. SERV. REV. 16, 37 (1993) (stating that when the consumers of a service are not the same people who purchase the service, "a link that is crucial for the market's operation" becomes severed).

vouchers, thus transferring purchasing power from the government to the consumer. Alternatively, some government agencies have used consumer reviews as measure of performance by contracting service providers. Under both strategies, the government mitigates the asymmetry problem by accessing consumer information about service quality. Unfortunately, these tactics are less useful in the child welfare context. The primary consumers of foster care are children, many of whom are quite young. Suffice it to say, children are not in a position to use vouchers to choose a foster care agency, nor does it seem likely that their participation in quality review efforts would yield much information. Although their parents are also in some sense consumers of foster care, their interests may be misaligned with the interests of their children or the government. Given the option, a parent whose child has been taken away will choose the foster care agency likely to return the child most quickly, regardless of other factors affecting the child's well-being. Similarly, a social worker with low reunification standards will always receive high quality marks from such parents, while those who attempt to terminate parental rights will be rated poorly, even when termination is appropriate.

Ultimately, the information asymmetry inherent in foster care contracts is exceptionally difficult to overcome, arguably more so than in any other area of government contracting. Section B examines the extent to which the government can avoid the quality problems associated with information asymmetry by contracting only with those agencies whose incentives are aligned with its own.

### B. *Incentive Alignment, the Nonprofit "Solution," and Adverse Selection*

As stated earlier, information asymmetry is only problematic if there is a misalignment of interests between the principal and agent. When the agent has the same objective as the principal, the principal can rest assured that the agent, in attempting to maximize her own individual utility, will also perform her contractual duties according to the principal's preferences.[15] Accordingly, a wise principal will seek, whenever possible, to contract with agents whose interests are compatible with her own.[16]

From this perspective, the government's task is to identify and select foster care service providers whose primary goal is child well-being, and who, like the government, place high value on kinship care, reunification,

---

[15] *See* Bruce Walker, *Monitoring and Motivation in Principal-Agent Relationships: Some Issues in the Case of Local Authority Services*, 47 SCOT. J. POL. ECON. 525, 527, 538 (2000) (observing that principal-agent theory "is based on the assumption that the objectives of self-seeking P[rincipal]s and A[gent]s are incompatible . . .").

[16] *See* Sappington, *supra* note 12, at 56 ("[A]n important component of the principal's task is to select the 'best' agent or agents.").

and permanency planning. The question, of course, is whether such agencies exist and, if so, how to distinguish them from agencies whose interests are not aligned with those of the government. As a starting point, Henry Hansmann argues that when, as in foster care, the principal is unable to police producers by ordinary contractual devices, nonprofit firms fulfill a crucial role in the functioning of the market.[17] Hansmann theorizes that because nonprofits face a nondistribution constraint—i.e., any net monetary gain must be reinvested in the mission of the organization rather than distributed among shareholders—they have far less incentive to sacrifice quality or misuse funds than their profit-focused counterparts.[18] Furthermore, many nonprofits are characterized not only by the absence of profit-seeking behavior, but also by "entrenched norms of professional, publicly oriented behavior" not typically found in for-profit firms.[19] In other words, a lack of interest in financial gain is often correlated with a strong interest in providing high-quality services. Hansmann observes that the nondistribution constraint thus serves as a screening device that allows principals to select as agents "precisely that class of individuals whose preferences are most in consonance with the fiduciary role that the [contract] is designed to serve."[20] Indeed, empirical evidence indicates that nonprofit organizations have a significantly better record than for-profits in providing socially complex services; they are more likely to go beyond specified contractual duties in serving clients,[21] and less likely to take advantage of information asymmetries to the detriment of their donors and patrons.[22]

From a principal-agent perspective, then, the nondistribution constraint of nonprofits serves as a useful signaling device to the government as it attempts to identify foster care agencies whose interests are compatible with its own. It is not, however, sufficient in itself to ensure high quality

---

17 Henry B. Hansmann, *The Role of Nonprofit Enterprise*, *in* THE ECONOMICS OF NONPROFIT INSTITUTIONS, *supra* note 15, at 57, 71. *See also* David Easley and Maureen O'Hara, *The Economic Role of the Nonprofit Firm*, 14 BELL J. ECON. 531, 538 (1983) (arguing that "nonprofits may be superior to for-profits if the output cannot be costlessly observed"); Evelyn Brody, *Agents Without Principals: The Economic Convergence of the Nonprofit and For-Profit Organizational Forms*, 40 N.Y.L. SCH. L. REV. 457, 462 (1996) (describing nonprofits as delivering "'trust goods,' the production of which the donor or patron cannot monitor.").

18 *See* Mangold, *supra* note 4, at 1315-18.

19 Janna J. Hansen, Note, *Limits of Competition: Accountability in Government Contracting*, 112 YALE L. J. 2465, 2477 (2003).

20 Hansmann, *supra* note 17, at 78.

21 *See* Michele Estrin Gilman, *Legal Accountability in an Era of Privatized Welfare*, 89 CAL. L. REV. 569, 600 (2001) (discussing research regarding the superiority of nonprofits in providing social services).

22 *See* Burton A. Weisbrod & Mark Schlesinger, *Public, Private, Nonprofit Ownership and the Response to Asymmetric Information: The Case of Nursing Homes*, *in* THE ECONOMICS OF NONPROFIT INSTITUTIONS, *supra* note 14, at 133, 147 (summarizing research regarding nonprofit involvement in the nursing home sector).

services; the absence of profit-seeking behavior does not always indicate incentive alignment, and even those agencies that are indeed dedicated to child well-being will not always operate as efficiently and effectively as possible. A frequently cited weakness of nonprofits is that they lack incentive to choose the least costly means of providing services and thus tend to operate with a higher degree of slack, or inefficiency, than for-profit firms.[23] Perhaps most importantly, nondistribution constraint theory does not provide the government with any mechanism for choosing between competing nonprofits—i.e., for identifying those agencies with missions most similar to its own and which are most qualified and capable of carrying out their contractual duties.[24]

It thus appears that even when the market for service providers consists only of nonprofits, the government faces an adverse selection problem; in other words, there is still the risk of what George Akerlof famously dubbed a "market for lemons."[25] As in the used car market, the contractors most willing to accept low payment for foster care services are often those that have the least to offer—i.e., those that will readily sacrifice service quality and are thus undeterred when the reimbursement rates offered by the government are insufficient to cover the cost of providing adequate services. However, an interesting and perhaps unique distinction of the social services market is that "lemon" contractors are not alone in their willingness to accept underpayment. Although it is widely agreed that very few foster care contracts provide compensation sufficient to cover the cost of providing high (or even mediocre) quality foster care,[26] some reputable nonprofits continue to enter these contracts by using private donations to subsidize their costs.[27] Still, it remains true that the services provided by nonprofit agencies vary widely in quality, and the government has no mechanism for separating out the "lemons."

In summary, the inability of the government to monitor the performance of foster care providers creates an information asymmetry that in turn functions as a significant barrier to effective contracting in the foster

---

23 *See* Hansmann, *supra* note 17, at 80; Brody, *supra* note 17, at 507.

24 *See* Brody, *supra* note 17, at 463-64 (arguing that the nondistribution constraint does not solve all principal-agent problems).

25 George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q. J. OF ECON. 488 (1970) (arguing that when information asymmetries lead to uncertainty on the part of the buyer about the quality of the product, this uncertainty will decrease the price the buyer is willing to pay and increase adverse selection among producers, ultimately resulting in poorer quality products in the market).

26 *See* Mangold, *supra* note 4, at 1316; JOEL F. HANDLER, DOWN FROM BUREAUCRACY: THE AMBIGUITY OF PRIVATIZATION AND EMPOWERMENT 96 (1996).

27 *See, e.g.*, ALFRED J. KAHN & SHEILA B. KAMERMAN, CONTRACTING FOR CHILD & FAMILY SERVICES: A MISSION-SENSITIVE GUIDE 116 (1999) (discussing an agency that continued to submit proposals for child welfare services despite inadequate reimbursement rates because its leadership decided the agency "had a contribution to make and a public responsibility" to accept such contracts).

care system. This asymmetry cannot be remedied through the use of vouchers or consumer review, and although efforts to ensure incentive alignment have some mitigating force, foster care contracts clearly continue to suffer from serious principal-agent problems. The picture painted here is thus quite bleak, raising an important question: why contract out foster care services at all? If it is really so difficult to create a high quality foster care system through privatization, wouldn't it be more effective for the government to provide these services in-house? Part III examines this question, discussing the various rationales offered for privatization and their application to contracting in the foster care system.

## III. PRIVATIZATION THEORY AND ITS APPLICATION TO FOSTER CARE SERVICES

In the government context, privatization refers to the delegation of public functions to private entities.[28] The Reagan administration initiated a broad trend towards privatizing government services that continues today, as evident in state and federal policies regarding the creation of school voucher programs, the selling off of public housing projects and public hospitals, the replacement of Social Security with individual retirement accounts,[29] and contracts with private managed care companies to administer Medicaid and Medicare services. Although private agencies have always provided foster care services, the decision of states and the federal government in the 1960s to take responsibility for child welfare services transformed foster care into a public function privatized right at inception.[30]

The justifications for privatization offered by economists and policymakers fall into five general categories: efficiency, flexibility, expertise, norms, and community relationships. This Part considers these rationales and discusses their relevance and application to the privatization of foster care.

### A. *Why Foster Care Privatization Fails to Promote Efficiency*

The most frequently cited benefit of privatization is that it promotes efficiency by exposing the provision of government goods and services to

---

28 *See* Mangold, *supra* note 4, at 1297 (stating that privatization "can involve delegation of policy making, regulation, and service delivery").

29 *See* Matthew Diller, *Form and Substance in the Privatization of Poverty Programs*, 49 U.C.L.A. L. REV. 1739, 1741 (2002).

30 *See* Mangold, *supra* note 4, at 1298 (observing that "foster care was always a 'privatized' system, never an exclusively public one").

the "disciplining forces of competition."[31] Whereas governmental agencies, as monopolies, lack incentive to provide services in the most efficient and effective manner possible, private firms compete with each other to obtain government contracts and must therefore offer the best possible product for the lowest possible price in order to win the government's business.[32] In other words, by subjecting public functions to market forces, privatization ensures that these functions are fulfilled efficiently and effectively.

This "competition-induced efficiency" explanation for privatization relies on two premises that are inapplicable to the privatization of foster care services. First, the efficiency rationale assumes the delegated tasks are such that the government can easily determine the quality of the product offered by private firms. If the government cannot observe compliance with contractual requirements or identify performance goals in advance, "the bids of competitors cannot be compared effectively and the performance of service providers cannot be evaluated meaningfully."[33] As discussed, although the primary goal for every foster child is well-being, the achievement of this goal is extremely complex and difficult to measure. Numerous economic theorists have noted that there is no reason to think the competitive aspects of privatization lead to increased efficiency or effectiveness in the provision of complex human services where, as in foster care, it is impossible to compare competitors.[34]

The second assumption of the efficiency rationale is that private firms will truly be competing with each other for government contracts. As an empirical matter, it is often the case that only a few agencies show interest in obtaining foster care contracts.[35] Although this is presumably due in part to low compensation, it also relates to entrenchment: in most "markets" for child welfare services, many of the private agencies that provide foster care today are the same agencies that engaged in child welfare efforts prior to government involvement in the provision of foster care services. These charitable organizations have close relationships with government officials, strong community ties, and a solid base of private donors to defray costs.[36] In addition, once the government contracts with an agency, ending that contractual relationship can be costly. From a child wellness perspective, continuity of services is crucial: moving to a new foster home can be traumatic for children, as can ending a relationship with a social worker,

---

31 Diller, *supra* note 29, at 1743.

32 *See* SCLAR, *supra* note 13, at 12.

33 Diller, *supra* note 29, at 1745.

34 *See, e.g., id.* at 1744-45; Gilman, *supra* note 21, at 600-02; Jody Freeman, *The Contracting State*, 28 FLA. ST. U. L. REV. 155, 170-71 (2000); HANDLER, *supra* note 26, at 83-86.

35 *See* KAHN AND KAMERMAN, *supra* note 27, at 34 ("In the social services field, one is much more likely to be dealing with oligopolistic situations than with competition."). *See also* Hansen, Note, *supra* note 19, at 2471 (discussing the lack of competition for child welfare contracts).

36 *See* Gilman, *supra* note 21, at 599-600.

therapist, or other important agency-employed adults.[37] As a result, once a child is placed with a particular service provider, the government will be very reluctant to terminate that contractual relationship.

### B. *Alternative Justifications for Privatization in Foster Care*

Although the "competition-induced efficiency" explanation of privatization has little application to foster care, not all justifications for privatization rely on assumptions about performance measurability or competitive bidding. At least four of these rationales arguably apply to foster care.

First, because private firms are not subject to the employee unionization and civil service rules that constrain government agencies, they are able to operate more efficiently and effectively than their public counterparts. In particular, the flexibility of private firms in hiring, firing, and compensating their employees allows them to "motivate and utilize workers" better than government departments.[38] This rationale assumes the government is not passing these constraints on to provider agencies through contract provisions, an assumption that appears valid in the context of foster care. Although states place some qualification and training requirements on social workers and foster parents, provider agencies presumably have more flexibility in their decision-making than public officials.

A second rationale for privatization "rests on the notion that the private sector can add real and measurable value to the public product, not merely duplicate what public employees can do."[39] In industries where private firms have more experience than the government in providing a good or service, privatization allows the government to take advantage of this expertise.[40] As discussed earlier, many of the private agencies that receive foster care contracts today have been providing services to abused and neglected children for fifty years or more, with some founded as far back as the late 1800s.[41] When federal and state legislatures began to allocate public money to the provision of child welfare services, it was simply more efficient and effective to use these funds to support and

---

37 *See* Peggy Cooper Davis, *The Good Mother: A New Look at Psychological Parent Theory*, 22 N.Y.U. REV. L. & SOC. CHANGE 347, 363 n.79 (1996).

38 *Id.* at 596.

39 SCLAR, *supra* note 13, at 68.

40 *See* KAHN & KAMERMAN, *supra* note 27, at 29 (observing that contracting in child welfare "provides access to expertise that public agencies may not have on staff"). Note that private sector expertise is primarily a product of institutional memory, not individual employee skill, which means the government cannot co-opt the benefits of expertise by hiring particular employees away from the private sector.

41 *See* Mangold, *supra* note 4, at 1301-02.

expand the work already being done in the private nonprofit sector than to attempt to provide these services through public agencies.

A third rationale for privatization is that in some sectors, private firms are inherently superior to the government in creating and maintaining a "quality-inducing" ethos.[42] This theory posits that the personal feelings of employees about the value of their work has an important influence on the quality of the provided service, and that these feelings are likewise influenced by the extent to which the employing agency establishes norms that encourage employee "buy-in" to the substantive goals of their organization. From this perspective, the religious and moral character of nonprofit child welfare agencies is an important asset that enhances the quality of their services beyond what the government could provide on its own. This reasoning clearly underlies the Bush Administration's efforts to increase government contracts with faith-based social service agencies, described by President Bush as "armies of compassion."[43] Indeed, as one commentator has noted, "When it comes to dispensing love, many would readily accept the argument that religious institutions have an edge over government agencies."[44]

The final rationale for privatization is that local organizations are better suited than the government to address the needs of their community.[45] This is particularly true with regard to social services, where the recipients of the service are often distrustful of the government and its employees. In foster care, successful reunifications and adoptions often depend on the ability of the social worker to gain the trust and respect of the families she serves. Because private agencies tend to be closely connected to particular communities, it is easier for their workers to establish effective relationships with clients. A closely related point is that local nonprofits often benefit from volunteer citizen participation that would be unlikely to occur if a government agency provided the services directly.[46] This is particularly true for religiously-affiliated service providers, which

---

[42] *See* Diller, *supra* note 29, at 1748 ("Local job training, childcare, or social service agencies may have an ethos and culture that is different than government agencies providing the same services and may appear more attentive to the needs of clients.").

[43] President George W. Bush, *Remarks by the President to the United States Conference of Mayors* (June 25, 2001), cited in Diller, *supra* note 29, at 1760.

[44] Diller, *supra* note 29, at 1761.

[45] *Id.* at 1747-48 ("Reliance on private nonprofit social service providers may strike political liberals as more responsive to community needs than services delivered through public agencies."); Gilman, *supra* note 21, at 596 ("[One] strand of the privatization movement sees privatization as a democratizing force that returns power from the government to local communities and their mediating institutions, such as churches, neighborhoods, and voluntary organizations, which are better situated to address a community's needs.").

[46] *See* KAHN & KAMERMAN, *supra* note 27, at 29 (noting that foster care contracting "may more readily promote volunteer citizen participation in program innovation, governance, and service delivery than most formal government bureaucracies").

frequently receive financial assistance and volunteer support from area churches.

In sum, although privatization is most often lauded as a mechanism for the government to reap the benefits of competition-induced efficiency, it is important to understand that this rationale has little application to foster care. The difficulty of monitoring foster care contracts and the absence of real competition among child welfare contractors make it unlikely that market forces have any impact on the efficiency or quality of foster care services. Nonetheless, private agencies may still be better qualified than the government to provide quality foster care services because of their freedom from bureaucratic constraints, their extensive experience and expertise in the child welfare field, their ability to create and maintain a service-oriented, quality-inducing ethos, and their strong community ties.

Given that private agencies are at least arguably superior to the government with regard to the provision of foster care, it is worth exploring whether foster care contracts can be designed and implemented so as to avoid—or at least mitigate—the principal-agent problems described in Part II. Part IV attempts to answer this question, arguing that an inputs-based approach to measuring foster care quality provides the government with the best opportunity to utilize the unique skill set of private agencies while also protecting against the pitfalls associated with information asymmetry. Part IV also suggests strategies for ensuring that the transaction costs of contract monitoring do not outweigh the benefits of privatization identified here.

## IV. WHY INPUT MONITORING IS SUPERIOR TO USING PROXY OUTCOMES IN FOSTER CARE CONTRACTS

A bedrock principle in privatization theory is that the benefits of producing goods and services through the private sector relate closely to the ability of private firms to operate without government interference. Indeed, all of the rationales in Part III assume that that the contracts offered by the government will be sufficiently flexible to allow contractors to utilize their strengths as they produce the desired output. This flexibility is important regardless of whether those strengths are competition-induced or are instead based on expertise, ethos, or other quality-related characteristics.

Conventional wisdom also holds that the best way to preserve agent flexibility is to design contracts that specify the agent's required output but allow the agent to choose the means by which she achieves that output. The problem with this strategy is that it is difficult to implement when the principal's contractual goals are not amenable to specification or measurement. In social service contracting, the government has attempted to overcome this problem through the use of proxy outcomes—i.e., easily measurable performance indicators that are believed to be closely associated with the actual (but unobservable) objectives of the contract. Section A of this Part describes government efforts to use proxy outcomes

in foster care contracts and explains why this strategy has not only failed to improve the foster care system, but actually undermines service quality by creating perverse incentives for contracting agencies.

Although economists and policymakers usually reject input monitoring as an undesirable and unnecessarily restrictive approach to government contracting, it is my contention that in foster care, the specification of inputs better serves the government's goal of child well-being than the traditional proxy outcome approach. Sections B and C lay out this argument in more detail and provide a framework for determining which inputs the government should regulate in the provision of foster care services.

### A. *The Problem with Proxy Outcomes*

In 1997, Congress ordered the Department of Health and Human Services to develop a set of outcome measures to assess state performance in operating child welfare programs.[47] The resulting indicator list is very similar to those frequently found in state contracts and has led to more outcome measurement as states modify their child welfare policies to account for the new assessment system.[48] As one policy researcher observes, "The current shift is away from inputs or process objectives (i.e., service units, quantity of specific services rendered) towards outputs or outcome objectives. Succinctly stated, public child welfare agencies are increasingly interested in purchasing outcomes as opposed to services."[49]

Among states that identify proxies for child well-being in their contracts, most focus on outcomes related to safety, permanency, and child and family functioning.[50] Florida, for example, requires that no more than ten percent of children in an agency's case load experience a confirmed incident of abuse or neglect during their foster care placement or the first twelve months after discharge.[51] Kansas requires its foster care agencies to place at least sixty-five percent of children with siblings, to discharge at least forty percent of children within six months of placement, and to return at least sixty percent of children to their biological families.[52] The purpose of these outcome requirements is understandable—if physical safety,

---

47 Adoptions and Safe Families Act of 1997, Pub. L. No. 105-89, 111 Stat. 2115 (1997). *See also* KAHN & KAMERMAN, *supra* note 27, at 105.

48 *See* KAHN & KAMERMAN, *supra* note 27, at 105 (discussing the set of child welfare outcome measures that DHHS developed).

49 FRED WULCZYN, PERFORMANCE-BASED CONTRACTING: THE BASICS 1 (July 2005), *available at* http://www.ffta.org/pbcpaper.pdf.

50 FRED H. WULCZYN & BRITANY ORLEBEKE, FOUR CASE STUDIES OF FISCAL REFORM AND MANAGED CARE IN CHILD WELFARE SERVICES 10 (1998).

51 *Id.*

52 KAHN & KAMERMAN, *supra* note 27, at 104-05.

permanency, and reunification promote child well-being, it seems sensible to focus on them as measurable indicators of contractual goal achievement. However, this proxy-based approach has three serious problems.

First, there's no reliable baseline data for determining the percentage targets.[53] One might decide as a normative matter that the majority of foster children should be returned to their biological parents rather than adopted, but this is much different than determining through empirical research that sixty percent of children in the foster care system would be better off with their biological parents than with an adopted family. Given the nebulous nature of well-being, it's difficult to imagine that any statistically-sound study could make such a determination. Indeed, policymakers in a number of states confess that their initial percentage requirements were "pulled out of a hat."[54]

Second, even if such data were available, the individualized nature of foster children's service goals makes the connection between these proxies and child well-being in any particular case quite tenuous. Although shorter placement may be beneficial for most children, it is certainly not a positive outcome in all cases.[55] If, for example, a child returns to his biological mother before she has overcome her drug addiction, the benefits of shorter placement are substantially outweighed by the danger of returning home. Similarly, although keeping sibling groups together is usually a good idea, in some situations the medical or mental health needs of particular children are such that separate placement is desirable.[56] Accordingly, even an agency that complies with a sixty-five percent sibling placement or sixty percent reunification requirement may not be achieving the right outcomes for the right children. Percentage-based performance measures provide no indication of whether the agency is actually meeting the needs of individual children and are thus quite useless as an indicator of service quality.

Finally, because the measured outcomes are only proxies for child well-being and do not fully reflect the government's contractual goals, they tend to create perverse incentives for agencies. As numerous economic theorists have noted, monitoring and rewarding only measurable aspects of an agent's multi-dimensional performance will induce the agent to neglect all indeterminate tasks, resulting in a suboptimal outcome for the principal.[57] This phenomenon is clearly observable in the foster care system, where "gaming the system" through case choices and internal

---

53 *See* WULCZYN & ORLEBEKE, *supra* note 50, at 10 (discussing the lack of data for determining outcome requirements in foster care contracts).

54 KAHN & KAMERMAN, *supra* note 27, at 109.

55 *See id.*

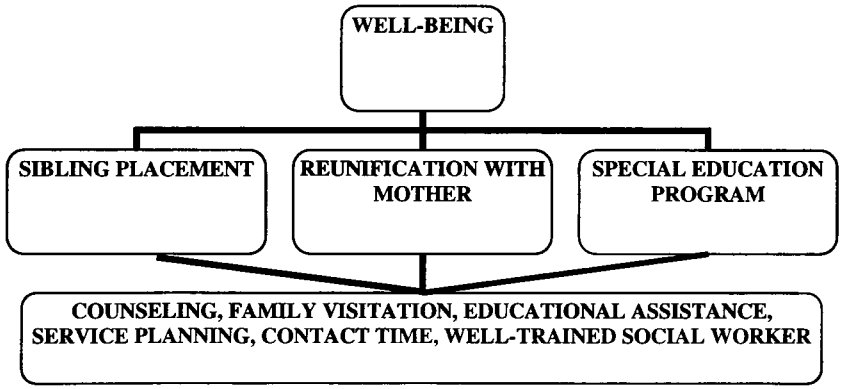
56 For example, a child with severe mental health problems may require placement in a residential treatment facility while his siblings are more likely to thrive in a less restrictive setting.

57 *See, e.g.*, Walker, *supra* note 15, at 526.

reporting procedures is a common occurrence.[58] As discussed above, a sixty percent reunification rate requirement creates no incentive to reunify the particular child for whom reunification is most beneficial, and may instead put pressure on contracting agencies to discharge children to parents who are not yet prepared to provide adequate care. Similarly, the most likely impact of Florida's ten percent cap on child abuse reports is that agencies are more reluctant to report incidents of abuse than they would be otherwise. In this way, proxy outcomes may actually undermine the very goals they are intended to promote.

### B. *An Inputs-Based Approach to Monitoring Foster Care Quality*

As mentioned earlier, input requirements are generally undesirable because they prevent agents from utilizing their own expertise and experience to determine the most efficient and effective means for achieving the desired outcome. However, the contracting problems presented by foster care are quite different than those usually found in the principal-agent relationship. For foster care providers, flexibility is most important in the determination of *output*. Consider the following diagram:




Although well-being is always the ultimate goal of a foster care contract, it has different meaning for each child. The second tier of this diagram represents the individualized service goals of a particular child—i.e., a child whose well-being is best served through placement in a foster

---

[58] KAHN & KAMERMAN, *supra* note 27, at 109. *See also* Hansen, Note, *supra* note 19, at 2493 (observing that in New York City, contracting agencies are so bogged down by "trivial" requirements that they cannot focus on overall outcomes for children and families).

home with his siblings, eventual reunification with his mother, and placement in a special education program. For a different child, these second tier goals might include temporary placement in a mental health facility, vocational training, and eventual adoption by a non-relative family. As discussed, policymakers frequently designate these secondary goals as proxy outcomes, thus restricting the ability of agencies to determine which secondary goals are appropriate for a particular child. Note, however, that many of the *inputs*, illustrated by the third tier, will be the same *regardless* of which secondary goals have been chosen. For example, all foster children need counseling to help them cope with the grief and trauma of foster care placement. Similarly, regular visitation with family members is always essential during a child's stay in foster care. Another universal input is contact time—time spent by the social worker getting to know the family, learning the child's needs, monitoring the foster home, and assisting the biological parents. Perhaps the most important input is a well-trained and highly capable social worker—someone with good personal skills, a strong work ethic, sound judgment, and the ability to connect with clients and help them achieve their goals.[59] The problem with the proxy outcome approach is that it restricts agency decisions where flexibility is needed most, i.e., in determining second tier goals, while leaving unrestricted the one area where monitoring and regulation are most appropriate—third tier universal inputs.

An inputs-based approach to foster care contracting imposes minimum requirements on agencies with regard to universally desirable inputs like counseling, contact time, and social worker qualifications. Although input monitoring is less popular than proxy outcome monitoring, many states have incorporated at least a few input requirements into their foster care contracts. Consider the following examples:

*Medical treatment:* Texas requires that all children receive physical, dental, and psychological exams within ten days of entering care.[60]

*Foster parent screening:* Colorado requires agencies to subject potential foster parents to background checks, medical exams, reference checks, and character and suitability assessments.[61]

*Foster home visits:* Texas requires caseworkers to visit children in their places of residence at least once every three months.[62]

---

59 *See* KAHN & KAMERMAN, *supra* note 27, at 34 (describing social workers as "the heart of the foster care system").

60 *See* CAROLE KEETON STRAYHORN, FORGOTTEN CHILDREN: A SPECIAL REPORT ON THE TEXAS CARE FOSTER SYSTEM 5 (2004), *available at* http://www.window.state.tx.us/forgottenchildren/.

61 *See* Colorado Association of Family and Children's Agencies, Inc., *A White Paper on Foster Care in Colorado* 14 (2002), *available at* http://cafca.net/fostecarewhitepaper.htm.

62 *See* STRAYHORN, *supra* note 60, at 11.

*Legal parent visits:* Michigan requires caseworkers to have face-to-face contact with legal parents monthly and to visit them at their residence every three months.[63]

*Foster home standards:* Utah requires foster parents to provide each child with her own bed, balanced meals, adequate space for recreational activities, and a safe physical environment.[64]

*Foster parent training:* Pennsylvania requires agencies to provide new foster parents with an orientation and to ensure that foster parents receive at least six hours of training annually.[65]

*Social worker qualifications:* Delaware requires that caseworkers have a bachelor's degree in social work or a related field.[66]

Perhaps the most common input measurement is the number of cases handled by each social worker. Despite the increasing popularity of outcome objectives, state policymakers continue to emphasize the correlation between lower caseloads and successful foster care programs,[67] and rightly so. As the Child Welfare League of America reports:

> Child welfare work is labor intensive. Caseworkers must be able to engage families through face-to-face contacts, assess the safety of children at risk of harm, monitor case progress, ensure that essential services and supports are provided, and facilitate the attainment of the desired permanency plan. This cannot be done if workers are unable to spend quality time with children, families, and caregivers.[68]

In other words, by limiting the number of cases handled by agency social workers, states increase the portion of the caseworker's time and effort which the worker "inputs" into each individual child in her care.

Although, as the examples above illustrate, states have successfully identified numerous universally-desirable foster care inputs, the universality

---

63 *See* Michigan's Children's Foster Care Manual, CFF 722-6, *available at* http://www.mfia.state.mi.us/olmweb/ex/cff/722-6.pdf.

64 *See* UTAH ADMIN. CODE r. 501-12-7 (2004), *available at* http://www.rules.utah.gov/publicat/code/r501.

65 *See* 55 PA. CODE § 3700.38, 3700.65 (2006).

66 *See* Delaware Requirements for Child Placing Agencies 4.12, *available at* http://www.state.de.us/ kids/pdfs/occl_regs_cpa.pdf.

67 *See, e.g.*, Jess McDonald, Illinois' Performance Contracting in Child Welfare, Testimony before the Government Management, Information and Technology Sub-Committee of the House Committee on Government Reform (Sept. 6, 2000), http://www.state.il.us/DCFS/docs/testimony.html (discussing Illinois' success in decreasing caseloads); Lisa Snell, Keys to Success in the Florida Child Welfare Privatization Effort, http://www.rppi.org/floridachildwelfare.html (citing high caseloads as a problem in Florida's foster care system); Texas Department of Family and Protective Services, *Commissioner announces historic changes for DFPS* (Oct. 22, 2005), *available at* http://www.dfps.state.tx.us/about/Releases_and_Newsletter/2005/2005-09-01_Commissioner_announces_historic_changes.asp (describing caseload reduction as a key goal in improving Texas' foster care system).

68 Child Welfare League of America, *Guidelines for Computing Caseload Standards*, http://www.cwla.org/programs/standards/caseloadstandards.htm.

requirement restricts the ability of input measurements to capture the extent to which agencies are serving children's individual needs. A number of states have mitigated this limitation by creating service categories which tailor input requirements to a child's level of need. The following service levels, used by a number of counties in Colorado, are typical:[69]

| Level of service | Input requirement: therapy | Input requirement: caseworker contact |
|---|---|---|
| Level one (minimal service needs) | 4 hours per month, group or individual | 2-3 contacts per month, by phone or in person |
| Level two (moderate service needs) | 5-8 hours individual and 4 hours group therapy per month | Weekly, with at least 1-2 in person contacts per month |
| Level three (high service needs) | Weekly, multiple-session therapy, 8-12 hours family therapy per month | Weekly, in person contact |
| Level four (very high service needs) | Weekly, multiple-session therapy, 8-12 hours family therapy | 2-3 in person contacts per week |

At the time of placement, the state assesses the child's medical, emotional, and educational needs to determine the appropriate service level, and then contracts with the agency accordingly. The payment rate for the child corresponds directly to the service level, so that agencies receive more money to care for needier children.[70] This sort of service level designation allows states to expand the number of inputs they measure without imposing service requirements which aren't necessary to the well-being of all children.

Although it is beyond the scope of this article to offer any detailed recommendations for foster care reform, this kind of input monitoring makes sense from a principal-agent perspective. As long as an input is universally desirable within a particular service level designation, the government can include it as a contract requirement without creating undue inflexibility or perverse incentives. Furthermore, in contrast to child well-being, the inputs identified here can be measured and are thus amenable to monitoring and enforcement mechanisms. An inputs-based approach may not solve the principal-agent problem in foster care contracting, but it does

---

69 *See* COLO. ASS'N OF FAMILY AND CHILDREN'S AGENCIES, A WHITE PAPER ON FOSTER CARE IN COLORADO app. a (2002), *available at* http://cafca.net/fostecarewhitepaper.htm.

70 *See generally* Paul Stratton, *Levels of Care and Treatment Foster Care* (2005), http://www.ffta.rog/locpaper.pdf (providing an overview of the use of care levels in foster care contracting); Lucy Beliner & David Fine, *Long-Term Foster Care in Washington: Children's Status and Placement Decision-Making* 7-8 (2001), http://depts.washington.edu/hcsats/pdf/research/longtermfc2001-6.pdf (describing Washington's care level system).

mitigate some of the more undesirable consequences of information asymmetry, particularly organizational slack. At the least, the analysis here illustrates that input monitoring better serves the government's goal of child well-being than the traditional proxy outcome approach.

### C. *Minimizing the Transaction Costs of Input Monitoring*

Regardless of whether the focus is on inputs or outputs, monitoring incurs costs. Principals will engage in monitoring activity only to the extent that its marginal benefits (i.e., higher agent performance) outweigh its marginal costs. Under some circumstances, the difficulty of monitoring will prevent a contract from arising. In other words, the transaction costs of contracting out a service will be so high that it becomes more efficient for the principal to perform the service herself. An important question in foster care is whether the transaction cost of monitoring private service providers outweighs the benefits of privatization discussed in Part III. If so, it might be preferable for the government to provide its own foster care services despite the expertise, desirable norms, and community relationships of private nonprofits. This is not, however, a question that can be answered by economic theory—a comprehensive cost-benefit analysis requires empirical evidence that is beyond the scope of this article.

Nonetheless, it is clear that it is prohibitively expensive to monitor all of the inputs that contractors invest in their foster care programs. This raises another important question: how do states decide which foster care inputs to monitor? Most principal-agent analysis focuses on the efficiency of monitoring input versus output,[71] without much discussion about the extent to which particular inputs (or outputs) are worth monitoring. This is primarily because a principal's choice of monitoring instruments requires cost-benefit analysis that is necessarily context-specific. Although I identified a number of universal foster care inputs listed in Part IV.B (such as grief counseling, family visitation, and social worker training), it is important for state policymakers to determine as an empirical matter whether the increased contractor compliance associated with this sort of input monitoring outweighs its costs. However, Fahad Khalil and Jacques Lawarrée have made an important theoretical contribution to this decision-making process with their model of ex post monitoring.[72] Khalil and Lawarrée argue that principals can minimize monitoring costs by announcing a large number of performance measures ex ante, but deciding

---

[71] *See, e.g.*, Eric Maskin & John Riley, *Input versus Output Incentive Schemes*, 28 J. PUB. ECON. 1 (1985); Fahad Khalil & Jaques Lawarée, *Input versus Output Monitoring: Who is the Residual Claimant?*, 66 J. ECON. THEORY 139 (1995).

[72] Fahad Khalil & Jacques Lawarrée, *Catching the Agent on the Wrong Foot: Ex Post Choice of Monitoring*, 82 J. PUB. ECON. 327 (2001).

ex post which ones to actually monitor. They observe that this approach "gives the principal the option of monitoring only a subset of the variables and save on monitoring cost while using the incentive power of a large number of variables."[73]

The Khalil-Lawarrée model is quite useful in the context of foster care contracts. Although it is prohibitively expensive for states to track the visitation history of every foster child or the educational qualifications of every social worker, states are able to achieve a certain amount of contractor compliance through random auditing of case files. Again, determining the optimal level of auditing is an empirical question, but this approach makes input monitoring a viable strategy for minimizing the costs of information asymmetry in foster care contracts.

## CONCLUSION

The difficulty of creating and maintaining a high quality foster care system may be frustrating, but it is not surprising. Although researchers and policymakers rarely consider economic principles in their reform efforts, this article contends that economic theory provides an important lens for understanding the obstacles faced by state governments in effectively contracting with foster care providers. Within this economic framework, I have attempted to make three related points. First, the poor quality of foster care services is best understood as a principal-agent problem created by the absence of meaningful performance measures, which in turn leads to an information asymmetry between the government and contracting agencies. Second, despite the difficulties inherent in foster care contracting, privatization may still be desirable because private agencies have certain strengths in provide foster care services which the government lacks. Third, an inputs-based approach to measuring foster care quality better serves the government's goal of child well-being than the traditional proxy outcome approach, and the costs of such monitoring can be minimized through ex post auditing decisions.

As the analysis in this article illustrates, successful foster care reform remains an illusive goal. The application of economic principles is unlikely to solve the foster care crisis, but it can make an important contribution to the efforts of policymakers and researchers to devise and implement strategies that improve the quality of foster care in the United States.

---

[73] *Id.* at 329.

# LIFE v. DEATH: WHO SHOULD CAPITAL PUNISHMENT MARGINALLY DETER?

*Charles N. W. Keckler*[*]

## INTRODUCTION

Econometric measures of the effect of capital punishment have increasingly provided evidence that it deters homicides. However, most researchers on both sides of the death penalty debate continue to rely on rather simple assumptions about criminal behavior. I attempt to provide a more nuanced and predictive rational choice model of the incentives and disincentives to kill, with the aim of assessing to what extent the statistical findings of deterrence are in line with theoretical expectations. In particular, I examine whether it is plausible to suppose there is a marginal increase in deterrence created by increasing the penalty from life imprisonment without parole to capital punishment. The marginal deterrence effect is shown to be a direct negative function of prison conditions as they are anticipated by the potential offender—the more tolerable someone perceives imprisonment to be, the less deterrent effect prison will have, and the greater the amount of marginal deterrence the threat of capital punishment will add. I then examine the empirical basis for believing there to be a subset of killers who are relatively unafraid of the prison environment, and who therefore may be deterred effectively only by the death penalty. Criminals, empirically, appear to fear a capital sentence, and are willing to sacrifice important procedural rights during plea-bargaining to avoid this risk. This has the additional effect of increasing the mean expected term of years attached to a murder conviction, and may generate a secondary deterrent effect of capital punishment. At least for some offenders, the death penalty should induce greater caution in their use of lethal violence, and the deterrent effect seen statistically is possibly derived from the change in the behavior of these individuals. This identification of a particular group on whom the death penalty has the

* Ph.D. (all but dissertation, Human Evolutionary Ecology), University of New Mexico; J.D., 1999, University of Michigan; M.A. (Anthropology), 1992, University of Michigan; A.B., 1990, Harvard College. Initial research for this paper occurred during my residence as John M. Olin Fellow at Northwestern University School of Law. Further support during the preparation of the Article was provided by George Mason University School of Law. My thanks to the assistance of the members of the Levy Workshop on Law and Liberty at George Mason University School of Law, who provided insightful comments during an initial presentation of these ideas. Charles Keckler may be contacted via email to: charles@keckler.net.

greatest marginal effect naturally suggests reforms in sentencing (and plea bargaining) which focus expensive capital prosecutions on those most insensitive to alternative criminal sanctions.

Driven by a renewed application of econometric tools to homicide statistics, the debate between death penalty "abolitionists" and "retentionists,"[1] has entered a phase of renewed vigor. As recently as 1996, symposia on the death penalty were hard-pressed to find anyone in the legal academy willing to advocate the practice—or even to tolerate it as an exercise of legislative discretion.[2] To the chagrin of abolitionist academicians, recent years have seen the development of a series of studies, which can, in comparison with the early work on the death penalty, draw on a much larger sample of executions, and a much longer time span in which to observe their effect. It is now over thirty years since the restoration of the death penalty, and numerous states have moved in and out of the practice, allowing for comparison and inference of causation.

This work has shown a substantial effect in terms of lives saved by each execution, although the coefficients of how many lives are saved range wildly, from eighteen to approximately four, the latter estimate being outside the margin of error of the former.[3] These results have led Cass

---

[1] Use of these terms is one of the few conventions of the death penalty debate used here. This terminology, however convenient, is dubious, as it evokes on the one hand William Lloyd Garrison and *The Liberator*, Underground Railroads and the like, while leaving death penalty proponents with some unpleasant Freudian connotations (at the same time as it underdescribes their program, which includes introducing the death penalty where it does not exist as well as retaining it where it is already present). Nevertheless, the logical alternatives such as "pro-life" seem to be semantically occupied, or, as to its opposite, "pro-death," giving no great improvement in neutral description.

[2] Daniel D. Polsby, *Recontextualizing the Context of the Death Penalty*, 44 BUFF. L. REV. 527, 527 (1996). Some commentators have even characterized such conferences as performing an essentially ritual rather than rational function, with the death penalty proponent as a stereotyped monster figure to be overcome by the forces of good. *See* Ronald J. Allen & Amy Shavell, *Further Reflections on the Guillotine*, 95 J. CRIM. L. & CRIMINOLOGY 625, 626 ("Conferences on the death penalty in American law schools typically are self-righteous displays of commitment to revealed truth, the truth being that opposition to the death penalty goes without saying and the only issue is how strongly its proponents can be tarnished with either their illogic or moral depravity. Indeed, the opposition (i.e., the proponents of the death penalty) are typically represented, if at all, by someone who is supposed to utter barely comprehensible rantings about victims and deterrence, but the real point of the display is to demonstrate the horrifying moral shortcomings of one who wishes deliberately to take another's life.").

[3] Hashem Dezhbakhsh, Paul H. Rubin, & Joanna M. Shepherd, *Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data*, 5 AM. LAW & ECON. REV. 344, 369 (2003) (providing estimate of 18, plus or minus 10); H. Naci Mocan & R. Kaj Gittings, *Getting off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment*, 46 J.L. & ECON. 453, 469 (2003) (coefficient of 5); Joannna M. Shepherd, *Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment*, 33 J. LEGAL STUD. 283, 309 (2004) (coefficient of 4.5); Paul R. Zimmerman, *State Executions, Deterrence, and the Incidence of Murder*, 7 J. APPLIED ECON. 163, 163 (2004) (19, with confidence interval between 7 and 31). *See also* Lawrence Katz, Steven D. Levitt, &

Sunstein and Adrian Vermeule to conclude that capital punishment is not only allowable, but in fact may be "morally required," because "[s]tates that choose life imprisonment, when they might choose capital punishment, are ensuring the deaths of a large number of innocent people."[4] Given the assumption of deterrence, this moral precept is hardly remarkable. It is notable, though, that the first of these authors characterizes himself as "skeptical of capital punishment for moral reasons,"[5] while the second had recently declared that the empirical debate over deterrence would be unresolved for the "foreseeable future," but they now imply this debate may well be resolved sooner rather than later.[6]

Because of the large number of variables that may affect the homicide rate, and the relatively small samples of most studies (whose data points are usually states in different years), it is unsurprising there have been difficulties in statistical inference.[7] In the most recent year for which statistics are available, there were 16,137 murders in the United States.[8] There were 59 people executed during this same year.[9] Aside from the

---

Ellen Shustorovich, *Prison Conditions, Capital Punishment, and Deterrence*, 5 AM. L. & ECON. REV. 318, 336, n.9 (2003) (although not considering the estimate to be reliable, calculating a coefficient for the latest period they study, years 1970-1990, that translates into 10.04 lives saved per execution). *Cf.* Joanna M. Shepherd, *Deterrence Versus Brutalization: Capital Punishment's Differing Impacts Among States*, 104 MICH. L. REV. 203, 214-219 (2005) (providing recent summary of empirical literature). *But see* John J. Donohue & Justin Wolfers, *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791, passim (2005) (attacking all aforementioned research as inconclusive).

[4] Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 706 (2005).

[5] *Id.*

[6] *Id.* at n.22.

[7] *See generally* Samuel Cameron, *A Review of the Econometric Evidence on the Effects of Capital Punishment*, 23 JOURNAL OF SOCIO-ECONOMICS 197 (1994). *See also* Katz, Levitt, & Shustorovich, *supra* note 3, at 319 ("Even if a substantial deterrent effect does exist, the amount of crime rate variation induced by executions may simply be too small to be detected. Assuming a reduction of seven homicides per execution (a number consistent with Ehrlich, 1975), observed levels of capital punishment in Texas since 1976 (a total of 144 executions through 1997) would have reduced the annual number of homicides in Texas by about fifty, or 2% of the overall rate. Given that the standard deviation in the annual number of homicides in Texas over this same time period is over 200, it is clearly a difficult challenge to extract the execution-related signal from the noise in homicide rates.") (*citing* to the early work of Isaac Ehrlich, *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, 65 AM. ECON. REV. 397 (1975)). I note only that there would be some value in extracting this signal, since if true, twenty years of executions would have saved the lives of a thousand people in Texas alone.

[8] *Crime in the United States 2004* (Fed. Bureau of Investigation, Washington, D.C.), at 15, *available at http://www.fbi.gov/ucr/cius_04/documents/CIUS2004.pdf.*

[9] Thomas P. Bonczar & Tracy L. Snell, *Capital Punishment, 2004*, STATISTICS BULLETIN (Bureau of Justice, Washington, D.C.), Nov. 2005, at 1, *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/cp04.pdf. These figures are similar to 2003, when there were 16,503 criminal homicides, *see supra* note 8, at 15, and 65 people were executed. Thomas P. Bonczar & Tracy L. Snell, *Capital Punishment, 2003*, STATISTICS BULLETIN (Bureau of Justice, Washington, D.C.), Nov. 2004, at 1, *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/cp03.pdf.

obvious fact that the number of people killed by other individuals swamps those killed by the state, the important point to keep in mind is that homicide rates have fluctuated substantially over the last few decades. Most notably, considering only those states that have the death penalty, murder has declined by more than 50% from its peaks in the early 1980s and the early 1990s.[10] The current national rate has declined by a third just since 1995.[11]

One simplistic way to frame the debate is therefore to start with the proposition that there are about eight thousand fewer people killed every year in this country than would have been the case if formerly prevalent homicide rates had continued into the present, and to inquire whether any of these eight thousand per annum were saved by the presence of the death penalty. It would be absurd to assume that all or even the majority of them were "saved" in this way, and even the most ardent retentionists have not made so grandiose a claim. Instead, the claim is that some amount of the variation in homicide rates (here using the informal example of a figure varying between sixteen and twenty-four thousand) is accounted for by the presence or absence of a death penalty. But how much? Assume, as some researchers have, that about seven fewer murders are generated by each execution. At the current rate of execution, that means about four hundred people are saved in some future year. However, since the yearly variation is eight thousand, the death penalty, assuming a quite meaningful and "important" deterrent effect, still only accounts for about five percent of the variation in homicide rates (reducing the absolute rate of homicide by about two percent). Presumably, therefore, if one follows the usual econometric techniques of least-squares regression or related methods, the issue will be the amount of additional explanation provided by a properly entered death penalty variable, which will vary between 0% (in which case all changes in homicide are accounted for by other factors) and perhaps, up to 10% (which might seem an extraordinary effect when there is only a 1/250 chance of any particular murder being punished by execution).

Much of the statistical confusion and uncertainty over the deterrent effect of the death penalty stems from the simple fact that researchers are arguing over the coefficient of an independent variable that everyone acknowledges is not a primary determinant of the dependent variable; the question is whether a statistical signal for an effect can be extracted from the noise of homicide rates, even though the more powerful determinants of homicide rates (which are primarily demographic in nature) and which are required as control variables, have not yet been agreed upon. In order to "know" there is "no deterrence" we need to have empirical evidence that the appropriate coefficient relating execution to a change in the rate of homicide is centered on zero, plus or minus some relatively small level of

---

10 *See* Donohoe & Wolfers, *supra* note 3, at 801.

11 *See Crime*, *supra* note 8, at 15.

error, and this requires a substantial amount of data.[12] Moreover, the coefficient should be valid among that small group of very bad men for whom it is relevant, and it is rare that research is focused on these people. There is a great difference, too often elided, between concluding one currently lacks the ability to distinguish a coefficient from zero, and concluding that the coefficient *is* zero. Further, an important distinction needs to be made if the mean of an estimate shows zero, yet is not statistically distinguishable from zero (i.e. zero may be within the range of error of an estimate). In such circumstances, if we are sure for other reasons there is an effect *of some kind*, it is *more likely* to be a deterrent one. It is therefore at least risky to assume the absence of an effect, based merely on a current lack of proof.

My impression of the literature is that much of it proceeds by assuming that the absence of proof has proved absence, which is mistaken not only as a matter of logic, but also highly problematic as a matter of policy. If there is no effect of the death penalty on deterrence, it would most likely be because there are countervailing forces that prevent it from reducing the behavior it sanctions. The causal relationship will not, however, resemble the relationship the murder rate has with the position of the planet Mars, bringer of war. With apologies to astrology, we would *not* be surprised to find Mars has "no effect" on murder; by contrast, it goes against expectations to find that the state's threat of death, directed against specified acts of citizens, has "no effect" on the frequency of such acts. Without more information, we do not know whether this threat would have a very small effect, or the perverse effect of increasing rates, yet it is counterintuitive to assume it would possess no connection at all.

Of course, the same conditions hold for proof of deterrence as hold for proof of nondeterrence, the difference being that the advocate of deterrence would be attempting to confirm a coefficient centered on a number *greater* than zero. The current variability in the empirical estimates of deterrence should, in my view, make one pause before accepting as more than preliminary those studies on which Sunstein and Vermeule, for instance, condition their appropriately provisional argument in favor of expanded use of the death penalty. But even without full confidence in nonzero deterrence, the empirical work should make us profoundly question any supposed proof we have for zero deterrence effects, an assumption on which too much policy discussion depends. The serious possibility of deterrence must be acknowledged, and standing alone, this has important normative consequences implicitly incorporated below.[13]

---

[12] The lack of sufficient data to assure statistical significance is one of the chief critiques of Donohoe & Wolfers, *supra* note 3, at 836-837.

[13] For a more explicitly normative approach, *see* Sunstein & Vermuele, *supra* note 4, at 715, noting, among other valuable points, that "a degree of reasonable doubt need not be taken as sufficient

The econometric work attempting to extract a death penalty effect is also inconclusive for another reason: the lack of a behavioral basis for why the effects detected are in fact predictable and understandable from a theoretical model. It is commonsensical to posit fear of death as a behavioral motivator. Donohue and Wolfers, for instance, express the logic of capital punishment as: "raise the price of murder for criminals, and you will get less of it."[14] But they say little more than this, and it is not often made explicit why criminals should be expected to behave as the models infer, and why they should do so at some particular rate. That this is asking a lot of data analysts is an appropriate rejoinder, but in the absence of a behavioral model one is able to test, as opposed to results one is able only to interpret, I think it is likely the intellectual stalemate over death penalty deterrence will continue.[15] This is not to say that deterrence *lacks* a theoretical foundation; the burden of what follows is to put together the pieces of such a foundation long present in the literature, stating them explicitly and coherently at a time when the legal academy has once again been compelled to take seriously, at minimum, the factual possibility of deterrence.

As a cautionary note, inconclusiveness regarding a deterrent effect would not necessarily warrant a refusal to enforce capital punishment. The executive, the legislators and the citizenry are not at liberty to enjoy the subtle pleasures of intellectual ambivalence. To not execute criminals is to

---

to doom a form of punishment if there is a significant possibility that it will save large numbers of lives."

[14] Donohue & Wolfers, *supra* note 3, at 795. They go on to discuss a countervailing "more sociological approach" that supposes "there *may be* social spillovers as state-sanctioned executions cheapen the value of life, *potentially* demonstrating that deadly retribution is socially acceptable." *Id.* at 796 (emphasis added). From this they deem the realm of "theory inconclusive." Yet one must question whether the theoretical premises here are actually in equipoise; on the one hand we have a basic principle that explains most other behavior whether human or non-human; on the other we have a wholly hypothesized psychological mechanism in which criminals begin to perceive murder as somehow okay because the state kills murderers. The reader's own intuitions must be their guide as to whether these have equivalent prima facie plausibility. Unlike the situation with price theory, a literature search will not find this "sociological" theory used much outside the death penalty context, although presumably it predicts, for instance, that a state that engages in high levels of taxation will thereby encourage theft in the populace.

[15] *See* Dan M. Kahan, *The Secret Ambition of Deterrence*, 113 HARV. L. REV. 413, 436 (1999) ("No issue of criminal justice has been subjected to greater empirical study than whether the death penalty is an effective deterrent, and on none is the evidence more ambiguous and conflicting."). It is interesting to compare this judgment—made prior to the latest round of studies favorable to deterrence—with the far less agnostic views of Professor Lempert: "Supporters of the death penalty, for example, cannot be faulted for resorting to retributive arguments for capital punishment when empirical evidence consistently fails to find a substantial deterrent effect, but they are wrong to suggest that deterrence is an important reason for the death penalty, given the overwhelmingly negative findings of numerous empirical studies." Richard O. Lempert, *Activist Scholarship*, 35 LAW & SOC. REV. 25 (2001) (praising the findings of no deterrence as one of the greatest practical and scholarly achievements of the law and society movement).

make a choice regarding the effect of the death penalty, just as to execute them reflects such a choice,[16] and there seems to be no way around the fact that the state will do one or the other. It is fair to say that this choice is made today in doubt, but as all must recognize, the standard of proof for making policy is not beyond reasonable doubt. It is not even "clear and convincing" evidence. Aside from their legal meaning, these levels of proof are qualitative expressions of standards appropriate in scholarship before a phenomenon is asserted to be an established fact. Action, however, must sometimes be taken even when these standards are not met, guided by the best collective guesses regarding the weight of evidence and theory; this means, inevitably, that one must on occasion proceed even though one is *not* "convinced" and the matter is *not* "clear."

Assuming a murky empirical picture is indeed what we have to guide us, my purpose is to proceed from the general premises of rational choice theory, together with some basic demographic methods, to explore what we should *expect* to see from empirical evidence, as well as where we should look for such evidence. It would be absurd to think a few lines of algebra could convince either way those who are at loggerheads over what is shown by the longitudinal data on homicide; but the present effort might serve to sharpen this debate by developing assumptions held in common or exposing those hidden differences in models of human behavior that contribute to interpretive dispute.

The stakes of this debate may not be high—this depends on which side of the deterrence question one takes.[17] The abolitionist position of no effect means that, although there is an incremental effect on the punishment meted

---

[16] Allen & Shavell, *supra* note 2 at 628, arguing that the death penalty debate should be reframed as a problem of social planning ("In a universe with finite resources, allocation decisions with real consequences must constantly be made, and one of the primary consequences invariably is who will live and who will die, if not tomorrow, then sometime in the future."). *Accord*, Sunstein & Vermeule, *supra* note 4, at 733 (arguing the importance of capital punishment is that it is a currently feasible policy alternative that can be adopted while other methods of reducing homicide are investigated).

[17] This refers only to the question of deterrence, and the "consequentialist" view of the social desirability of the death penalty. I acknowledge the cultural observation that many people seem to be attached or opposed to the death penalty for so-called "moral" reasons, some of them related to what the criminal law literature is pleased to call "desert." To be specific, the quantitative implication of an extreme form of this claim is that any coefficient relating number of executions to number of murders would be irrelevant to the desirability of the death penalty as a matter of public policy. *See* Sunstein & Vermeule, *supra* note 4, at 719. The presence of such views is interesting from an anthropological perspective, but accounting for them is beyond the scope of the present study. If it were shown that indeed the proper measure of the deterrent effect of capital punishment was near zero (or negative), it would then seem appropriate to introduce secondary considerations of the public weal such as feelings of vindication or anxiety, international opprobrium, and the like. Possibly the death penalty would then not be worth the social expenditures made upon it, although given that the number of innocent lives at stake would be very small (involving only those wrongfully convicted and executed, itself a quantity that studies have failed to distinguish from zero), the interest in the policy decision would necessarily be diminished.

out every year to several dozen convicted murderers who are not allowed to live out their span of years within the confines of the penitentiary, the public at large is neither helped nor hurt by the presence of the death penalty. On the other hand, to the retentionist asserting a deterrent effect, these matters should be of considerable importance. The ending of the not-very-happy existence of the murderer, according to this hypothesis, redounds on the populace by saving some number of citizens who would perhaps otherwise be killed. According to the coefficients of the models noted above and multiplied over the number of executions, several hundred citizens might be saved. The death penalty is often noted for being a particularly emotional issue, involving strongly held and asserted beliefs, yet it is surprising how mild the reaction actually is to public behavior that, if one held to even the *possibility* of substantial deterrence, *might* amount to lethal activity on a scale far exceeding that which would usually be of concern (with regard to unknown hazards of pollutant release, say).

As an example, the then-governor of Illinois, George Ryan, shortly before leaving office in January 2003, commuted the sentences of all 156 murderers on Illinois's death row. This expanded the moratorium on executions Ryan had put into effect three years earlier in January of 2000. Governor Ryan was sufficiently praised by the abolitionist community so as to be nominated for the Nobel Peace Prize.[18] The community in favor of the death penalty reacted with a certain amount of grumbling, and there were legal challenges to the propriety of this action, but these were largely focused on the process by which the commutation had occurred, rather than any substantive effect of Ryan's magnanimous gesture upon the public welfare.[19] Nonetheless, the result of a statistical inquiry into death penalty practice with regard to commuting sentences and releasing condemned prisoners, using data up to 1997, indicates that "an additional execution generates a reduction in homicide by five, an additional commutation increases homicides by four to five, and an additional removal brings about one additional murder."[20] The same data failed to show any effect of executions or lack thereof on the rates of other violent crime (robberies,

---

[18] *See* Kevin McDermott, *He's Nobel Finalist, Sources Say*, THE ST. LOUIS POST-DISPATCH, October 5, 2003 ("'Out of 6 billion people in the world, it's down to (an estimated) five people. (Ryan's) in there with the pope,'" said Francis Boyle, the University of Illinois at Champaign-Urbana law professor who has spearheaded the Ryan-Nobel drive."). This acclaim preceded Ryan's felony indictment by a federal grand jury for being the head of a wide-ranging criminal conspiracy, something that might have complicated international travel to pick up his prize. *See* U.S. v. Warner, 396 F.Supp.2d 924, 942 (N.D.Ill. 2005) (detailing charges and rejecting Ryan's efforts to introduce in his defense "good character" evidence of his death penalty opposition and commutation).

[19] *See* People *ex. rel.* Madigan v. Snyder, 208 Ill.2d 457, 479 (finding Governor's action "unreviewable" but expressing hope that governors would use this power in the future in "*individual* cases") (emphasis added).

[20] For a review of this evidence, *see* Mocan & Gittings, *supra* note 3, at 465-466.

burglaries, rapes, or motor vehicle thefts), indicating "capital punishment is a murder-specific deterrent."[21]

It is interesting to apply this model to the natural experiment provided us by Governor Ryan, whose declarations of moratorium and then clemency were responding in part to a widely publicized anti-death penalty campaign led by significant segments of the legal community and the largest Illinois newspaper, the *Chicago Tribune* (awarded the Pulitzer Prize for its coverage). This publicity is relevant because information about a change in law (this one being made through the exercise of purely executive authority) needs to be disseminated in order to induce a change in behavior. To study how criminals *might* have responded—in a purely illustrative spirit—I compared the violent crime statistics provided by the FBI for Illinois over the period 1999-2004 (the six years for which they are readily available online). These list both the total number and rate of all violent crimes, and of homicides.[22]

*Table 1. Illinois Crime Rates 1997-2004*

| | Violent Crimes | Rate per 100,000 | Homicides | Rate per 100,000 | % Violence that is Lethal |
|---|---|---|---|---|---|
| 2004 | 69,026 | 542.9 | 776 | 6.1 | 1.12 |
| 2003 | 70,456 | 556.8 | 896 | 7.1 | 1.28 |
| 2002 | 78,214 | 620.7 | 949 | 7.5 | 1.21 |
| 2001 | 79,504 | 636.9 | 986 | 7.9 | 1.24 |
| 2000 | 81,567 | 656.8 | 891 | 7.2 | 1.10 |
| 1999 | 88,838 | 732.5 | 937 | 7.7 | 1.05 |

The most apparent feature from Table 1 is that, although violent crime has been declining both in absolute and population terms, the number of murders remained relatively steady, at least in absolute numbers, until the most recent year. Consequently, for three years after 2000, murder became a more important component of the overall victimization, and this ratio peaked in 2003, before falling back. Whatever had been driving down the rate of violence in Illinois was having less of an effect (if any) on the rate of homicide. (The recent reduction in homicide is apparently due to a specific effort to reduce the murder rate in Chicago). Naturally, there are many possible hypotheses for this. Perhaps the murder rate was already so low little could be done about it (but the recent reduction belies this). Or, perhaps the factors reducing violent crime are specific to them and not to killing. Nevertheless, the statistics are certainly consistent with the view

[21] *See id.* at 473.

[22] CRIME IN THE UNITED STATES reports, *available at* http://www.fbi.gov/ucr/ucr.htm#cius.

that some form of a *countervailing factor specifically increasing the homicide rate* is present in certain years.

Presuming for the moment the presence of this factor, we can calculate its practical consequences. If homicides had fallen at the same rate as other violent crime in those three years, the percent of homicides would have remained at, taking the higher of the first two observations, 1.1% of such crimes. This would have resulted in a predicted number of murders in 2003, 2002, and 2001, respectively, of 775, 860, and 875. This is 121, 89, and 111 fewer than what is actually observed or, in aggregate, 321 more people turn up dead than we might have thought. Using alternative methods involving changes in the relationship between the state and national homicide rates, Cloninger and Marchesini tied approximately 150 homicides to these events.[23] These are very rough estimates, but they are in line with, although possibly more hopeful than, the statistical prediction of Mocan and Gittings that 624 people (4.0 more murders/commutation x 156 commutations) would be killed by the commutation alone.

A simple calculation, applied to a single case, is naturally inconclusive, and is presented only—complementing the econometric evidence—as an inquiry into why policymakers do not consider more troubling the issue of whether or not such effects are real. It is admittedly difficult for econometrics to pick out an effect on behavior undeniably influenced by many social phenomena. However, this argument cuts both ways because, if one is unable to specify the right model for a complex dependent variable, one cannot be certain a particular independent variable of interest *does not* exert some influence.

Whatever ill might be thought of Governor Ryan (or his erstwhile allies among the bar, the press, and the academy), it would be unrealistic and cruel to suppose that he in any way intended 150-600 innocent people to die as a consequence of his policy. Indeed, because the anti-death penalty campaign was frequently couched in terms of a concern to eliminate the *mere possibility* that innocent lives (by wrongful conviction) might be lost, and there is no reason to doubt the sincerity of this concern, the people involved must have at a fundamental level assumed that the deterrence argument was not simply *unproven*, but *implausible*, perhaps *impossible*. Otherwise, it is very difficult to account rationally for their actions—for if they were *in doubt* about the criminological consequences of commutation, a blanket commutation would have been recognized as

---

[23] The key element was something also visible in the basic data I have presented here: the relative insensitivity of the Illinois murder rate to trends tending to reduce this rate. In my example, this insensitivity was to intrastate factors acting generally on violent crime. In their study, the insensitivity was judged against generally declining national murder rates. Dale O. Cloninger & Roberto Marchesini, *Execution Moratoriums, Commutations and Deterrence: The Case of Illinois* (Economics Working Paper Archive, Working Paper No. 0507002). *But see* Donohue & Wolfers, *supra* note 3, at 819-820 (critiquing this study).

extremely *reckless*, rather than being touted as an essentially costless act of mercy.

Consequently, it seems appropriate that greater attention be paid to examining whether it is at least plausible to believe that at least some substantial portion of potential murderers take seriously the difference between life in prison and death. In Part I, as a necessary preliminary, the model of marginal deterrence is made specifically applicable to the special circumstances where the crime and its perceived benefits involve killing another person, and the possible but not certain social response to the crime involves ending the offender's life. As discussed in Part II, there is good reason to believe that, when the death penalty is a jurisdiction's maximum penalty for first-degree murder, the presence of this sanction *ought* to marginally deter *some* potential murderers from committing at least *some* of the murders they would otherwise be prone to commit in a jurisdiction where the maximum criminal liability is life in prison without parole.

In particular, the most fundamental assumption of the deterrence model, that death is perceived by murderers to be worse than life in prison, appears to hold for the great majority of murderers, although a small but identifiable fraction is either basically indifferent or actually prefers death. In addition, there is inferential but strong evidence that there is wide variation among prisoners in the utility cost per unit time incarcerated. Consequently, certain offenders will experience a relatively lighter deterrent effect from threat of prison, and these will disproportionately consist of violent felons previously incarcerated for long periods, a characteristic of many but not all of those who may become murderers. Finally, for all those charged with murder, but especially for those for whom death represents a *substantially* worse outcome than long-term imprisonment, avoiding any risk of a death sentence will cause them to accept longer terms of imprisonment in lieu of trial, raising the expected mean prison term, and incorporating into the effect of the death penalty any marginal effect on crime of longer sentences. Part III briefly considers a few of the implications of the foregoing conclusions for research and policy.

## I. THE UTILITY OF VICTIMS

Justice Holmes is credited with articulating the "bad man" theory of the law.[24] To put it in its essential form, Holmes proposed that law exists because there are bad people who are selfish, and who will do bad (i.e.,

---

[24] Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 462 (1897) ("If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict, not as a good one, who finds his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience.").

socially harmful) things for their personal gain, unless and insofar as society punishes those bad things with imposed costs in excess of their potential for gain. This view has often been quibbled with, because it is an incomplete theory of the various functions and purposes of the law, and also because the populace is sufficiently self-regarding, often unconsciously, that most of us must qualify as "bad men" who at times require guidance and incentive to keep in accord with social norms. Holmes no doubt intended at least this latter inference, for his view of mankind was not a rosy one, and he pointed to the nature of the law as evidence of human incorrigibility.[25] Regardless of this point, one need not accept wholeheartedly the Holmesian vision to acknowledge in a more limited fashion that certain species of law seem to be directed at individuals whose preferences are unusual as a statistical matter, and moreover, if such persons were given full opportunity to express these preferences, much of the rest of the population would consider itself worse off.[26] Hence, the incentive exists on the part of the majority to put in place mechanisms that will stymie the expression of those harmful desires of the criminalized few. For want of a better word, I will restrict use of the term "bad" to this frustrated subset of the greater population.

Due to the grave difficulty many people—even many sophisticated people—have in understanding the apparently extremely complex and mysterious term "bad," I will offer an example. In February 1992, the State of Wisconsin sentenced Mr. Jeffrey Dahmer to fifteen consecutive life sentences (although he only lasted two years in prison before being killed by a fellow convict). It might be objected, probably correctly, that Jeffrey Dahmer is not quite what Justice Holmes had in mind by the "bad man." No doubt Dahmer represents an extreme case, chosen specifically to illustrate an extension of Holmes's reasoning applicable to the law of capital murder. For, if the law in general is directed at the behavior of the bad man, the law of capital murder is directed at the *very* bad man,[27] and the very bad man—unlike Holmes's bad man, perhaps—differs from you and me in important ways.

---

25 *See* J. W. Burrow, *Holmes in His Intellectual Milieu, in* THE LEGACY OF OLIVER WENDELL HOLMES, JR. 17, 25 (Robert W. Gordon ed., 1992) (explaining the origins of Holmes's Darwinian attitudes).

26 *Cf.* Kenneth G. Dau-Schmidt, *Economic Analysis of the Criminal Law as a Preference-Shaping Policy*, 1990 DUKE L. J. 1, 15-16 (1990) (characterizing criminals as those whose preferences impose negative externalities on others).

27 Since the population subset of interest is in fact almost exclusively male—itself a highly relevant fact, although it seems to be only uncomfortably accommodated in the criminal law literature—I feel no compunction about maintaining the nineteenth-century flavor of this phrase. *See* MARTIN DALY & MARGO WILSON, HOMICIDE 178 (1988) (reporting that for homicides occurring in connection with a crime—which are the ones usually eligible for capital murder—more than 97% of perpetrators were male). *See also Capital Punishment, 2004*, *supra* note 9, reporting 52/3314 capital inmates are female, meaning death row is greater than 98% male.

Prior to his detention, Mr. Dahmer had been busy since the late 1970's gratifying his needs for a peculiar sort of companionship, resulting in, among other activities, attempts to transform acquaintances into docile helpmates by "drilling holes in his living victims' heads, [and] pouring in chemicals to 'zombify' them."[28] Dahmer's involuntary patients died, however, because he could never perfect his psychosurgery technique. One should observe—as the Wisconsin jury in this matter certainly did—that the social norms of Milwaukee appear to have compelled Mr. Dahmer to pursue his activities surreptitiously, hampering them by, for instance, requiring him to reject potential subjects who had automobiles, as abandonment of these would create problematic inquiries upon the zombification and disappearance of their drivers.[29] He was also forced to expend extra labor in concealment, for instance, pulverizing bones or dissolving them with acid. Dahmer correctly perceived that the authorities would prevent his activities, as indeed they did after 13 years and 17 victims. Obviously, the legal regime and its associated threat of incapacitation and punishment failed to *fully* deter Dahmer. In certain circumstances over those years (opportunities partly contrived by him) his subjective expected gains were still larger than his expected costs, yet the law did alter the *rate* of his crimes by increasing the labor and time investment necessary to reduce the chance of capture[30] and, thus, the expected cost. The detection-avoidance investments were sufficiently onerous so that only occasionally was it "rational" for him to violate the legal rules against killing and eating people. On most days, Dahmer could not, for fear of the authorities, *act as he otherwise would have*, a phenomenon we usually call deterrence. Dahmer, a very bad man, *was* deterred, and in consequence, lives were saved.

To abstract a bit, Dahmer had certain tastes in his set of preferences that created negative externalities for his victims and society at large; the satisfaction of these preferences is banned by the law of murder and is of no account in most versions of social utility, although by hypothesis, it forms a major component of Dahmer's personal utility, which he pursued to some extent rationally. Although not as lurid in their desires, other criminals can likewise be characterized as possessed of socially inutile preferences, such

---

28 Anne C. Gresham, *The Insanity Plea: A Futile Defense for Serial Killers*, 17 LAW & PSYCHOL. REV. 193, 200 (1993).

29 *See* Gresham, *id.* at 205 (the calculated manner by which victims were selected and evidence concealed was emphasized during trial, in the successful effort to defeat Dahmer's insanity plea).

30 This is a direct implication of Becker's original economic model of crime, the primary progenitor of most subsequent models (including, of course, this one). *See* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968). For a blessedly concise treatment of the economic theory of criminal deterrence, *see* Steven Shavell *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 COLUM. L. REV. 1232, 1234-35 (1985); for a more extended and recent review of Shavell's thinking, *see* STEVEN SHAVELL, FOUNDATIONS OF THE ECONOMIC ANALYSIS OF LAW, 471-568 (2004).

as wishing for a monopoly on the affections of their inamorata despite the presence of a sexual rival, or desiring the use of an expensive automobile without the inconvenience of paying for it. As such, there is nothing "irrational" or socially harmful about possessing preferences for love and luxury. The problem with criminals, for the most part, even murderers, with whom we are most concerned, is simply that their desires are not constrained by their budgets, and they therefore resort to force and fraud in order to obtain them more cheaply.

Only rarely is capital murder done for its own sake, à la Leopold and Loeb; there is generally an instrumental purpose to violence, and if sufficient resources were at hand, and one was not miserly about spending them, most ends can be achieved. Even as to criminals who desire goods and services for which there exists no legal market, i.e. cocaine or sex with minors, this does not imply violence, for if the criminal is willing and able to invest time and money, such goals can often be satisfied illicitly (but without physical force). Most criminals are not millionaires, however, in part because most people are not millionaires, but also because the criminal mindset is not conducive to gaining or retaining large amounts of wealth, and even more so because millionaires have ways of getting what they want through legal exchange, even if they want rather a lot.

Consider a potential killer who thinks a female relative has "dishonored" the family, and in order to remove the stain on him and the rest of the woman's family—that is, to gain a perceived benefit—the woman must be killed; this is an abbreviated description of the phenomenon of so-called honor killing, a problem in several areas of the world.[31] Of course, the woman's offense could involve several degrees of violation of the local code of sexual morality. She could shame the family more or less, and thus the "benefit" of removing this shame would be more or less. Therefore, an honor killing, although it is bound up with a variety of what we could call unreasonable emotions, is likely to respond to rational incentives. If we increase the sanction for taking a woman's life, a violent male relative will probably find it within himself to tolerate certain low-grade "lapses" in behavior or at least require stronger proof of "immorality" before acting. And, indeed, the global effort to rein in honor killings focuses on the excessive tolerance some countries show for such murders and betrays a deterrence rationale along with its obvious point about upholding a woman's right to life and personal autonomy.[32] It is

---

[31] Kathryn Christine Arnold, *Are the Perpetrators of Honor Killings Getting Away With Murder? Article 340 of the Jordanian Penal Code Analyzed Under the Convention of the Elimination of All Forms of Discrimination Against Women*, 16 AM. U. INT'L L. REV. 1343, 1358-1359 (2001).

[32] *See, e.g., id.* at 1373, complaining that "Jordanian law [on honor killing] is particularly discriminatory to women because of the lack of legal deterrent and the wide scope of males who will benefit from it." *See also* Marie D. Castetter, Note, *Taking Law Into Their Own Hands: Unofficial and Illegal Sanctions By The Pakistani Tribal Councils*, 13 IND. INT'L & COMP. L. REV. 543, 552 (2003)

generally thought the low penalties attached to "honor killings" produce more of them than would otherwise occur and, indeed, permit murders not truly motivated by family honor, but which can be disguised as such.[33] Most of these murderers—unlike Dahmer—have only one person they are prone to kill, but they too can be deterred because, whether or not they think a woman "ought" (in their system of belief) to be killed, they will not want to bear a high cost of doing so.

The usual conception of deterrence in the literature, that our goal is to raise the cost of the forbidden act beyond the value set upon it by the potential actor, can therefore be slightly misleading. The value of the act varies among individuals, and even as to individuals, its net value (taking into account, for instance, the possibility of resistance by the victim) varies between occasions. It is simply not feasible to deter the most extreme individuals, when they are their most desperate, by means of a generally applicable penalty, and it can be inefficient (as well as draconian) to do so.[34] What deterrence can do is to reduce the number of individuals for whom a particular act is ever rational (the not so very bad men) and to reduce the number of times acts are rational for those very bad persons who will have motives for crime that are not fully deterrable. This necessarily presumes a focus on the particular mentality of potential criminals as a distinct class of economic actors, together with the cost-benefit context in which they make their decisions.[35]

Introducing a certain level of formalism may now be appropriate in order to clarify how the view expressed here diverges from previous considerations of deterrence. The basic deterrence model can be phrased as stating that D commits crime Q if and only if, as to crime Q, it is profitable as defined by (1) below, where B is the benefit D derives from successful commission of the crime; C is the intrinsic non-legal cost of the crime, incorporating such risks as victim resistance and expenditures for

---

(reporting that with regard to rural Pakistan: "The killings are on the rise because the murderers in honor killings are rarely punished.").

33 *See* Arnold, *supra* note 31, at 1370 (honor killing is a pretext).

34 *See* Shavell, *supra* note 30, at 1242. Shavell's model of designing penalties is based on judicial uncertainty about what the correct penalty is for a particular person, knowing that those who commit a crime, or who are tempted to commit one, vary in those characteristics relevant to the fixing of an optimal penalty *for them*, and this creates an informational difficulty for society, which is obliged to attach to crimes a generic penalty *for everybody*. In this way it clearly foreshadows the current analysis, wherein an attempt is made to get greater clarity about the level and distribution of the relevant variation.

35 Or, as others have defined it, the concern here is with "the cost-benefit analysis potential offenders perceive—which is the only cost benefit analysis that matters." Paul H. Robinson & John M. Darley, *Does Criminal Law Deter?: A Behavioural Science Investigation*, 24 OXFORD J. LEGAL STUD. 173, 174 (2004).

equipment;[36] A is the probability that a legal sanction will be imposed (melding here for simplicity chances of arrest, prosecution and conviction); and P is the level of that sanction.

(1) $B-C > A*P$

In a world without law, or as to activities not legally discouraged, A and P are set to 0, and (1) reduces to a simple comparison of cost and benefit. In order to make subsequent discussion more concrete, assume the act under consideration is the murder of another human being, Q, and the law will take some concern with it, meaning $A > 0$, and $P > 0$.

If B is the benefit one derives from killing another person, it clearly must vary not only with the defendant but with the victim. For example, Dudley kills Quigley to what gain? Even if Quigley couldn't fight back and C is very low, and even if Dudley could "get away with it" (A is very low), there has to be something Dudley gets out of the transaction.[37] This deserves significant emphasis. With six billion people in this world, it is notable they do not kill each other nearly as often as they could, because merely harming another person (as opposed to robbing them, for instance) can only help a killer in peculiar circumstances. Whether or not the death penalty ought to deter does not depend on how we, or "people," or even other members of the animal kingdom (whose actions, by and large, are governed by a notable aversion to mortality) set a value on our respective

---

[36] C is a variable that is not always separately considered. *See, e.g.*, Shavell, *supra* note 30, at 1238. This is a mistake, since important policy concerns depend on it, like those involving the desirability of an armed population, and individual differences in this variable help account for why some people (e.g. elderly unarmed ladies) are more likely to be found as victims rather than perpetrators. According to Holmes's speculation, C for some people would also include something called "the conscience" that imposes an increment of subjective cost to actors, over and above the cost to their material well-being, but Holmes is not very interested in such hypothetical persons.

[37] What we would all do if the cost of murder were to fall to near-zero we need not address. This is emphatically *not* the case in pre-legal societies, since the potential vengeance by surviving relatives, as well as victim resistance, usually keep the effective cost of murder well above zero. *See* MARTIN DALY & MARGO WILSON, HOMICIDE 224-227 (1988) (discussing among other groups, the Ifugao of the Philippines, who have but "one general law . . . *a life must be paid with a life*") (internal quotation marks omitted, italics in original). In societies with tradable goods, homicide compensation in the form of a reciprocating transfer of chattels or valuables to the victim's heirs places an enumerated "cost" for the crime. *See, e.g.*, MERVYN MEGGITT, BLOOD IS THEIR ARGUMENT 137-143 (1977) (describing negotiated compensation in pigs paid in 57/76 homicides in Highland New Guinea group). This contrasts with the theoretical costs of which we are speaking here, created by the probabilistic and heterogeneous collection of non-monetary harms inflicted by the state in "advanced" societies. The only circumstances where costs for murder fall to near-zero for a potential murderer occur when there is not only an *absence of state punishment* for the crime, but also *approval by a state* with monopoly of force, guaranteeing the perpetrator will act without consequences or resistance. Hence, this cost structure is a feature of genocide, mass murder, massacres, death squads, lynching and related phenomena now all too familiar.

lives vis-à-vis any value we might place on the socially unsanctioned elimination of the lives of other people. Rather, it depends entirely on the behavioral niceties of a small, distinct and identifiable group of males who will at one or more times in their lives seriously contemplate murder as either a means to some end or as an end in itself.[38] Studies of this group show they have an average of four major felony arrests, undoubtedly an underestimate since it fails to account for their usually lengthy juvenile records.[39]

Disregarding group enemies, a vanishingly small number of other individuals will be known by most actors to satisfy the equation B-C > 0, given any particular pair of persons d and q. Formally, for the set of people $D_1, D_2, D_3 \ldots D_{six\ billion}$, and potential victims, $Q_1, Q_2, Q_3 \ldots Q_{six\ billion}$, $C_{d,q} > B_{d,q}$ is much more likely than $B_{d,q} > C_{d,q}$. The law of murder is simply not concerned with that subset of D for whom C > B for all Q. Rather, it must focus purely on the complementary subset, call it D*, for whom there exists a pair (d,q) where $B_{d,q} > C_{d,q}$, where, in other words, a murder will occur in the absence of some probability (A) of a meaningful sanction (P).

Although A certainly can be adjusted by legal rules, especially when it includes as it does here issues of the likelihood of conviction of guilty parties, A is basically an issue of law enforcement, and the main concern here is P, the level of the sanction. Probability of arrest and conviction certainly matter, but since we are focusing on murder, they are already relatively high.[40] The reason Dahmer was deterred from pursuing victims

---

[38] Sir James Stephen, the great nineteenth-century deterrence theorist and legal reformer, was among the many who have no doubt anticipated what *should be* an obvious point: "Some men, probably, abstain from murder because they fear that, if they committed murder, they would be hung. Hundreds of thousands abstain from it because they regard it with horror." JAMES FITZJAMES STEPHEN, A GENERAL VIEW OF THE CRIMINAL LAW OF ENGLAND 99 (1863). Stephen's view is perhaps overly optimistic; the real restraint on murder is the lack of the usual police triumvirate of means, motive and opportunity, and the most important of these is lack of motive. Killing other people is not like killing flies—it is an inherently dangerous and risky activity with which almost nobody has any experience. To murder someone without a very good reason to do so is supremely irrational, and most of us simply lack the necessary very good reasons. Moreover, because lethal violence has probably *always* been costly, including in human ancestral environments, discussed *supra* note 37, it is extremely unlikely that evolution would have produced a "killer instinct." Instead, there should have been strong selective pressure for a mechanism that could at some level take account of the benefits and risks of such perilous behavior. *See generally*, Owen D. Jones, *Time-Shifted Rationality and the Law of Law's Leverage: Behavioral Economics Meets Behavioral Biology*, 95 NW. U. L. REV. 1141, 1173 (2001) (discussing reference to ancestral problems for the development of human decision making apparatus).

[39] Daniel D. Polsby & Don B. Kates, Jr., *American Homicide Exceptionalism*, 69 U. COLO. L. REV. 969, 996 (1998) (noting that that any pretended similarity of murderers to the general population is an "extravagant falsification of reality").

[40] The "clearance rate" for murder, or chance of an arrest, is in excess of 60%. *See Crime*, *supra* note 8, at 264, (reporting rate of 62.6%). Subsequent prosecution and conviction are lower, but never fall to such an insignificant probability that a potential murderer would ignore the potential of suffering legal repercussions. It does mean that individuals who would derive a benefit from a killing that is a

with cars was, at the simplest level, because a Q(with car) was associated with a high A, one that resulted in (1) not being satisfied, whereas a Q(without car) sufficiently lowered A so that (1) was satisfied. It would be dangerous to assume, though, that only the probability of arrest and not its multiplier, P, mattered in these cases. If Dahmer had known that, upon capture, he would merely have been ticketed and ordered to attend a "safe sex" class, the frequency of circumstances where (1) is satisfied would have been higher. From his behavior we know that:

(2) $A_{CAR} * P > B\text{-}C > A_{NO\ CAR} * P$

We do not know the level of A precisely, but assume that B, C, and P are the same between the two situations, and that P is life in prison (because Wisconsin is a non-death penalty state). Suppose $A_{CAR}$ is approximately equal to 60% (a normal clearance rate for a murder) and $A_{NO\ CAR}$ is equal to 10% (a rate commensurate with the number of people Dahmer killed). This would mean an incremental difference of .5 P provided deterrence, because B-C was somewhere between .6P and .1P.

Reducing the penalty therefore is mathematically equivalent to some level of reducing the arrest rate.[41] What we can tell (given these assumptions) is that, if we reduced the penalty from life imprisonment to something lower, and the expected cost of a sanction fell below that associated with a 10% chance of life imprisonment, then people with cars would have been at risk from Dahmer. In this case, the equivalent reduction would be somewhere between P and P/6, whatever is $1/6^{th}$ as bad as life in prison. There is no *a priori* function relating this, mainly because individual criminals will vary in how quickly they discount future losses, as well as in how they perceive the fixed cost of any arrest and imprisonment; P/6 might mean 1 year's imprisonment or 5 years. However, P/6 is almost certainly greater than the perceived sanction of probation or a suspended sentence because, looked at from the other side, it must be that the murderer perceives the net benefits of killing as a meaningful fraction of the murderer's lifetime utility, and it is most unlikely that lifetime utility is so

---

significant percentage of their own expected lifetime utility cannot be deterred by the criminal justice system, because, as is discussed below, the most that can be achieved by either lifetime imprisonment or capital punishment is the elimination of all future utility. Since this is discounted by A, the maximum expected penalty is lower than this total loss. *See* Shavell, *supra* note 30, at 1244 & n.45.

41 These interact, however, in a multiplicative way. Therefore, if one increases the arrest rate in a state with a higher P, this has a greater effect than where P is lower. Liu has recently tested this theoretical result, finding, in line with this prediction, that the deterrent effect of an increase in the conviction rate is stronger in states with capital punishment. *See generally*, Zhiqiang Liu, *Capital Punishment and the Deterrence Hypothesis: Some New Insights and Empirical Evidence*, 30 E. ECON. J. 237 (2004).

affected by noncustodial sanctions that an individual would forego what he perceives to be such important benefits.

So, the level of the penalty matters, even if some people are not deterred, and some people are deterred only some of the time. The question for deterrence theory is not what will deter all crime but what the "optimal sanction" will be, given the reduction in crime, balanced against the various costs associated with application of the sanction. In Shavell's formulation, "As the level of sanctions rises, more undesirable acts will be deterred, but the social cost of imposing sanctions in a given instance becomes greater, as does the problem of discouraging socially desirable acts. The optimal level of the sanction will be that which makes the best compromise between these competing effects."[42] Applying this basic theory to murder adds considerable richness but also certain complications to the model. The primary focus of this article concerns how many "more undesirable acts" (murders) will be deterred when the level of sanction increases from life imprisonment to capital punishment, a reasonably concrete question. Nevertheless, deterrence theory instructs that, in order to justify such a focus, we must first justify why the social costs, which differ between these two punishments, can largely be ignored in the present discussion.

To begin with, the number of "socially desirable acts" we might confuse with murder, most especially first-degree murder, is rather small.[43] This is not a problem of mistaken identity, but of a circumstance when the actor did indeed do the act, but the act is not actually a murder, although it so appears. There are certainly circumstances, such as euthanasia or abortion, where somebody can end up dead at the hands of someone else. Likewise, risky activity such as firefighting, coalmining, or surgery brings with it the risk of death by misadventure and potential liability for someone who caused the death. The important point to note is that almost none of these cases fulfill the conditions for first-degree murder, and, where they are criminalized at all, they have much lower penalties. In the current context, therefore, the availability of the death penalty is not relevant to the rate at which such acts are performed.

The most likely socially desirable act confused with first degree murder is probably self-defense or defense of others, wherein socially acknowledged negatives such as injury or death are forestalled by killing another (truly bad) person before he can commit them. However, any confusion that will occur in the justice system will be between cases where

---

42 SHAVELL, *supra* note 30, at 1243-44.

43 From a strictly utilitarian view, it is arguable that some murder victims are themselves "undesirables" whose existence imposes positive harm on the communities in which they live. For instance, the murder rate is in part determined by criminals killing one another. If we increase the sanction for killing generally, deterrence will save "bad" people from an untimely death along with "good" people. Nevertheless, vigilantism is a strained interpretation of what we consider socially desirable—the optimal level of this activity is usually thought to be near zero.

lethal force was justified and where some force was justified, but not lethal force. If the system confuses the former with the latter, somebody may be convicted who committed a socially desirable act, and the possibility of this confusion will presumably cause people to be more cautious (than they should be) about defending themselves with lethal force. However, in many states such persons have recourse to an imperfect self-defense plea and will not be eligible for first-degree murder (or at least, it will serve as a mitigating factor to prevent their being eligible for the death penalty). The rate, therefore, of socially desirable acts of this type lost to excessive precaution is more dependent on the sanctions applied to manslaughter, not to murder.

There are also non-killing acts that are deterred by sanctions for intentional killing with malice aforethought, but they are generally socially undesirable as well, both intrinsically and because they carry the risk of death. For instance, individuals on trial may argue that they did not "intend" to kill their victims—they only intended to beat them "within an inch of their lives" and the victims "accidentally" died when the extra "inch" was taken. Or, rapists may partially strangle their victims in order to control them during the commission of the sexual assault, but the victims struggle and "accidentally" end up with broken necks. Or, a robber may not intend to use a weapon to carry out the robbery, but because the victim had a gun, the robber "had to" shoot to defend his own life. The risk that a death will in fact result from these activities, and that the perpetrator will therefore be punishable by the sanction associated with first-degree murder, means these activities will be engaged in at a rate that falls with the sanction for first-degree murder. Since these activities are harmful, though, this is a simply a social benefit, and it does not decrease the severity of the optimal sanction.[44]

More problematic is the additional social cost required to carry out a capital sentence. For most ordinary crimes, social cost scales naturally with the severity of a sanction. If we imprison someone for one year, society is on the hook for a year's worth of room, board, guarding and so on, and there is a loss of one year's worth of economic productivity (if any) of the inmate. If we increase the sentence to five years, multiplying the one-year

---

[44] Not all the foregoing examples would actually be prosecuted according to the theory of felony murder. However, the reasoning about the absence of an overdeterrence problem applies *a fortiori* to circumstances where the predicate offense involved an actual intentional assault. Therefore, the analysis of felony-murder is generally relevant: "In the felony-murder context, we need not be concerned with the overdeterrence problem. We would regard as a benefit any effect the felony-murder rule had on deterring individuals from committing the underlying felony. Because the felony-murder rule does not operate against those who are engaged in activities that 'are closely allied with and easily confused" with noncriminal conduct, the overdeterrence problems . . . are absent." Kevin Cole, *Killings During Crime: Toward a Discriminating Theory of Strict Liability*, 28 AM. CRIM. L. REV. 73, 102 (1990) (quoting Louis Michael Seidman, *Soldiers, Martyrs, and Criminals, Utilitarian Theory and the Problem of Crime Control*, 94 YALE L.J. 315, 325 (1984)).

cost by five is at least a good first approximation of social cost. Since it costs many thousands of dollars to imprison people, long prison terms for petty crimes are usually inefficient. Death, as they say, is different; it is very cheap to kill someone, but very expensive, even for the State, to secure the *right* to kill someone. As far as costs intrinsic to the penalty are concerned, death sentences are cheaper than life imprisonment, a unique inflection in our scale of sanctions.

As a practical matter, it may be that the social cost of a death sentence, associated as it is with intense expenditure of legal resources, is in fact more costly than life imprisonment. Estimates vary between one and two million dollars per death penalty case;[45] this is the price society pays for putting somebody on death row. In addition, because most death row inmates serve long prison sentences under more expensive conditions of confinement, only a small amount of savings are realized relative to a life sentence, and it is generally thought this cannot compensate for the fixed cost of obtaining a death conviction.[46] Assuming that some additional procedure is appropriate when moving between a life sentence and a death sentence, it is unclear to what extent this fixed cost would be balanced by the decreased cost which occurs when a convict only spends one or two years on death row rather than the current several or more years, and in comparison with the decades of confinement and care (including medical care in old age) that are associated with life without parole. On the other side of the ledger, the presence of the death penalty may induce many murder defendants to agree to an early plea bargain involving a non-capital sentence, a phenomenon to which I return in Part II.B, but which is relevant here because of the costs it may save the justice system. Whatever the precise accounting, what follows will assume that, although it is no cheaper to apply a death sentence (which might encourage us to resort to it more readily than a life sentence), it is also not prohibitively more expensive so long as there is a meaningful level of deterrence associated with it.

This appears justified because even a one or two-million dollar increment is usually deemed to be insufficient in social policy to justify the loss of a life.[47] If the coefficient of deterrence is therefore above 1, and each execution results in saving one life, then execution is socially

---

45 *See* Alex Kozinski & Sean Gallager, *Death: The Ultimate Run-On Sentence*, 46 CASE W. RES. L. REV. 1, 12-14 (1995).

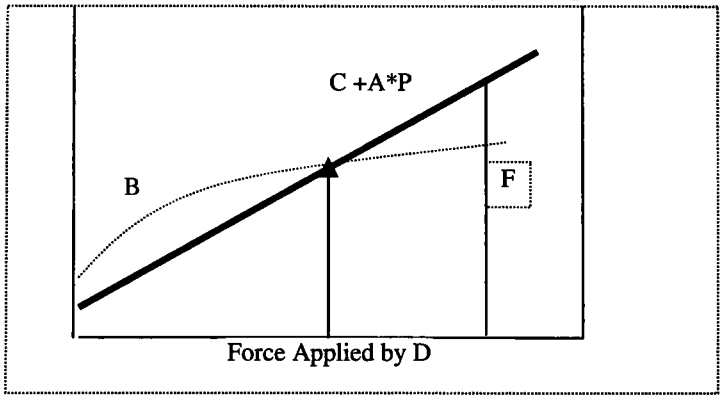
46 These cost factors, though, are pure products of a legal system of capital adjudication widely considered to be bizarrely convoluted, drawn out and inefficient. It is very hard to know what the additional net cost would be if this system were adequately reformed. *See id.*

47 The current estimates used by the federal government in making regulatory decisions are between 5 and 6.5 million. *See* Eric A. Posner & Cass R. Sunstein, *Dollars and Death*, 72 U. CHI. L. REV. 537, 549-550 (2005). A "market" based evaluation, compiled by the same authors through examination of jury awards for wrongful death, and excluding outliers, indicates that the price assigned is somewhat under $3 million. *See id.* at 548. Even this lower figure is well above the incremental cost of a death trial.

desirable.[48] If the coefficient of deterrence is below 1, we probably cannot distinguish the coefficient from zero, so the issue of whether or not the death penalty is implemented loses much of its urgency. Because human lives are sufficiently valuable—following the practice of assigning commonplace values of one million dollars or more even to the economically less valuable members of society—the debate shifts from *cost* toward what is the true coefficient in lives saved by capital punishment's deterrent effect. Assuming we pay extra to obtain a death sentence, that factor alone is not determinative because, if capital punishment has a significant deterrent effect, it is deterring the loss of something valuable enough to justify the extra cost.

The central inquiry therefore returns to the extent to which an increase in sanction reduces the number of murders committed. In the way I will shortly define, this amount of reduction represents the "marginal deterrence" of capital punishment. "Marginal deterrence" has come to mean two distinct, if related things, in the criminal deterrence literature. Following Stigler, Shavell and others have defined this to mean the effect on the *level* or intensity of some ongoing criminal activity.[49] To take the classic example, the more violent and brutal someone is while carrying out a robbery or rape, the higher the sanction will normally be if he is caught.[50]

*Figure 1 – Marginal Deterrence After Stigler*



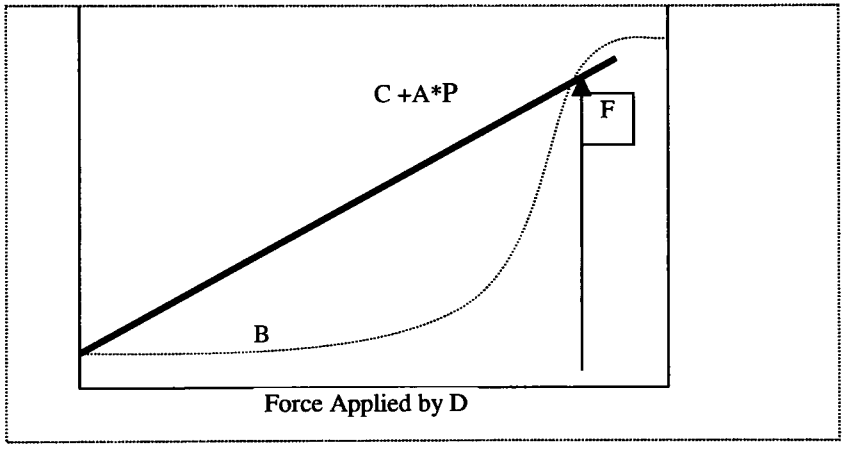

[48] This measure slightly varies from the one employed by Donahue & Wolfers, *supra* note 3, at n.57, because they deduct 1 from their figures on the net gain in lives from each execution to account for the loss of the murderer's life. Without attempting to defend one of these choices as morally superior to another, the use of "innocent lives saved" seems more in keeping with common practice. Moreover, as discussed *infra*, the *marginal* loss to an individual convict by increasing his penalty to death is not the same as the loss experienced by a murder victim.

[49] George J. Stigler, *The Optimum Enforcement of Laws*, 78 J. POL. ECON. 526-527 (1970).

[50] *See* SHAVELL, *supra* note 30, at 1232.

According to this view of "marginal deterrence," the scaling of sanctions in this way has an efficiency justification, because even if B increases with the force applied, it may increase more slowly than C + A*P, and those not totally deterred may still have an incentive to do less harm. In Figure 1, this usual model has been sketched for a hypothetical crime. The criminal uses force up to the level after which (1) is no longer satisfied and continued force would be unprofitable. This is sometime before point F, where force becomes lethal. If, it is usually argued, the curve for sanctions was flat and all crimes of this type had similar punishments, most particularly if one followed the traditional law and punished all felonies with capital punishment, then there would be no "marginal deterrence," causing criminals to desist at some pre-lethal level of violent force (although if C rose with force, there would be some of this).

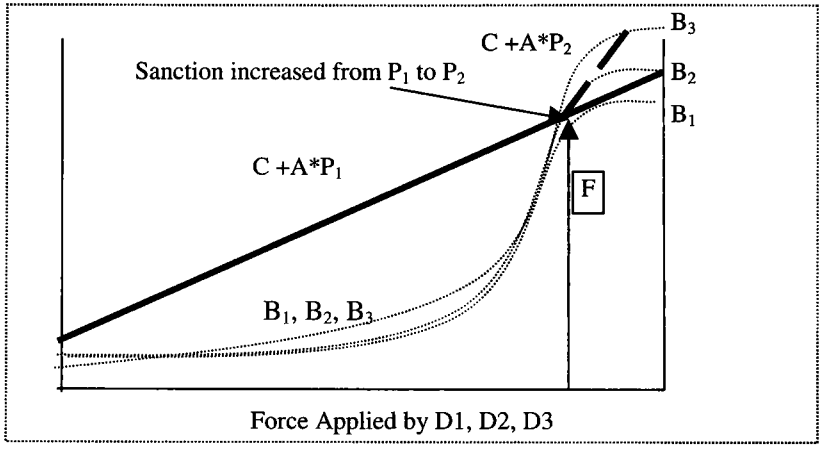
*Figure 2 – The Limits to Stigler Deterrence For Murder*




The model illustrated in Figure 1 is useful for explaining certain features of the criminal justice system, but it seems less useful for many types of murder, where the benefit that accrues to the killer is wholly derived from his application of maximum force to the victim. This creates a step function and a solution illustrated in Figure 2. There is no "Stigler deterrence" with this assumption about the benefit curve to the perpetrator, where we can assume that optimal force is above point F, the level of lethal force. In addition, even as to murders that take place within the context of other felonies, there seems to be a behavioral error with regard to the Stigler model because the goods produced by additional criminal activity are generally distinct from those generated by earlier activity. Consider, for instance, a rapist who applies a certain level of violence to perform the rape and who thereby gets whatever satisfaction might be obtained from coerced sex. The criminal is then at the choice point of whether to kill his victim.

Although there are, of course, psychosexual sadists whose satisfaction is directly proportionate to the harm inflicted, these are not the norm even among sex criminals.[51] (For such persons, there is presumably Stigler deterrence). The usual rapist, when considering whether to kill, is considering a quite separate problem from continuing the sexual assault, namely whether he needs to silence the victim in order to avoid apprehension. The benefits and costs of force *at this decision juncture* are more likely to resemble Figure 2 than Figure 1.

*Figure 3 – Altering Step Functions For Deterrence*




One does not have to abandon an individual rational choice perspective to take this view. A potential murderer will face a series of varying incentives and risks for each killing opportunity. The question will be whether these ever reach the level of profitability (or for multiple murderers, how often they do). By altering the level of P, we deter undesirable acts by making it less likely that (1) is satisfied; how much less likely is reflected in a reduction in the murder rate. This can be illustrated in Figure 3, where three different defendants face a choice, and where an increased sanction for murder, going from $P_1$ to $P_2$, applied at F, deters Defendant 2. Alternatively, the three different benefit curves, $B_1$, $B_2$ and $B_3$, might characterize three different sets of circumstances for a defendant repeatedly presented with a choice of whether to kill. Defendant 1 is deterred by the ordinary regime of sanctions. Defendant 3 gains sufficient

[51] *See* Christina E. Wells & Erin Elliott Motley, *Reinforcing the Myth of the Crazed Rapist: A Feminist Critique of Recent Rape Legislation*, 81 B. U. L. REV. 127, 157 (2001) (reviewing evidence of the non-psychopathic nature of rapists).

benefits from applying lethal force so that even the additional increment in sanctions applied to murder does not deter him.

All three of these defendants are very bad people who are restrained from killing, if at all, by the legal system we construct with them in mind. We know there are people like Defendant 1—people who would gain a benefit from killing another person but who are deterred by what we might call a standard punitive regime consisting of a lengthy prison term. We know there are people like Defendant 3, because murders keep occurring in every jurisdiction and some people are neither deterrable by a standard punitive regime, nor by an increased sanction for murder, $P_2$ (not to hide the ball here, let's call this increased sanction "capital punishment"). Another way to rephrase the basic question in the death penalty debate is whether there is anybody like Defendant 2, or alternatively, whether a potential killer is ever describable as Defendant 2.

The number of people like Defendant 2 who commit the undesirable act of murder when $P = P_1$, but not when $P = P_2$, measured at the same violent activity level, F, are the individuals I will consider as *marginally deterred* by the increase in sanction. In some cases these individuals will have been committing other crimes and their unwillingness to kill will appear as a "limit" on the intensity of their criminality. In other cases, probably most cases, "marginal deterrence," as I am using the term, will involve people who commit no crime at all (at least that day). By contrast, the *absolute deterrence* effect when $P = P_2$ encompasses both Defendant 1 and Defendant 2.[52] Finally, the number of *undeterrable* persons is represented by those like Defendant 3. To the extent a choice to kill is represented by the middle curve, $B_2$, outlined on Figure 3, the outcome of that choice will be determined by whether or not capital punishment is the expected sanction. The number of choices characterized by the bottom curve is largely unknown to us and must be inferred—it is the benefit we get from penalizing murder at all. The expected marginal reduction in the murder rate is dependent on the relative ratio of the number of choices best characterized by the middle curve in Figure 3 and the top curve in Figure 3. If nobody—or more strictly, no decision—is characterized by the middle curve, then there will be no reduction in the murder rate. That the set of defendants like Defendant 2 is an empty set is an implicit assumption of people assuming no marginal deterrence effect for capital punishment.

If deterrence is admitted at all with regard to criminal sanctions, it is difficult to deny theoretically the plausibility of a death penalty sanction deterring serious murders, at least premeditated deaths or felony murders. In order to do so, it would seem that one must demonstrate that a different type of decision process is occurring for murders—as opposed to other premeditated acts, and for felonies that resulted in death—as opposed to

---

52 For a similar usage, *see* Donald S. Nagin, *Criminal Deterrence Research at the Outset of the Twenty-First Century*, 23 CRIME & JUSTICE 1, 3-4 (1998).

those felonies that, adventitiously, did not result in death. On the face of it, these distinctions seem unlikely to be generally true. In particular, although deliberation is not the strong suit of violent felons, the choice to murder will usually involve a serious contemplation of risks and benefits, scaling at least with the amount of time spent in determining how one can best go about it.[53]

To be sure, not everyone in criminology or law would be willing to stipulate to the existence of general deterrence, whether for crime generally, or for murder particularly. While admitting a certain level of means-end rationality in criminals, some argue: "It would be a mistake, however, to infer from this that they are aware of and sensitive to even substantial variation or changes in the schedule of threatened punishments. Most often they are not."[54] Again, because we are concerned about only a select subset of the population, the fact that this subset might be peculiarly present-oriented, impulsive and driven by emotion cannot be ignored in the design of the penal sanction. This hypothetical subpopulation's insensitivity is presumably the behavioral theory underpinning the argument for no deterrent effect of capital punishment. On the other hand, there is a completely contrary claim made by other criminologists, who "hold that the threat of punishment would be most salient for, and thus have the greatest impact on, individuals most prone to crime."[55]

Since these clashing views of behavioral distinctiveness both have surface plausibility, but opposite predictions, deciding between them is best done empirically. Based on limited evidence, it appears to be the case that "sanction threats inhibit the criminal activity of those most at risk of

---

53 Sunstein & Vermuele, *supra* note 4, at 714, hypothesize cognitive variation among "boundedly rational" criminals such that some of them will overestimate the likelihood of execution or will base their behavior on the badness of an uncertain outcome (death) rather than its likelihood. In keeping with the goal here to focus initially on developing a more standard rational choice model, I leave these reasonable suppositions to later research. One point to be made in passing is that to the extent there is deviation from rational choice, it is at least as likely to accentuate deterrence as to blunt it, so that merely pointing to the incompleteness of rational choice theory is inadequate to refute deterrence, and indeed makes it more likely that *some fraction* of "irrational" criminals will be deterred by their exaggerated estimates of the possibility of a capital sentence.

54 Bradley R. E. Wright, et al., *Does the Perceived Risk of Punishment Deter Criminally Prone Individuals? Rational Choice, Self-Control, and Crime*, 41 J. RES. IN CRIME & DELINQUENCY 180, 186 (2004) (quoting as typical NEAL SHOVER, GREAT PRETENDERS: PURSUITS AND CAREERS OF PERSISTENT THIEVES 162 (1996)). *See also* Robinson & Darley, *supra* note 35, at 173, arguing that the preconditions necessary to influence behavior ex ante are rarely met, at least for specific criminal law formulations.

55 *Id.* at 205. One reason for possibly greater level of deterrence is that criminals may act in a more thoroughly egoistic and "cold" manner in pursuit of self-interest. *See, e.g.*, Linda Mealey, *The Sociobiology of Sociopathy: An Integrated Evolutionary Approach*, 18 BEH.& BRAIN SCI., 523, 530-542 (1995) (presenting a rationale for the existence of a behavioral type that combines close attentiveness to personal interest with short-term thinking).

offending" despite their greater impulsivity.[56] Rather than take a position on this particular controversy, however, the model here relies only on the commonsense notion that criminals are deterrable, not necessarily that as a group they are *particularly* deterrable.[57] This position is a distillation of extensive empirical evidence for the basic proposition that the presence of a cost, attached to a crime, causes fewer of such crimes to be committed.[58] Generally, the more reasonable participants in the death penalty debate admit "a" penalty upon murder acts as a deterrent to its commission, so that if the choice is between the death penalty and *no penalty at all*, the death penalty deters.[59]

Even if one is quite skeptical of the importance of criminal deterrence overall, capital punishment probably satisfies the behavioral conditions for potential effectiveness. Unlike most other crimes, the odds a murderer will be caught are usually better than even, so that data showing a psychological tendency to disregard low probability threats of punishment does not apply.[60] In addition, in contrast with more abstruse rules of conduct, it is widely understood that the police and courts will do *something* rather aversive to you if you engage in murder, with the death penalty a salient possibility in certain jurisdictions. With these barriers to deterrence surmounted, remaining doubts center around the more basic rational-choice question of whether there could be someone who, if he knew something was a crime, thought he might get punished, and had a rough estimate of the costs of this punishment, would perceive the benefits of killing as not worth this perceived risk.[61] That is, in terms of Figure 3, deterrence skeptics admit the existence of Defendant 1, but doubt the existence of Defendant 2.

---

56 Wright et al., *supra* note 54, at 206 (relating the results of a long-term longitudinal study relating perceived risk and severity of criminal sanctions to criminal conduct).

57 Indeed, as I will discuss shortly, there is good reason to believe *some* criminals are less deterrable by prison sentences than other criminals are (or than the general population would be), a result with importance for the question of the value of capital punishment as a marginal deterrent for these offenders.

58 For a review of this evidence, including time-series, perceptual and ecological studies, see Nagin, *supra* note 52, at 3 and *passim* (concluding that "the evidence for a substantial deterrent is much firmer than it was fifteen years ago" and "the collective actions of the criminal justice system exert a very substantial deterrent effect").

59 Although Robinson and Darley specifically eschew any discussion of death penalty deterrence, *supra*, note 35, at 200, this would seem to approximate their position, since they temper their skepticism of deterrence by noting at the outset that "[h]aving a criminal justice system that imposes sanctions no doubt does deter criminal conduct." *Id.* at 173.

60 *See id.* at 183-184 (relying on laboratory data to suggest that a 50 percent chance of punishment is effective in reducing behavior, while a 10 per cent chance may often be disregarded entirely).

61 *See, e.g., id.* at 196, proposing that if rules are known and punishment perceived as possible, there will still be no deterrent effect if the potential offender "does not see the overall costs as outweighing the overall benefits . . . , or perceives an overall net cost but is unable or unwilling to bring this information to bear on his conduct choices"). The discussion here centers on the first aspect of Robinson and Darley's critique, which remains within the domain of rational choice. The second

Consequently, a claim the death penalty does not "deter" may be taken to mean there is a lack of proof it "deters better than" some alternative lesser sanction. Likewise, for those who claim that the death penalty *does* provide deterrence for crime—they are not making (what ought to be) the trivial claim that potential criminals may be forestalled by fear of death.[62] Both sides instead are speaking of the marginal deterrence supposed by one side to occur upon the "increase" of the penalty for certain kinds of murder from life imprisonment (which for simplicity's sake I will assume to be without the possibility of parole) to execution, as this is applied to the decision of the potential perpetrator.

Slightly more formally, and put in behavioral rather than aggregate terms, the baseline of discussion is the hazard that a person, D, will commit a murder, of Q, all else equal, given that the cost to the offender, upon arrest and conviction for murder, is a function of the imposition of life imprisonment, $P_1$. This is based on the chance, given that $P = P_1$, circumstances will arise such that (1) is satisfied, where his optimal response is lethal force. Translated to Figure 3, this includes the aggregate probability that D's benefit curve will be above the corresponding cost curve, the chance that either $B_3$ or $B_2$ accurately describes D's choice. This gives us the conditional probability:

$$(3) \quad H_1 = \text{prob}(Q)_D | f(P_1)$$

Here, murder is defined as the presence or absence of an event in some time period over which we observe the behavior of D, a potential murderer. The corresponding hazard, where total cost is a function of the imposition of an execution punishment, is represented by the chance that D's benefit curve is $B_3$ and that he is undeterrable. Cost is given by $f(P_2)$, and combined with the chance of being above the cost curve, gives the conditional probability:

---

critique, that criminals behave irrationally, references larger questions, but two observations about it should be made. As already discussed *supra* notes 37-41 and the accompanying text, even highly abnormal offenders act rationally to avoid arrest, and the level of this detection avoidance behavior is likely to be conditioned by the intensity of the associated punishment, which will inevitably result in foregone opportunities for crime. Moreover, although some murders occur over a very short time span (even if the law considers them "premeditated"), many do not, leaving time for some amount of deliberation over means, and in almost all cases, over whether the action is advisable.

62 *See* Allan D. Johnson, Note, *The Illusory Death Penalty: Why America's Death Penalty Process Fails to Support the Economic Theories of Criminal Sanctions and Deterrence*, 52 HASTINGS L. J. 1101, 1112 (2001) (admitting the basic logic of deterrence but arguing capital punishment does not provide any additional effective deterrence).

(4) $H_2 = \text{prob}(Q)_D | f(P_2)$

There seems to be some consensus that, although $P_1$ and $P_2$ are denominated quite differently, $f(P_2) > f(P_1)$ such that a change in $f(P)$ over the range from $P_1$ to $P_2$ is positive, $\Delta f(P) > 0$. This view is reflected by the greater slope in the modified line at point F in Figure 3. For an individual, the question of whether the death penalty deters is whether $H_2$ is less than $H_1$, or if one prefers, whether:

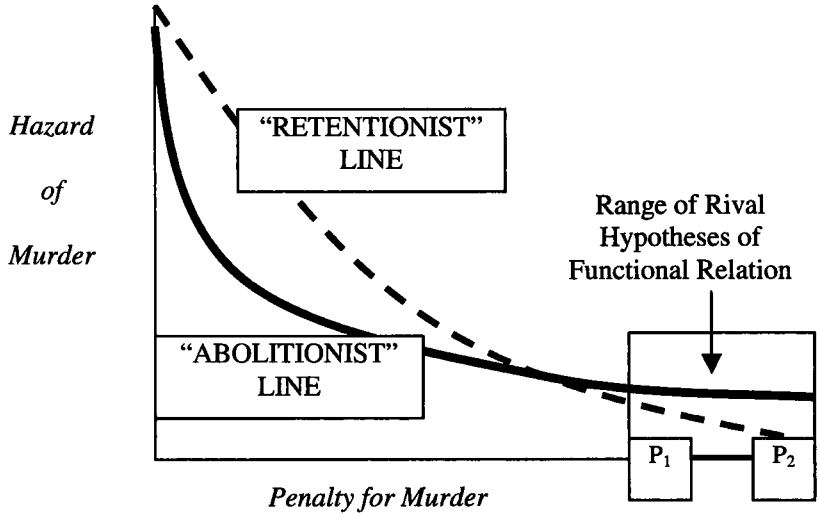
(5)(a) $\Delta H/\Delta f(P) < 0 | P_1 \leq P \leq P_2$ Deterrence

(5)(b) $\Delta H/\Delta f(P) \approx 0 | P_1 \leq P \leq P_2$ No Deterrence

Collectively, for the death penalty to show substantial deterrence there must be enough killing decisions where 5(a) is true, where capital punishment is the difference between life and death (for the potential victim), or to restate the graphical formulation, where curve $B_2$ represents the most accurate description of the costs and benefits of the actor. Retentionists maintain that (5)(a) is satisfied in the jurisdiction they are analyzing, while abolitionists maintain that (5)(b) is the correct form[63] or, perhaps more often, that the retentionists have failed to show change in H significantly different from zero (which of course, is not quite the same thing as empirical validation of (5)(b)).

[63] At one time, and still occasionally, abolitionists promoted the view that, in fact, capital punishment actually caused more people to kill, usually explaining this counter-intuitive result by arguing that the death penalty "brutalized" people, making them more comfortable with killing. *See* Shepherd, *supra* note 3, at 240 (claiming to find evidence of such an effect). Brutalization theorists did not usually deny that, on a direct trade-off basis, execution was perceived as worse than imprisonment. Rather, they supposed that the desensitized, lawless population would react less to *all* kinds of legal and non-legal (conscience-based) restraints on behavior, and thus have its entire cost curve for murder lowered. In economic terms, this discounting of murder's costs would presumably act as a subsidy, raising the probability of crime. The theory of how this can apply to decision making is not well explicated, however, *see supra* note 14, and ignored here in favor of a more thorough discussion of the main lines of the present debate, characterized here as between those who advocate 5(a), and those who advocate 5(b), as the correct assumption for legal policy-making.

*Figure 4 – Two Theories of Sanction Sensitivity*




Represented in the above Figure 4, and ignoring as irrelevant any variant hypotheses about deterrence prior to $P_1$, the debate is over the slope of the line from $P_1$ to $P_2$. If an individual decision maker is sensitive to the difference between $P_1$ and $P_2$, then there will be fewer circumstances in which the benefits of lethal force exceed its intrinsic and socially imposed costs, reducing the hazard that someone will kill. If the number of lethal force decisions is held constant, then this reduced individual hazard will translate into a reduced social hazard, i.e., there will be a lower murder rate per capita.

## II. TRADING LIFE FOR DEATH

If the problem of the death penalty has now been restated as a debate about the slope of the line in an implicit model of sanction, this is hardly surprising, but to what extent has it advanced our understanding, given that we seem to be left once again only with rival intuitions? One way to assess the relative plausibility of these intuitions is to look more closely at the actual difference in cost between life and death, which the model leaves unnecessarily and damagingly vague. As mentioned above, the difference between a life sentence and a death sentence is not the same as between a one-year sentence and a five-year sentence, or even between a five-year sentence and a life sentence. Do we think that with regard to the imposed cost, $f(P_2) > f(P_1)$, the cost imposed by execution is *much* greater than the cost imposed by life imprisonment? Or are we to suppose that it is only

*slightly* greater? Any deterrent effect, $\Delta H$, will scale with increasing differential between $f(P_1)$ and $f(P_2)$, which places demands on specifying them more precisely.

(6) $(\Delta H' > \Delta H) \rightarrow (\Delta f(P)' > \Delta f(P))$

At the limit, where there was only infinitesimal difference between life in prison and execution, 5(b) would almost certainly be correct, because a miniscule change in the x-axis would be matched by a minuscule change in the y-axis, *regardless of the true slope of the line.* By contrast, if there is a great difference in perceived utility loss between life in prison and death, and one takes seriously the idea of deterrence generally, then the absence of some detectable slope—and a corresponding deterrent effect—becomes harder to accept. Assuming deterrence holds in principle, this implies a monotonically negative function relates sanctions to the likelihood of committing an act. The abolitionist position seems to require not only that the slope of the sanction function has become not just fairly flat, but that it should be very close to zero over an extended portion of the range of the function.

### A. *The Intensity of Revealed Preference For Life By Those Sentenced To Death*

Those who are convicted of capital murder provide an approximation of the set of very bad people for whom we design the law punishing this act. Of course, there is a selection bias here, because capital convicts consist only of the subset of very bad people who did, after all, decide to kill, who then proceeded to get caught, and who had the misfortune to be in a jurisdiction and before a jury that handed them a death sentence. Still, it is inarguable they are *among* those we are concerned to regulate, and their opinion about how much they value their own skin can tell us something about whether execution is a meaningful addition to the social tax we impose on killing.

In general, those convicted of capital murder have the option of choosing to abandon their appeals and accept execution—those who do are referred to as "volunteers" and constitute a significant fraction of the people executed in those states where very few people are executed, but a much lower fraction where executions are routine and the process does not require the acquiescence of the defendant.[64] The most obvious evidence the death penalty is perceived as worse by the relevant population is provided

---

[64] *See* John H. Blume, *Killing the Willing: "Volunteers," Suicide, and Competency*, 103 MICH L. REV. 939, 959 (2005).

by the fact that the vast majority of death row inmates *do not* become "volunteers."[65] According to Blume's recent tabulation, there were 106 successful "volunteers" from 1977 to 2003.[66] The relevant comparative population is *not* the number of people involuntarily executed, which is only seven times as large (786).[67] Rather, it more closely approximates those who are continuing to "fight" their death sentence, which surely constitutes a fair proportion of those people who have ever been sent to death row, 7,061 between 1977 and 2003.[68]

Some of these individuals remain in the process of appealing their *convictions*, so their relevant choice was freedom versus death, and they did not reach the point of contesting directly between life in prison and a death sentence; we do not know if they would be "volunteers." So, 106/7061 would be an overestimate, but not a large one. Blume's data show a maximum of twelve volunteers per year (in 1999)[69]—this in a year when 272 death sentences were handed down.[70] Taking this as a conservative estimate of the relative preference for life over death, 4.4% of the relevant population will "volunteer," but in excess of 95% of death row inmates would prefer to stay alive.

Gary Gilmore revealed an understandable preference when he told the Supreme Court "he did not 'care to languish in prison for another day.'"[71] If all potential murderers had the preference set of a Gary Gilmore, the general deterrence rationale for capital punishment would simply collapse, because its core assumption of an incremental disincentive would be violated. This would, incidentally, create at least a rhetorical problem for the abolitionist cause, as arguments used against the death penalty frequently portray it as being more severe,[72] a position contradicted by its

---

65 This point is made by Judge Posner, in discussing the "evidence to support the common-sense proposition that there is indeed an incremental" deterrent effect of capital punishment. Richard A. Posner, *Foreword: A Political Court*, 119 HARV. L. REV. 31, 64 (2005) ("When was the last time a death-row prisoner declined to have his death sentence commuted?").

66 *See* Blume, *supra* note 64, at 959.

67 *See id.* at 961.

68 *Capital Punishment, 2004*, *supra* note 9 at 1.

69 Blume, *supra* note 64, at 959.

70 Tracy L. Snell, *Capital Punishment, 1999*, STATISTICS BULLETIN (Bureau of Justice, Washington, D.C.), Nov. 2000, at 1, *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/cp99.pdf.

71 Gilmore v. Utah, 429 U.S. 1012, 1015 n.4 (1976) (rejecting the appeal of Gilmore's mother to stop his execution by firing squad). The law on voluntary relinquishment supposedly prevents such choices being made due to the "duress" caused by conditions of imprisonment, which is understandable as a legal rule to avoid incentives for recreating medieval prisons. Nonetheless, if prisoners were treated like Napoleon after Waterloo, and given their own Mediterranean island to rule as emperor, they would not prefer death.

72 As discussed above, this is not logically inconsistent with the rejection of substantial deterrence. The abolitionist point is that, under a non-deterrence hypothesis, this increased imposed cost is especially egregious because it brings about no corresponding social benefit. Although the convict

being chosen voluntarily. But Gilmore's set of preferences, although not *per se* insane or inconsistent,[73] appears to be quite unusual.

This is by no means an obvious result. Despite the overwhelming confinement emphasis of our penal system, it is not dictated by necessity, much less by the unanimous assent of convicts. The Romans, for instance, thought of life in prison as worse than the death penalty,[74] and few pre-modern civilizations even had prisons. Where there were prisons, the abysmal conditions of confinement (usually leading to an unpleasant slow death through disease, malnutrition, overwork and exposure) and the omnipresent shame and indignity could easily cause expectations of the future utility derivable from any future period to dip below zero. Such negative prospects make rational any action that will avoid these costs, including suicide.[75] Even under current prison conditions, a "life sentence" adds considerable extra risks of mortality in comparison with living on "the outside,"[76] whereas a "death sentence" de facto consists of a long stretch of imprisonment combined with an extra mortality risk created by an unpredictable legal process that might choose to actually execute in some particular year. In other words, the penalties are not nearly as distinct as one might suppose—yet capital convicts normally perceive them as sharply distinguishable.

In order to determine whether to choose life imprisonment over the death penalty, a convict will assess the expected utility he will receive in the prison environment over the period of his natural life. For sake of simplicity, suppose the decision is just whether or not to pursue a legal

---

pays the cost, murderers are supposedly sufficiently insensitive to cost such that they will not be deterred.

[73] *See, e.g.*, Harper v. Parker, 177 F.3d 567, 570 (6th Cir. 1999) (rejecting attempt of Kentucky Department of Public Advocacy to intervene and stop execution when testimony showed the convict was "competent and that his determination to have counsel discharged and not to proceed with filing a habeas petition is based on his desire not to have to live in prison for the rest of his natural life").

[74] *See* Russ VerSteeg, *Law and Justice in Caesar's Gallic Wars*, 33 HOFSTRA L. REV. 571, 575 (2004) (noting that Julius Caesar, when acting as magistrate, expressed this inverse view of the relative harshness).

[75] A distinction that may be noted between Roman and common law practice is that in the former, actual suicide was frequent, if conviction or a lighter sentence (such as exile) was unavoidable. Influenced by Christian tradition, individuals in the common law were ingrained with prohibitions against suicide (prohibitions enforced to some extent by post-mortem humiliation of the body and family of the suicide). Therefore, even if an individual preferred death, there were certain barriers to implementing this desire.

[76] *See* Katz, Levitt, & Shustorovich, *supra* note 3, at 323. Although Katz, Levitt and Shustorovich argue that death rates in prison provide a stronger deterrent than execution, they do not deny that execution might have some effect; they claim to have been unable to come to an empirical conclusion about it. As Sunstein & Vermeule imply, this result is behaviorally odd, because the capacity of prison death to deter crime would activate the very same mechanism on which capital punishment is supposed to work, and the additional mortality risk of execution would only enhance any such effect. *See* Cass R. Sunstein & Adrian Vermeule, *Deterring Murder: A Reply*, 58 STAN. L. REV. 847, 847 (2005).

strategy to delay execution and allow the prisoner to survive into the next time period, x+1. Using conventional demographic form, an individual has an exogenous probability $p_x$ of surviving from his current age x into the next time period x+1. During that period he will experience (or rather, expect) a certain amount of utility, $U_{x+1}$. Apart from any other discount rate applied to this future utility stream, the value expected from this period, E, must be discounted for the possibility that the potential recipient will die of some other cause than execution.[77]

$$(7) \quad E = p_x \bullet U_{x+1}$$

Execution prior to x+1 reduces $p_x = 0$, and therefore (E| execution) = 0. An individual would therefore prefer execution only when (E| ~ execution) < 0, and since $p_x$ is a probability that can vary only between 0 and 1, this can occur only when $U_{x+1} < 0$. In other words, people choose to die when life, even if you can get it, is not worth living. This may be rather obvious, but is worth establishing at the beginning of developing a model of punishment cost.

The total amount of expected utility in the person's remaining lifespan is determined by the potential to survive to some age, $l_x$ (which is the product of the joint probabilities of surviving all preceding periods, $p_0 \bullet p_1 \ldots \bullet p_{x-1}$). This chance of living to age x is multiplied by the utility one gains at that age if in fact you live long enough to enjoy it, $U_x$. The utility times the survivorship, $l_x \bullet U_x$, is then totaled by summing across all future time periods (x varies from current age to extreme old age) to give an expected value of the remaining lifespan, V.[78]

---

[77] During 2004, there were 19 deaths from natural causes among death row inmates, indicating a background yearly mortality risk of about one-third of that stemming from execution. *See Capital Punishment, 2004*, *supra* note 9, at tbl. 4.

[78] The above formula adapts the reproductive value equation generally used in biology, although it is also related to various models of human capital. A biological equation would divide the summation in Eq. 8, by the probability of surviving to the individual's current age, to account for the fact that he has, actually, beaten the odds so far and has greater expected value in the future years than was the case at his birth. Second, utility would be directly replaced by biological fertility ($m_x$) as the measure of the utility productivity of a future time period. Since it is not necessary or possible here to discuss the extent to which utility tracks this objective measure in human beings, the conventions of economics are usually adhered to instead. I would note that punishment by confinement, like punishment by death, successfully suppresses reproduction by preventing sexual contact with the opposite sex, and more generally, activity that would allow accumulation of resources. I need not defend that this is *why*, at a fundamental level, prison is experienced as unpleasant. Such motivation is *usually* unconscious; cases such as Gerber v. Hickman, where the prisoner, a "lifer," sought to assert a constitutional right to impregnate his wife through artificial insemination, are rare. The analogy nevertheless goes slightly beyond the formal one, and suggests that the en banc opinion in *Gerber* was on the right track when it held procreation to be inconsistent with incarceration, given "the nature and goals of the correctional system, including isolating prisoners, deterring crime, punishing offenders, and providing

$$(8) \quad V = \sum (l_x \bullet U_x)$$

The current form of (8) is useful insofar as it shows there is a fairly specific meaning to lifetime utility and what one sacrifices at death. For the sake of simplicity, we can ignore the changes in $l_x$ and $U_x$ that may occur at different ages, as well as the structure of the equation, and simply note that V is a positive function of l and U.

$$(9) \quad V = f(l, U) \;\; \Delta V/\Delta l > 0 \;\; \Delta V/\Delta U > 0$$

Presumably the problem with incarceration is that the utility of a period spent in prison is less than the utility derivable from a period unincarcerated, or else prison would not be as undesirable as it is.[79] In addition, a lowered life expectancy within prison may make it less attractive. If V is the lifetime expected utility as a free man, we may set V′ equal to the utility derived from an imprisoned natural life, based on the altered prison variables, l′ and U′. We could also assess V given execution, but we know it already as a generalization from (7): all future periods produce zero utility if you are dead, so V upon application of capital punishment falls to zero. With these definitions we are now ready to specify how much is taken away by punishment levels $f(P_1)$ and $f(P_2)$, and thereby find the marginal increase in cost $\Delta f(P)$.

$$(10)(a) \quad f(P_1) = V - V'$$

$$(10)(b) \quad f(P_2) = V$$

$$(10)(c) \quad \Delta f(P) = V'$$

We arrive at the conclusion that the unique deterrent effect of capital punishment *depends entirely on the length and quality of a life sentence.* The value of one's life as a free person is irrelevant at the margin—although highly relevant to the deterrent effect of life imprisonment.[80] To

---

rehabilitation." Gerber v. Hickman, 291 F.3d 617, 622 (9th Cir. 2002) (reversing panel opinion in 264 F.3d 882, 890 (2001), which would have found the Constitution guaranteed a right to ship sperm out of the penitentiary for reproductive purposes).

[79] It is possible that some people might actually prefer prison to the world outside, although I would suspect them to be a smaller group than "volunteers," I discuss this problem briefly, below, and suggest the amount of utility such people are likely to be deriving in prison is quite low. It not being so much that they like prison, but that their unincarcerated life would be yet more unpleasant.

[80] Thus, someone who has a relatively shortened expectation of life, and/or low expectations for the possibility of utility generation at future ages, will be less deterred by the threat of jail, and will

the extent existence in prison becomes progressively more lethal or odious, the additional deterrence provided by a threat of execution goes away. This creates a behavioral critique of the study of Katz, Levitt and Shustorovich (KLS) on prison conditions.[81] According to the theory developed here, bad prison conditions should relate negatively to the marginal deterrent effect of execution, so the two primary regressors of interest used by KLS are causally related in a way they do not discuss. Indirectly, however, their study does suggest such a relation, since they report a significant deterrent effect of capital punishment on murder for later decades (post-1971) of more than 1 per 100,000 in population.[82] Earlier decades, however, show a much smaller (if any) effect on deterrence, in line with the theoretical structure of equation 10(c), since prison conditions as measured by mortality were much worse during this early period.[83]

Instead of comparing different time periods, it is more crucial to identify those persons within a particular time on which capital punishment might have most effect, and for this we should look to those who still have something significant to lose, despite being put in prison. That there are such people is again suggested indirectly by volunteers, whose presence signals a certain level of variation in V′. We can infer that, if there are persons within the death-sentenced population who are relatively less able to bear incarceration, *there should be another group, who are relatively more able to bear incarceration.* For this latter group, V′ remains significantly different from zero, and they therefore have the most to fear from death; consequently they are predicted to be most deterred by capital punishment. Naturally, there are general trends that affect the difference between life in prison and execution, the general nature of prison conditions being the most obvious one—the worse these are, the better life in prison serves as a deterrent, and the worse execution will serve as a *marginal* deterrent.

---

commit more crimes. For instance, Jack Ruby, who was dying, shot Lee Harvey Oswald, and the fact that Ruby was dying no doubt lowered the expected loss from the social sanction, raising the benefit of the killing over its costs. The commission of interpersonal crime, especially homicide, will have localized effects that reduce the expected survivorship and economic security of surrounding individuals, thereby reducing their *own* incentives against criminal conduct. The existence of hypothetical environments containing such feedbacks may warrant sociological investigation. If the term "brutalization" were not co-opted for a hypothesized effect of capital punishment, it might appropriately designate what could occur—as pure consequence of rationality—to people in these violent environments.

81 *See* Katz, Levitt, & Shustorovich, *supra* note 3.

82 *See* Katz, Levitt, & Shustorovich, *supra* note 3, at 335 (cautioning this effect, -1.15 murders/execution per 100,000 population, is highly dependent on the econometric model employed and highly variable in the sample).

83 *See id.* at 324 (general secular trend of improving conditions) and at 335 (pre-1971 coefficient is estimated at approximately -.1 murder/execution per 100,000 population, ten times lower than the mean of the fluctuating coefficient detected from the modern death penalty era).

The current model, in part because it is designed to focus the death penalty debate more precisely, simply takes this one step further, to consider the *individual variation* in the relative difference between life in prison. The policy implication is that, while it is pointless to threaten a Gary Gilmore with execution, and the usefulness of this threat is debatable as to some other percentage of killers, there should exist a group for whom the death penalty might be a particularly effective deterrent, regardless of the background cost of prison conditions for the average inmate. Moreover, V′ is negatively correlated with V - V′ (unless one assumes that these individuals are just naturally happy, adaptable people who have better lives—higher V—outside the stir as well as in it). Therefore (10)(a) will be lower than normal, as will be the deterrence provided by a life sentence; they are people for whom execution is a particularly serious threat.[84]

Presumably, we would be looking for a group of individuals—let's call them "Toughs"—willing and able to use violence, and who are relatively comfortable in the prison environment. The most well-established marker of such persons is prior experience in prison, accompanied by gang membership.[85] In DeLisi's large scale study, he found such persons to commit most offenses behind bars, essentially continuing the criminal "career" they had established prior to incarceration and aggressively bullying other inmates.[86] From an ex ante perspective, previous time in prison appears likely to "condition" individuals to experience prison as less painful. Indeed, Robinson and Darley argue that this psychological "adaptation" influences post-release decision-making because the experience of prison "remembered after the fact, [ ] has taught these people that prison 'isn't so bad after all' and risking it is not an important consideration in one's thinking in deciding whether to offend."[87]

Some further information about variation in prison experience is provided by a four-year field study conducted in Lorton Prison in the late

---

84 The "desert" model of punishment would track this conclusion, since putting people into what amounts to a "second home" for them is hardly commensurate with the punitive notion of desert.

85 *See* Matt DeLisi, *Criminal Careers Behind Bars*, 21 BEHAV. SCI. LAW, 653, 656-657 (2003).

86 *Id.* at 663. A small group (8% of the sample) of "extreme career criminals" committed 100% of the homicides, 75% of the rapes, 80% of the arsons and 50% of the assaults in his study. *Id.* at 662. The skewed nature of activity within prisons is relevant to the extent the offender finds satisfaction in his continued crimes. It also assumes he to some extent anticipates such satisfactions will be available, prior to making a decision about whether to commit an offense—such as a murder—that threatens to land him in prison.

87 *Supra* note 35, at 191. Robinson and Darley raise this point in support of their argument as to why prison sentences generally, and marginal increases in sentence length in particular, are unlikely to be effective deterrents, especially against recidivist offenders. All else equal, however, anything that decreases "the punitive bite" of long prison sentences raises the potential marginal deterrence of the death penalty, which is distinctive in cost structure.

1980s.[88] Although this research was qualitative in nature, one of its chief conclusions was that the aversive nature of the prison experience varied widely, with many individuals suffering serious mental distress due to the high level of risk of injury and death and the "fear, uncertainty and boredom" involved in incarceration; others, however, perceived the prison as at least a tolerable place where they could "play basketball, get high, dress well, work out, have sex and watch color television . . ."[89] Generally, those having the easiest time are, naturally, the "toughest criminals who committed the worst crimes" because they usually have established contacts for contraband goods and services, obtain supervisory positions in the prison work details, and can effectively terrorize the "timid, short-term first offender[.]"[90] For our purposes here, and assuming this scale to be approximately correct, those experiencing a relatively non-aversive environment would often include murderers in for long sentences, especially those with prior prison experience, although other murderers, for whom their crime is not an outgrowth of a criminal career, may lack prison experience and suffer greater than average costs for at least a certain period.

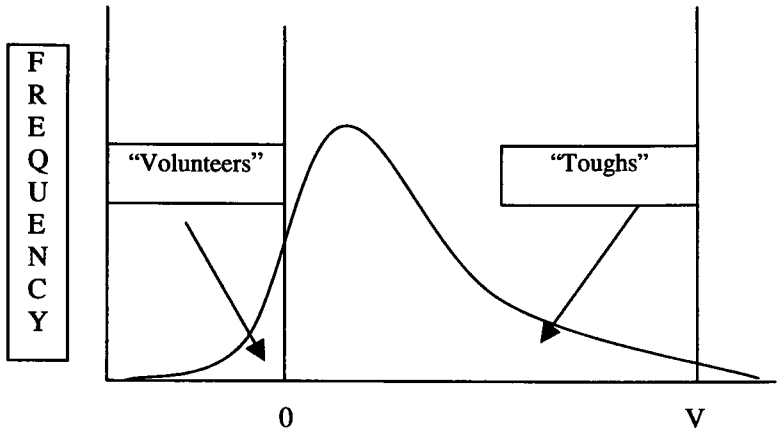
---

[88] *See* Robert Blecker, *Haven or Hell? Inside Lorton's Central Prison: Experiences of Punishment Justified*, 42 STAN. L. REV. 1149, 1154 (1990). Lorton prison was in Virginia, but housed inmates convicted of serious crimes within the District of Columbia, such as robbery, murder, rape, and narcotics offenses. *Id.* The District of Columbia has (and had at the time of the study) no death penalty.

[89] *Id.* at 1216. One revealing inmate comment is that, for some younger prisoners, "life in prison is 'pretty much the same as on the street. Their little honeys can come and see them; they can sneak off and have a little sex here and there. Drugs when they want, get drunk when they want. They can have personal clothing so they can dress similar [sic] to the way they did in the streets. Get up late. Pretty much the same thing as at home. Somebody to look after them." *Id.* at 1172. If this was in any way accurate, it obviously greatly reduces the deterrent effect of the prison, with V' converging toward V.

[90] *Id.* at 1173. As noted by Blecker, who is sympathetic to the retributivist viewpoint, a moral desert model of punishment would reject this pattern of prison suffering. It also provides an additional justification (from either the retributive or deterrence view) for application of the death penalty if (and only if) a death-eligible individual is likely to experience an abnormally easy time serving a life sentence.

*Figure 5 – A Hypothetical Distribution of Prison Quality of Life*




Noting again the selection bias (out of all potential murderers) that comes with doing so, I will make do with relevant observations on two further sub-samples for which there is significant data: (1) death row inmates and (2) murder defendants more generally. Keeping in mind the caveats expressed above about who enters these groups, it seems useful to investigate whether it is possible to identify—as a preliminary to other potential empiricism, perhaps—groups who have particularly high marginal costs for death. These individuals, in their cost structure, are essentially the reverse of death penalty "volunteers." For volunteers, $V' < 0$, so death is preferable, because the cost of life in prison, $f(P_1)$, is equal to $V - V'$, and this is greater than V for them. Yet we can safely assume they represent one extreme end of a distribution in Figure 5, the bulk of which is greater than zero, and whose mean is centered between zero and V. The right tail of the distribution can be taken as V,[91] while the left tail is some arbitrary negative number. The greater the skew toward the right of the distribution

[91] Hypothetically, the tail in Figure 5 extends out past V, to include those who may actually prefer long-term imprisonment to life on "the outside." There is some anecdotal evidence for this among persons whose V is relatively low, but the percentage for whom $V' > V$ is likely small. Individuals of this type, because *both* V and V' are probably low, should not show an especially strong specific aversion to the death penalty because they are insensitive to penalties generally. *See* Blecker, *supra* note 88, at 1177, who reports inmate (and rational calculator) Leo Simms as musing that [sic] "'A person have to have something that you take away from them, . . . but if you don't got nothing, or consider what you have of any value or any importance, then you ain't took nothing from them.'"

in Figure 5, the greater predicted incentive for avoiding conduct that could result in the death penalty.

One sample of death row inmates, already discussed, has been characterized at a sufficient level to perhaps make certain distinctions. The 156 killers given indiscriminate clemency by George Ryan were all given capsule profiles by the Chicago Tribune.[92] Although these narratives are brief, they allow for more precise coding of the activities of the offenders that ultimately led to their capital crimes, coding that can be crosschecked for most of them by reference to state appellate opinions on their cases. Within this sample, there were no "volunteers" who complained about the commutation. There was, however, a blind spree-killer whose preference reversal between life and death is recorded in his attempt to withdraw his unbargained-for guilty plea to capital murder.[93] This capital convict, Ernest Jamison, had a busy June 19, 1995. He began by killing a man in Memphis, stealing his car, and heading north. On his way, Jamison killed a gas station attendant in Missouri, and, when his first car broke down in Illinois, he shot a woman in the head for her Honda. The Honda went into the ditch during a high-speed chase with the local sheriff, and, when approached, Jamison turned the gun on himself. This indicates at least a momentary preference against incarceration in favor of death.[94]

Jamison failed in his suicide attempt, but his gunshot to the head rendered him blind and depressed him. He voluntarily changed his plea to guilty and waived his right to a trial. A subsequent hearing found him eligible for the death penalty. Apparently, the "defendant became sad when he spoke about the future and maintained he would rather die than go to prison because he was afraid of being victimized in prison due to his blindness."[95] There is no doubt that Jamison, at this point, can be located at

---

92 *Ryan Issues Blanket Clemency: Death Row Inmates Receive Life*, CHI. TRIB., Jan. 12, 2003, § 1 at 18, *available at* http://www.chicagotribune.com/media/flash/2003-01/6205019.pdf and http://www.chicagotribune.com/media/flash/2003-01/6205191.pdf.

93 *See* People v. Jamison, 756 N.E.2d 788, 790 (Ill. 2001).

94 *See* Blume, *supra* note 64, at 968 ("there are important similarities between persons who commit suicide and those who volunteer for execution"). Of course, those who do not let the cops take them alive are special cases—they are making, in effect, a penal choice that indicates their expectations of V', rather than the normal suicide, whose action reveals expectations about $V < 0$. Nevertheless, it is not surprising if these were correlated, since, if somebody already has a low expected value of V—little to live for—a lifetime of incarceration is unlikely to improve his outlook on life. Formally speaking, a low V indicative of those at risk for suicide makes people less resilient to the subtraction effected in (10)(a) as punishment and makes it more likely that $V' < 0$, producing a potential volunteer. Blume also appears puzzled by the racial disparity in volunteering, in that, although 42% of death row inmates are black, only 3% of volunteers are. *See id.* at 961-62 & n.120 (reporting that this phenomenon has been unanalyzed in the legal literature). Although there are no doubt many factors involved here, one implication of the present model is that, if greater organization of African-American gangs in prisons buffers the effects of incarceration for members, and thereby reduces its cost, volunteering would become less likely for those who are (or who are, at least, eligible to be) members of such a gang.

95 People v. Jamison, 756 N.E.2d at 793.

the left hand of the hypothetical distribution shown in Figure 5. However, after a healthy dose of antidepressants, Jamison's mood improved; his boredom was relieved by provision of books-on-tape and music, and he "attended Bible study sessions, wrote rap songs, and enjoyed witnessing his faith to his fellow inmates."[96] It is plausible to conclude that Jamison now had something to live for, had subjectively determined $V' > 0$, no longer wished to die and, hence, attempted to withdraw his plea. Jamison's expected utility can now be shifted to the right of the zero line—but, perhaps not very far to the right.

Other inmates can be inferred to be further out to the right on the distribution hypothesized in Figure 5. For instance, consider Victor Ganus, convicted in 1990: "While serving a life term for murder in 1988 at Menard Correctional Center, he stabbed and strangled fellow inmate Lucas Gonzales in a gang dispute." Next, in alphabetical order, on the mercy list is Oasby Gilliam: "Kidnapped and robbed Aileen D'Elia, 79, of Chicago in 1992, then beat her with a tire iron and dumped her body in Downstate Jefferson County."[97] Neither the law that condemned them, nor the blanket executive clemency that saved them, made any particular distinction between Mr. Ganus or Mr. Gilliam, both of whom, I think it is not disparaging to remark, are very bad people. Nor did Ganus or Gilliam abandon their appeals and show a preference for volunteerism, as Jamison initially did.

Nevertheless, as revealed by examining their cases, there is an important difference between these men. Gilliam appears to be a typical economically motivated criminal with no particular brief for killing old ladies, but also no particular compunction against doing so.[98] There is no reason to believe he would find a life sentence particularly easy or hard, and therefore, at a guess, we can place him at the middle of the cost distribution and probable deterrability.[99] Mr. Ganus, on the other hand, seems to be a bit of harder nut—Blecker's "tough criminal" or DeLisi's "extreme career criminal"—and the murder for which he received his death sentence was

---

[96] *Id.* at 798.

[97] *Death Row Inmates Receive Life*, Chi. Trib., Jan. 12, 2003, § 1, at 18, *available at* http://www.chicagotribune.com/media/flash/2003-01/6205019.pdf.

[98] *See* People v. Gilliam, 670 N.E.2d 606, 612 (Ill. 1996) (victim was murdered as a potential witness after Gilliam carjacked her in order to escape from a botched attempt to rob a liquor store). He drove with her (locked in the trunk) for several hours as "he considered what to do with the victim." *Id.* This is notable as a counterexample to any claim that criminals lack the mental capacity or time to consider whether or not to escalate their criminal activity to the level governed by maximum penalties.

[99] Gilliam fled to Mississippi where he holed up with relatives, but appeared "restless and nervous" and refused to specify what was wrong, saying, "If you only knew." *Id.* at 611. After two months, he returned to Chicago by bus and requested that his girlfriend arrange for his surrender to the authorities, to whom he then confessed his crime. *See id. Cf.* FYODOR DOSTOYEVSKY, CRIME AND PUNISHMENT (1866).

"part of his duty as cellhouse security chief for the Latin Kings gang."[100] Ganus was already serving a natural life sentence at the time for a 1985 murder that he committed just forty-one days after getting out of a previous stretch in prison.[101]

There are 152 men (and four women) formerly on Illinois death row in this sample. The women are too small a group to make much of, but, of the 152 men, I coded 15 as committing their principal crimes as part of gang activity, in pursuit of their commercial activities as drug dealers, or both. Gang members and drug dealers, in most American penitentiaries, have the opportunity to continue to pursue the same type of economic activities they did while not incarcerated, thereby providing compensable services and having a greater opportunity for utility generation than would be the case for the average inmate. Because a substantial portion of the gang is imprisoned at any particular time, and its rivalry with economically competitive gangs is positively affected by people willing to act in capacities such as "cellhouse security chief," the imprisoned gang member—even the "lifer"—remains integral to overall economic life of the criminal organization. *Ceteris paribus*, $U'_{GANUS} > U'_{GILLIAM}$, generating $V'_{GANUS} > V'_{GILLIAM}$. The protection afforded by the gang in terms of survival probability may or may not be offset by the greater exposure of the individual as a target from rivals, so we perhaps can ignore that factor until more evidence is gathered on it, and keep the focus on individual differences in retained utility. At least 10% of the sample of capital convicts is therefore shifted right on the distribution by a retained capacity to continue valuable economic activity.

Another subgroup from this sample that *may* be partially immunized by the effect of the prison environment is the subgroup of killers who are also rapists—at least those who, while at liberty, satisfied desires using adult victims.[102] Including some marginal cases who assaulted and killed teenagers, I identify 27 members of this category, or about 20% of the sample. Arguably, because rape is unfortunately prevalent in many correctional institutions, it is possible for these individuals to continue, to some extent, an activity for which they revealed a preference while not incarcerated. It is true that the majority of these men raped women,

---

[100] People v. Ganus, 594 N.E.2d 211, 212 (Ill. 1992) (Ganus's victim supposedly had "raped the wrong [pejorative term for female person] out on the street" and Ganus found it necessary to increase the penalty that had been imposed by the justice system).

[101] People v. Ganus, 706 N.E.2d 875, 878 (Ill. 1998). Expert psychological testimony established that Mr. Ganus "had a tendency to be excessively absorbed with his own needs and insufficiently aware of the needs of others. Defendant is impulsive, shows poor judgment and has difficulty controlling himself." *Id.* at 877.

[102] Those who rape and kill children, anecdotally, have lower survivorship in prison and may be subject to disutility generated by fellow inmates. *See* Blecker, *supra* note 88, at 1172 (noting that "child molestation or crimes against the elderly" may result in a prisoner being "ostracized or physically attacked").

whereas in prison they will have access only to other males as victims. However, particularly if the criminal preference is for coercive sex *per se*, there should be some substitutability of one "good" for another in the prison environment.[103] More speculation is required here, but again, if those who have the motivation and opportunity to engage in highly desired behavior have a higher retained utility upon imposition of a life sentence, they will be more susceptible to the next step in penalization.

### B. *Revealed Preference For Life, By Those Eligible For Death*

Those who are convicted of—or more often, plead guilty to—a murder *eligible* for the death penalty, but who do not actually receive it, are no doubt more representative of the relevant regulated population than are death row inmates. However, offenders in this group receive much less legal treatment (or attention from legal reformers and the bar), in part because their sentences are generally not separately appealable, or they may have waived appeal of their conviction and sentence as part of the exchange for avoiding a possible death penalty. It is therefore more difficult to estimate to what extent they mirror the distributions estimated above, or how they might reveal the variability in immunity to prison costs at the heart of the deterrence question.

Nevertheless, studies of them reinforce the essential point that life is preferred to death. The option of the death penalty gives prosecutors an important amount of leverage in plea negotiations. With the exception of New York, where there are legal restrictions on the use of this threat, it is said to be "the virtually universal day-to-day practice in every other American death-penalty jurisdiction."[104] The actual *amount* of leverage of this threat, however, is specified by the same formal relations as have been sketched above, since the amount a bargaining defendant would be willing to exchange (in terms of possibility of acquittal or legal recourse) is highly dependent on the additional amount of cost imposed upon them by a death sentence.[105] This, in turn, will vary according to the individual's expected quality and length of life in prison. Gary Gilmore, presumably, could

---

103 *See* RICHARD A. POSNER, SEX AND REASON 121 (1992).

104 Joseph L. Hoffmann, Marcy L. Kahn & Steven W. Fisher, *Plea Bargaining in the Shadow of Death*, 69 FORDHAM L. REV. 2313, 2316 (2001). In any event, the death penalty in New York has once again been put in abeyance by the state courts. *See* People v. LaValle, 3 N.Y.3d 88 (NY 2004) (striking down death penalty procedure on state constitutional grounds).

105 It is not wholly dependent because such negotiations take place prior to trial and the "death is different" legal regime may allow certain prosecutorial advantages during the trial that could increase the chance of conviction. *See* James S. Liebman, *The Overproduction of Death*, 100 COLUM. L. REV. 2030, 2097 (2000) (describing the rationale for, as well as sharply criticizing, the practice of "overcharging" murder defendants).

assert, with Patrick Henry, "Give me liberty or give me death!" There would be no prosecutorial leverage over such a defendant.

Over the average defendant, however, the influence of the death penalty is apparently quite substantial. In order to achieve a *plea* of life in prison without the possibility of parole, where the trial outcome is uncertain, it is patent that there must be some penalty that is perceived to be substantially in excess of life imprisonment, such that a rational defendant chooses with certainty a result of $f(P_1)$. Indeed, only those who are death-eligible will accept, without adjudication, costs of $f(P_1)$.[106] To simplify matters somewhat by ignoring the possibility of conviction on a lesser charge than death-eligible murder, and of avoiding the death penalty at the sentencing phase, we get the following decision, which, when true, will result in accepting a plea bargain:

$$(11)\ V' > Prob_{acquittal} * V$$

We can see that, as prison conditions (V') improve from the defendant's subjective viewpoint, the attractiveness of taking a deal increases. The problem with this is that the "Toughs" and most experienced inmates have the greatest incentive to take a deal, although the ex ante optimal penalty (from society's point of view) for them is *more likely* to have been execution than it would have been for other prisoners who have a sufficient fear of incarceration. Obviously, as the chance of avoiding a death sentence falls, a plea becomes more likely. The potential for acquittal (and the right side of (11)) would rarely fall to zero, however, given the vagaries of the justice system, and this suggests that defendants who take a "deal" of a natural life sentence are giving up something of importance. The only reason why they would act this way (and *be professionally counseled* to act this way by their lawyers) is that the fear of even a relatively small chance of a death sentence operates as a prime determinant of their post-arrest litigation behavior.

In a study of Nebraska capital punishment, Baldus and his colleagues analyzed the results of 185 "death-eligible" murder cases.[107] For 96 of these cases, the death sentence was not sought, often because of pretrial plea bargain. Of the remaining 89, seventeen pled guilty at the guilt phase of the trial. This largely eliminated the risk of a death sentence (only two of the seventeen received such a sentence), and, so, might be thought of as an implicit equitable plea bargain, i.e., "throwing one's self on the mercy of the court." Of the seventy-two contested cases, however, 37% resulted in a

---

106 *See* Hoffman, et al., *supra* note 104, at 2350.

107 David C. Baldus et al., *Arbitrariness and Discrimination in the Administration of the Death Penalty: A Legal and Empirical Analysis of the Nebraska Experience (1973- 1999)*, 81 NEB. L. REV. 486, 496 (2002).

death sentence. Presumably, those who are engaged in plea bargaining have some sense of their chance for acquittal or for receiving a death sentence and behave accordingly.

For those who fear a death sentence, a strong incentive is created that shifts the bargaining outcome of the negotiation between the criminal defendant and the state, resulting in a greater length of agreed-upon imprisonment than would be the case were the death penalty not present in the background. The relevance of this phenomenon for the deterrence debate is manifold. Most obviously, it strongly suggests that, controlling for likelihood of acquittal, the sample of those who actually get on death row is biased towards those killers who have the lowest marginal cost of death, and the lowest "quality of life" in prison. Retained utility in prison correlates with the incentive to conclude a plea bargain, which avoids a death sentence, in exchange for a longer period of term imprisonment or for sacrificing the possibility of parole. This is perverse from both a deterrence and desert perspective.

A perhaps more important consequence of considering bargaining is that, quite against the intent of the critics who have generally examined these practices, the death penalty is shown to have a far more pervasive effect among murderers than is commonly assumed. Statistics showing a very small coterie of murderers actually receive the death penalty radically underestimate the total amount of additional cost imposed by the presence of the death penalty on the population of all convicted murderers. The aggregate additional cost capital punishment attaches to murder comprises not merely a small probability of offender death, but also the de facto increase in prison sentences implemented as a prosecutor-favorable shift in bargained-for sentences, which determine the penal outcome in most murder cases.

For murderers generally, therefore, the probability of receiving $P_1$ (life without parole) is a positive function of the probability of receiving $P_2$ (a death sentence). Only if the probability of receiving $P_1$ (rather than some penalty less than life without parole) is *also* irrelevant to the hazard of committing murder would a reduction of the probability of receiving $P_2$ to zero—as required by a death penalty moratorium or abolition—have no net effect on murder rates. So long as the individual is at least plausibly eligible for the death penalty—and this may include all, or nearly all, first-degree murderers[108]—there is an expected increased cost from the death

---

[108] *See* Steven F. Shatz & Nina Rivkind, *The California Death Penalty Scheme: Requiem for Furman*? 72 N.Y.U. L. REV. 1283, 1330 (1997) (85% are death eligible); *see also id.* at 1340 ("the prosecutor can use the death penalty threat against almost any defendant charged with first-degree murder"). In Georgia, Baldus and his colleagues indicate, the comparable percentage is a near-identical 86%. DAVID C. BALDUS, ET AL., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS 268 (1990). The Florida system is more straightforward and simply makes first degree murder a potentially capital crime, with the question of life imprisonment or a death sentence

penalty, whether or not it is actually sought, obtained, upheld, or carried out. Formally, this allows relaxation of the range conditions applied *supra* to the competing functional hypotheses in (5)(a) and (5)(b). These can now be restated in a form that requires more stringent assumptions by proponents of abolition.

$$(12)(a) \quad \Delta H/\Delta f(P) < 0 \quad \text{Deterrence}$$

$$(12)(b) \quad \Delta H/\Delta f(P) \approx 0 \quad \text{No Deterrence}$$

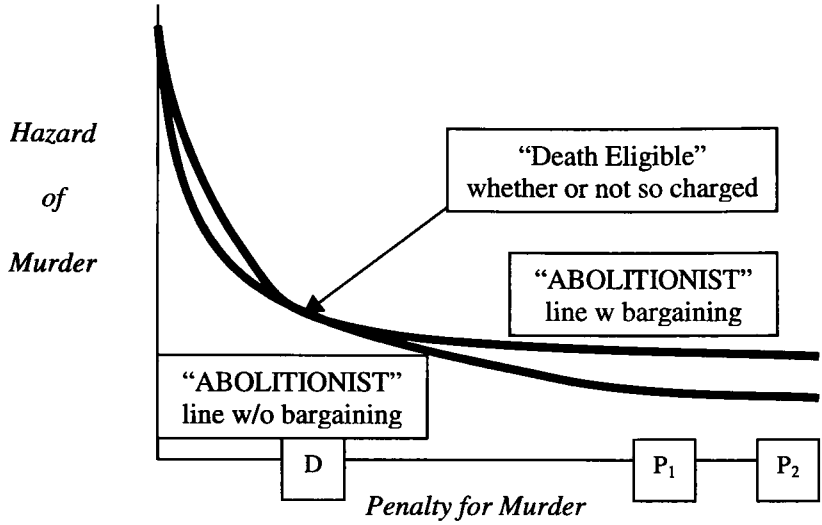
The anti-deterrence argument therefore becomes that much more difficult to sustain, at least on a theoretical level. The change in expected costs for murder wrought by capital punishment, it is now apparent, should be relevant to *most* murderers: there is much greater certainty than is commonly recognized that the death penalty will apply in *some cost-increasing fashion* to any particular person contemplating murder. In addition, the predicted extra years added on to sentences indicates the aggregate cost differential between death penalty and non-death penalty jurisdictions would seem to be higher than that attributable to the mere elevation of mortality risk among a small fraction of the most heinous offenders. Moreover, for there to be "no effect," the non-deterrence thesis now requires that murderers not only be insensitive to the difference between death and life, but also that they do not respond behaviorally to the difference between life with parole and life without, nor to the difference between life and a term of years, nor distinguish between a longer term of years and a shorter term of years, because the bargaining solution will be shifted in the defendant's favor when the chance of a death sentence falls to zero. All of these cost-shifts, and not just execution, are consequences of having the death penalty available.[109]

---

determined after conviction in the guilt phase of the trial. *See* Hildwin v. Florida, 490 U.S. 638, 638-39 (1989) (describing the system in Florida).

[109] This has, interestingly, created the potential for an indirect *incapacitation* effect of capital punishment, beyond alleviation of the rather minor risks posed by prison escapees or for murders internal to the prison. By hypothesis, the convict is unable to commit crimes during the years effectively added on to his sentence by the threat of death. Any theoretical effect involving the later years of a murder sentence should become increasingly detectable with time, but could easily be masked (or mimicked) by restrictions on parole, for instance.

*Figure 6 – Effects of Death Penalty, Including Bargaining*




Therefore, instead of merely asserting a zero slope at one point at the extremity of the cost curve, as described *supra* in Figure 4, a thoroughgoing rejection of capital punishment deterrence seems to require a much more general assertion of cost insensitivity along the complete length of the curve relating costs of murder to the likelihood of its commission. This is illustrated in Figure 6, where the implicit "abolitionist line" from Figure 4 is reproduced for comparison with a new implicit line that takes account of the bargaining process and therefore flattens out at a much earlier stage. Because penalties to the left of point D, where death eligibility begins, will be applied with a greater average severity in a capital jurisdiction, one is forced to assume that this greater average severity is irrelevant.

There are three other, more qualitative, consequences of incorporating life and death bargaining into the deterrence debate. First, it calls into doubt any attempt to fall back upon the claim that a life sentence without parole is "just as good as" the death penalty[110] because, even if this were true, the possibility of getting such sentences is not independent of the death penalty. It seems necessary for death penalty abolitionists, insofar as murder is concerned at least, to claim that the rational model of crime is

110 *See, e.g.*, Johnson, *supra* note 62, at 1125, ("[i]t is hard (if not impossible) to argue that the death penalty, as currently administered, could possibly provide a significant enough marginal deterrent over life imprisonment without parole to outweigh the death penalty's exorbitant deadweight losses.").

without meaning and penalties simply do not deter. Second, and related, is that the death penalty's effect during plea bargaining cannot be discounted as psychologically unlikely because the rational faculties are overwhelmed by any "instinctive" reactions or unreasonable optimism about success thought to accompany crime offenses. Rather, the increased threat available to a death penalty jurisdiction is more "real" because the person is already arrested, and it occurs during a negotiation conducted over an extensive period, with the assistance of trained professionals, on an obviously important matter, all circumstances conducive to rational deliberation. Third, it provides another plausible mechanism by which the death penalty—although rarely applied—could conceivably enter the calculations of the rational criminal and cause him to refrain from, or at least be more circumspect about, the choice to kill.[111] Although, empirically, this effect might be difficult to note, since capital jurisdictions are generally "tough on crime" and have higher *de jure* sub-capital penalties as well as *de facto* ones negotiated through leverage, it might be worth further research to investigate whether the lack of a credible threat of the death penalty, and/or plea bargaining practices that restrain the use of such a threat, may influence variation in the deterrent effect of the death penalty *among* capital jurisdictions.[112]

---

[111] *See, e.g.*, Johnson, *supra* note 62, at 1123 (arguing the "credibility of the death penalty cannot possibly be maintained when so few persons who are sentenced eventually receive the punishment"). As long as there is a realistic chance of actually receiving the death sentence if one goes to trial, and of then being executed, it should influence the calculations of murderers who take a plea deal and have it within their control to avoid being one of the unlucky ones on death row. In fairness, however, there are probably some "death penalty states" where the sanction is so rarely applied or where judicial resistance to its execution is so strong, that practically speaking, only "volunteers" or possibly individuals whose cases are a *cause célèbre* among the public stand any realistic chance of being executed, *regardless* of their procedural choices. The question a potential murderer would ask is whether someone "like me" has been executed in the state. If the answer is "yes" then he would be expected to factor this mortality risk into his decision.

[112] *See* Shepherd, *supra* note 4. Shepherd finds a "brutalization effect" in several putative death penalty states where few executions actually take place, but a deterrent effect in most of the states where nine or more people have been executed. (She finds an overall national deterrent effect.) *See supra* note 63, at 240. A brutalization effect anywhere is contrary to the current theory. One possible hypothesis to consider might be that, in terms of signaling, if it becomes widely known and publicized, in the context of the rare and drawn out capital case, that a particular state will not, without great difficulty, kill anyone, this might send a signal encouraging lethal violence. The potential killer might reason that a prosecutorial death penalty threat is not only without any credibility, but that, if the death penalty were actually obtained at trial, it would mobilize judicial and public sympathy and procedural scrutiny such that his prison conditions might improve and his underlying conviction as well as sentence might be overturned. Thus, it would not truly be "brutalization" that increases murder in these states, but rather the criminal response to the perceived sympathy and public uproar generated by introduction of the death penalty, which sends signals to killers of a pro-defendant orientation within a jurisdiction.

## III. APPLICATIONS

Because we should *expect* to see an additional deterrent effect among hardened (in the way described above) and potentially lethal criminals, econometric results showing a general deterrent effect of executions are consistent with theoretical expectations. Those studies showing no such effect would seem to require a behavioral explanation for why the general process of cost-benefit decision making does not apply for murderers before they commit their crimes, despite the fact that it influences their decisions *after* they are apprehended, during the trial process, and during incarceration. It seems implausible, for instance, that criminals simply assume, unrealistically, they will get away with their crimes, and that they only become cognizant of the risk of death penalty when caught, given that the majority of murders in the United States do in fact result in arrest.

A fuller behavioral model may assist econometric modeling, most particularly by indicating that the primary source of deterrence will be located among a particular subset of potential murderers. It further predicts that efficient communication of the fear of death to this group may determine how effective the death penalty will be. This should focus both research and the judicial system on identifying the hypothesized group of individuals resistant to incarceration. In addition, since deterrence works through the threat of death by law—rather than directly by the executions that keep this threat credible—a better independent predictor for the murder rate may be how often this threat is credibly made by the law to the relevant persons, rather than how many executions are carried out.

As discussed, plea-bargaining converts the threat of death into more years of imprisonment, and this shifts the focus of the death penalty debate back to questions of the efficacy of deterrence generally, at least in part. Some element of death penalty deterrence must arise from its effect in plea-bargaining, but we will not understand the extent of this secondary effect until we know in more detail *to what extent* a death threat increases the mean penalty. Specifically, we need to know how many additional years it adds, and in addition, how costly these additional years are to potential murderers. Such knowledge would have the side benefit of separating out the deterrent effects present in a "tough on crime" state that mandates longer sentences generally for crimes, and which also uses the death penalty. Although it can be clarifying to direct attention to the precise difference in imposed cost represented by a life incarcerated and death, ultimately the debate over the deterrence of the death penalty is bound up with the question of the deterrent effect (on potential murderers) of criminal penalties generally.

The primary legal method of distinguishing murderers who receive the death penalty from those who do not is a state-by-state system of aggravating and mitigating factors that has been created and regulated by the United States Supreme Court, based on its understanding that the

Constitution requires certain special adjudicative procedures in a capital case.[113] This "system" is not generally thought to be based upon any single theoretical basis, such as a rational program of criminal deterrence.[114] Indeed, a deterrence rationale is sometimes specifically disclaimed.[115] Nevertheless, taking the current system of regulation as a constraint, the definition of aggravating and mitigating factors (or of capital murder) is susceptible to rational reform in light of the foregoing model. These factors could be reconceived as informational tools to distinguish the relative costliness of prison to certain offenders and, through the logic of marginal deterrence, focus the death penalty on the subcategory of persons to whom it is most salient.

To generalize, a more rational system of selecting persons for capital punishment from the set of all murderers would base this decision on empirical studies of those who are most likely to be unfazed by imprisonment. Each capital sentencing "factor" would be considered a point of similarity with the "target group" for the death penalty. Alternatively, it could be thought a piece of information that shifts the sentencing authorities' estimate of the marginal effect a death sentence would have on *potential* offenders *like the one* under consideration. This estimate, and the ultimate decision of whether to apply the death penalty, would be guided by defining aggravating factors as those increasing the probability of being over that legally-defined threshold of undeterrability for which first-degree murderers are to be executed. Likewise, each "mitigating" factor would be defined as a datum decreasing this probability. This system is probably most applicable in a procedural context like that of Florida. In Florida, all first-degree murder is, by definition, a capital crime, but, unless a sentencing hearing is held, life imprisonment is imposed as an alternative punishment. The jury hears evidence on aggravation and mitigation and makes a recommendation regarding the applicability of the

---

113 *See* LINDA E. CARTER & ELLEN KREITZBERG, UNDERSTANDING CAPITAL PUNISHMENT LAW, Ch. 3 (2004). Discussing this fascinatingly intricate body of jurisprudence, derived primarily from a five-word restriction against "cruel and unusual punishments inflicted[,]" is beyond the scope of the present analysis.

114 *See, e.g.*, Atkins v. Virginia, 536 U.S. 304, 342-352 (2003) (Scalia, J., dissenting) (ridiculing "death-is-different jurisprudence" as consisting of "lip service to precedents," "empty talk" and a "grab bag of reasons" put forth in support of the case-specific "*feelings* and *intuitions* of a majority of the Justices" regarding decency, penology and mercy, and any attention they give to considerations of incapacitation, deterrence or retribution "does not bear analysis") (emphasis in original). *See also* Posner, *supra* note 65, at 64-65 (pointing to an apparent willful ignorance of countervailing evidence in recent death penalty cases).

115 *See* Ring v. Arizona, 536 U.S. 584, 614-615 (2002) (Breyer, J., concurring) ("Studies of deterrence are, at most, inconclusive") (citing to one study of capital punishment's effect in one state, one "special report" by the NEW YORK TIMES and one survey soliciting the "expert views" on the topic given by seventy past and current presidents of two criminological associations and the Law and Society Association).

death penalty. However, the trial judge can ignore this recommendation. The positive and negative factors are "weighed" by both the judge and jury.[116] Florida's process seems to possess the legal potential for a judge to rationally update beliefs based on evidence relevant to an estimate of the offender's deterrability by prison (his probability of being a "Tough"). I have in mind a novel application of Bayes's Theorem in the sentencing context.[117]

Actual deterrence, of course, requires such potential killers to know they are under enhanced risk of death. For some possible factors implied by the current model, this would be relatively easy, since prior prison experience would likely harden someone against future imprisonment and shift him toward the target group of a rationally applied death penalty.[118] At the same time, however, extended prison time (say, for a violent felony) offers plenty of time and opportunity for the criminal justice system to educate the offender that, should he kill upon his release, the death sentence is waiting.

Aggravating factors or capital murder specifications[119] often include elements such as participation in organized crime,[120] gang membership,[121] directing, hiring, or compelling others to kill,[122] or killing as part of the drug trade.[123] All of these could be interpreted to reflect the likelihood the

---

116 *See* Bottoson v. Moore, 833 So.2d 693, 713-714 (Fla. 2002) (Shaw, J., concurring in result only) (describing and criticizing system).

117 For a description of this rule, and other applications, *see* Stephen E. Fienberg & Mark J. Schervish, *Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U. L. REV. 771, 774 (1986).

118 *See* CARTER & KREITZBERG, *supra* note 113, at 117 (reporting that 32 of 38 death penalty states make prior convictions an aggravating circumstance). Most of these involve a requirement of *violent* felony, however. Although this makes some sense, given that these individuals will more likely face a decision whether or not to murder in the future, any felony giving a long sentence could adapt the felon to make him resistant to the sanction of further incarceration. *But see* James Robertson, *Closing the Circle: When Prior Imprisonment Ought to Mitigate Capital Murder*, 11 KAN. J.L. & PUB. POL'Y 415, 419 (2002) (arguing that prison harms inmates psychologically and makes them less culpable for post release crimes). Comparison of the present analysis and Robertson's, which reach diametrically opposite conclusions, may be a good example of the distinction between a consequentialist and a desert approach to capital sentencing.

119 Alternatively, certain kinds of murders may be defined as including "special circumstances." For purposes of the current analysis, the two are not distinguished.

120 *See, e.g.*, 21 U.S.C. § 848 (2006).

121 *See, e.g.*, CAL. PENAL CODE § 190.2(a) (West Supp. 2004) (making potentially capital "murder carried out to further the activities of a criminal street gang").

122 *See, e.g.*, 21 U.S.C. § 848(e)(1)(A) (2006), creating a possible death sentence for someone who "intentionally . . . counsels, commands, induces, procures, or causes the intentional killing of an individual and such killing results[.]"

123 *See, e.g.*, VA. CODE ANN. § 18.2-31(9) (2004) (making potentially capital "willful, deliberate, and premeditated killing of any person in the commission of or attempted commission of a violation . . . involving a Schedule I or II controlled substance, when such killing is for the purpose of furthering the commission or attempted commission of such violation.").

individual involved is, or has the potential to be, part of a continuing criminal network with operations inside and outside a prison. Therefore, it is more likely the convict will have continued economic capacity as well as some preferred status during incarceration.[124] Although it may be factually true that these features were attached to an increased penalty because the underlying activity (organized crime, drugs, etc.) was thought especially socially harmful in itself and, therefore, killings attached to it are especially "heinous," these bases can be justified in light of a deterrence rationale.

This is not true of factors like the killing of children or the killing of pregnant women,[125] or, more generally, murders "especially heinous, atrocious or cruel,"[126] which are also considered "deserving" of the death penalty. Of course, even from the deterrence perspective taken here, assigning the death penalty to heinous killings may serve the public good. The great majority of all murderers fear death and, at least plausibly, are deterred by it. Therefore, providing for potential capital punishment of all first-degree murder (essentially true already in several states) might be justified. Further, people who commit heinous crimes could be particularly dangerous. This is an almost definitional point, since the types of killings they commit are designated by society as especially harmful. Given plea-bargaining, putting them under fear of a death sentence makes it more likely a life sentence or long term of years can be obtained with certainty, and it thereby serves an important incapacitation function.

Realistically, however, the amount of time and money a particular state will allocate to pursue the death penalty will not be limitless, and, therefore, it is advisable that these resources be directed where they can achieve the greatest deterrent effect. Many current death sentences, handed down for grotesque spasms of violence committed by borderline retarded or brain damaged offenders, often intoxicated during the crime, are probably suboptimal as social messages. The potential offenders who might recognize "themselves" as similar to such killers are poor deliberators; even if they were dimly aware of an increased punishment for murder, the circumstances of their potential offense is likely to involve only a short sequence of a behavioral trigger followed by a lethal rage response, leaving no time for deliberation even if their thinking were not chemically distorted. So the deterrent signal will be quite weak for them, while other more canny

---

124 *See, e.g.*, United States v. Shyrock, 342 F. 3d 948, 961 (9th Cir. 2003) (describing one hybrid prison street gang, the "Mexican Mafia") ("By using violence, the Mexican Mafia eventually gained significant power and control over illegal activities in the California prison system. As members were released from state custody, they extended their influence outside the prison system to control drug distribution—principally by 'taxing' drug dealers—in parts of Southern California."). The rest of this extended case describes several lethal operations of this group, and also describes specific instances where members received narcotics in jail, sold them and collected money for their suppliers. *See id.* at 969.

125 VA. CODE ANN. § 18.2-31 (2004).

126 CAL. PENAL CODE § 190.2(a)(14) (West Supp. 2004).

criminals, may perceive the deterrent as not applicable to their own circumstances. These potentially deterrable offenders may assume that only subnormal individuals, whose killings are lurid and psychotic (and who are unable to afford counsel or assist their defense), are, as a matter of fact, going to receive the death penalty, and that the *effective* sanction for any murder *they* commit will remain mere imprisonment. Consequently, the widespread use of "aggravating factors" unrelated to targeting those offenders most susceptible to marginal deterrence probably represents a misallocation. Assuming capital punishment could be having a deterrent effect, using it preferentially on undeterrable types of killers blunts whatever effectiveness it might have.[127]

As discussed in Part II, who actually is sentenced to death is partly a product of the process of plea-bargaining, and there is the potential for a critique here of current prosecutorial practice. Individuals who expect a relatively "easy" time in prison have every incentive to save themselves and avoid the death penalty, and the most obvious such people are those with command authority over other criminals inside and outside of prison, such as organized crime bosses, terrorist leaders, or senior members of street gangs.[128] But, if caught, these people will normally successfully bargain for their lives. Probably, and understandably, many prosecutors find appealing the prospect of getting prominent public enemies "off the street" with certainty and without an expensive trial.

However, the above analysis suggests there are serious dangers to giving in to this temptation, particularly since, unlike the murderer who is faced only once with a decision about whether to kill, a "kingpin" will be faced with multiple lethal decisions, and any deterrent effect on such a person will be multiplied over the number of these decisions. This trait is shared by serial killers, who are also implied by the present model to be an especially appropriate target of death penalty prosecutions and, thus, the use of aggravating factors likely to cover them.[129] To a greater extent even than

---

127 A further practical problem with the focus on "heinous" crimes is that such crimes are committed by extremely abnormal persons, who often have severe emotional and mental deficits, frequently caused by horrific developmental experiences. These are usually statutory "mitigating" factors that require extensive efforts to litigate and overcome in the sentencing phase and then uphold over multiple levels of appeal. Thus, it is not at all clear that the net cost of prosecuting such persons is less than that of more "normal" killers who commit more "normal" murders—and who are probably more ex ante deterrable.

128 Jeff Fort, the head of the Chicago street gang once known as the Blackstone Rangers, and later as the El Rukns, provides a well-documented example. Although Fort was put in federal prison for a long sentence, "Fort remained the leader and mastermind of the El Rukns." United States v. McAnderson, 914 F.2d 934, 939 (7th Cir. 1990) (describing Fort's planning from prison to have the gang hire itself out as a terrorist group for the Libyan government in exchange for millions of dollars, planning murders to impress the Libyans and acquiring light anti-tank missiles).

129 The death penalty might be more effective on these categories, for instance, than it would be on individuals who commit multiple murders at the *same* time (a very common aggravating factor). As

serial killers, moreover, kingpins (and some terrorists) are the criminals most likely to deliberate and be knowledgeable about the personal risks and rewards of killing. A targeted message to criminal leaders—that they will be selected for capital sentencing if a death can be attributed to them—is not likely to stop their activities, of course, but it could make them more cautious, killing fewer people or taking costly precautionary measures. This would have an additional salutary effect of decisively showing "crime does not pay" for prominent killers whose criminal success prior to arrest encourages others within their communities and organizations to engage in criminality and lethal violence.[130] If someone like this continues to wield power and authority from prison, this message is not sent.

Assuming that a prominent crime figure is tied to murder, there is at least theoretical merit in the policy (and message) of "no deals" and seeking the death sentence whenever possible. If there is indeed death penalty deterrence, these people will most likely and most often be the ones whose behavior is influenced by it, and, therefore, despite the increased cost of trying them even in comparison with other capital defendants, any currently existing deterrent effect should be enhanced by identifying, prosecuting—and publicizing the prosecution of—calculating criminals who are unafraid of prison. Such persons can hardly complain of inequity in this attempt at welfare-maximization, given that they have chosen to treat the life of *everyone else* according to calculations of *their own* utility. In different ways, gangland figures, serial killers, and terrorists all make killing their "business"; it is therefore sensible, as well as singularly appropriate, to selectively focus the use of capital punishment on raising the price of their crimes.

---

was illustrated with the case of Jeffrey Dahmer, capital punishment may not prevent serial killers entirely, but it has a good chance of slowing them down. This distinction points out a difference between the present theory and the alternative view of marginal deterrence as described by Stigler. When marginal deterrence is viewed as targeted toward the intensity (rather than the frequency) of criminal activity, we might imagine that a rule of "life for one murder, death for two" would result in the criminal only killing one person when he could have killed two. While this might happen occasionally, it seems somewhat unrealistic in most murder contexts. For instance, if there are two witnesses and you kill one but not the other, you have saved some expected cost, but you have gained nothing of value. For any benefit, both witnesses must die. So, we should not see much "Stigler deterrence" due to the death penalty. I hypothesize more lives are likely to be saved by reducing the number of occasions when people do *any* killing, as well as the number of felonies engaged in at all, when offenders know there will be a risk they may have to kill in order to implement the crime.

130 *Cf.* Blecker, *supra* note 88, at 1223: "A general deterrence advocate would urge the D.C. police to concentrate on the ostentatious criminals—perverse role models who parade the streets poisoning the minds of the kids. These criminals, who most undermine deterrence, would not be permitted to hustle openly and flaunt their lifestyle."

# *VERIZON v. TRINKO*: FROM POST-CHICAGO ANTITRUST TO RESOURCE-ADVANTAGE COMPETITION

*Christopher M. Grengs*[*]

## INTRODUCTION

In January, 2004 the Supreme Court decided the case of *Verizon Communications, Inc. v. Law of Office of Curtis V. Trinko, LLP* ("*Trinko*").[1] There, customers who received local telephone service from a competing local exchange carrier ("LEC") brought an action against incumbent LEC Verizon, alleging violations of Section 2 of the Sherman Antitrust Act and the Telecommunications Act of 1996 ("Telecommunications Act"). Principally, the Supreme Court held that: (1) the Telecommunications Act of 1996 had no effect on the application of traditional antitrust principles, due to the existence of an antitrust-specific savings clause that precluded the finding of implied immunity; (2) traditional antitrust principles did not justify adding an exception to the general rule that a refusal to cooperate with rivals is not anticompetitive conduct; and (3) plaintiff's complaint alleging Verizon breached a duty to share its network with competitors did not state a monopolization claim under § 2 of the Sherman Act.

The *Trinko* decision was widely anticipated for its implications regarding the interplay between antitrust law and telecommunications law.[2]

---

* Attorney Advisor, Federal Trade Commission Office of Policy Planning. The views expressed in this article are those of the author and do not necessarily represent the views of the FTC or any of its commissioners. The author would like to thank Jerry Ellig, Shelby D. Hunt, Frank M. Machovec and James C. Cooper for reviewing this article and providing useful comments.

[1] Verizon Commc'ns., Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398 (2004) ("*Trinko*").

[2] In particular, the Court had the opportunity to resolve a split among the circuit courts as to whether an alleged violation of the Telecommunications Act of 1996 also states a monopolization claim under Section 2 of the Sherman Act. *See* Goldwasser v. Ameritech Corp., 222 F.3d 390, 400 (7th Cir. 2000) ("it would be undesirable here to assume that a violation of a 1996 Act requirement automatically counts as exclusionary behavior for purposes of the Sherman Act § 2."). *Compare* Law Offices of Curtis V. Trinko v. Bell Atlantic Corp., 305 F.3d 89 (2d Cir. 2002) (violation of the Telecommunications Act of 1996 may constitute Section 2 monopolization claim); Covad Commc'ns. Co. v. BellSouth Corp., 299 F.3d 1272, 1282 (11th Cir. 2002) (rejecting *Goldwasser* "to the extent that it is read to say that a Sherman Act antitrust claim cannot be brought as a matter of law on the basis of an allegation of anti-competitive conduct that happens to be 'intertwined' with obligations established by the 1996 Act."); MetroBet Servs. Corp. v. U.S. West Commc'ns., 325 F.3d 1086 (9th Cir. 2003) (rejecting a *Goldwasser*-type defense and holding that even "[w]here the conduct challenged under the antitrust laws 'is the product of the regulated business' independent initiative and choice, [that conduct] is properly subject to antitrust scrutiny.").

The case also represents the Supreme Court's most important monopolization decision in a decade.[3] Commentators have largely focused on the telecommunications and facial antitrust aspects of the case.[4] This Article explains that *Trinko* also represents a profound change in the Supreme Court's jurisprudence on microeconomic competition. Specifically, *Trinko* represents the first Supreme Court case to break distinctly with both the well-defined "Structure-Conduct-Performance" and "Chicago" schools of microeconomic analysis, as well as the vaguely defined "post-Chicago" school of microeconomic analysis. In *Trinko*, the Supreme Court articulated a classical, rivalrous process view of competition, as refined through the corollary insights of the "Resource-Advantage" theory of competition, consistent with the 1890 enactment of the Sherman Act. In doing so, the Court rejected the neoclassical construct of "perfect" competition as a welfare ideal. Section I. of this Article provides an overview of the development of microeconomic competition law and theory. Section II. analyzes the *Verizon v. Trinko* case. Section III. discusses *Trinko*'s implications for the concepts of competition, monopoly, and entry.

## I. MICROECONOMIC COMPETITION LAW AND THEORY—AN OVERVIEW

From at least the time of *Darcy v. Allein* in 1602, it was well-recognized at English common law "[t]hat every person of that society has been used and accustomed to buy, sell, and trade freely all merchantable property within this realm of England from whatsoever person or persons, etc."[5] In contrast to such "free" microeconomic activity, government "monopoly" grants for "the sole making" of certain products were condemned because they excluded others from trade, caused their impoverishment, made their goods worthless, and, thus, increased prices due to artificial product scarcity.[6]

---

*See generally Antitrust, Telecom Issues Take Center Stage at Supreme Court, Verizon Communications v. Law Offices of Curtis V. Trinko*, 11 No. 11 ANDREWS ANTITRUST LITIG. REP. 11 (Feb. 16, 2004); Matthew L. Cantor, *Is 'Trinko' the Last Word on a Telephone Monopolist's Duty to Deal?*, 2004 N.Y.L.J. 4 (col. 4) (2004).

[3] *See* Eastman Kodak Co. v. Image Technical Services, Inc., 504 U.S. 451 (1992); Spectrum Sports v. McQuillan, 506 U.S. 447 (1993).

[4] *See supra* note 2.

[5] SIR EDWARD COKE, REPORTS CH. 11 (summarizing "The Case of Monopolies" (1602) Trinity Term, 44 Elizabeth I In the Court of King's Bench). There, the Queen gave to Edward Darcy a monopoly for the sole right to stamp playing cards in exchange for an annual payment. Darcy sued Allein, a London haberdasher, for selling playing cards without paying him for the privilege or for the use of his stamp. The King's Bench ruled the monopoly void as against common law, which protects the freedom of trade and liberty against such exclusive grants.

[6] *Id.*

European and American economists from the time of Adam Smith's *The Wealth of Nations* (1776) through Alfred Marshall's *Principles of Economics* (1890/1920) viewed competition as being fundamentally a dynamic, rough-hewn, and open-ended process animated by rivalrous, entrepreneurial activity.[7] Whether or not a particular economic activity was

---

[7] FRANK M. MACHOVEC, PERFECT COMPETITION AND THE TRANSFORMATION OF ECONOMICS 2, 9-10, 111, 47-49, 96-158, 244-45, 292 (1995). "From Smith to Marshall, market activity had been suitably framed within a process perspective" in both Europe and America. *Id.* at 111. "[T]he equilibrium model became the dominant paradigm . . . less than eighty years ago." *Id.*

Adam Smith's famous description of the "Invisible Hand" in the WEALTH OF NATIONS is a metaphor that describes a process in which dispersed, self-interested economic activity is spontaneously coordinated without the need for a central authority. "By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it." ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS Bk. IV, Ch. 2, P. IV, 2.9 (Edward Cannan, ed., 5th ed. 1904) (1776). Smith viewed microeconomic competition as a process of rivalry. "Upon most occasions [a seller] can hope to jostle that other out of his employment by no other means but by dealing upon more reasonable terms . . . ." *Id.* at Bk. II, Ch. IV, P. II, 4.8. "Rivalship and emulation render excellency, . . . and frequently occasion the very greatest exertions." *Id.* at Bk. V, Ch. I, P. V, 1.133.

Capitalism's most extensive critic, Karl Marx, also shared this process perspective. "Constant revolutionizing of production, uninterrupted disturbance of all social conditions . . . distinguish the bourgeois epoch from all earlier ones. . . . All fixed, fast, frozen relations . . . are swept away, all new-formed ones become antiquated before they can ossify. All that is solid melts into air; all that is holy is profaned . . . ." KARL MARX & FREDERIC ENGELS, THE MANIFESTO OF THE COMMUNIST PARTY (1848).

For Marshall "economics, like biology, deals with a matter, of which the inner nature and constitution, as well as the outer form, are constantly changing." ALFRED MARSHAL, PRINCIPLES OF ECONOMICS, Appendix C, Bk. IV, Ch. XIII, P. 4 (8th ed. 1920) (1890). Marshall observed "that selective influence of struggle and competition which in the earlier stages of civilization caused those who were strongest and most vigorous to leave the largest progeny behind them; and to which, more than any other single cause, the progress of the human race is due." *Id.* at Bk. IV, Ch. V, P. 22. Further, "we know that an animal or a vegetable species may differ from its competitors by having two qualities, one of which is of great advantage to it; while the other is unimportant, perhaps even slightly injurious, and that the former of these qualities will make the species succeed in spite of its having the latter . . . ." *Id.* at Bk. IV, Ch. VIII, P. 11. Perhaps most famously, Marshall compared the growth of businesses to the growth of trees:

> We may read a lesson from the young trees of the forest as they struggle upwards through the benumbing shade of their older rivals. Many succumb on the way, and a few only survive; those few become stronger with every year, they get a larger share of light and air with every increase of their height, and at last in their turn they tower above their neighbors. . . . One tree will last longer in full vigour and attain a greater size than another; but sooner or later age tells on them all. Though the taller ones have a better access to light and air than their rivals, they gradually lose vitality; and one after another they give place to others, which, though of less material strength have on their vigour of youth.

*Id.* at Bk. IV, Ch. XIII, P. 4-5. Marshall's *Principles* "was also the Bible to most early American students"—the last great bookend to the classical era. MACHOVEC, *supra* note 7, at 244-45. So, "[i]n[to]1920, the centre of gravity in the US, the UK, and even more so on the Continent, was grounded in process thinking, not static analysis." *Id.* at 29.

deemed to be "competitive" was generally based on whether freedom of entry into that activity existed.[8]

In this "classical" era it was widely recognized that robust microeconomic competition required that a seller be allowed to seek out an economic space where it could create a competitive advantage for itself, as compared to other actual or potential competitors.[9] By the end of the nineteenth century, theories of competition as a biological, evolutionary process of advantage selection were prominent.[10] For example, although he is considered to be the "father" of marginal mathematical economic analysis, Marshall still viewed economics as being fundamentally a biological, evolutionary process. For him, the notion of a static, mathematical equilibrium state was merely a useful hypothetical.[11] In the classical view, then, competition can be defined as: "the process over time of one or more parties acting to secure an exchange with a second party from a third party or potential third party by offering terms that are more favorable than those of the third party."[12]

By contrast, the term "monopoly" was typically used to describe a seller who faces absolutely no actual or potential competition, due to the exclusion of all competitors in a manner distinct from rivalrous competition. Initially, the term was closely associated with the protectionist grant of an exclusive franchise by government.[13] Therefore, "the nature of

---

[8] MACHOVEC, *supra* note 7, at 2, 11, 16-17, 119. *See also* Edward S. Mason, *Monopoly in Law and Economics*, 47 YALE L. J. 34, 36 (1937). "[T]he antithesis of the [classical] legal conception of monopoly is *free* competition, understood to be a situation in which the freedom of any individual or firm to engage in legitimate economic activity is not restrained by the state, by agreements between competitors or by the predatory practices of a rival." *Id. See also infra* note 202.

[9] THOMAS CARL SPELLING, A TREATISE ON TRUSTS AND MONOPOLIES 37 (1893) (described by Areeda and Hovenkamp as "an influential treatise on trusts and monopolies . . . ." PHILLIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW 102 (2d ed. 2002)). "[P]ublic policy requires that every man shall be at liberty to work for himself, and shall not deprive himself or the State of his labor, skill, or talent, it is equally a principle of public policy that a man shall be enabled to sell *to the best advantage* anything that he has acquired by his labor, skill, or talent," subject to prohibitions against unreasonable restraints on trade. *Id.* (summarizing Leather Cloth Co. v. Lorsont, 9 L. R. Eq. 345 (1869)) (emphasis added).

[10] *See supra* note 7.

[11] "The main concern of economics is with human beings who are impelled, for good and evil, to change and progress. Fragmentary statical hypotheses are used as temporary auxiliaries to dynamical—or rather biological—conceptions; but the central idea of economics, even when its Foundations alone are under discussion, must be that of living force and movement." MARSHALL, *supra* note 7, at Preface, P. 22.

[12] "The strict meaning of competition seems to be the racing of one person against another, with special reference to bidding for the sale or purchase of anything." *Id.* at Bk. I, Ch. I, P. 13. *See also* MIRRIAM WEBSTER'S NEW COLLEGIATE DICTIONARY (2006), defining competition as "the effort of two or more parties acting independently . . . to secure the business of a third party by offering the most favorable terms." *See also supra* note 8 and *infra* note 19.

[13] "[C]lassical economists did not associate monopoly damage with a particular state of affairs at equilibrium (such as P > MC), but rather saw it as an impediment to the will to compete, a condition

monopoly was to be found mainly in restrictions on trade, and its remedy was . . . 'a fair field with no favor.' "[14]

English and American courts and "influential" secondary-source legal authorities generally shared these classical-era views of "competition" as a rivalrous process[15] and (in diametric opposition) of "monopoly" as a seller facing absolutely no actual or potential competition, due to the exclusion of all competitors from that process.[16] "The [classical] economists' emphasis

---

spawned by exclusive production rights, usually bestowed by a government." MACHOVEC, *supra* note 7, at 16. Although classical parlance sometimes used ambiguous language, as for example by confusing land ownership with monopoly, "[i]n general . . . the classical concept of monopoly was tightly linked to the permanent profits garnered from mercantilist franchising rights which insulated favoured firms from fear of entry by rivals." *Id.*

For example, Adam Smith's *Wealth of Nations* is the first major critique of the European mercantile practice of imposing such exclusionary monopolies on commerce. "Every European nation has endeavored more or less to *monopolize* to itself the commerce of its colonies, and, upon that account, *has prohibited* the ships of foreign nations from trading with them, *and has prohibited* them from importing European goods from any foreign nation." SMITH, *supra* note 7, at Bk. IV, Ch. 7, P. 43 (emphasis added). "Monopoly of one kind or another, indeed, seems to be the sole engine of the mercantile system." *Id.* at Bk. IV, Ch. 7, P. 175. Thus, for Smith, "monopolization" is the exclusion of competition, typically by law, and a "monopolist" is one who engages in such exclusion.

Similarly, a century later, in outlining "Some Fundamental Notions" of economics, Marshall described "material goods" as "includ[ing] the physical gifts of nature, land and water, air and climate; the products of agriculture, mining, fishing, and manufacture, buildings, machinery, and implements; mortgages and other bonds; shares in public and private companies, *all kinds of monopolies, patent-rights, copyrights*; also rights of way and other rights of usage." MARSHALL, *supra* note 7, at Bk. II, Ch. II, P. 3 (emphasis added). That is, Marshall linked "monopolies" with exclusive legal forms such as patents and copyrights. At the same time he distinguished "monopolies" from the mere ownership of a mortgage, bond, or company, or the mere existence of naturally occurring heterogeneous resources.

[14] Mason, *supra* note 8, at 35.

[15] "*Rivalry is the life of trade.* The thrift and welfare of the people depend upon it. Monopoly is opposed to it all along the line." SPELLING, *supra* note 9, at 155 (emphasis added).

[16] "A monopoly is described by Lord Coke to be an institution or allowance by the king by his grant, commission, or otherwise, to person or persons, bodies political or corporate, of or for the *sole* buying, selling, making, or using of anything, whereby any person or persons, bodies politic or corporate are sought to be *restrained* of any freedom of liberty they had before, or hindered in their lawful trade." 7 MATTHEW BACON, A NEW ABRIDGEMENT OF THE LAW 22 (1768) (quoted in SPELLING, *supra* note 9, at 160-61 n.1, emphasis added). *See* also 4 WILLIAM BLACKSTONE, COMMENTARIES Bk. IV. Ch. 12 (1765-69) (quoted in SPELLING, *supra* note 9, at 160-61 n.1) (making the same point using almost identical language).

An 1839 entry in the English *Penny Cyclopedia* is also instructive on the legal meaning of "monopoly" in this era. The entry acknowledges that in the classical era, "[i]f a number of individuals were to unite for the purpose of producing any particular article or commodity, and if they should succeed in selling such article very extensively, and almost solely, such individuals in popular language would said to have a monopoly." Reprinted in GEORGE J. STIGLER, THE ECONOMIST AS PREACHER, AND OTHER ESSAYS 40-41 (1982). But this did not mean that such individuals had escaped the competitive process, that it had been inhibited. "Now, as these individuals have *no advantages given them by the law over other persons*, it is clear they can only sell more of their commodity than other persons by producing the commodity cheaper and better." *Id.* (emphasis added). The entry explains that: "[i]t seems . . . that the word monopoly was never used in English law, except when there was a

on free entry into the industry as characteristic of competition and restriction of entry into the industry as the *differentia specifica* of monopoly was in complete harmony with the judicial predilection."[17]

Through the 1930s, "the courts have found monopoly because of conspiracy and the exclusion of others from the market rather than control of the market" through the sale of a product at a price above marginal cost or because of a market's particular structure.[18] That is, "[t]he original meaning of monopoly, as an *exclusion* of others from the market by a sovereign dispensation in favor of one seller, has continued to mean exclusion in the broad sense of restriction of competition."[19] Similarly,

---

*royal grant* authorizing some one or more persons only to deal in or sell a certain commodity or article." *Id.* (emphasis added).

This English definition of "monopoly" was also generally adopted into American law and recognized by classical-era American courts. *See generally* Charles River Bridge v. Warren Bridge, 36 U.S. 420 (1837):

> A monopoly, then, is an *exclusive privilege* conferred *on one, or a company*, to trade or traffick in some particular article; such as buying or seller sugar or coffee, or cotton, in derogation of a common right. Every man has a natural right to buy and sell these articles; but when this right, which is common to all, *is conferred on one*, it is a monopoly, and as such, is justly odious. It is, then, something carved out of the common possession and enjoyment of all, and equally belonging to all, and *given exclusively to one*.

*Id.* at 451 (emphasis added). *See also* Memphis v. Memphis Water Co., 52 Tenn. 495 (1871), 1871 WL 3872. "We know of no better definition of a monopoly, than that given by Lord Coke, and adopted by the Supreme Court in the case of *Charles River Bridge v. Warren Bridge* . . . ." *Id.* at *17. *See also* Patterson v. Wollman, 5 N.D. 608 (1896), 67 N.W. 1040. "In the celebrated case of [*Charles River Bridge v. Warren Bridge*] Justice Story has set forth this matter in very clear language . . . ." *Id.* at 1044. Thus, American legal commentators also viewed "monopoly" as being harmful because it could "*exclude rivalry*, monopolize business, and engross the market." SPELLING, *supra* note 9, at 5 (emphasis added). *See also id.* at 160-61 n.1 (citing Charles River Bridge v. Warren Bridge). *See also supra* notes 13 and 14 and related text and *infra* note 68 and accompanying text.

[17] Mason, *supra* note 8, at 35. "In both countries there is substantial harmony between the definitions of monopoly . . . ." SPELLING, *supra* note 9, at 181.

[18] Mason, *supra* note 8, at 39. Thus, in this era, "[i]t is doubtful whether the act of engrossing itself [(increasing market share)], in the absence of a conspiracy to exclude competitors, would carry any monopoly connotation in the law." *Id.*

[19] *Id.* at 44 (emphasis added). "[I]n the earliest development of the law in the Elizabethan period in England monopoly came to be identified with an exclusive grant by the crown to individuals for the conduct of particular businesses, and the idea of exclusion, in the broad sense of restriction of competition, has been retained in the development of the law." Robert W. Harbeson, *The Present Status of the Sherman Act*, 39 MICH. L. REV. 190, 204 (1940).

Mason elaborates:

> The disposition of American courts has been, at least until very recently, to hold all contracts for division of territory, pooling, fixing of prices, common marketing control of supply, or which restrict the freedom of the contractors to compete in other ways, unenforceable and, since the Sherman Act, illegal. The opinions of the court in these cases constantly refer to control of the market, but little examination of evidence pertinent to the question of market control [through the sale of a product a price above marginal cost] is ever undertaken. The test of monopoly, or attempt of monopoly, is here restriction of competition. American courts have in this class of cases been willing to accept the contract itself as evidence of restriction and, consequently, of an attempt to monopolize, without inquiring further into the question of how great a control of the market is secured to the contracting parties . . . .

"nor is control of the market to be inferred merely from the number of existing competitors. Potential competition must be considered."[20] "Indeed, the dicta of many trust cases might be interpreted as indicating a judicial opinion that in the absence of legal restraints or overt predatory acts against potential competitors, free entry to the market precludes any element of control."[21] Therefore, "a monopolistic situation, or an attempt to monopolize, is evidenced to the courts primarily, if not exclusively, by a limitation of the freedom to compete,"[22] due to either a government-imposed exclusion of competition or private conduct having the same effect.[23]

"The [classical] legal view of monopoly is epitomized in . . . the *United States Steel Corporation* case in 1920, in which it was held that the corporation was not a monopoly within the meaning of the Sherman Act, primarily on the ground that it was not at the time of the suit guilty of predatory tactics toward competitors, and that 'the law does not make mere size an offence or the existence of un-exerted power an offence.' "[24] The

---

> The British courts . . . have returned a somewhat different answer. They have tended to accept every contract designed to limit contracting competitors as reasonable in the absence of intention or actual attempt to injure or destroy a competitor. . . . In neither case has the rule of reason been given any intelligible content in terms of control of the market [through the sale of a product at a price above marginal cost] despite the frequency with which this phrase has graced judicial utterances.
>
> Cases involving a union between competitors accomplished by amalgamation or fusion or merger have in this country most frequently involved the application of the rule of reason, and it is in these cases that the characteristic legal conception of monopoly is most evident. An amalgamation of competing firms may, and ordinarily does, take place for reasons other than to secure control of the price of the articles produced or sold by these firms. The courts could not, therefore, plausibly assume, as they did in the case of contracts to limit competition, that all amalgamations were *prima facie* evidence of an attempt to monopolize.
> . . . .
> By monopoly, however, the courts did not mean control of the market [though the sale of a product at a price above marginal cost] but restriction on competition . . . . If the manifestation of the intention to limit the competition of outsiders took the form of overt acts such as local price discrimination, espionage, or securing of railway rebates, the courts could find evidence of restrictions directly relevant to their conception of monopoly. As a matter of fact it is clear that this was the direction taken in the judicial application of the rule of reason. The size of the combination or its share of the total output of a product became important only when accompanied with predatory practices affecting the freedom of others to compete.

Mason, *supra* note 8, at 41-43.

20 Mason, *supra* note 8, at 47.

21 *Id.*

22 *Id.*

23 "In view of the [classical] legal definition of monopoly, intent to monopolize would in such cases be indicated almost solely by predatory tactics against present or potential competitors." Harbeson, *supra* note 19, at 205-06.

24 Robert W. Harbeson, *A New Phase of the Antitrust Law*, 45 MICH. L. REV. 977, 978-79 (1947) (*citing* United States v. U.S. Steel Corp., 251 U.S. 417, 451 (1920) (affirming dismissal of suit to dissolve defendant and related companies)).

Supreme Court re-emphasized this point in its 1927 *U.S. v. Int'l Harvester Co.* decision, saying "[t]he law does not make the mere size of a corporation, however, impressive, or the existence of unexerted power on its part, an offense, when unaccompanied by unlawful conduct in the exercise of its power."[25] Similarly, classical-era English and American courts generally defined an illegal restraint on competition as consisting of conduct intended to, and having the tendency directly of: (1) creating scarcity in, or (2) increasing the price of some product, irrespective a business's size or whether it sold a product at a price above its marginal cost.[26]

Frank H. Knight, professor of economics at the University of Iowa and later the University of Chicago, crystallized the mathematical model of what has come to be known as "perfect" competition in his 1921 *Risk, Opportunity, and Profit*.[27] Knight's original formation used eleven premises. Today, those eleven premises are typically condensed into four: (1) all sellers offer identical, perfectly homogenous products in response to perfectly homogenous consumer demand; (2) there are numerous small sellers and customers; (3) sellers face no "barriers" to entry or exit; and (4) each seller and buyer is perfectly informed about available products and their prices with costless information. In turn, these four premises require two additional, often unstated, premises: (5) that all sellers have identical,

---

[25] United States v. Int'l Harvester Co., 274 U.S. 693, 708 (1927) (affirming dismissal of suit seeking further relief related to previous suit alleging restraint of trade and monopolization).

[26] As Spelling explains:

> It has been seen that the courts are not governed by any hard and fast rule in determining whether a particular contract is in restraint of trade and amenable to the rule of public policy rendering such contracts invalid, the test being whether the restriction is reasonable and necessary to the party's protection, the public interest being constantly kept in view. The nearest to a definite proposition which may be advanced is that any compact between two or more persons or corporations affecting any article or commodity of which the public must have a constant supply, the sole intent and direct tendency of such arrangement being the creation of a scarcity or the enhancement of the price, will be nullified by the courts, or specific enforcement refused.

SPELLING, *supra* note 9, at 76.

Also, as Mason explains:

> The development in the 17th and 18th century of the doctrine of 'reasonable restraints,' as applied to restrictive covenants in connection with the sale of a business, does not seem to have involved any closer consideration of the [neoclassical] monopoly problem.
>
> . . . .
>
> This protection of the public interest was leveled primarily not against [neoclassical] monopolistic control of the market [through the sale of a product at a price above marginal cost] but against the loss to the common weal of the services of a productive agent. There is no evidence that the courts examined the data relevant to the question whether such a contract might lead to control of the market [through the sale of a product at a price above marginal cost]. If any monopoly consideration was involved, it was monopoly in the sense of restriction of competition, not of control of the market [through the sale of a product at a price above marginal cost].

Mason, *supra* note 8, at 39-40.

[27] F.H. KNIGHT, RISK, OPPORTUNITY, AND PROFIT (Houghton Mifflin 1921) (defining perfect competition in terms of eleven premises). *See also* MACHOVEC, *supra* note 7, at 238.

perfectly homogenous production functions; (6) which are created from resources that are perfectly homogenous and perfectly mobile.

In puritanical fashion, consumers are expected to desire a single, perfectly homogenous product that, in turn, is supplied by indistinguishable, perfectly homogenous sellers. Therefore, the products of all sellers are equally good substitutes for each other. Indistinguishable atomistic sellers passively take the same prices from consumers and, in exchange, merely produce and sell the maximum amount possible. As a result, there is no seller market power, where a product's price is greater than its marginal production cost. Likewise, there is no active rivalry between sellers. No seller has any advantage over any other seller, in terms of resources, information, understanding, or initiative. Thus, all sellers are limited to maximizing profits at the same, average returns. No above-average economic profits may be obtained. There is no incentive for a potential entrant to enter a perfectly competitive market over some other market, where earning an above-average profit might be possible. Thus, there is no role for new entrepreneurs[28] or for the introduction of new and innovative products. Imperfections in economic actors' information have been assumed away. Under these exacting conditions, without new entry, the result is a static equilibrium outcome (or at least a tendency toward that outcome). There is no economic growth. "Competition" is reduced to an ethereal, mathematical abstraction where all product prices are equal to their marginal production cost in a world where sellers face infinitely elastic demand as represented by a horizontal demand curve.

Mistakenly, Knight believed the classicals had been incipient perfect competition theorists.[29] In reality, the widespread, post-1921 acceptance of "perfect" competition as the appropriate welfare ideal set off a dramatic conceptual revolution in the study of microeconomics.[30] The bulk of microeconomic theory, scholarship, and law since that time have used this construct as a welfare ideal. Importantly, though, as Knight himself duly pointed out, the use of the term "perfect competition" to describe his model is an unfortunate misnomer. It is not competition at all. Actually, as Knight freely admitted, it is simply a passive, non-rivalrous, mediocre "atomistic" state—completely unlike competitive activity in the real

---

28 "The entrepreneur became a eunuch in neoclassical economies . . . ." MACHOVEC, *supra* note 7 at 10.

29 Knight believed that "the historic body of economic theory rests upon the assumption of perfect competition, but that the precise character of this assumption has been partially implicit and never adequately formulated." KNIGHT, *supra* note 27, at P. II, Ch. III.

30 "The perfectly competitive model did not make its real debut as an analytical tool until the 1920s—after the profession had digested Frank Knight and after the influence of Alfred Marshall had waned." MACHOVEC, *supra* note 7, at 12; SHELBY D. HUNT, A GENERAL THEORY OF COMPETITION, RESOURCES, COMPETENCES, PRODUCTIVITY, ECONOMIC GROWTH 249 (Naresh K. Malhotra ed., 2000).

world.[31] Nonetheless, as the study of economics became increasingly mathematical, economists increasingly adopted this abstract, non-competitive mathematical stasis as the proper "competitive" welfare ideal for analyzing microeconomic performance. Failing to realize that "perfect competition" is not competition at all,[32] most economists began to be guided by a false idol.

In this "neoclassical" paradigm ideal, perfect "competition" performs best under these exacting conditions.[33] From this "radically different,"[34] upside-down perspective, classical-era concepts were replaced through a

---

[31] F.H. Knight, *Immutable Law in Economics: Its Reality and Limitations*, 36 AM. ECON. REV. 93, 102 (1946). As Knight observed:

> The "perfect" market, of theory at its highest level of generality, is conventionally described as perfectly or purely "competitive." But use of this word is one of our worst misfortunes of terminology. There is no presumption of psychological competition, emulation, or rivalry, and this is rather contrary to the definition of economic behavior. Market relations are impersonal, between persons and goods; and persuasion or "bargaining" is also excluded. The nature of a market is simply effective intercommunication among buyers and sellers (actual or potential) with freedom to make and to accept or decline offers to trade. The meaning of "competition" is that they are numerous and act individually; "atomistic" is a better word for the idea.

*Id. See also* F.A. Hayek, *The Meaning of Competition*, *in* INDIVIDUALISM AND ECONOMIC ORDER (1996):

> [T]his theory [of perfect competition] throughout assumes that *state of affairs* already to exist which, according to the truer view of the older theory, the *process of competition* tends to bring about (or to approximate) . . . .
>
> . . . .
>
> The modern theory of competition deals almost exclusively with a state of what is called "competitive equilibrium" in which it is assumed that the data for the different individuals are fully adjusted to each other, while the problem which requires explanation is the nature of the process by which the data are adjusted. . . . [C]ompetition is by its nature a dynamic process whose essential characteristics are *assumed away* by the assumptions underlying static analysis.

*Id.* at 92-94 (emphasis added).

This grave misunderstanding has persisted until today. For example, in one of the most widely-used introductory texts, WILLIAM J. BAUMOL & ALAN S. BLINDER, MICROECONOMICS: PRINCIPLES AND POLICY (1994), the authors state that "[p]erfect competition . . . is a benchmark of ideal performance against which other market structures can be judged." *Id.* at 276-77.

[32] "[T]he antithesis of the [neoclassical] economic conception of monopoly is not free but *pure* competition, understood to be a situation in which no seller or buyer has any control over the price of his product." Mason, *supra* note 8, at 36. "In a market from which control is completely absent every seller and buyer, acting independently, could increase or decrease his purchases or sales without appreciable effect on the price. Such markets . . . may be said to be purely competitive . . . ." *Id.* at 46. "Competitive markets are those of many sellers, no one of whom sells an appreciable fraction of the total supply of the commodity or service on the market, and no one of whom has any discretion as to the price he receives." Eugene V. Rostow, *The New Sherman Act: A Positive Instrument of Progress*, 14 U. CHI. L. REV. 567, 576 (1947). "It is under perfect competition that the market mechanism performs best. So, if we want to learn what markets do well, we can put the market's best foot forward by beginning with perfect competition." BAUMOL & BLINDER, *supra* note 31, at 223.

[33] *See supra* note 32.

[34] Mason, *supra* note 8, at 35.

process of semantic infiltration[35] with the erroneous notion that perfect competition is the only true[36] form of competition. "In mainstream economics, the term 'competition' lost its Adam Smith roots concerning rivalry. Indeed, 'competition' became so synonymous with *perfect* competition that 'perfect' was dropped in most academic discourse as redundant."[37] Because consumers are assumed to have perfectly homogenous demands, product differentiation is viewed as wasteful or harmful. Because imperfections in economic actors' information have been assumed away, advertising and marketing activities designed to provide information to consumers about heterogeneous products are viewed suspiciously as a means to create harmful market power through artificial product differentiation.[38] Above-average economic profits resulting from market power are pejoratively and suspiciously termed "abnormal" or "supracompetitive" profits or "rents," or "wealth transfers" from consumers to sellers. To the extent that microeconomic activity does not resemble this idealized touchstone and heterogeneous demand results in a downward sloping demand curve where price is greater than marginal cost, it is seen as a "monopoly" or "monopolistic" or "non-competitive" problem. "[H]ence, the term 'monopoly' morphed into 'monopoly power' and 'market power.' "[39]

Informed by their misguided conviction in this elegant new construct, prominent, self-styled neoclassical economists and lawyers deliberately set

---

35 As Senator Daniel Patrick Moynihan once put it, "In diplomacy this is known as semantic infiltration: if the other fellow can get you to use his words, he wins." MACHOVEC, *supra* note 7, at 276.

36 *See supra* note 32.

37 HUNT, *supra* note 30, at 249.

38 "The more successful this differentiation the greater the control of the market it is possible for the seller to achieve, and, consequently, the more entrenched his monopoly position. Extensive advertising expenditures may successfully differentiate in the minds of buyers the product of a given seller from those of his rivals." Mason, *supra* note 8, at 48.

39 HUNT, *supra* note 30, at 249. Harbeson provides a contemporary account of the semantic infiltration and replacement of the classical concepts of "competition" and "monopoly" with their neoclassical counterparts.

> [I]t is necessary to understand thoroughly the difference between the [neoclassical] economic and [classical] legal meanings of monopoly. In [neoclassical] economics the term monopoly means control of the market [where a product is sold at a price above marginal cost]; that is, the ability of a seller, by increasing or decreasing his output, to affect the price of the product sold. *Until recent years* monopoly was regarded as the antithesis of competition; the two were conceived to be qualitatively distinct.

*Id.* at 201 (emphasis added).

Thus, it came to be that "monopoly is identified with control of the market," through the sale of a product at a price above marginal cost. Mason, *supra* note 8, at 47. "It is easy enough to present evidence of monopoly situations, which, to [neoclassical] economics, is merely the absence of pure competition." *Id.* at 49. "[T]he crucial element of . . . monopolistic power is a degree of control over the prices they charge." Rostow, *supra* note 32, at 575. "[A] monopolist is not a *price taker* who must simply adapt to whatever price the forces of supply and demand decree. Rather, a monopolist is a *price maker* who can, if so inclined, raise the product price." BAUMOL & BLINDER, *supra* note 31, at 272.

out to have the Sherman Act revised away from its classical-era origin and transformed into a document that would uphold perfect competition as a normative welfare ideal. In his seminal 1937 *Yale Law Review* article, *Monopoly in Law and Economics*, Harvard economics professor Edward S. Mason expressly recognized that "[r]estriction on competition is the [classical] legal content of monopoly; control of the market [through sale of a product at a price above marginal cost] is its [neoclassical] economic substance. And these realities are by no means equivalent."[40] From his neoclassical viewpoint, "our modern law embraces an antiquated and inadequate conception of the monopoly problem."[41] Mason made no pretense about his ultimate aim: "if [neoclassical] economics is to be put in a position to contribute to the formulation of public policy, it must conceive the monopoly problem in a more extensive way than is at present customary."[42] He expressly recognized the Sherman Act's original, classical-era enactment and lamented as unlikely the possibility that the courts would ever shift away from their classical interpretation of it. For Mason, "[t]he weakness of our public policy is not the result of judicial interpretation, but of inadequacy of legislation. It can only be corrected by legislation which will *re-define* the monopoly and trade practice problem in the shaping of a more satisfactory public policy," consistent with the neoclassical paradigm.[43]

Rutgers economics professor Robert W. Harbeson also recognized in his 1940 *Michigan Law Review* article *The Present Status of the Sherman Act* that "[t]he Court's tolerance of market control [where a product is sold at a price above marginal cost] . . . is a result of the [classical] legal definition of monopoly, and that this in turn is consistent with the conception of the monopoly problem held by the framers of the Sherman Act."[44] Like Mason, Harbeson expressly recognized that, "[f]rom the [neoclassical] economic standpoint the [classical] legal conception of monopoly appears grossly inadequate."[45] This is so because "[t]he [neoclassical] problem of industrial control is thus *very different* and far more difficult than it was conceived to be by the framers of the Sherman Act."[46] Specifically, "[i]n their view, monopoly outside the public utility field was abnormal and exceptional, and the public interested could be

---

40 Mason, *supra* note 8, at 36. "The trend has led to a split between the approach to the monopoly problems in the [classical] law and [neoclassical] economics which requires bridging by interpretative work of a high order." *Id.* at 35.

41 *Id.* at 46.

42 *Id.* at 49.

43 *Id.* at 46 (emphasis added).

44 Harbeson, *supra* note 19, at 210. Subsequently, Harbeson went on to serve as Principal Economist for the Interstate Commerce Commission and Professor of Economics at the University of Illinois.

45 *Id.* at 204.

46 *Id.* at 212 (emphasis added).

adequately protected by maintaining freedom to compete by means of an anti-trust law."[47] So, like Mason, Harbeson expressly called for the Sherman Act to be amended to incorporate neoclassical concepts. For him, "it is clear that the Sherman Act must be revised in such a way as to give monopoly a [neoclassical] legal meaning in terms of control of the market [where a product is sold at a price above marginal cost] rather than in terms of [classical] restriction on freedom to compete."[48] The result would be that "the protection of the public interest in matters of prices, costs and profits requires some type of supervision and control over virtually the entire price system."[49]

Only ten years after Mason's article, though, as the influence of the neoclassical paradigm spread the Sherman Act's interpretation underwent a "remarkable" change, as Yale law professor Eugene V. Rostow recognized in his June 1947 *University of Chicago Law Review* article, *The New Sherman Act: A Positive Instrument of Progress*.[50] "Ten years ago Professor Edward S. Mason quite accurately concluded that the conceptions of monopoly in [classical] law and [neoclassical] economics were different, and were growing more different."[51] But "[w]ith revolutionary speed . . . the doctrine of the Sherman Act has lately been transformed."[52] No new legislation had been needed. The democratic process had not been enlisted to pass this "new Sherman Act." Instead, the Act's transformation into a "new" document had been accomplished through judicial decree. This "*new judicial view*"[53] in which "our antitrust doctrine has been reoriented"[54] followed from a 1945 decision in *U.S. v. Alcoa*[55] and a 1946 decision in

---

47 *Id.*

48 *Id.* at 210.

49 *Id.* at 203.

50 Rostow, *supra* note 32, at 575.

51 *Id.* at 574 (citing to Mason, *supra* note 8).

52 Rostow, *supra* note 32, at 574.

53 *Id.* at 577 (emphasis added).

54 *Id.* at 589.

55 *See* United States v. Aluminum Co. of Am., 148 F.2d 416, 427 (2d Cir. 1945) (Hand, J.). There, Judge Learned Hand held that a principal purpose of the Sherman Act is to preserve a state of atomized, homogenous resources, even if the result is less efficient than a more heterogeneous arrangement. "Throughout the history of these statutes it has been constantly assumed that one of their purposes was to perpetuate and preserve, for its own sake and in spite of possible cost, an organization of industry in small units that can effectively compete with each other." *Id.* at 429. Thus, under Section 2 of the Sherman Act, Judge Hand condemned Alcoa's acquisition of heterogenous business resources and the competitive advantage that resulted therefrom:

> It was not inevitable that it should always anticipate increases in the demand for ingot and be prepared to supply them. Nothing compelled it to keep doubling and doubling its capacity before others entered the field. It insists that it never excluded competitors; but we can think of no more effective exclusion than progressively to embrace each new opportunity as it opened, and to face every newcomer with capacity already geared into a great organization, having the advantage of experience, trade connections and the elite of personnel.

*American Tobacco Co. v. U.S.*[56] "In the *Aluminum* case Judge Hand finally interred and reversed the old dictum that size is not an offense under the Sherman Act. Size, he concluded, was not only evidence of violation, or a potential offense . . . it was the essence of the offense."[57] The Supreme Court confirmed this result in *American Tobacco*. There, "[t]he Supreme Court [held that] the power to exclude actual or potential competition, not the actual exclusion of competitors, is a hallmark of the offense prohibited by Section 2."[58]

Triumphantly, Rostow observed that "[m]arket control [through the ability to sell a product at a price above marginal cost] is now a far more important theme in Sherman Act cases than handicaps upon an individual's power to do business."[59] Thus, "[t]he old pre-occupation of the judges with evidence of business tactics they regarded as ruthless, predatory, and immoral has all but disappeared . . . . We are close to the point of regarding as illegal the kind of economic power which the [neoclassical] economist regards as monopolistic."[60] As a result, "[r]ecent decisions have given the Department of Justice its greatest opportunity since the act was passed to seek the enforcement of the law *on a grand scale,* and in ways which might produce *not piddling changes* in the detail of trade practice, *but long strides* towards the great social purposes of the statute."[61] In Rostow's technoractic view, these great social purposes were the "reorganization"[62] of society through the forced "dispersal of economic power and opportunity . . ." by judges.[63] Therefore, it was now a "crime" for a seller to be anything other than a trembling, palsied atom.[64] The appropriate "punishment"[65] for the offense of being anything other than inconsequential was to be reduced down to size in the name of "social virtue."[66]

Simultaneously, Harbeson echoed Rostow's observations in his own June 1947 *Michigan Law Review* article, *A New Phase of the Antitrust Law.* Like Rostow, Harbeson emphasized the dramatic differences between the Supreme Court's *U.S. Steel* and *International Harvester* precedents and the subsequent decisions in *Alcoa* and *American Tobacco*. "The importance of

---

*Id.* at 431-32. *Compare* with United States v. U.S. Steel Corp., 251 U.S. 417, 451 (1920); United States v. Int'l Harvester Co., 274 U.S. 693, 708 (1927); Mason, *supra* note 8, at 39.

56 *See* Am. Tobacco Co. v. United States, 328 U.S. 781, 809 (1946) (affirming convictions for conspiracy in restraint of trade, monopolization, and attempt to monopolize).

57 Rostow, *supra* note 32, at 578.

58 *Id.* at 583.

59 *Id.* at 575.

60 *Id.*

61 *Id.* at 574 (emphasis added).

62 *Id.* at 586.

63 *Id.* at 573.

64 *Id.* at 589.

65 *Id.*

66 *Id.* at 589, 573.

these two decisions lies . . . in the fact that they represent a significant step by the federal courts toward making market control [where a product is sold at a price above marginal cost] the test of monopoly *and thus toward assimilating the [classical] legal concept of monopoly to the [neoclassical] economic.*"[67] Harbeson expressly recognized that a process of semantic infiltration and replacement of the classical concepts of "competition" and "monopoly" with their neoclassical counterparts was taking place. "Although the decisions are not clear-cut on all points and leave some unanswered questions, taken together they represent *a very considerable departure* from the position announced in the *Steel* and *Harvester* cases."[68] In the *Tobacco* case, "the Court not only adopted the [neoclassical] economic test of monopoly but also made it possible effectively to apply the Sherman Act in the case of large-scale enterprises exercising varying degrees of market control *short of compete monopoly.*"[69]

Although he applauded the results of these cases as a general matter of policy, Harbeson openly recognized that this "new phase" of Sherman Act cases directly violated the original, classical-era meaning of the Sherman Act, as it was enacted in 1890. "Until recent years monopoly was regarded as the antithesis of competition. The two were regarded as qualitatively separate and distinct . . . . *This is the concept of monopoly implicit in the Sherman Act.*"[70] Originally, "[t]he framers of this legislation regarded public utilities as 'natural monopolies' while the remainder of industry, with rare exceptions, was regarded as essentially 'competitive.' "[71] Whereas, "by contrast, modern [neoclassical] economic theory regards monopoly as being a matter of degree, depending upon the numbers and buyers and sellers of a commodity and the availability of adequate substitutes."[72] As a result, in the neoclassical view there exists a "ubiquity of monopolistic elements in the industrial organization."[73] So, Harbeson openly recognized that the Sherman Act is a classical-era document that cannot be legitimately interpreted using neoclassical concepts. But, wanting to complete the replacement of classical-era legal and economic concepts with neoclassical ones, Harbeson once again renewed his call "for a revision of the Sherman Act which would *redefine the monopoly and trade practice problem*" consistent with the neoclassical paradigm.[74]

---

67 Harbeson, *supra* note 24, at 980 (emphasis added).

68 *Id.* (emphasis added).

69 *Id.* at 997 (emphasis added).

70 *Id.* (emphasis added).

71 *Id.* at 997-98.

72 *Id.* at 998.

73 *Id.* at 1000.

74 *Id.* (emphasis added).

In a follow-up 1949 *Illinois Law Review* article, *Monopoly Under the Sherman Act: Power or Purpose?*,[75] Rostow examined the Supreme Court's trio of 1948 monopolization decisions in *U.S. v. Griffith*,[76] *Schine Chain Theaters v. U.S.*,[77] and *U.S. v. Paramount Pictures*.[78] Rostow observed that "[t]he recent cases have *radically altered* the scope of the idea of monopoly—or, more precisely, of 'monopolizing,'. . . ."[79] As a result, "[t]he *legal redefinition* of monopoly" had occurred.[80] "Like much of our present day law, the newer Sherman Act cases seem to reject the basic attitudes which prevailed during the 'twenties, and derive in considerable part from the philosophy of an older generation."[81] So, "[w]hatever the subconscious source, the fact remains that in interpreting the Sherman Act, the Supreme Court seems to have discarded one set of its ancestors in favor of another."[82] As a result, "[i]t is not too much to say that the current revision of Section 2 presents the central issue of doctrine in the entire field of anti-trust law—the concept of market control [through the ability to sell a product at a price above marginal cost], which has increasingly become the chief, and often the only issue in anti-trust litigation."[83] Indeed, "it is indisputable after these cases that the existence of what the court will classify as monopoly power, coupled with a perfunctory and implied intent to use it, is illegal without reference to the techniques by which it was obtained."[84] Thus, the Court had engaged in a generalized re-orientation of its entire microeconomic and antitrust jurisprudence.

Importantly, "[t]hese were not the idle musings of an armchair academic. Rostow [as the new Dean of the Yale Law School] led the move in 1955 to redirect the Yale Law School curriculum. Armed with a sizeable grant from the Ford Foundation ($8,000,000 in 1995 purchasing power), Rostow and his departmental colleagues fashioned a legal program that was 'unique in the world.' "[85] Its "objective was to graduate lawyers 'whose

---

75 Eugene V. Rostow, *Monopoly Under the Sherman Act: Power or Purpose?*, 43 ILL. L. REV. 745 (1949).

76 United States v. Griffith, 334 U.S. 100 (1948) (reversing district court finding that defendants had not restrained trade, monopolized, or attempted to monopolize).

77 Schine Chain Theaters v. United States, 334 U.S. 110 (1948) (affirming in part, reversing in part, and remanding judgment for plaintiff alleging defendants had restrained trade, monopolized, and attempted to monopolize).

78 United States v. Paramount Pictures, 334 U.S. 131 (1948) (affirming in part, reversing in part, and remanding judgment denying leave to intervene but granting injunction and other relief related to suit alleging restraint of trade, monopolization, and attempt to monopolize).

79 Rostow, *supra* note 75, at 745 (emphasis added).

80 *Id.* (emphasis added).

81 *Id.* at 746.

82 *Id.*

83 *Id.* at 745-46.

84 *Id.* at 776.

85 MACHOVEC, *supra* note 7, at 197-98 (quoting CURRENT BIOGRAPHY YEARBOOK 395 (C. Moritz, ed., 1961)).

command of law is rooted in a sure knowledge of the historic . . . and economic sources and purposes of law.' "[86] Over the next generation, "Dean Rostow's influence on ambitious antitrust was immense."[87]

During the 1950s, the publication of Joe Bain's *Barriers to New Competition*[88] and *Industrial Organization*[89] and Carl Kaysen and Donald Turner's *Antitrust Policy*[90] further crystallized the framework for what came to be known as the "Structure-Conduct-Performance" ("SCP") approach to industrial organization. Using perfect competition as a welfare ideal, in this approach the structure of a market is the principal determinant of its sellers' conduct and, thus, is also the principal determinant of that market's performance. As such, generally, "big is bad; small is good" as far as sellers and market structure are concerned.[91] But, because the real world rarely resembles the perfectly competitive vision of atomistic price taking sellers, virtually every seller is considered to be a worrisome "monopolist" possessing dangerous market power simply because it faces a downward-sloping demand curve that naturally results from consumer demand being heterogeneous, and not perfectly homogenous. That is, virtually every seller is deemed to be a problematic monopolist simply because consumers have demands that are more varied than the puritanical homogeneity assumed by perfect competition and sellers actually dare to meet that demand at a corresponding price that is greater than marginal cost. "Barriers" to entry are no longer limited to classical-era restrictions on entry, and can result simply from the large capital resource requirements inherent in certain businesses.[92] As a result, narrowly defined true

---

86 *Id.*

87 Frederick M. Rowe, *The Decline of Antitrust and the Delusions of Models: The Faustian Pact of Law and Economics*, 72 GEO. L. J. 1511, 1522 n.64 (1984). "From 1953 to 1973, three of his former Yale law students led the [Department of Justice] Antitrust Division (Robert A. Bicks, Donald F. Turner, Richard McLaren)." *Id.* In addition, "Rostow was an intellectual dynamo of the REPORT OF THE ATTORNEY GENERAL'S NATIONAL COMMITTEE TO STUDY THE ANTITRUST LAWS (1955), to which Bicks and Turner contributed much. In turn, professor Turner's lasting imprint stems from his many influential law-review writings, as well as C. KAYSEN & D. TURNER, ANTITRUST POLICY (1959) and P. AREEDA & D. TURNER, ANTITRUST LAW (5 vols. 1978-1980)." *Id.*

88 JOE S. BAIN, BARRIERS TO NEW COMPETITION: THEIR CHARACTER AND CONSEQUENCES IN MANUFACTURING INDUSTRIES (1956).

89 JOE S. BAIN, INDUSTRIAL ORGANIZATION (1959).

90 KAYSEN & TURNER, *supra* note 87 (1959).

91 *See generally* HUNT, *supra* note 30, at 250.

92 Mason, *supra* note 8, at 47-48. Hovenkamp explains that:

> For example, Harvard economist Joe S. Bain, who exercised a strong influence on federal antitrust policy in the 1960s and 1970s, based his relatively prointerventionist theories on three important economic premises. The first was that economies of scale were not substantial in most markets and dictated truly anticompetitive concentration levels in only a small number of industries. As a result, many industries contained larger firms and were more concentrated than necessary to achieve optimal productive efficiency. The second was that barriers to entry by new firms were very large and could easily be manipulated by dominant firms. The third was that the noncompetitive performance (pricing above marginal cost) associated with oligopoly began to occur at relatively low concentration levels.

"competition" exists almost nowhere, and some type of "monopoly" problem or imperfection exists virtually everywhere.[93]

For SCP, then, the primary goal of antitrust was to prevent sellers from acquiring and exercising market power[94] and, thereby, extracting so-called pernicious "wealth transfers" from consumers.[95] The Supreme Court further articulated this anti-market power viewpoint in its1956 *U.S. v. Du Pont* decision.[96] In doing so, the Court cited approvingly to Rostow's 1949 article and his treatment of the *Alcoa* and *American Tobacco* decisions, including his observation that the Court had "*redefined monopoly power*" in the latter.[97] Thus, the Court openly acknowledged it had re-defined the Sherman Act's terms without the legislative process ever having been engaged. Under this re-defined version of the Act, "[m]onopoly power is the power to control prices or exclude competition . . . . [W]hen an alleged monopolist has power over price and competition, an intention to monopolize . . . may be assumed."[98] Instead of first examining a business's actual conduct, the mere possession of market power was now deemed to be the legal equivalent of exclusionary behavior. That is, the court reversed the classical-era approach to monopolization by pre-supposing an equation of market power to exclusionary conduct. Consistent with this viewpoint, SCP antitrust policy also focused on scrutinizing the conduct of "large" sellers and preventing the "dangerous" structural concentration of

---

Herbert Hovenkamp, *Antitrust Policy After Chicago*, 84 MICH. L. REV. 213, 219-20 (1985).

[93] Mason made this point expressly:

> Some of the consequences of this [neoclassical] trend have been . . . a recognition of monopoly elements in the practices of almost every firm, a recognition of the impossibility of using the fact of [classical] monopoly as a test for public policy, and a growing awareness of the necessity of making distinctions between market situations *all of which have monopoly elements*.

Mason, *supra* note 8, at 35 (emphasis added).

That is, "[s]ince, outside the sphere of agricultural and a few other products, almost every seller is in this position, it is easy to see that if monopoly is identified with control of the market [through the sale of a product at a price above marginal cost], *monopolistic elements are practically omnipresent*." *Id.* at 37. Thus, "nothing is more 'normal' than monopolistic market conditions . . . ." *Id.* at 44. "Markets . . . which may be said to be purely competitive in the sense of being completely devoid of any element of control over price, are comparatively rare." *Id.* at 46. "All markets, practically speaking, exhibit a fusion of monopoly and competitive elements." *Id.* at 47.

[94] *See* Oliver E. Williamson, *Economics and Antitrust Enforcement: Transition Years* 17 ANTITRUST 61 (2003).

[95] *See* Robert H. Lande, *Chicago's False Foundation: Wealth Transfers (Not Just Efficiency) Should Guide Antitrust*, 58 ANTITRUST L. J. 631 (1989); HUNT, *supra* note 30, at 250.

[96] United States v. E.I. DuPont de Nemours and Co., 351 U.S. 377, 392 n.19 (1956).

[97] *Id.* (citing to "Rostow, 43 Ill. L. Rev. 745, 753-763"). *See also* Rostow, *supra* note 75, at 763 (emphasis added).

[98] 351 U.S. 377, 392 (affirming district court finding of no unlawful monopolization by defendant).

industries[99] through the prohibition of mergers between horizontal competitors, even when they resulted in only slight increases in a market's concentration. For example, in its 1962 *Brown Shoe Co. v. U.S.* decision, the Supreme Court interpreted the Clayton Act's purpose as to "preserve" industry "structure[s]" having "numerous independent units." [100] Often, this "big is bad; small is good" view of microeconomics continued to be intermixed with "other" social and political goals, such as the dispersal of economic power.[101]

In its 1966 *U.S. v. Grinnell* decision, however, the Supreme Court retreated somewhat from the position of the *American Tobacco* line of cases, that market power equals exclusionary conduct.[102] In *Grinnell*, the Court re-defined the offense of monopolization as having two elements: "(1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business, acumen, or historic accident."[103] In this formulation, the Court formally distinguished between the mere state of possessing market power and a competitor's conscious conduct. The Court, however, did not elaborate on this broad formula and, at the same time, contradictorily cited *Du Pont*'s holding that market power equals exclusionary conduct and may be inferred merely from a large market share.[104]

In the 1970s and early 1980s, SCP was supplanted as the leading theory of industrial organization by the so-called "Chicago" school of antitrust associated with members of the law, economics, and business

---

99 HUNT, *supra* note 30, at 250. Typically, SCP theorists viewed a four-firm concentration ratio of greater than 75% as "dangerously concentrated," where government should prohibit horizontal mergers between competitors and consider corporate dissolutions. *Id.* An industry with a four-firm concentration ratio of less than 50% was considered to be "close enough" to the model of perfect competition "and, thus, no cause for concern." *Id.* In between, "[a] ratio between 50% and 75% was 'worrisome' and firms in the industry should be watched carefully." *Id.*

100 *See* Brown Shoe Co. v. United States, 370 U.S. 294, 333 (1962) (horizontal aspects of merger found to violate Section 7 of the Clayton Act because combination exceeded 5% of relevant market and vertical aspects of merger found to also constitute a violation because it would foreclose .0948% of relevant market). *See also* United States v. Von's Grocery Co., 384 U.S. 270 (1966) (merger found to violate Section 7 of the Clayton Act where horizontal aspects of combination amounted to 7.9% of relevant market); United States v. Philadelphia Nat'l Bank, 374 U.S. 321, 363 (1963) ("[A] merger which produces a firm controlling an undue percentage share of the relevant market, and results in a significant increase in the concentration of firms in that market is so inherently likely to lessen competition substantially that it must be enjoined in the absence of evidence clearly showing that the merger is not likely to have such anticompetitive effects.").

101 *See, e.g., Brown Shoe*, 370 U.S. at 316. *See also* Robert Pitofsky, *The Political Content of Antitrust*, U. PA. L. REV. 1051 (1979).

102 United States v. Grinnell 384 U.S. 563 (affirming in part and remanding district court decision that defendant violated Section 2 of the Sherman Act).

103 *Id.* at 571.

104 *Id.*

faculty at the University of Chicago. The principal foundational works of this approach include Judge Robert Bork's *The Antitrust Paradox*,[105] Judge Richard Posner's *Antitrust Law*[106] and *Economic Analysis of Law*,[107] and Yale Brozen's *Concentration, Mergers, and Public Policy*.[108] For Chicago the "exclusive goal" of antitrust is the maximization of consumer welfare, a position the Supreme Court expressly adopted for the first time in its 1979 *Reiter v. Sonotone Corp.* decision, citing to Bork.[109] The antitrust laws are exclusively economic statutes that leave no room for non-economic social and political considerations.[110] Consumer welfare is to be analyzed through the use of rigorous economic analysis perceived to be largely lacking in then-existing antitrust precedent.[111]

A majority strand of Chicago analysis retained SCP's use of perfectly competitive equilibrium as a guide to reasoning about consumer welfare.[112] In this view, "the whole task of antitrust can be summed up as the effort to improve allocative efficiency without impairing productive efficiency so greatly as to produce either no gain or a net loss in consumer welfare."[113] SCP and majority Chicago, then, are at bottom conceptual cousins. Although majority Chicago recognized that real world markets deviate from the ideal of perfect competition, its "basic notion is that the world is close enough to perfectly competitive equilibrium that the model serves as a framework for understanding real-world phenomena."[114] Although majority

---

105 ROBERT H. BORK, THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF (1978/1993).

106 RICHARD A. POSNER, ANTITRUST LAW (1979).

107 RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW (1986).

108 YALE BROZEN, CONCENTRATION, MERGERS, AND PUBLIC POLICY (1982).

109 Reiter v. Sonotone Corp., 442 U.S. 330, 343 (1979) (holding a consumer whose money has been diminished by a manufacturer's antitrust violation has been injured in his property within the meaning of the Clayton Act). "[The Sherman Act] floor debates . . . suggest that Congress designed the Sherman Act as a 'consumer welfare prescription.' " *Id.* (citing BORK, *supra* note 105, at 66). *See also* United States v. General Dynamics Corp., 415 U.S. 486 (1974) (rejecting argument that statistical evidence of increased concentration necessarily indicates a reduction in competition); Continental T.V. v. GTE Sylvania, 433 U.S. 36 (1977) (vertical agreements should be judged under rule-of-reason analysis); Copperweld Corp. v. Independence Tube Corp., 467 U.S. 752 (1984) (describing the benefits of economic efficiency to consumers); and Matsushita Electric Indus. Co. v. Zenith Radio, 475 U.S. 574 (1986) (treating predatory pricing claim with skepticism). *But see* Aspen Skiing Co. v. Aspen Highlands Corp., 472 U.S. 585 (1985) (finding lack of any efficiency "business justification" for defendant's unilateral conduct toward competitor).

110 "The basic tenet of the Chicago school [is] that problems of competition and monopoly should be analyzed using the tools of general economic theory rather than those of traditional industrial organization . . . ." Richard A. Posner, *The Chicago School of Antitrust Analysis*, 127 U. PA. L. REV. 933-34 (1979).

111 BORK, *supra* note 105, at xi (1993 version).

112 Jerome Ellig, *Untwisting the Strands of Chicago Antitrust*, 37 ANTITRUST BULLETIN 864 (1992).

113 BORK, *supra* note 105, at 91.

114 Ellig, *supra* note 112, at 865, 869-70. As Ellig points out:

Chicago did not develop a definitive definition of a "barrier" to entry, the fact that certain resources may be required to enter and compete in a particular economic space does not necessarily mean those resources constitute such a barrier.[115] Likewise, neither product differentiation designed to meet consumers' varied demands, nor advertising and marketing practices designed to reach imperfectly informed consumers are inherently suspect.[116] But because the real world looks substantially like perfect competition and "monopoly" is typically transitory, antitrust policy should generally restrict itself to policing cartel activity, such as price-fixing; mergers to very large market shares that plainly do not resemble perfect competition; and narrowly defined predatory activity.[117]

However, a minority strand of Chicago theorists viewed competition as a rivalrous process, consistent with the views of classical, Schumpeterian,[118] and "Austrian"[119] economists such as Ludwig Von Mises and F.A. Hayek.[120] Recognizing that imperfect information in the real world prevents the realization of a perfectly competitive equilibrium, this minority strand relaxes perfect competition's extreme assumptions about

---

> The Chicago approach to industrial organization represents an amalgam of traditional [neoclassical] price theory, which takes perfect competition as a normative ideal, and market rivalry theories.
>
> . . . .
>
> The pull of real world theory is just too strong to prevent Chicago theorists from ignoring the problems of imperfect information that generates rivalrous competition. At the same time, the lure of rigorous theory is too strong to let them abandon the concept of perfectly competitive equilibrium as a fundamental theoretical tool for analyzing the real world.

*Id.*

Judge Bork readily acknowledges this point:

> The economic model of perfect competition was never intended as a policy prescription, and it is a basic, though extremely common error to suppose that markets do not work efficiently if they depart from the model. . . . *[A]ntitrust must use the model and its implications as a guide to reasoning about actual markets*, but the pure model must never be mistaken for that "competition" we wish to preserve.

BORK, *supra* note 105, at 60 (emphasis added).

115 BORK, *supra* note 105, at 310-11 ("We may begin by asking what a 'barrier to entry' is. There appears to be no precise definition . . . .").

116 *Id.*

117 *Id.* at 405-06; Ellig, *supra* note 112, at 873.

118 In the view of Joseph Schumpeter:

> [I]n capitalist reality as distinguished from the textbook picture, it is not that kind of competition which counts [producing known products within known constraints] but the competition from the new commodity, the new technology, the new source of supply, the new type of organization . . . competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of existing firms but at their foundations and their very lives.

JOSEPH SCHUMPETER, CAPITALISM, SOCIALISM, AND DEMOCRACY 84 (1942).

119 *See* F.A. HAYEK, INDIVIDUALISM AND ECONOMIC ORDER 101 (1948); Ludwig Von Mises, *Profit and Loss*, *in* PLANNING FOR FREEDOM AND SIXTEEN OTHER ESSAYS AND ADDRESSES 108 (4th ed., 1995); ISRAEL M. KIRZNER, DISCOVERY AND THE CAPITALIST PROCESS (1985) and COMPETITION AND ENTREPRENEURSHIP (1973).

120 Ellig, *supra* note 112, at 864-65, 870-71.

economic actors' knowledge and, thus, views competition as a rivalrous information discovery process.[121] In other words, this perspective views competition as a process of overcoming ignorance through the entrepreneurial acquisition, interpretation, and application of information. Faced with these two contradictory views of competition, some theorists have attempted to reconcile their differences by postulating a tradeoff between perfect competition's "allocative" efficiency and the "productive" or "innovative" efficiency that results from entrepreneurial activity or the formation of a new company through a merger or acquisition.[122]

Into the 1980s, "both [the SCP and majority Chicago] sides agreed that the neoclassical research tradition is the appropriate starting point for analysis."[123] However, in the 1980s something of a counter-revolution was launched against the Chicago school.[124] The counter-revolutionaries generally conceded that SCP's "big is bad; small is good" view had been discredited.[125] But they challenged majority Chicago's single-minded focus on efficiency and re-emphasized the equilibrium argument that seller market power amounts to an anticompetitive wealth transfer from

---

121 *Id.* at 865-69.

122 *Id.* at 871-72. *See* Oliver Williamson, *Economies as an Antitrust Defense: The Welfare Tradeoffs*, 58 AM. ECON. REV. 18 (1968); BORK, *supra* note 105, at 107; HUNT, *supra* note 30, at 247-48.

123 HUNT, *supra* note 30, at 253. That is:

> The striking aspect of the antitrust debate is not the points of disagreement but the commonalities between advocates of the [SCP] wealth-transfer and [majority Chicago] efficiency views. Both sides agree that partial equilibrium analysis is the appropriate method of analysis; neither side argues from an evolutionary economics or Austrian economics perspective. Both sides agree "competition" means *perfect* competition; neither side argues for an evolutionary, process view. Both sides agree that "monopoly" means simply a downward-sloping demand curve; neither side argues that "monopoly" should be restricted to the control of the supply of some generic commodity. Both sides agree that market power (downward-sloping demand curves) results in a misallocation of resources and constitutes a major problem for society; neither side argues that downward-sloping demand curves are natural and contribute to the efficient allocation of resources.

*Id. See also* BAUMOL & BLINDER, *supra* note 31, at 276-77.

124 *See* Robert H. Lande, *Wealth Transfers as the Original and Primary Concern of Antitrust: The Efficiency Interpretation Challenged*, 34 HASTINGS L. J. 65-85 (1982); Herbert Hovenkamp, *Antitrust After Chicago*, 84 MICH. L. REV. 213-45 (1985); Eleanor M. Fox and Lawrence A. Sullivan, *Antitrust Retrospective and Perspective: Where Are We Coming From? Where are We Going?*, 62 N.Y.U. L. REV. 936-88 (1987); Robert H. Lande, *The Rise and (Coming) Fall of Efficiency as the Ruler of Antitrust*, 33 ANTITRUST BULLETIN 429-65 (1988); Robert H. Lande, *Chicago's False Foundation: Wealth Transfers (Not Just Efficiency) Should Guide Antitrust*, 58 ANTITRUST L. J. 631 (1989); Robert H. Lande, *Chicago Takes It on the Chin: Imperfect Information Could Play a Crucial Role in the Post-*Kodak *World*, 33 ANTITRUST L. J. 193-201 (1993); Oliver E. Williamson, *Economics and Antitrust Enforcement: Transition Years*, 17 ANTITRUST 61 (2003).

125 Lande (1989), *supra* note 124, at 631 ("[M]arket power leads to a major economic effect in addition to allocative inefficiency: a transfer of wealth from consumers to the firm or firms with market power. . . . The 'big is bad/small is good' school of antitrust has been thoroughly defeated, and I will not attempt to defend or resurrect it.").

consumers to sellers.[126] Yet, at the same time, some of these critics expressed disillusionment with static neoclassical price theory.[127] In addition, some critics introduced new game-theoretic approaches to competition which postulated that market "imperfections" may allow sellers to engage in strategic anticompetitive behavior over time.[128] Some observers questioned whether the Supreme Court deviated from the majority Chicago approach in its 1992 *Eastman Kodak Co. v. Image Technical Serv., Inc.* decision,[129] and whether it reversed course again the following year in *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*[130] By the mid-1990s, some commentators viewed the competition debate as having moved into a loosely defined "post-Chicago" era. But these post-Chicago theories have had little impact in the courts, due to the difficulty of developing operational rules based on game-theoretic approaches.

At about the same time, "multiple and eclectic groups of scholars in economics and business strategy [began] developing theories of competition that place innovation and change at the heart of the competitive

---

126 *Id.*

127 Hovenkamp, *supra* note 124, at 256-57.

> The neoclassical price theory model is static. This means that it measures the effects of certain practices on price or output given that the market is unaffected by external events. Unfortunately, antitrust policy must deal with real world markets, and real world markets are always affected by a complex array of external influences. Application of a static model to a real world market often causes a court to ignore the obvious.

*Id.*

128 *See generally* Michael S. Jacobs, *The New Sophistication in Antitrust*, 79 MINN. L. REV. 1, 37-38 (1994) ("post-Chicago scholars have applied insights from game theory to develop models of strategic behavior describing how small and moderate-sized firms can use market imperfections to disadvantage competitors. . . . In general, post-Chicagoans emphasize the capacity of market impefections to create market power, even for firms with small market shares."); Jonathan B. Baker, *Recent Developments in Economics That Challenge Chicago School Views*, 58 ANTITRUST L. J. 645 (1989) (describing theories of: (1) vertical foreclosure that raises rivals' costs; (2) rational price predation; (3) collusion with occasional price wars; (4) a rehabilitated version of conglomerate forbearance; (5) a rehabilitated version of scale economies as barriers to entry; and (6) new empirical approaches to industrial organization); Herbert Hovenkamp, *Post-Chicago Antitrust: A Review and Critique*, 2001 COLUM. BUS. L. REV. 257 (2001); Malcolm B. Coat & Jeffrey H. Fischer, *Can Post-Chicago Economics Survive Daubert?*, 34 AKRON L. REV. 795 (2001).

129 Eastman Kodak Co. v. Image Technical Services, Inc., 504 U.S. 451 (1992) (holding that respondents presented genuine issues for trial as to whether Kodak monopolized or attempted to monopolize the service and parts markets for its equipment). "[T]he Supreme Court decision . . . reflects a deep-seated suspicion about dominant firms in differentiated markets—largely the consequence of a structural approach." E. THOMAS SULLIVAN & HERBERT HOVENKAMP, ANTITRUST LAW, POLICY AND PROCEDURE: CASES, MATERIALS, PROBLEMS 65 (5th ed. 2003).

130 Brooke Group Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209 (1993) (holding that defendant's alleged below-cost pricing did not result in competitive injury). "By contrast, the Supreme Court's decision only a year later . . . reflects the Chicago view that even highly concentrated markets are likely to reflect substantial amounts of competition." SULLIVAN & HOVENKAMP, *supra* note 129, at 61.

process."[131] Principal strands of these research programs include Schumpeterian,[132] evolutionary,[133] Austrian,[134] and resource-based[135] perspectives. These research programs are still a work in progress. To date, the most complete theoretical statement drawing on these research programs is the "Resource-Advantage" ("R-A") paradigm articulated by Shelby D. Hunt in his work, *A General Theory of Competition, Resources, Competences, Productivity, Economic Growth.*[136]

In the R-A perspective, competition "is the disequilibrating process that consists of the constant struggle among firms for comparative advantages in resources that will yield marketplace positions of competitive advantage for some market segment(s) and, thereby, superior financial performance."[137] Thus, R-A is a corollary refinement of the classical rivalrous, evolutionary process view of microeconomic competition, in that it identifies a seller's resources as a source of its comparative competitive advantage. In this R-A view, seller firms are combiners of resources that are naturally heterogeneous and imperfectly mobile.[138] Hence, resources are

---

131 DYNAMIC COMPETITION AND PUBLIC POLICY 5 (Jerry Ellig ed., 2001). *See* Gary Hamel and C.K. Prahalad, *Strategic Intent*, HARV. BUS. REV., May-Jun. 1989, at 67; Bruce D. Henderson, *The Origin of Strategy*, HARV. BUS. REV., Nov.-Dec. 1989, at 139; C.K. Prahalad and Gary Hamel, *The Core Competence of the Corporation*, HARV. BUS. REV., May-Jun. 1990, at 79; George Stalk et al., *Competing on Capabilities: The New Rules of Corporate Strategy*, HARV. BUS. REV., Mar.-Apr. 1992, at 57; David J. Collins and Cynthia A. Montgomery, *Competing on Resources: Strategy in the 1990s*, HARV. BUS. REV., Jul.-Aug. 1995, at 118.

132 In the Schumpeterian process perspective: "Firms compete not on the margins of price and output, but by offering new products, new technologies, new sources of supply, and new forms of organization." DYNAMIC COMPETITION AND PUBLIC POLICY, *supra* note 131, at 6. Therefore, "[p]osession of market power is consistent with vigorous competition . . . ." *Id.*

133 In the evolutionary process perspective: "Firms develop different 'routines' for doing things, and competition among firms selects for survival the bundles of routines that best allow the firm to grow and prosper in its environment." *Id.*

134 In the Austrian process perspective:

> The future is unknowable, information about production methods and consumer desires is seriously incomplete, and much economically relevant knowledge is highly specific and difficult to communicate. In this kind of world, competition is a process by which firms discover new resources and better ways of satisfying consumers. Especially alert innovators may acquire market power, but the resulting profits are a reward for discovering things that would otherwise go unnoticed.

*Id.*

135 In the resource-based perspective:

> Firms compete by assembling heterogeneous combinations of resources to meet consumer desires. Key determinants of a firm's competitiveness are its capabilities to the transform resources into valuable outputs. Capabilities that are rare and difficult to imitate lead to superior profitability. Empirical research suggests that firms' unique capabilities, rather than market power, account for most of the supranormal profits that firms earn.

*Id.*

136 HUNT, *supra* note 30.

137 *Id.* at 12.

138 *Id.* at 10.

distributed asymmetrically among rival seller firms.[139] This heterogeneity results in a diversity of firm sizes, scopes, levels of profitability and unique firm "core competences" and "capabilities,"[140] not just across industries, but also within industries.[141]

Consumer demand is viewed as being significantly heterogeneous.[142] That is, consumers view certain products to be more desirable than others. Variety is the "spice of life," so to speak. Downward sloping consumer demand curves naturally follow from such diverse demand. As a result, sellers within the same industry typically provide an array of different product offerings to satisfy these varying consumer preferences.[143] Therefore, relevant product markets and industries should be defined broadly to recognize that competition occurs not only on the price variable, but also on a variety of other qualitative, non-price dimensions. Drawing on differential advantage theory, heterogeneous sellers seek out unique marketplace positions where they have a comparative advantage in meeting heterogeneous consumer desires, for which they earn a superior financial return.[144]

Drawing on evolutionary economics, competition is viewed as a process of rivalrous selection, a struggle, which produces the Schumpeterian innovations that increase productivity and economic growth.[145] Drawing on Austrian economics, R-A views competition as also being a process of entrepreneurial knowledge-discovery in a world where information is imperfect, dispersed, often tacit, and, therefore, naturally asymmetric.[146] In addition, consistent with classical economic views, R-A recognizes that societal institutions influence the functioning of the market process.[147] The result is that "the process of . . . resource-advantage competition . . . produces gains in static efficiency and increases in dynamic efficiency, that is, economic growth."[148]

---

139 *Id.* at 54.

140 "Competences are viewed as socially complex, interconnected combinations of tangible (e.g., specific machinery) and intangible (e.g., specific organizational policies and procedures and the skills and knowledge of specific employees) basic resources that fit coherently together in a synergistic manner." *Id.* at 24; *see also* Prahalad and Hamel, *supra* note 131. "A capability is a set of business processes strategically understood." Stalk et al., *supra* note 131, at 62.

141 HUNT, *supra* note 30, at 137.

142 *Id.*

143 *Id.*

144 "[F]irms can have (1) an efficiency advantage, that is, *more* efficiently producing value, (2) an effectiveness advantage, that is, efficiently producing *more* value, or (3) an efficiency-effectiveness advantage, that is, *more* efficiently producing *more* value . . . ." *Id.*

145 *Id.* at 10.

146 *Id.*

147 *Id.*

148 *Id.* at 9.

R-A theory does not quarrel with the model of perfect competition, per se. Instead, R-A views perfect competition as a special atomistic state of passivity that might occasionally be approached at some point within a *general*, rivalrous competitive process. "That is," in most instances, "compared with perfect competition theory, R-A theory's premises are *closer* approximations of real-world conditions."[149]

R-A theory also has particular policy implications. "For R-A theory, 'competition' implies a particular kind of evolutionary process, not a particular kind of market structure. Likewise, for R-A theory, the term 'monopoly' has its classical meaning . . . ."[150] The existence of downward sloping demand curves are merely the result of naturally occurring heterogeneous demand, and not the "monopolistic" absence of competition or a competitive "imperfection" that government should worry over or attempt to "solve."[151] Contrary to the model of perfect competition, because resources are naturally heterogeneous, imperfectly mobile, and, therefore, naturally result in comparative advantages for sellers, "one *expects* competition to produce price differentials that are often long-lasting."[152] Thus, instead of trying to achieve a particular, idealized passive end-state that looks like perfect competition (which may only be fleeting anyway) government should safeguard the incentives for the freewheeling *process* of competition that creates innovation and economic progress over time.[153]

## II. *VERIZON V. TRINKO*

The Telecommunications Act of 1996 imposed upon local telephone companies certain duties designed to facilitate market entry by competitors.[154] Most importantly, the Telecommunications Act requires that an incumbent local exchange carrier share its network with competitors.[155] New entrant LECs (so-called "competitive" LECs) may then resell this access to their customers.[156] Pursuant to the Act, Verizon signed interconnection agreements with competitors such as AT&T.[157] In late 1999 competitive LECs complained to regulators that many orders for network access pursuant to such interconnection agreements were going

149 *Id.* at 3.

150 *Id.* at 256.

151 *Id.*

152 *Id.* (emphasis added).

153 *Id.* at 256-57.

154 The Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56. *See Trinko*, 540 U.S. 398, 401-05 (2004).

155 540 U.S. at 401-02.

156 *See id.*

157 *Id.*

unfilled, in violation of Verizon's obligation to provide open access.[158] Subsequently, the New York Public Service Commission and Federal Communications Commission ("FCC") opened parallel investigations, which ended with various consent decrees and orders.

The day after Verizon entered its consent decree with the FCC, The Law Offices of Curtis V. Trinko, LLP filed a complaint in the Southern District of New York, on behalf of itself and a class of similarly situated customers. Trinko's complaint alleged that Verizon had filled rivals' orders on a discriminatory basis as part of an anticompetitive scheme to discourage customers from becoming or remaining customers of competitive LECs, thus impeding the competitive LECs' ability to enter the market for local telephone service.[159] According to the Trinko class, these actions violated the Telecommunications Act of 1996 and, therefore, also constituted monopolization under Section 2 of the Sherman Antitrust Act. The district court dismissed the complaint in its entirety. In particular, it concluded that Trinko's complaint failed to satisfy the requirements of Section 2 of the Sherman Act.[160] On appeal, the Second Circuit reinstated the complaint in part, including the antitrust claim.[161] The Supreme Court granted certiorari to consider whether the Second Circuit erred in reversing the Southern District of New York's dismissal of respondent's antitrust claims.[162] Thus, the question before the Supreme Court was whether a complaint alleging a breach of the incumbent LEC's duty under the 1996 Telecommunications Act to share its network with competitors states a claim under Section 2 of the Sherman Act.[163]

Writing for the Court, Justice Scalia considered whether, as a preliminary matter, the 1996 Telecommunications Act had any effect on the application of traditional antitrust principles in this instance. The Court noted the Act imposed certain duties upon LECs. But it concluded the Act's antitrust-specific savings clause precluded a finding of implied immunity from the antitrust laws.[164]

Then, the Court considered the question of whether Trinko's allegation that Verizon denied interconnection services to rivals in order to limit entry, in fact, stated a monopolization claim under Section 2 of the Sherman Act. The Court held that it did not. In doing so, it reaffirmed *U.S. v. Grinnell Corp.*'s[165] holding that "[i]t is settled law that this offense requires, in addition to the possession of monopoly power in the relevant market, 'the

158 *Id.*

159 *Id.*

160 *Trinko*, 540 U.S. at 402.

161 Law Offices of Curtis V. Trinko, LLP v. Bell Atlantic Corp., 305 F.3d 89 (2d Cir. 2002).

162 *Trinko*, 540 U.S. at 403.

163 *Id.* at 401.

164 *Id.* (analyzing Section 601(b)(1) of the 1996 Act).

165 United States v. Grinnell Corp., 384 U.S. 563 (1966).

willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident.' "[166] However, unlike *Grinnell*, the Court then went on to elaborate on this standard. Breaking sharply with the perfectly competitive ideal of passive, atomistic price-taking firms, as articulated in the *Alcoa, American Tobacco,* and *Du Pont* line of cases, the Court held that:

> The mere possession of monopoly power, and the concomitant charging of monopoly prices, is not only not unlawful; it is an important element of the free-market system. The opportunity to charge monopoly prices—at least for a short period—is what attracts "business acumen" in the first place; it induces risk taking that produces innovation and economic growth.[167]

Contrary to the static, passive outcome of perfect competition, the Court views the free-market system as an *active process*, something that *moves* dynamically. This market process spotlights a Schumpeterian risk-taking entrepreneur as the disequilibrating prime mover of economic progress.[168] The entrepreneur is incentivized by the prospect of above-average profits to deploy his "business acumen" in a particular economic space where consumers are under-served, forgoing other opportunities perceived to be less potentially lucrative.

Directly contrary to the neoclassical viewpoint, the Court does not view these above-average profits suspiciously, as anticompetitive "wealth transfers" from consumers to sellers. Instead, the Court views them positively, as a Misesian informational signal[169] that attracts the

---

166 *Trinko*, 540 U.S. at 406-07.

167 *Id.* at 407. Indeed, notably absent from the decision is any type of *Alcoa, American Tobacco Co.*, or *DuPont* anti-market power language. *Compare* Eastman Kodak Co. v. Image Technical Services, Inc., 504 U.S. 451 (1992). "Market power is the power 'to force a purchaser to do something that he would not do in a competitive market.' " *Id.* at 464 (quoting Jefferson Parish Hospital Dist. No. 2 v. Hyde, 466 U.S. 2, 14 (1984)).

168 *See* SCHUMPETER, *supra* note 118, at 81-84 (1950).

169 In Mises' view:

> [Economic p]rofits are never normal. They appear only where there is a . . . divergence between actual production as it should be in order to utilize the available material and mental resources for the best possible satisfaction of the wishes of the public. They are the prize of those who remove this maladjustment; they disappear as soon as the maladjustment is entirely removed. In the imaginary construction of an evenly rotating economy there are no profits. The sum of the prices of the complementary factors of production, due allowance being made for time preference, coincides with the price of the product . . . .
>
> If all people were to anticipate correctly the future state of the market, the entrepreneurs would neither earn any [economic] profits nor suffer any losses. . . . No room would be left either for profit or for loss . . . .
>
> Profit and loss are ever-present features only on account of the fact that ceaseless change in the *economic data* makes again and again new discrepancies, and consequently the need for new adjustments originate.

entrepreneur to enter the competitive arena and satisfy an unmet consumer desire, or to better satisfy a consumer desire which is being imperfectly met.

Said another way, the prospect of a superior economic return is the incentivizing spark that lights the fire of competition. Market power and above-average profits, without more, are not "anticompetitive" or "supracompetitive" "rents" or "wealth transfers," at least in the short run. They are not even merely benign. Instead, they are the key data signal that alerts entrepreneurs to the fact that consumer desires in a particular economic space are not being adequately satisfied. Market power and above-average profits are the rightful prize of the risk-bearing trailblazer that successfully introduces a new product desired by consumers or better satisfies consumer desires by improving an existing product through, for example, a lower price, better quality, a new feature, better service, or lower transaction costs. The Court, however, could not find its way to break free from its habit of using suspicious, pejorative neoclassical terms like "monopoly power" and "monopoly prices" to describe these important and beneficial data signals.

Thus, consistent with classical-era law and economics and R-A theory's corollary incentive-based policy implications, the Court held that "[t]o safeguard the incentive to innovate, the possession of monopoly power will not be found unlawful unless it is accompanied by an element of anticompetitive *conduct*."[170] But the Court also retained *Grinnell*'s requirement that a product's price be greater than its marginal production cost, apparently as a prophylactic screen for such anticompetitive conduct. The Court's "at least for a short period" caveat, however, suggests that it might potentially consider durable, long-term market power to be an indicia of some type of competitive problem, although it does not indicate what, exactly, such a problem may be. But there is no inherent reason that durable market power, without more, necessarily indicates a competitive problem. Durable market power is the natural result of meeting heterogeneous demand with a comparative advantage derived from heterogeneous, imperfectly mobile resources,[171] as where a seller continues to "reinvent it itself" by internalizing what Schumpeter called "the process of Creative Destruction"[172] and, thus, stays ahead of its competitors' abilities to serve consumers.

The Court appears to anticipate that, generally, any market power and above-average profits accruing to a particular entrepreneur will tend to be competed away over time by other sellers attracted by that very same

---

Mises, *supra* note 119, at 108-19 (emphasis added).

170 Verizon Commc'ns. Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. at 407.

171 HUNT, *supra* note 30, at 256.

172 SCHUMPETER, *supra* note 118, at 82. *See also* ANDREW S. GROVE, ONLY THE PARANOID SURVIVE: HOW TO EXPLOIT THE CRISIS POINTS THAT CHALLENGE EVERY COMPANY (1999).

market power and profit. In the Court's view, a healthy free market system is an *active, moving process*. The disequilibrating risk-bearing entrepreneur is rewarded with superior economic returns for serving consumers in a superior fashion. In turn, the prospect of winning some portion of this superior economic return serves as a data signal to induce additional sellers to enter the same economic space, or existing sellers to better serve consumers. This competing away of economic profits, standing alone in the abstract, could, hypothetically, be described as an equilibrating tendency. But the real world process described by the Court cannot rightly be said to be closed-ended, as it remains continually subject to the endogenous, disequilibrating stimulus of entrepreneurial activity continually seeking superior economic returns. The Court does not appear to expect this process to come to rest at some final state. Rather, the Court appears to view the market process as being an open-ended phenomenon that never "ends." The market process is one of "becoming."

The Court then again broke sharply with the perfectly competitive ideal of homogenous, mediocre sellers having homogenous production functions created from plentiful, perfectly homogenous and perfectly mobile resources. "Firms may acquire monopoly power by establishing an infrastructure that renders them uniquely suited to serve their customers."[173] The logical implication of this statement that sellers may be uniquely suited to serve certain customers is that those customers have unique, heterogeneous desires that they want to have met in a particular manner. The Court also emphasized the central importance of heterogenous resources and comparative advantages in meeting this demand. "Compelling such firms to share the *source of their advantage* is in some tension with the underlying purpose of antitrust law, since it may lessen the *incentive* for the monopolist, the rival, or both to invest in those *economically beneficial facilities*." [174]

In contrast to the neoclassical model of perfect competition and its corollary structural presumption, the Court does not describe such heterogeneous resources as being inherently suspect "barriers" competition, or even merely benign. Instead, the Court describes heterogeneous resources as being "*beneficial*" to a *rivalrous*, competitive process, as distinct from institutional barriers imposed by government force that artificially restrict competition by law or private conduct that, similarly, excludes actual or potential competitors in a manner distinct from growth or development as a consequence of a superior product, business acumen, or historic accident.[175] Said another way, the Court recognizes that in the real world heterogeneous and imperfectly mobile (i.e., "unique") resources (i.e.,

---

173 Verizon Commc'ns., Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. at 407.

174 *Id.* at 407-08 (emphasis added).

175 *Id.* at 407-12. *See especially id.* at 412 (describing "the statutory restrictions upon Verizon's entry into the potentially lucrative market for long-distance service.").

"infrastructure" or "facilities") allow sellers to best serve their customers (i.e., with a comparative "advantage"). Indeed, consistent with Resource-Advantage theory, the terms "heterogeneous resources," or "core competences" and "capabilities," could just as easily be substituted for the Court's "unique infrastructure" language, and the meaning would be exactly the same.

The *Trinko* Court's positive view of heterogeneous resources is directly contrary to the overt hostility of *U.S. v. Alcoa* and its progeny[176] toward heterogeneous resources that yield comparative advantages and the central thesis of the SCP school: that the intensity of competition in a particular industry (its "performance") is essentially determined mechanically by its structure. Instead, the Court holds that competition must be analyzed on a case-by-case basis in a fact-specific manner that includes non-structural factors. "Antitrust analysis must always be attuned to particular structure *and circumstances* of the industry at issue."[177] By placing no limit on the meaning of the additional term "circumstances," the Court indicates, consistent with R-A theory, that structure does not presumptively indicate the intensity of competition. For the Court, non-structural factors may be of equal, or even greater, importance than structure to an industry's performance.

Thus, in these passages the Court practically articulated a "Resource-Advantage" theory by name, in that it held that: (1) the natural existence of heterogeneous resources, or "unique[] . . . infrastructure[s]," (2) are "the source" of naturally resulting comparative "advantage[s]" among sellers. In doing so the Court, consistent with the classical tradition, also clarified the issue of entry by firmly distinguishing between the mere natural, beneficial existence of heterogeneous resources and the artificial imposition of institutional "barriers" to entry by government or private conduct that, similarly, restricts the competitive process.

The Court emphasized that it had "been very cautious in recognizing such exceptions" to the general rule that a seller's possession of private, heterogeneous resources is inviolate, even though those resources may be extremely rare and, therefore, difficult or even impossible to obtain by others.[178] As such, the refusal of a seller possessing a heterogeneous resource to cooperate with a rival such as the case of "*Aspen* is at or near the outer boundary of § 2 liability."[179] The Court characterized *Aspen Skiing* as a unique situation where "the defendant's termination of a voluntary agreement with the plaintiff suggested a willingness to forsake

---

176 *See supra* note 167.

177 Verizon Commc'ns., Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. at 412 (emphasis added).

178 *Id.* at 409.

179 *Id.*

short-term profits to achieve an anticompetitive end."[180] The refusal to cooperate alleged in the case at hand did not fall within *Aspen*'s "limited exception."[181] The Court emphasized that, in contrast to the *Aspen Skiing* scenario, where the product in question was already sold externally to consumers, in the present case "the services allegedly withheld are not otherwise marketed or available to the public." [182] But it did not expand further on this factual distinction. Importantly, though, the Court pointedly refused to either recognize or repudiate the so-called "essential facilities" judicial doctrine of forced access to a heterogeneous resource adopted by some lower courts.[183]

Furthermore, contrary to perfect competition's theoretical postulate of perfect, costless information, the Court recognized, in the tradition of Smith and Hayek, that in the real world information is imperfect, decentralized, and may be costly to acquire—especially for generalist courts.[184] Therefore, the Court concluded that "[t]he costs of false positives," from imperfect

---

180 *Id.*

181 *Id.* at 410.

182 *Id.*

183 *Id.* at 410-11.

184 *See* SMITH, *supra* note 7; F.A. Hayek, *The Use of Knowledge in Society*, 4 AM. ECON. REV. XXXV 519-30 (Sept. 1945), *reprinted in* INDIVIDUALISM AND ECONOMIC ORDER 77 (1948). Hayek states:

> [T]he 'data' from which the economic starts are never for the whole society 'given' to a single mind which could work out the implications and can never be so given.
>
> The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess. The economic problem of society is thus not merely a problem of how to allocate 'given' resources—if 'given' is taken to mean given to a single mind which deliberately solves the problem set by these data. It is rather a problem of how to secure the best of use of resources known to any of the members of society, for ends whose relative importance only these individuals know. Or, to put it briefly, it is a problem for the utilization of knowledge which is not given to anyone in its totality.

*Id.* at 77-78. *See also* New State Ice Co. v. Liebmann, 285 U.S. 262 (1932) (J. Brandeis dissenting), recognizing that our federal system of decentralized state governments allows States having only imperfect information to engage in legislative experiments designed to meet changing social conditions, without imposing the consequences of failed experiments on the entire country. As Justice Brandeis explains:

> Merely to acquire the knowledge essential as a basis for the exercise of this multitude of judgments would be a formidable task; and each of the thousands of these judgments would call for some measure of prophecy. Even more serious are the obstacles to success inherent in the demands which execution of the project would make upon human intelligence and upon the character of men. Man is weak and his judgment is at best fallible.
>
> . . . .
>
> The discoveries in physical science, the triumphs of invention, attest the value of the process of trial and error. In large measure, these advances have been due to experimentation. . . . It is one of the happy incidents of the federal system that a single courageous state may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.

*Id. See also* THOMAS SOWELL, KNOWLEDGE AND DECISIONS 3-20 (Basic Books 1980).

information, "counsels against an undue expansion of § 2 liability."[185] Quoting Areeda, the Court adopted the rule that: "[n]o court should impose a duty to deal that it cannot explain or *adequately and reasonably supervise*. The problem should be deemed irremedia[ble] by antitrust law when compulsory access requires the court to assume the day-to-day controls characteristic of a regulatory agency."[186] The Court reiterated the difficulty and cost of assembling the decentralized knowledge necessary to monitor, or "supervise," a particular industry on a daily basis. That is to say, courts, as an initial matter, are likely to be substantially ignorant about the changing details of the market process in any given industry, just like any other real world actor trying to interact with the economy.[187] In the case of forced sharing, courts "are ill-suited . . . to act as central planners, identifying the proper price, quantity, and other terms of dealing,"[188] details which, in the market process, are spontaneously coordinated through what Smith called "The Invisible Hand."

The Court then distinguished between the 1996 Telecommunication Act's *positive* purpose of eliminating institutional monopolies enjoyed by the successors to AT&T's local franchises and the Sherman Antitrust Act's *negative* proscription against certain conduct.[189] According to the Court, "[i]t would be a serious mistake to conflate the two goals. The Sherman Act . . . does not give judges *carte blanche* to insist that a monopolist alter its way of doing business whenever some other approach might yield greater competition."[190] As such, the Court concluded that the Trinko complaint failed to state a claim under the Sherman Act, and reversed and remanded for further proceedings consistent with its opinion.[191]

In sum, the *Trinko* Court articulated a classical view of competition, as refined through the corollary insights of R-A competition, consistent with the 1890 enactment of the Sherman Act. Indeed, the Court practically articulated an R-A theory by name. It is unclear, however, to what extent the Court is actually familiar with the multiple research programs that R-A theory incorporates, or whether the Court has simply refined a classical-era understanding of competition through its own consideration. The only potential caveat contrary to R-A competition made by the Court is its suggestion that durable market power might indicate some kind of problem with the competitive process. Importantly, the Court did more than narrowly implement a set of *negative* proscriptions *against* certain

185 Verizon Commc'ns., Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. at 414.

186 *Id.* at 415 (quoting Areeda, *Essential Facilities: An Epithet in Need of Limiting Principles*, 58 ANTITRUST L. J. 841, 853 (1989)).

187 *Id.*

188 *Id.* at 408. *Compare* with *supra* notes 48-49 and related text.

189 *Id.* at 415-16.

190 *Id.*

191 *Id.*

activities.[192] The Court did much more than that. It explained in *positive* terms a generalized, highly coherent theory of how microeconomic competition *works*—its most expansive and precise such theory to date. Thus, the Court firmly broke with the SCP, Chicago, and post-Chicago schools of microeconomic theory. In doing so, it rejected perfect competition as a welfare ideal.

## III. IMPLICATIONS FOR THE CONCEPTS OF COMPETITION, MONOPOLY, AND ENTRY

The *Trinko* court's articulation of a classical, rivalrous process view of microeconomic competition, as refined through the corollary insights of R-A theory, is highly coherent. However, the Court could not break free of its habit of using suspicious, pejorative neoclassical terms like "monopoly power" and "monopoly prices" to describe features that the Court correctly views as being beneficial to the competitive process. For example, the Court used the neoclassical pejoratives of "monopoly power" and "monopoly prices" in place of the neutral, descriptive term "market power" to describe what it correctly views as "an important element of the free-market system."[193]

As explained above, the *Trinko* court distinctly articulated a rivalrous, evolutionary process view of microeconomic "competition" in its decision and rejected all of the premises of perfect competition. This view of competition is consistent with a rivalrous definition of competition as "the process over time of one or more parties acting to secure an exchange with a second party from a third party or potential third party by offering terms that are more favorable than those of the third party."[194] Importantly, the court's articulation of a rivalrous, process theory of competition, as refined through the corollary insights of R-A theory, is also consistent with the classical-era 1890 enactment of the Sherman Act. By contrast, neither SCP nor its cousin the majority Chicago school strand can claim that their views of "competition" or "monopoly," are consistent with the original, classical-era enactment of the Sherman Act. For a welfare ideal, both perspectives rely on the mathematical abstraction of the state of perfect competition, which did not come into existence until 1921 and required additional time to become popularized. Thus, both SCP and majority Chicago engage in conceptual time travel to interpret the Sherman Act as if it had been enacted thirty or more years after it actually became law.

This point is critical. The text of the Sherman Act protects the human action of competition and punishes acts contrary to it. Section 1 of the

---

192 *Id.* at 409 (describing the negative prohibitions of the Sherman Act).

193 *Id.* at 406.

194 *See supra* note 12.

Sherman Act "declared to be illegal . . . every . . . restraint of trade or commerce . . . ." The Act's text is famous for its broad, sparse phraseology. But, the inescapable implication of the Act's ban on conduct *restraining* trade or commerce is that the Act protects "unrestrained" or "free" human action in trade and commercial exchange—otherwise known as "competition." This conclusion is buttressed by Section 2 of the Act, which punishes persons who do not engage in competitive conduct, but instead "monopolize or attempt to monopolize, or combine or conspire . . . to monopolize any part of . . . trade or commerce . . . ." This conclusion is further reinforced by the Clayton Act's multiple "supplement[al]" provisions protecting "competition."[195] The Sherman Act and its supplements did not adopt or incorporate any particular pre-existing unitary body of federal, state, or common law relating to "restraints on trade," "competition," or "monopoly"—as no such body of law existed in the first place.[196] Thus—for better or worse—the only logical way that the Act (and its supplements) can be interpreted is as a classical-era statute that governs microeconomic activity using broad, generalized terms, consistent with how they were generally understood by American lawyers and economists in the year 1890.[197]

Given that the Sherman Act's 1890 enactment occurred at a time when Americans were informed by classical-era law and economics, as a legislative matter the Act's broad text can only be logically interpreted to protect microeconomic "competition" as a dynamic, rough-hewn, and open-ended process animated by rivalrous, entrepreneurial activity.[198] The same conclusion follows for the 1914 Clayton Act and its amendments, which are entitled: "An Act to *supplement* existing laws against unlawful restraints and monopolies, and for other purposes," and which specifically refers to the Sherman Act as "the Act . . . approved July second, *eighteen hundred and ninety* . . . ."[199] No other law contradicts this

---

195 Clayton Act § 2(a), 15 U.S.C. § 13 (2000) ("It shall be unlawful for any person . . . to discriminate in price . . . where the effect of such discrimination may be substantially to lessen competition or tend to create a monopoly . . . or to injure, destroy, or prevent competition . . . ."); § 3, 15 U.S.C. § 14 (making unlawful sales, etc. on agreement not to use the goods of a competitor "where the effect of such lease, sale, or contract for sale or such condition, agreement, or understanding may be to substantially lessen competition or tend to create a monopoly . . . ."); and § 7, 15 U.S.C. § 18 ("No person . . . shall acquire, directly, or indirectly, the whole or any part of the stock or other share capital . . . where . . . the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly.").

196 AREEDA & HOVENKAMP, *supra* note 9; BORK, *supra* note 105, at 20 ("there was no unitary body of common law doctrine that could give meaning to the statute.").

197 "It is apparent that the definition of restraint of trade or monopoly to be contained in the [Sherman Act] statute, must be more comprehensive than the common-law definitions of these terms; that is, it must declare some agreements to be illegally restrictive, and some acts to be monopolistic, which were not so at the common law . . . ." SPELLING, *supra* note 9, at 225.

198 *See* Mason, *supra* note 8; Harbeson, *supra* note 19; Harbeson, *supra* note 24.

199 Clayton Act, *supra* note 195 (emphasis added).

conclusion.[200] As a matter of basic chronology, the antitrust laws cannot be logically interpreted to protect "competition" as a passive, atomistic state of "perfect competition" which would only be conceived three decades later.[201]

---

[200] *E.g.*, the Federal Trade Commission Act, 15 U.S.C. § 45; the Antitrust Civil Process Act, 15 U.S.C. § 1311 *et seq.*; the Local Government Antitrust Act, 15 U.S.C. § 34 *et seq.*

[201] Notably, in THE ANTITRUST PARADOX, perhaps the most extensive and influential analysis of the antitrust laws in the second half of the twentieth century (it has been cited by the Supreme Court in fifteen cases), Judge Robert Bork erroneously rejected the idea that the Sherman Act protects competition as "a process of rivalry." According to Bork, "[c]ompetition may be read as the process of rivalry. This is a natural mode of speech. . . . Yet it is a loose usage and invites the further, wholly erroneous conclusion that the elimination of rivalry must always be illegal . . . ." BORK, *supra* note 105, at 58-59. "But this identification of competition with rivalry will not do for antitrust purposes. . . . No firm, no partnership, no corporation, no economic unit containing more than a single person could exist without the elimination of some kinds of rivalry between persons . . . ." *Id.* "[W]e must be on guard against the easy and analytically disastrous identification of competition with rivalry." *Id.*

Bork's error, however, is that he engages in a linguistic switch between the phrase "process of rivalry" and the mere numerical counting of rivals. Bork does not consider rivalry as a process involving *action over time*, in which one or more parties *acts* to secure an exchange with a second party from a third party *or a potentially entering third party* by offering more favorable terms. Instead, Bork merely discusses various *states* in which rivals exist in different numbers. The mere addition or subtraction of a certain number of rivals to arrive at different states, however, says nothing about any *conduct* intended to secure an exchange—or any *act of response or potential act of response* that might occur at a later point in time.

Of course, a reduction in the number of rivals, without more, is perfectly consistent with competition as a classical, rivalrous process. As Adam Smith explained, sellers "hope to jostle that other out of his employment. . . . Rivalship and emulation render excellency, . . . and frequently occasion the very greatest exertions." SMITH, *supra* note 7. *See also* STIGLER, *supra* note 16, at 40-41. Although Bork attempts to distinguish his discussion of the *"process* of rivalry" from a *state* of "fragmentation" of sellers, the result is that he erroneously conflates the two concepts and, therefore, never actually considers the former. BORK, *supra* note 105, at 60-61 (emphasis added).

Bork declares that "[w]e must stick to the words themselves *as if they were newly minted statutes.*" *Id.* at 58 (emphasis added). Ironically, though, Bork does not do this. With notable imprecision, Bork declares that "competition" should be read as "a shorthand expression, a term of art, designating *any state of affairs* in which consumer welfare cannot be increased by moving to an alternative state of affairs through judicial decree." *Id.* at 61 (emphasis added). The imprecision of this "term of art" is compounded further by the fact that just a few pages earlier Bork also describes competition as "inherently a *process* in which rivals seek to exclude one another." *Id.* at 49 (emphasis added). Thus, it is impossible to determine what, exactly, it is that Bork means by the word "competition."

Bork's welfare ideal for analyzing "competition" and "monopoly" is the neoclassical state of perfect competition, concepts which he discusses using neoclassical terminology. *E.g, id.* at 99 ("Since [a] monopolist is the entire industry, it faces the sloping industry demand curve rather than the competitive firm's flat demand curve."). Thus, Bork erroneously imports into an 1890 statute concepts that were not popularized until after 1921, contrary to his own originalism.

Bork's sometimes-definition of "competition" as a "state of affairs" rather than as a "process" particularly reflects this erroneous importation. "The state of affairs ultimately created by competition was certainly discussed by every classical writer, but to apply a magnifying glass to the price-equals-cost (equilibrium) condition, as if it were the heart of the classical analysis, is a case of mistaking 'the shadow for the substance.' " MACHOVEC, *supra* note 7, at 100. For example, "Adam Smith's most emphatic and recurring thematic point—his explanation of the invisible hand—had nothing to do with

Moreover, as an economic matter the antitrust laws cannot be logically interpreted to protect "competition" as a passive, atomistic state of "perfect competition" because—as Knight freely admitted—perfect competition is not actually competition at all. Thus, neither SCP's goal of preventing allegedly pernicious "wealth transfers" due to the mere existence of market power, nor its cousin majority Chicago's goal of improving a state of "allocative efficiency" without greatly impairing "productive efficiency," is consistent with the text and chronology of the Sherman Act and it supplements. These acts can only be logically interpreted to protect the classical *process* of rivalrous competition—a process that results in economic growth in the form of both static, allocative efficiency and dynamic efficiency.

Consistent with the Court's restoration of the classical distinction between mere natural, economically beneficial heterogeneous resources and the exclusion of competitors from the competitive process, the negative, pejorative term "monopoly" should be used only to denote a seller that faces absolutely no actual or potential competition in a broadly-defined product market, due to the exclusion of all competitors in a manner distinct from rivalrous competition.[202] That is, the term "monopoly" should be used only to denote a seller that receives all the sales in some broadly-defined product market due to either a government grant that excludes all other competitors or private conduct that, similarly, excludes all other competitors in a manner distinct from growth or development as a consequence of a superior product, business acumen, or historic accident.[203] The terms "monopoly price" and "monopoly profit" should be used only to

---

the final *results* of the process and had everything to do with the role of incentives, i.e., the *nature* of the process." *Id.* So, Bork engages in the same conceptual time travel committed by the predecessor SCP school.

Notably, the Supreme Court rejected Bork's "term of art" definition of competition in *Trinko*. *See Trinko*, 540 U.S. at 415-16 and text accompanying *supra* note 190. This is not to say, however, that particular critiques of the predecessor SCP school made by Bork and majority Chicago are, therefore, incorrect to the extent that they do not rely on perfect competition as a welfare ideal.

202 "It will be noticed that, in all the foregoing [classical-era] definitions of 'monopoly,' there is embraced two leading elements; viz., an exclusive right or privilege, on the one side, and a restriction or restraint on the other, which will operate to prevent the exercise of a right or liberty open to the public before the monopoly was secured." SPELLING, *supra* note 9, at 235. *See also* Mason, *supra* note 8, at 35.

203 United States v. Grinnell, 384 U.S. 563 (1966); Verizon Commc'ns., Inc. v. Law Offices of Curtis V. Trinko, *LLP*, 540 U.S. 398, 407 (2004). Consistent with *Grinnell*'s emphasis on "willful" action and *Trinko*'s emphasis on "conduct," some modern commentary generally in accord with the classical view allows that "it is not inconceivable that . . . exceptional conditions . . . may . . . evolve within a free market that could result in monopolized restrictions on output *not* instigated by state protection against competition . . . ." Frank M. Machovec, *Mises, Monopoly, and the Market Process*, Vol. 19. No. 2 CATO J. 247, 252-55 (1999). "Cartelization to fix prices *does* occur, especially where collusion to rig bidding is easy, as in highway construction contracts and used-car auctions . . . ." MACHOVEC, *supra* note 7, at 232.

denote the price and profit associated with a product offered for sale by such a monopoly. The term "monopoly power" should be used only to denote the market power possessed by such a monopoly. By contrast, the neutral, descriptive term "market power," simpliciter, should be used to denote market power where it merely results from heterogeneous demand that is met at a corresponding price above marginal cost by a seller possessing a comparative competitive advantage derived from heterogeneous resources.[204]

Likewise, recognizing its derivation from the term "monopoly," the negative, pejorative term "monopolize" should be used only to denote conduct by a government or private actor that excludes existing or potential competitors in a manner distinct from growth or development as a consequence of a superior product, business acumen, or historic accident. The term "monopolist" should be used only to denote an actor that engages in the conduct of monopolization. Therefore, the negative, pejorative term "monopolist" should be used only to denote a seller that is partially—but not completely—insulated from actual or potential competition in a broadly defined product market, due to the exclusion by it of one or more competitors in a manner distinct from growth or development as a consequence of a superior product, business acumen, or historic accident.[205] The terms "monopolistic price" and "monopolistic profit" should be used only to denote the price and profit associated with a product offered for sale by a monopolist. The term "monopolistic power" should be used only to denote market power possessed by a monopolist.

None of the negative, pejorative terms "monopoly," "monopolist," or "monopolistic," nor any variant thereof, should be used to describe a seller who receives a "large" or even a total share of the sales of a particular product, simply because it is the natural result of a process of rivalry.[206] For purposes of the Sherman Act and its supplements, a seller receiving a large or total share of the sales of some product, simpliciter, should simply be called that—and nothing more. Neither a "world-beater" seller that simply serves consumers better than all of its other competitors, nor a seller that is first to market with a new, unique product, deserves to be saddled with the pejorative moniker of being called a "monopoly," or some variant thereof. To do so, as through the casual use of the term "monopoly" or "monopolist," is to illogically equate the epitome of competition with its exact opposite. That is, such usage erroneously conflates the *active process*

---

204 Classical "free competition thus understood is quite compatible with the presence of monopoly elements in the [neoclassical] *economic* sense of the word monopoly." Mason, *supra* note 8, at 36. "Before the Sherman Act monopoly actions were brought, with but few exceptions, before the courts on the suit of private interests. These interests were more likely to be directly affected adversely by predatory practices or attempts at exclusion from the market than by control of prices." *Id.* at 46.

205 *See* Mason, *supra* note 8, at 44; *see also Grinnell*, 384 U.S. at 563.

206 *See supra* notes 15-26 and related text.

of competition with its inhibition, due to the mere existence of a certain *state* of relative numerousness at a particular time.[207] The effect is to declare that a seller should "compete," but that it should not dare to be "too successful" or to "win" the favor every potential customer—lest it suffer the fate of being viewed with suspicion or outright hostility.[208]

Likewise, none of these "monopoly"-type terms should be used to describe a seller that merely faces a downward-sloping demand curve that naturally results from heterogeneous demand and then meets that consumer demand at a corresponding price above marginal cost using a comparative competitive advantage derived naturally from its possession of heterogeneous resources. To do so is to illogically saddle a seller with suspicion simply because it dares to satisfy diverse consumer desires that deviate from the perfect competition model's assumption of puritanical, perfectly homogenous consumer demand.

In addition, the negative, pejorative term "barrier to entry" should be used only to denote a restriction on entry resulting from either the imposition of government force or private conduct that, similarly, excludes competitors in a manner distinct from rivalrous competition. That is, the term "barrier to entry" should be used only to denote either a government grant that excludes other competitors or the result of private conduct that, similarly, excludes other competitors in a manner distinct from growth or development as a consequence of a superior product, business acumen, or historic accident.[209] Therefore, the term "barrier to entry" should not be applied to a resource that is naturally heterogeneous merely as a neutral matter of fact, or to a comparative advantage that results therefrom. Instead, if a particular heterogeneous resource is necessary to compete in a particular economic space, the neutral, descriptive term "requirement for entry" should be applied to such a resource. Thus, "monopoly" and "monopolistic" markets can exist only where some such "barrier to entry" exists—but do not result merely because some particular heterogenous resource may naturally be "required" to enter and compete in a particular economic space.

## CONCLUSION

*Verizon v. Trinko* represents the Supreme Court's most important monopolization decision in a decade. Commentators have largely focused on the telecommunications and facial antitrust aspects of the case. However, *Trinko* also represents a profound change in the Supreme Court's jurisprudence on microeconomic competition. Specifically, *Trinko*

---

207 This is precisely the structural error that Knight warned against. *See* Knight, *supra* note 31.

208 *See e.g.* JOSEPH HELLER, CATCH-22 (Simon & Schuster Inc. 1955).

209 *See* Mason, *supra* note 8, at 47.

represents the first Supreme Court case to break distinctly with both the well-defined "Structure-Conduct-Performance" and "Chicago" schools of microeconomics analysis, as well as the vaguely defined "post-Chicago" school of microeconomic analysis. In *Trinko*, the Supreme Court articulated a classical, rivalrous process view of competition, as refined through the corollary insights of the "Resource-Advantage" theory of competition, consistent with the 1890 enactment of the Sherman Act. In doing so, it rejected "perfect" competition as a welfare ideal. Contradictorily, however, while the *Trinko* court's articulation of an R-A theory is highly coherent, the Court could not break free from its habit of using suspicious, pejorative neoclassical terms like "monopoly power" and "monopoly prices" to describe features which the Court correctly views as being beneficial to the competitive process. Therefore, the Supreme Court and lower courts should refine their use of terms relating to competition, monopoly, and entry to conform to the Court's classical, rivalrous process view of competition, as refined through its articulation of the key premises of R-A theory.

# DO CAMPAIGN CONTRIBUTIONS AND LOBBYING CORRUPT? EVIDENCE FROM PUBLIC FINANCE

*Gajan Retnasaba*[*]

## ABSTRACT

Private interests expend great amounts of resources attempting to influence government decisions using tools such as campaign contributions and lobbying. Yet, little is known about whether they use means fair or foul to achieve their goals. A better understanding of how private interests influence government, specifically whether they use corruption, is vital for informing debate on how such activities should be regulated.

This paper presents two empirical attempts to measure the presence of corruption in state public finance. In the first it investigates whether campaign contributions caused corruption in the public finance industry of the early 1990s using an event study methodology. In the second it investigates the current controversy surrounding the use of lobbyists in public finance and whether their use is linked to corruption using a methodology that exploits state heterogeneity in the supply of corrupt decision makers. In both cases it finds strong evidence of corruption.

## INTRODUCTION

Private interests channeled over $300 million of political contributions to federal candidates alone during the last election cycle in 2004.[1] A further $3 billion was spent lobbying the winners of these elections.[2] Why do they

---

[*] Gajan Retnasaba graduated from Harvard Law School in 2004, and currently works at McKinsey & Company. Gajan would like to thank Vicki Frank, Greg Weston and the editors of the Journal of Law, Economics & Policy for their helpful comments. The data is available from the author on request. Comments are welcome at gajanr@gmail.com.

[1] Corporations cannot donate directly to political candidates. Corporations can however set up and administer Political Action Committees (PAC's). These committees raise funds from individuals and then use these funds to help candidates gain or retain office primarily through making campaign contributions. Most PAC's are set up to represent a corporation, industry or interest group. *E.g.*, FEDERAL ELECTION COMMISSION, CAMPAIGN GUIDE FOR CORPORATIONS AND LABOR ORGANIZATIONS, 7 (2001); Federal Election Commission, *PAC Activity Increases for 2004 Elections*, April 13, 2005.

[2] Center for Public Integrity, *available at* http://www.publicintegrity.com (aggregating data from disclosures made by lobbyists to the Senate Office of Public Records).

spend this money? Are these expenditures thinly veiled bribes or do they represent legitimate participation in the political process?

This paper attempts to provide an answer to this question for one segment of government. Examining public finance, first it looks at the extensive campaign contributions made by investment banks in the early 1990s and whether they corrupted the decision-making of government. Second, it looks at the current controversy surrounding whether the widespread use of lobbyists by investment banks is related to corruption.

## I. CORRUPTION

### I.1. *Corruption and Its Effects*

Corruption, broadly defined, is the misuse of public office for private gain.[3] Corruption is an enemy of good governance for two primary reasons.

First, corruption transfers wealth from society to the corrupt. Corrupt officials divert resources from the public treasury to their own pocket, effectively creating a tax on society.[4] Consider countries such as Nigeria where poverty prevails despite an abundance of natural resources.[5] The wealth generated from natural resources is siphoned, through bribes, kickbacks and outright theft, away from the public purse into the wallets of corrupt government officials. This allows only a small fraction of each government dollar to be spent on bettering society.

Second, corruption causes mismanagement.[6] The corrupt make decisions based on what will increase their own welfare rather than what is best for greater society and consequently misallocate resources.[7] For example, corrupt officials tend to build new, large, and complex infrastructure projects over maintaining existing infrastructure because the awarding of large construction contracts provides considerably better opportunities for kickbacks. This mismanagement leads to society paying

---

[3] *See, e.g.*, WORLD BANK, HELPING COUNTRIES COMBAT CORRUPTION: THE ROLE OF THE WORLD BANK 8 (1997).

[4] Shang Jin Wei, *How Taxing is Corruption on Institutional Investors*, 82 REV. ECON. & STATS. 1 (2000) (estimating that an increase in corruption from the level of Singapore to Mexico is equivalent, in terms of deterring foreign direct investment, to raising the level of taxation by 50%).

[5] Nuhu Ribau, *Implications of Economic and Financial Crimes on the National Economy*, ECONOMIC AND FINANCIAL CRIMES COMMISSION OF NIGERIA (2004); TRANSPARENCY INTERNATIONAL, *Global Corruption Report 2004* (estimating that Former Nigerian President Sani Abacha stole up to $5 Billion from the country during his five year reign).

[6] *See, e.g.*, Andrei Schliefer & Robert W. Vishny, *Corruption*, 108 Q. J. ECON. 599 (1993); *Contra* Nathanial Leff, *Economic Development Through Bureaucratic Corruption*, 8 AM. BEHAV. SCIENTIST 8 (1964) (arguing that corruption gives incentives to government officials to work faster and helps industry to mediate the effect of bad rules).

[7] *Id.*

for infrastructure it does not need while critical infrastructure deteriorates.[8] It is estimated that for every $1 of corruption, society incurs a cost of $1.67 due to the mismanagement effect.[9]

And left untreated, corruption is contagious.[10] Once corruption takes root in one area, unnoticed or unpunished, it gains acceptance.[11] This creates a greater likelihood it will increase, spread, and intensify.[12] Further, corruption is self-entrenching.[13] Once in power, corrupt incumbents can raise more funds than honest challengers, leaving challengers at a disadvantage. Honest challengers then either face an uphill battle to unseat corrupt incumbents who can outspend them or must resort to corruption in order to gain sufficient financial support.[14]

### I.2. *The Need to Measure Corruption*

So, responsible governments seek to curtail corruption by implementing regulatory, operational, and enforcement measures that reduce opportunities for corruption and make corruption less attractive.[15]

But as bad a problem as corruption is, the solutions can be worse. Fighting corruption inevitably imposes costs. Firstly, direct financial costs are imposed on both government and the private sector. Government pays to conceive, administer, and implement new rules and the private sector pays to comply with these rules.[16]

Secondly, opportunity costs are incurred from the diminished flexibility of government. Anticorruption laws typically blunt the discretionary power of officials, diminishing their ability to respond appropriately when circumstances require.[17]

---

[8] Vito Tanzini & Hamid Davoodi, *Corruption, Public Investment and Growth*, (Int'l Monetary Fund Working Paper No. 97/139, 1997) (using cross-sectional data to show that corrupt nations spend more on infrastructure, less on maintenance and have poor quality infrastructure).

[9] Vinod D. Hrishiikesh, *Statistical Analysis of Corruption Data and Using the Internet to Reduce Corruption*, 10 J. ASIAN ECON. 591 (1999).

[10] *See, e.g.*, NORWEGIAN AGENCY FOR DEVELOPMENT CO-OPERATION, RESEARCH ON CORRUPTION. A POLICY ORIENTED SURVEY (2000); WORLD BANK *supra* note 3 at 10-11, 20-21, n.3.

[11] *Id.*

[12] *Id.*

[13] Eric C. C. Chang, *Electoral Incentives for Political Corruption under Open-List Proportional Representation*, 67 J. POL. 716 (2005).

[14] *Id.*

[15] Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

[16] Philip K. Howard, THE DEATH OF COMMON SENSE: HOW LAW IS SUFFOCATING AMERICA 55-111 (1994) (giving a series of anecdotal examples of laws designed to reduce corruption that have administrative costs exceeding their benefit).

[17] *Id.* at 1-55 (giving a series of anecdotal examples of laws that have constrained the discretionary powers of public officials leading to absurd results).

Because of these costs anticorruption policies should be intelligently targeted. A first step in doing so, is to measure corruption in the various parts of government. Measurement provides a more rigorous basis for the imposition of anticorruption policies. Further, once anticorruption policies are implemented, measurement allows us to monitor how effective the policies are in the actual reduction of corruption, allowing remedial action to be taken where necessary.[18] Finally, measurement allows government to credibly demonstrate progress to the public, creating greater support, deterring those who remain corrupt,[19] and providing a more compelling justification for the costs and restrictions that anticorruption policies impose on the public.[20]

However, a method to more objectively measure corruption is not obvious.[21] Corruption is typically illegal, giving the corrupt strong

---

[18] *See, e.g.*, JEREMY POPE, TRANSPARENCY INTERNATIONAL, TRANSPARENCY INTERNATIONAL SOURCEBOOK: CONFRONTING CORRUPTION: THE ELEMENTS OF A NATIONAL INTEGRITY SYSTEM 287 (2nd ed. 2000). *See also* Rafael Di Tella & Ernesto Schargrodsky, *The Role of Wages and Auditing During a Crack-Down on Corruption in the City of Buenos Aires*, 46 J.L. & ECON. 269 (2003) (using a measure of corruption to determine whether an anti-corruption measure, payment of above market wages to those vulnerable to corruption, is effective).

[19] *See, e.g.*, Michael Johnston & Sahr J. Kpundeh, *The Measurement Problem: a Focus on Governance*, 2 F. ON CRIME & SOC'Y 33, 33 (2002).

[20] For example anticorruption measures that restrict the free speech rights of a campaign donor require a showing of a "compelling [government] interest" if they are to avoid being struck as unconstitutional. *See, e.g.*, Nixon v. Shrink Mo. Gov't PAC, 528 U.S. 377 (2000). In determining whether anticorruption campaign laws serve a "compelling interest" the courts have noted that the academic literature has been inconclusive on the issue. *Id.* at 394. More rigorous measures of corruption could provide a more solid platform for showing a compelling government interest, providing a justification for further reaching anticorruption measures.

[21] Currently, the most widely cited measures of corruption are subjective indices such as Transparency International's Corruption Perception Index and the World Bank's Control of Corruption Indicator which measure the perceived levels of corruption. *E.g.*, TRANSPARENCY INTERNATIONAL, CORRUPTION PERCEPTIONS INDEX 2004 (2003), *available at* http://ww1.transparency.org/pressreleases_archive/2004/2004.10.20.cpi.en.html; Daniel Kaufmann et al., WORLD BANK, GOVERNANCE MATTERS IV: GOVERNANCE INDICATORS FOR 1996-2004 (World Bank, Policy Research Working Paper No. 3630, 2005). These indices are essentially based upon surveys of people thought to be knowledgeable about the business environment in their country. The surveys typically record the participants' perception of corruption in their country and are used to produce indexes of the relative perceived corruption between countries. While these surveys allow us to broadly compare corruption levels between nations, they have several issues limiting their usefulness in monitoring actual corruption. Firstly, these surveys are usually done at the country level, leaving the question remaining about specifically where corruption is present and how it operates. Secondly, it is uncertain whether survey respondents accurately perceive the level of corruption. Since corruption is usually hidden and varies significantly by institution it is not obvious that even the most eagle-eyed observer is capable of accurately gauging the general level of corruption. Rather perceptions are likely to be greatly influenced by factors such as media reporting. Thirdly, the survey population may not be representative. Surveys are only distributed to selected persons, usually senior executives at corporations and NGO's, about half of whom respond, thus respondents may not be representative of the general population. And fourthly, it is unknown how truthfully respondents report levels of corruption. Particularly in industries where

incentives to keep their activities hidden from the public and making the direct collection of reliable data on corruption difficult. Further, indirect evidence of corruption is often sparse. Unlike traditional property or personal crimes, there is typically no smashed glass, empty vault, or hospital visit to mark that an act of corruption has taken place. Typically, even victims of corruption are unaware a crime has taken place. Contrasting traditional crimes where the impact is direct and concentrated, the impact of corruption is dispersed throughout entire communities with each individual victim feeling only a small indirect effect through higher taxation or lesser provision of services.[22]

And even with greater transparency, corruption may not be easy to spot. Acts of corruption are frequently entangled in highly subjective decisions, such that it may not be possible to detect individual cases of corruption even with perfect transparency.[23] Even participants in corruption may be unaware that their actions are corrupt. A large and robust psychology literature shows that in interpreting ambiguity, people tend to form an interpretation that serves their self-interest. The process is an unconscious one, caused by decision-makers focusing on favorable information and evaluating such information uncritically[24] while disregarding and being overly critical of contradictory information.[25]

This psychological bias occurs even among well-meaning professionals.[26] For example, Moore *et al.* show that even professional auditors, a group with the utmost professional and ethical duty to remain unbiased, have their judgment swayed by the assigning to them of a 'client' despite conscious efforts to remain neutral.[27] This mental shift occurs

---

participants seek to keep corruption concealed, we should be concerned about whether respondents report corruption levels truthfully. *E.g.*, Tina Søreidel, IS IT RIGHT TO RANK? LIMITATIONS, IMPLICATIONS AND POTENTIAL IMPROVEMENTS OF CORRUPTION INDICES 5-6 (2005), *available at* http://www.cmi.no/publications/publication.cfm?pubid=1973; Johan Graf Lambsdorff, *Measuring Corruption—The Validity and Precision of Subjective Indicators, in* MEASURING CORRUPTION 95 (Charles Sampford et al. eds., 2006).

22 *See* Johnston, *supra* note 19 at 35.

23 For example, government officials often have to select the 'best' from among several well qualified and similar goods or service providers. This is particularly so when the attributes of the providers being evaluated are subjective. In such cases, many different decisions could be justified making it hard to determine if corruption was a factor in the decision.

24 *See, e.g.*, Keith J. Holyoak & Dan Simon. *Bidirectional Reasoning in Decision Making by Constraint Satisfaction*, 128 J. EXPERIMENTAL PSYCHOL.: GEN. 3 (1999); Derek J. Koehler, *Explanation, Imagination, and Confidence in Judgment*, 110 (3) PSYCHOL. BULL. 499 (1991).

25 P. H. Ditto & David F. Lopez, *Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions*, 63 J. PERSONALITY & SOC. PSYCHOL. 568 (1992).

26 Don A. Moore et al., *Auditor Independence, Conflict of Interest and the Unconscious Intrusion of Bias* (Harv. Bus. Sch. Working Paper No. 03-116, 2003).

27 *Id.*

despite subjects having no knowledge of their bias, nor any intent to act in a biased manner.[28]

Therefore we should anticipate that public officials may be biased toward those actions that can confer them with benefits and we should anticipate this bias may be unknown even to the officials themselves. Knowledge that benefits, such as campaign contributions, travel junkets, or lobbying business will accrue to a public official, their family, friends, or staff, may unconsciously if not consciously cause the official to act in accordance with the wishes of donors. And irrespective of any lack of mal intent, such acts are still corrupt and are still damaging to the public interest.

This corruption that remains hidden on the individual level may prove more visible when viewed at the aggregate level using statistical techniques. Indeed, a small literature has emerged, primarily from the social sciences, which utilizes statistical techniques to illuminate impropriety in areas where direct detection is difficult.[29] Following the lead of this literature, this paper proceeds with two attempts to measure corruption in government decisions related to public finance.

## II. CAMPAIGN CONTRIBUTIONS AND CORRUPTION

### II.1. *Campaign Contributions*

Campaign contributions are monetary or in-kind gifts given to candidates, or associated organizations, to aid the candidate in winning or retaining office. Campaign contributions assist candidates in financing the

---

[28] *Id.*

[29] A number of studies have examined procurement auctions to find evidence of bid rigging between suppliers of goods and services. *See, e.g.*, Robert H. Porter & J. Douglas Zona, *Detection of Bid Rigging in Procurement Auctions,* 101 J. POL. ECON. 518 (1993) (providing evidence of bid rigging in contests to supply construction services in Long Island, New York by comparing the expected pattern of bidding, that companies with underutilized crews should bid lower for jobs than companies with busy crews, to the actual pattern of bidding). For an overview of this literature, *see* Patrick Bajari & Garrett Summers, *Detecting Collusion in Procurement Auctions*; 70 ANTITRUST L.J. 143 (2002).

A system of bout rigging, among elite sumo wrestlers has been illuminated. *See* Mark Duggan & Steven D. Levitt, *Winning Isn't Everything: Corruption in Sumo Wrestling*, 92 AM. ECON. REV. 1594 (2002) (noting that wrestlers 'on the bubble,' the point where they receive great rewards for winning one additional match, win significantly more often than would otherwise be expected when their opponents have little incentive to win; and that they lose significantly more often than expected the next time the pair meet, indicative of a quid pro quo).

Most closely related to this paper there have been a set of studies that investigate the effect of campaign contributions on politicians' voting patterns. These studies have had varying levels of success and have produced conflicting results. We will examine them in greater detail below in the campaign contributions section of this paper.

expenses associated with running for office. It is widely believed that financial resources play a significant role in determining the eventual winner of election contests.[30] So the receipt of campaign contributions is very important to most candidates.

However, there is widespread concern that candidates become beholden to donors and once elected grant donors favors to repay debts and to ensure future contributions are forthcoming.[31] To combat such concerns, a myriad of regulations govern campaign contributions.[32]

Campaign finance regulations have simultaneously attracted criticism in the literature for being too strict, stifling free speech,[33] and for being too lax, forcing candidates to become corrupt and captive to donors.[34] But this debate has been largely uninformed by any rigorous evidence, perhaps because research in the area is still inconclusive.

A number of researchers have looked at the question of whether campaign contributions buy votes. An abundant literature shows that politicians who receive contributions for a particular cause tend to vote in favor of legislation favoring that cause.[35] Such findings, however, are subject to a simultaneity problem. That is, contributions may influence political decisions, but political decisions may also influence contributions. So these findings may simply reflect the fact that politicians who honestly support a cause will tend to attract more financial support from donors who support the same cause.

Three clever methods have emerged to avoid this simultaneity problem.

The first method, employed by a number of researchers, is to use instrumental variables regression analysis.[36] Essentially they attempt to predict a legislator's true opinion on bills using variables such as the

---

[30] *See, e.g.*, Alan Gerber, *Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables*, 92 AM. POL. SCI. REV. 401 (1998); Gary C. Jacobson, *The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments*, 34 AM. J. POL. SCI. 334 (1990). *But see* Steven D. Levitt, *Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House*, 102 J. POL. ECON., 777 (1994) (looking at instances where the same two candidates repeatedly faced each other, finds that campaign spending has little effect on the outcome of the election).

[31] *See, e.g.*, James M. Snyder, Jr., *Long-Term Investing in Politicians; or, Give Early, Give Often*, 35 J.L. & ECON. 15 (1992).

[32] *See, e.g.*, Federal Election Campaign Act of 1971, as amended, Pub. L. No. 92-225, 2 U.S.C. §§ 431-42, 451-6.

[33] Stephen E. Gottlieb, *The Dilemma of Election Campaign Finance Reform*, 18 HOFSTRA L. REV. 213 (1989).

[34] Fred Wertheimer & Susan Weiss Manes, Campaign *Finance Reform: A Key to Restoring the Health of Our Democracy*, 94 COLUM. L. REV. 1126 (1994).

[35] For a summary of this literature see Stephen Ansolabehere et al., *Why is There so Little Money in U.S. Politics?*, 17 J. ECON. PERSP. 105, Table 2 (2003).

[36] For a summary of this literature, *see id.*

legislator's constituency, political affiliation, and voting history.[37] They then compare the predictions generated to a legislator's actual voting record to determine if campaign contributions induced the legislator to change their vote. But the effectiveness of these predictors in capturing personal differences between legislators is questionable, leaving it unclear whether the simultaneity problem has actually been overcome.[38] Indeed, largely due to differences in the selection of predictors, this method has produced very mixed results ranging from findings that campaign contributions have a strong effect to campaign contributions having no effect.[39]

A second method, devised by Bronars and Lott, is to look at retiring federal Congressmen. Bronars and Lott assume that retiring Congressmen have less reason to act corruptly since they do not need to raise funds for reelection, thus will reveal their true opinion in their final term. Comparing voting records during a Congressman's earlier days of holding office to their voting records for their final term should reveal the impact of campaign contributions. Using this method they find no change in the voting patterns of retiring Congressmen and conclude that vote buying does not take place.[40] But the assumption that retiring Congressmen are more likely to express their true opinion in their final votes rather than following the wishes of past donors is questionable. Firstly, retiring Congressmen still have incentives to follow the wishes of donors. Some Congressmen during the time of the study were able to keep unspent campaign contributions upon retirement.[41] And those that could not retain contributions could use unspent contributions to donate to other members of their party, an act that may be motivated by altruism or a desire to gain favor with colleagues.[42] Further, retiring Congressmen may seek continued relationships with donors, such as board positions or consulting engagements. Secondly, retiring Congressmen might feel duty-bound to continue to support past donors. The purchase of votes may not be a purely economic transaction. Over time a donor and a Congressman may engage in repeat interactions and build up social ties, which will influence their decisions even when economic incentives are removed. Thirdly, retiring Congressmen might wish to appear consistent with their past actions. Finally, even if retiring politicians have less incentive to swap votes for

---

37 *Id.*

38 Thomas Stratmann, *Can Special Interests Buy Congressional Votes? Evidence from Financial Services Legislation*, 45 J.L & ECON. 345 (2002).

39 Ansolabehere, *supra* note 35, at Table 1.

40 Stephen G. Bronars & John R. Lott Jr., *Do Campaign Contributions Alter How a Politician Votes? Or Do Donors Support Who Value the Same Things They Do?*, 40 J.L. & ECON.317 (1997).

41 *See, e.g.*, Tim Groseclose & Jeff Milyo, *Buying the Bums Out: What's the Dollar Value of a Seat in Congress*, Research Paper No. 1601, Graduate School of Business, Stanford University (1999), 1-2.

42 Steven Weiss, *Legacy War Chests: what happens to all that money when a law maker leaves office or quits a race, available at* http://www.opensecrets.org/newsletter/ce71/04warchests.asp.

contributions, this may be counterbalanced by their being less subject to the discipline of their electorate as the Congressmen do not face reelection.[43]

The third method, devised by Stratman, looks at changes in voting patterns between two votes on similar financial services reform bills seven years apart. Stratman finds that those receiving campaign contributions, particularly younger Congressmen, tend to change their position in line with their donor's ideology. Stratman suggests that this change was due to the Congressmen being bought, and that younger Congressmen, who have a greater need for funds, are bought more easily.[44] However, these results are also compatible with an interpretation that donors are giving to the Congressmen most open to changing their position on the matter, and that younger Congressmen tend to be more open to reversing their position than older Congressmen.

Thus, results in the area have been mixed, and no clear consensus has emerged from the literature as to whether campaign contributions induce politicians to act corruptly and alter their decisions.[45] Further efforts have been concentrated on the federal legislature, and there have been no attempts to measure whether political contributions cause corruption in the executive branch of government.

### II.2. *Public Finance*

Campaign contributions' corrupting influence has been of particular concern in the public finance industry, particularly in the underwriting of municipal bonds. Municipal bonds are debt instruments whereby the issuer, a state or municipal entity, raises money by selling investors the right to receive some greater sum of money in the future.[46] Currently, about $2 trillion of these bonds are outstanding in the United States, with an additional $300 billion being issued every year.[47]

Underwriters act as middlemen between issuers and investors. The underwriter purchases the bonds from the issuer and in turn resells the bonds to investors, charging fees to both the issuer and investors.[48] The

---

43 This may be best illustrated the anecdotal example of the pardoning of fugitive tax evader Marc Rich by President Bill Clinton in the last hours of his presidency. Clinton is widely thought to have improperly pardoned Rich because of large contributions made by Rich's wife to Clinton, his party and his wife. While Clinton, as an outgoing president, had no need to solicit further contributions and might have been expected to act impartially, circumstances suggest that this might not have been the case. Jessica Reeves, *The Marc Rich Case: A Primer*, TIME, Feb. 13, 2001.

44 Stratmann, *supra* note 38, at 361-62.

45 Stratmann, *supra* note 38, at 346; Nixon v. Shrink Mo. Gov't PAC, 528 U.S. 377, 394 (U.S. 2000) (Souter J, finding that finding no academic consensus has coalesced on this issue).

46 JUDY WESALO TEMEL, THE FUNDAMENTALS OF MUNICIPAL BONDS, 1-2 (5th Ed. 2001).

47 THOMSON FINANCIAL.

48 TEMEL, *supra* note 46, at 2.

business is highly profitable, last year earning underwriters over $2 billion in fees.[49] And although services provided by the large investment banks that dominate the business are largely homogenous[50], the underwriters are reluctant to engage in price competition.[51] Consequently, when price is not the only criteria, the choice between the leading underwriters is largely based on additional highly subjective criteria, such as perceived knowledge of the issuer, reputation, and relationship with the issuer. Thus although there is strong competition for work,[52] the issuer makes highly discretionary decisions, creating great opportunity and temptation for corruption.

With so much discretion given to issuers in selecting underwriters, rumors of corruption have swirled around the industry, coupled with the occasional corruption scandal.[53] During the mid 1990s, many believed that that *pay-to-play*, a practice whereby underwriters made campaign contributions[54] to politicians in return for underwriting business, was widespread.[55] This suspicion of campaign contributions was not universal, with many believing that contributions by underwriters were made for legitimate purposes and were not designed to improperly influence government. These defenders of campaign contributions dismissed the occasional reports of corruption as being isolated incidents and unrepresentative of the general state of the industry.[56]

---

[49] Eddie Baeb, *Bond Advisors Irk Wall Street, Curb Fees on Public Finance*, BLOOMBERG, Jun. 30, 2005.

[50] Product innovation is difficult in investment banking since successful innovations can be copied instantly. Anand Bharat & Alexander Galetovic, *Relationships, Competition, and the Structure of Investment Banking Markets*, J. INDUST. ECON. (forthcoming).

[51] Bharat Anand & Alexander Galetovic, *Strategies That Work When Property Rights Don't*, in INTELLECTUAL PROPERTY AND ENTREPRENEURSHIP 23, (Gary Libecap ed., 2004).

[52] For example, 50 underwriters bid on a recent Illinois pension fund issue. Justin Pope, *With Bond Business Lobbying Thriving, Regulators Mull Ban*, ASSOCIATED PRESS, Mar. 29, 2004.

[53] John Racine, *Politicians, Money, and Getting Business: Schwartz Case Opens a Door on Prudential*, BOND BUYER, Jul. 31, 1992 at 1; John Racine, *Cash Is Politically Nourishing, and Bond Firms Deliver*, BOND BUYER, Feb. 27, 1992 at 12.A.

[54] Underwriters were thought to be among the largest municipal politics donors, although the actual magnitude of contributions is unknown. Due to haphazard state disclosure requirements and the fact that contributions could be made in several forms including in the name of a investment banker, the name of a family member or through a PAC, systematically determining what portion of contributions came from underwriters is difficult and prone to underreporting. The only study known to the author that investigates this area looks at the state of Louisiana. The study finds that more than 10% of contributions made to the State Treasurer, the Senate President (whom subsequently became Treasurer) and a member of the Louisiana State Bond Council came from underwriters. Steven Filling, *Influence Peddling and the Municipal Market: An Investigation of Campaign Contributions And Accounting Disclosures* (1994) (unpublished Ph.D. dissertation, Louisiana State University), *available at* http://panopticon.csustan.edu/filling/dissertation.toc.htm.

[55] *See, e.g.*, Stephen J. Hedges & Warren Cohen, *The Politics of Money - How Underwriters of Municipal Bonds Win their Business*, U.S. NEWS AND WORLD REPORT, Sept. 30, 1993.

[56] *See* Blount v. SEC, 61 F. 3d 938 (D.C. Cir. 1995), cert. denied, 116 S. Ct. 1351 (1996) (petitioner arguing that the prohibition of campaign contributions by underwriters to municipal officials

We now look at whether corruption via campaign contributions was present in the public finance world of the early 1990's and attempt to measure the magnitude of corruption. We determine this by gauging the response of issuers to the sudden cessation in the making of campaign contributions that occurred following the banning of campaign contribution in 1994. In particular, we focus on the movement between the two primary bond issuing processes available to issuers: the competitive issue, and the negotiated issue.[57]

Competitive issues involve bonds being sold by the issuer to the underwriter via an auction, with the bidder offering the best terms winning the right to buy the bonds.[58] This method is recommended for most issues,[59] primarily because the direct competition between underwriters leads to better terms for the issuer.[60] Since selection is purely based on price, the issuer has little discretion in selecting an underwriter and there is little opportunity for corruption.[61]

Negotiated bonds involve the issuer first selecting an underwriter and then negotiating bond price and fees with them.[62] Here the decision as to which underwriter to select is highly discretionary, being based on subjective criteria such as the reputation of the underwriter, perceived expertise, and the relationship between the underwriter and the issuer.[63] It is generally accepted that the lack of competition results in higher fees being paid to the winning underwriter, higher interest rates, and

---

from whom they solicit business did meet the "strict scrutiny" requirements since there was no systematic evidence of campaign contributions causing corruption).

57 Note that there is a third method of selling bonds, private placement which accounts for a negligible 2% of bonds issued.

58 TEMEL, *supra* note 46, at 2.

59 *See, e.g.*, GOVERNMENT FINANCE OFFICERS ASSOCIATION, *Selecting and Managing the Method of Sale of State and Local Government Bonds* (1994), *available at* http://www.gfoa.org/services/rp/debt/debt-selecting-managing.pdf.

60 *See, e.g.*, David S. Kidwell & Michael D. Joehnk, *Comparative Costs of Competitive and Negotiated Underwritings in the State and Local Bond Market*, 34 J. FIN. 725 (1979) (finding that interest rates for negotiated bond issues are between 0.15% and 0.35% higher than for competitive issues for RB's); William Simonsen & Mark D. Robbins, *Does It Make Any Difference Anymore? Competitive Versus Negotiated Municipal Bond Issuance*, 56 PUB. ADMIN. REV. 57 (1996) (finding that negotiated bond issues have an interest rate 0.30% higher than competitive issues). Note there is a minority view suggesting no difference; *see, e.g.*, Jun Peng & Peter F. Brucato, Jr., *Another Look at the Effect of Method of Sale on the Interest Cost in the Municipal Bond Market: A Certification Model*, *34* PUB. BUDGETING & FIN. 73 (2003).

61 *See, e.g.*, Kenneth A. Kriz , *Comparative Costs of Negotiated Versus Competitive Bond Sales: New Evidence from State General Obligation Bonds*, 43 Q. REV. FIN. & ECON. 191, 192 (2003); Joe Mysak, *It's Time for Philly to Try a Bond Auction Sale*, BLOOMBERG, Jul. 20, 2005, *available at* http://www.bloomberg.com/apps/news?pid=10000039&sid=atZoGH.vobh4&refer=columnist_mysak.

62 TEMEL, *supra* note 46, at 7.

63 *See id.*

consequently higher total cost of debt.[64] Consequently, negotiated bond issuances are only recommended in a few limited situations such as where the issue is particularly large or complex.[65] However, given the great discretionary power placed in the hands of the issuers it provides a good opportunity for acting corruptly.

While competitive bonds were historically the dominant mode of issue, there has been a general trend toward the use of a negotiated process, a trend which has been widely criticized.[66] The trend has been attributed to three factors. Firstly, it is thought issuers are becoming lazy and opting for negotiated issues as they involve less work for the issuer.[67] Secondly, complex bond issues are becoming more common (perhaps unnecessarily so[68]) and these complex issues are perceived to require a competitive issue.[69] And thirdly and most worryingly, it is suspected that some issuer's agents choose to use a negotiated underwriting in order to gain the opportunity to extract some personal benefit from the competing underwriters.[70] Since the issuer has greater discretion in selecting an underwriter, use of negotiated bonds enables politicians to 'reward' any underwriter who has benefited or will benefit them in the future and to 'punish' those who do not,[71] ensuring that a stream of contributions from the bidding underwriters will be forthcoming.

Following a series of scandals, the fifteen largest underwriters agreed to cease making campaign contributions in late 1993.[72] This cessation led

---

64 *See, e.g.*, Rueben Kessel, *A Study of the Effects of Competition in the Tax Exempt Bond Market*; 79 J. POL. ECON. 706 (1971) (analyzing historical data to show that similarly situated counties that used competitive offerings had lower costs of debt over those that used negotiated offerings); Mark D. Robbins, *Testing the Effect of Sale Method Restrictions in Municipal Bond Issuance: The Case of New Jersey*, 22 PUB. BUDGETING & FIN 40 (2002) (using a natural experiment following the restriction of negotiated offerings in New Jersey to show that cost savings are achieved by switching to competitive issues).

65 GOVERNMENT FINANCE OFFICERS ASSOCIATION, *supra* note 59, at 1-2; Darrell Preston, *Cities Shun Finance Competition, Victimizing Taxpayers*; BLOOMBERG, Feb. 24 2005, *available at* http://www.saberpartners.com/press/articlepages/BN_1_3_05.html.

66 *E.g.*, Preston, *supra* note 69; Joe Mysak, *Municipal Auction Sale Nears Extinction in 17 States*, BLOOMBERG, Oct. 8, 2004, *available at* http://quote.bloomberg.com/apps/news?pid=10000039&refer=columnist_mysak&sid=aAoWx4iei7Rc.

67 Mysak, *supra* note 66.

68 Lawrence Harris & Michael S. Piwowar, *Municipal Bond Liquidity*, American Finance Association 2005 Philadelphia Meetings, February 19, 2004 (suggesting that municipalities would benefit from issuing simpler bonds).

69 Mysak, *supra* note 66.

70 *Id.*

71 *See, e.g.*, Preston, *supra* note 69; Joe Mysak, *Underwriting Bonds Moves Into the Political Circle*, BLOOMBERG, Mar. 25, 2005.

72 Arthur Levitt, *The State of the Municipal Securities Market*, 9 GOV'T FIN. REV. 33, 33-35 (1993).

to the promulgation of Municipal Securities Rulemaking Board[73] (hereafter MSRB) General Rule 37 (hereafter G-37) in early 1994;[74] which prohibits all underwriters and their employees from conducting business in states where they have made campaign contributions in the past two years[75] and prohibits contributions in the two years following winning of bond business. This rule effectively ended campaign contributions by underwriters.

Thus, the enactment of G-37 provides a natural experiment. Before G-37 there were incentives for issuers to select negotiated bond offerings so that they could personally profit from higher contributions. And after G-37 there was less incentive for issuers to do so as they would not receive contributions from the competing underwriters. Thus, comparing the period before and the period after the enactment of G-37 provides an opportunity to observe if the making of campaign contributions altered the decisions of issuers and consequently, whether they were acting corruptly.

### II.3. *A Model of Corruption Through Campaign Contributions*

Formalizing the above thought, if we assume that $NEGOTnat_t\%$ of issuers have a natural preference for negotiated sales at time $t$, that is they would choose a negotiated sale without any inducement of the possibility of a campaign contribution from an underwriter. Hence, $(1\text{-}NEGOTnat_t)\%$ have a natural preference for competitive bonds at time $t$.

Now if $CORRUPT\%$ of issuers are corrupt, such that they will select a negotiated sale irrespective of their natural preference in order to gain the opportunity to extract campaign contributions from competing underwriters, and assuming these corrupt issuers are randomly distributed among the population of issuers, then the proportion of negotiated offering that will be observed ($NEGOTobs$) at time $t$ is:

$$(1) \quad NEGOTobs_t = NEGOTnat_t + (1 - NEGOTnat_t) \times CORRUPT \times LAW_t + \epsilon$$

---

[73] The Municipal Securities Rulemaking Board is a congressionally established board that develops rules to regulate the underwriting, trading and selling of municipal securities. The majority of the board is composed of members of the municipal securities industry.

[74] Order Approving Proposed Rule Change Relating to Political Contributions and Prohibitions on Municipal Securities Business, Exchange Act Release No. 33,868, 56 SEC Docket 1045 (Apr. 7, 1997).

[75] Note there is a de minimus exception for contributions up to $250 made by employees to candidates for whom the employee was eligible to vote.

Where $LAW_t$ is a dummy variable equal to 1 when laws permit campaign contributions to be made and equal to 0 when laws prohibit these practices. And where $\varepsilon$ is an error term.

Now since there has been a general increase in the use of negotiated bonds over time, we make the assumption that $NEGOTnat_t$ can be modeled as a linear function of time. i.e. $NEGOTnat_t = B_1 + B_2 \times t$, where $B_1$ and $B_2$ are constants. Thus:

$$(2) \quad NEGOTobs_t = B_1 + B_2 \times t + (1 - (B_1 + B_2 \times t)) \times CORRUPT \times LAW_t + \in$$

II.4. *Data and Summary Statistics*

Aggregate data for all long-term competitive and negotiated municipal bond issues for each year from 1982 to 2003 as collected by Thompson Financial is used.

Data from the years 1985 and 1986 is excluded from the sample. In mid-1986, new laws came into effect that changed the way in which municipal bonds were taxed.[76] Consequently, there was a surge in municipal bond issues prior to this date as issuers rushed to take advantage of the old more favorable regime.[77] These extra issues were disproportionately negotiated issues. Thus, these years are excluded from the sample in order to avoid this anomaly's affecting results. Including these two years still gives statistically significant results, but with lower coefficient and $R^2$ values. (See Appendix 1)

Time $t$ is measured in years, and the year 1994 is set as t=0. The dummy variable $L_t$ is set as 1 for years prior to 1994; and set as 0 for the year 1994 and beyond.

This yields a data set of 20 years, comprising over $4 trillion of bonds.

II.5. *Results*

The regression results were calculated using an iterative ordinary least squares procedure and a three, five, and ten year event window and are shown below in Table 1. The results for the five year window are also shown below visually in Figure 1.

---

[76] Tax Reform Act of 1986, Pub. L. 99-514.

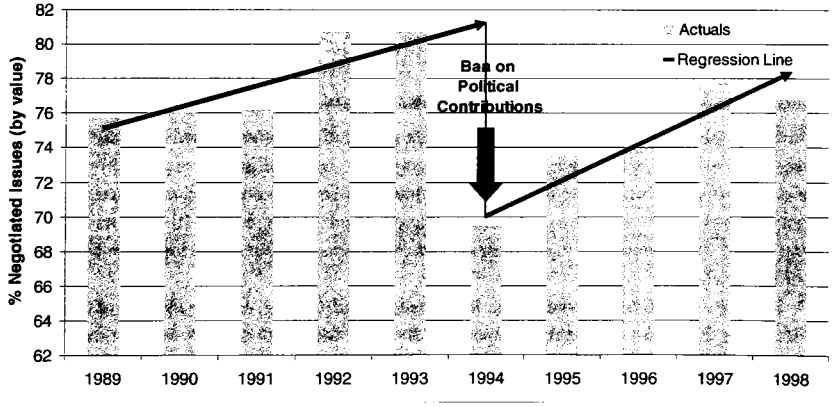
[77] John M. Quigly & Daniel L. Rubinfeld, *Private Guarantees for Municipal Bonds: Evidence for the Aftermarket*, 44 NAT. TAX J. 29, 30 (1991).

TABLE 1—IMPACT OF CAMPAIGN CONTRIBUTIONS ON MODE OF ISSUE

| | Three Year Window | Five Year Window | Ten Year Window |
|---|---|---|---|
| $B_1$ | 0.70*** (0.0082) | 0.70*** (0.009) | 0.72*** (0.010) |
| $B_2$ | 0.027* (0.0082) | 0.0204*** (0.0035) | 0.008*** (0.0018) |
| CORRUPT | 0.43* (0.080) | 0.382*** (0.055) | 0.30*** (0.066) |
| Observations | 6 | 10 | 20 |
| Adjusted $R^2$ | 0.89 | 0.83 | 0.50 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parenthesis)

FIGURE 1 - IMPACT OF CAMPAIGN CONTRIBUTIONS ON MODE OF ISSUE (FIVE YEAR WINDOW)




The results show that, as would be expected in the presence of corruption, the use of negotiated bonds dropped suddenly following the banning of campaign contributions. Results imply that about one-third of

municipal bond issuers (measured by value) acted corruptly, willing to switch from their natural preference for a competitive issue to a negotiated issue in order to gain the opportunity to realize a private gain in the form of campaign contributions. The results display a high degree of statistical significance and are robust to the selection of the event window.

The results suggest the prohibition of campaign contributions was effective in reducing a large portion of the corruption in the industry. A rough estimate suggests that the enacting of G-37 by reducing corruption, saved municipalities $500 million in real interest costs for bonds sold in the first year it was enacted alone.[78]

Results also confirm that use of negotiated bonds has been generally growing. The results suggest that negotiated bonds have been increasing in market share by 1-2% annually.

### II.6. *Alternative (Non-corruption) Hypotheses*

#### II.6.1. Underwriters Support Candidates With Similar Ideologies

An alternative explanation of the result is that underwriters make campaign contributions to candidates who have ideologies similar to their own and that these contributions influence elections. By giving to candidates who favor the use of negotiated offerings these candidates have more resources than counterparts who support competitive offerings and thus are more likely to win and retain office. But if this was the only mechanism by which donors influence government, the drop in negotiated offerings following the banning of campaign contributions should have been gradual. Given that municipal and state elections are typically held only every two to five years, and given the tendency of incumbents to retain office,[79] we would expect any results via this mechanism to become

---

[78] The regression results suggest that the share of negotiated offerings fell by about $18 billion because of G-37. If we assume that negotiated offerings have a rate of interest that is 0.3% higher than that of competitive bonds (see note 64), then the higher annual interest costs due associated with the use of negotiated bonds are $54 million. Assuming that the bonds had a term of 20 years and applying a discount rate of 10%, the present value of these extra interest payments over the life of the bond is in the order of $500 million.

[79] *See, e.g.*, Alan I. Abramowitz et al., *Incumbency, Redistricting, and the Decline of Competition in U.S. House Elections*, Conference Paper for ANNUAL MEETING OF THE SOUTHERN POLITICAL SCIENCE ASSOCIATION (2005) (noting that U.S. House incumbents had a 99% re-election rate in the 2004 house elections); Stephen Ansolabehere & James M. Snyder, Jr., *Using Term Limits to Estimate Incumbency Advantages When Officeholders Retire Strategically*, ELECTION L.J. (forthcoming) (estimating that federal incumbents have a 10% advantage over challengers); Stephen Ansolabehere and James M. Snyder, Jr., *The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Office*, forthcoming ELECTION L.J. (finding that incumbent state legislative and executive candidates enjoy an advantage of 4-10% of the vote over non-incumbents); ZOLTAN L. HAJNAL, PAUL G. LEWIS &

apparent only over many election cycles. Instead, we see a sudden drop in the fraction of negotiated offerings that is more compatible with the corruption hypothesis.

### II.6.2. Bond Volatility

Some observers of municipal bonds have attributed the sudden drop in the fraction of negotiated bonds to an increase in the volatility of bonds in 1994. There are some suggestions that the fraction of negotiated bonds drops during times of interest rate volatility as negotiated issuers have more flexibility in when to issue their bonds and attempt to time the market. [80] This explanation is not consistent with the data. While there was an increase in volatility in 1994, the increase was relatively small as compared with the drop in negotiated offerings. Adding a bond volatility term to the regression does not significantly affect the corruption result. It does, however, diminish the $B_2$ coefficient result, since bond volatility and time are substantially collinear. (See Appendix 2).

### II.6.3. Access

It has been argued extensively that campaign contributions do not buy favorable decisions, but rather buy access to decision makers so donors can present their case. Firstly, charging for access is in itself a form of corruption. Selectively awarding or denying access to the ear of government on the basis of payments is surely a misuse of public office for private gain, fitting squarely within our definition of corruption. This is particularly the case when government is awarding contracts, as they should be providing access to all bidders, not just those whom provide payments.[81]

Secondly, if all that campaign contributions bought was access, the voice of underwriters should not have significantly changed the decisions of government officials. Officials already utilize independent financial advisors for most transactions. These independent advisors should provide sufficient information to the issuer for them to make intelligent decisions as to what mode of issue is optimum. Additional comments from underwriters, particularly when they are known to have a vested interest in

---

HUGH LOUCH, MUNICIPAL ELECTIONS IN CALIFORNIA: TURNOUT, TIMING AND COMPETITION, 55-60 (2002) (finding that municipal incumbents had a 80% in California retained office); Timothy B. Krebs, *The Determinants of Candidate Vote Share and the Advantage of Incumbency in City Council Elections*; 42 AM. J. OF POL. SCI. 921 (1999).

[80] *See, e.g.*, Jun Peng & Kenneth A. Kriz, *Do Municipal Bond Issuers Use Negotiated Offerings to Time the Market?* (2003), *available at* http://kkriz.lunarpages.com/Files/Vita%20-%20Long.doc.

[81] *See Reformers Question Ethics of 'Paying for Access*, N.Y. L.J., June 15, 1995, at 5.

suggesting negotiated bonds would be expected to have minimal value to officials, given the opposing interest of the underwriters.[82]

And finally, as discussed previously, the use of negotiated bonds is widely considered the inferior choice in most situations due to the higher underwriters fees and interest for negotiated issues. Thus, greater access alone should not cause government officials to so heavily favor negotiated issues.

### II.7. *Normative Implications*

This finding, that corruption via campaign contributions was widespread in public finance in the mid 1990s, suggests some normative implications.

#### II.7.1. Public Finance

Since a significant fraction of state government officials involved in underwriting seem willing to act corruptly when choosing between underwriters, it should raise suspicions over the awarding of contracts to other public finance service providers. For example, bond lawyers, financial advisors, and financial printers are also hired by municipal entities in the underwriting process. As with underwriters of negotiated bonds, the selection of these service providers is highly discretionary and as with underwriting there have been rumors of corruption and the occasional scandal.[83] But, unlike bond underwriters, the professional groups governing bond lawyers, financial advisors, and financial printers undertook no G-37 type reforms to prevent the making of political contributors to government officials from whom they solicit business.[84] Given the evidence of corruption via campaign contributions by these same government officials in the underwriter context it would be surprising if a similar level of corruption was not associated with the choice of legal, financial advisory

---

82 GENE M. GROSSMAN & ELHANAN HELPMAN, SPECIAL INTEREST POLITICS, 103-18 (2001).

83 *See, e.g.*, Joe Mysak, *Philadelphia Pay to Play Trial Ends, Fallout Begins*, BLOOMBERG, May 10, 2005 (reporting conviction of the Philadelphia State Treasurer for awarding work to selected municipal bond lawyer, financial advisers and printers in return for an assortment of goods and services).

84 Bond lawyers have strongly opposed such a ban. *See, e.g.*, Comment Letter, National Association of Bond Lawyers, Re NYC Bar's Proposed "Pay-to-Play" Rule (Feb. 12, 1998). For a summary of the failed attempts to ban campaign contributions by bond lawyers see Brian C. Buescher, *ABA Model Rule 7.6: The ABA Pleases the SEC, But Does Not Solve Pay to Play*, 14 GEO. J. LEGAL. ETHICS 139 (2000); Pauline A. Schneider, *Attempts to Limit Campaign Contributions by Municipal Finance Professionals and Lawyers in* PRACTICING LAW INSTITUTE - CORPORATE LAW AND PRACTICE COURSE HANDBOOK SERIES (Sept. 2004).

and printing services in the municipal bond industry. Consequently, a ban on campaign contributions by all public finance professionals to government officials from whom they seek business seems justified. Certainly further investigation of these industries is warranted.

### II.7.2. Outside Public Finance

Outside of public finance, these results should add some empirical weight to the chorus of concerns about campaign contributions. These results are particularly relevant in considering situations where those with strong interests in the outcome of decisions make contributions to government officials who have highly discretionary decision making power over these decisions. Some examples of such situations include congressional committee members' receipt of contributions from the PACs of the industries they are responsible for regulating and the executive's receipt of political contributions from contractors competing to provide services to government.

## II.8. *Epilogue*

Despite the initial success of G-37, the results show that following the brief reversal of the trend away from competitive issues, the decrease in competitive issues has continued (See Table 1 and Figure 1), and currently sits at pre G-37 levels. Since we know that post G-37 direct campaign contributions have been closed off as an avenue for corruption, we may harbor suspicions that corruption has reemerged in other forms.

Indeed, it is reported that, with direct campaign contributions blocked off as a means by which underwriters could influence politicians, a number of indirect opportunities have become more attractive. Anecdotal reports suggest that post G-37 underwriters began to donate to politicians favorite charities,[85] sponsor quasi-political conferences,[86] make campaign contributions in the names of family members not bound by G-37,[87] and contribute to bond referendum campaigns[88] in order to curry favor. However, the practice of greatest concern has been the hiring of lobbyists.

The next part of this paper examines the current controversy surrounding the use of lobbyists in the post G-37 era.

---

[85] Charles Gasparino & Josh P. Hamilton, *Cash Flow: 'Pay to Play' Is Banned, But Muni-Bond Firms Keep the Game Going*, WALL ST. J., May 13, 1998, A1.

[86] *Id.*

[87] Notice of Proposed Rule Change Concerning Indirect Violations, Exchange Act Release No. 52,235, 70 Fed. Reg. 48214-02 (Aug. 16, 2005).

[88] Bill Ainsworth, *Governor's Charm Playing Well in Capitol*, THE SAN DIEGO UNION-TRIB., Feb 22, 2004.

## III. LOBBYISTS AND CORRUPTION

### III.1. *Lobbying*

Lobbying, broadly defined, is "any attempt by an individual or group to influence governmental decisions."[89] While lobbyists and lobbying are most associated with the legislative branch of government,[90] lobbying of the executive[91] and judicial branches[92] occurs as well.

Lobbying in its ideal embodiment serves a socially valuable role bridging the gap between industry and government. Due to the vastness of government, elected officials often make decisions on subjects outside their realm of expertise. Lobbyists can supply the expertise and diversity of opinion that help officials better understand the implications of their actions. And lobbyists, who are typically former government officials, may also provide industry with a better understanding of the workings of government so that industry can better respond to government's needs. Thus, lobbyists may aid both industry and government by enhancing the flow of information between the two, thereby increasing the quality of law-making and the quality of industry's offerings to government.[93]

While lobbying can be socially valuable, there is a concern that lobbyists may be conduits for corruption, offering government officials private benefits in return for the rendering of favorable decisions. These private benefits could take a number of forms, including money; fundraising; goods and services; influence over others; promises of future employment; provisions for allies, friends, or family; or forbearance from supporting political enemies.

While an interested party may directly engage in corruption, the use of a lobbyist presents a number of structural advantages, making it a particularly attractive method for engaging in corrupt practices. First, the

---

89 Based on the definition provided by *Improving the Legislative Process: Federal Regulation of Lobbying*, 56 YALE L. J. 304, 306 (1947).

90 In 2004 over 17,000 firms lobbied members of Congress. Center For Public Integrity LobbyWatch, http://www.publicintegrity.org/lobby/ (last visited Aug. 29, 2006) (aggregating data from disclosures made by lobbyists to the Senate Office of Public Records).

91 In 2004 over 2,000 firms lobbied various departments of the federal executive, aggregating data from disclosures made by lobbyists to the Senate Office of Public Records. *Id.*

92 Professional rules prohibit lawyers from lobbying judges regarding pending cases. *E.g.*, Rule 3.5, *Model Rules of Professional Conduct*. Further, even the appearance of a conflict of interest should lead to a judge removing herself from the judicial process. *E.g.*, Canon 3(A)(4), *Code of Conduct for United States Judges*. So unsurprisingly lobbying of judges is not greatly publicized. But there are anecdotal reports of lobbying of judges. *E.g.*, David McKean, TOMMY THE CORK—WASHINGTON'S ULTIMATE INSIDER FROM ROOSEVELT TO REAGAN 267-72 (2004) (describing an attempt by prominent Washington lawyer and lobbyist Tommy Corcoran, to lobby Supreme Court Justices Brennan and Black to rehear a case in which his firm represented the petitioner).

93 BRUCE C. WOLPE ET AL., LOBBYING CONGRESS: HOW THE SYSTEM WORKS, 1 (1996).

rules covering third party lobbyists' conduct are often less restrictive than those covering their employers. Both legal rules and societal norms prevent corporations deploying lobbying resources in the most effective manner, these restrictions are often relaxed or non-existent for third party lobbyists. These relaxed restrictions enable lobbyists acting on behalf of clients to use resources more effectively, in a manner which may be of greater concern to the public welfare than corporations acting on their own behalf. For example, in the municipal finance industry, underwriters currently cannot make campaign contributions to those from whom they solicit business. But an underwriter is free to hire lobbyists who face no such restrictions.[94]

Second, lobbyists are subject to less critical monitoring than their principals. Not only are lobbyists subject to less restrictive rules, they are less likely to be detected when they break the rules. The large corporations that typically use lobbyists are often publicly owned and have numerous internal and external reporting requirements. Further, they are often subject to a great deal of public scrutiny. For example, modern accounting requirements for publicly listed companies make it difficult for renegade employees to pay large bribes using corporate funds without securing the cooperation of multiple levels of the corporate bureaucracy, increasing the chance of detection. Further, since a corporation's individual employees typically capture only a small fraction of the rewards of any one transaction, they are unlikely to have sufficient incentives to use personal funds to bribe a member of government. Conversely, the typical lobbyist, a solo practitioner, is subject to no internal scrutiny and considerably less external scrutiny, allowing them to flout rules with less fear of detection.

Third, lobbyists can utilize economies of scale. Since a single lobbyist can assist multiple clients, they may spread the cost of influencing decision makers. Further, a corrupt decision maker may prefer to 'wholesale' influence to a single lobbyist rather than cut individual deals with a series of interest parties as this is both time consuming and increases the chance of detection.[95]

Fourthly, using a lobbyist decreases the costs if corruption is detected. The consequences for firms that act corruptly can be severe, as the firms incur massive financial and reputation costs. The costs of using a lobbyist who is subsequently revealed as corrupt are much less severe. When a lobbyist is detected as being corrupt, a firm may immediately terminate and distance themselves from the lobbyist, claiming total ignorance of their lobbyist's malfeasance. Indeed, it is possible that many firms truly do not know that the lobbyists they hire use corruption to influence government.[96]

---

94 *See* Lynn Hume, *Muni Consultants Stepping Up Contributions to Issuer Officials*, BOND BUYER, Sep. 30, 2003, at 1.

95 *E.g.* Mysak, *supra* note 84.

96 If corrupt lobbyists are more effective than other lobbyists, then underwriters will tend to hire corrupt lobbyists simply by selecting on the basis of effectiveness.

Such separation cannot be achieved when an individual in a corporation is found to be using corruption to influence governments, as there is a greater presumption that the corporation knew or should have exercised greater supervision.

This combination of greater potency, lower costs, and lower risks make third-party lobbyists a more effective option for corruptly influencing government decisions than directly engaging in corruption for many firms. Thus, we should expect that, if corporations are using corruption to have favorable decisions rendered, many will use lobbyists. Indeed, in a competitive marketplace, corporations that do not use corrupt lobbyists may not survive.[97]

Despite the importance of lobbying and its ubiquity, there appear to be no previous attempts to measure or detect whether lobbyists use corruption to achieve their ends. The closest antecedent is the following set of studies that use event studies to show close connections between publicly listed corporations and government are perceived to be of value to the corporation.

For example, Gely and Zardkoohi show that the market price of firms whom retain law firms as lobbyists show abnormal gains when one of the partners at that firm obtains a federal cabinet position. This effect disappears when anti-lobbying laws are enacted.[98]

Roberts shows that companies located in the state of the ranking Democrat on the Senate Armed Services Committee, experience abnormal losses in their stock price following the senator's sudden, unexpected death. Conversely companies in the state of the Senator's political heir experienced abnormal gains.[99]

Finally, Fisman shows that shares of Indonesian companies perceived as connected to the corrupt President Suharto rise and fall with reports of Suharto's improving or deteriorating health.[100]

But these studies do not distinguish between gains realized for legitimate reasons such as better information and communication between corporations and government and gains realized for corruption. There is also a lesser concern that these studies rely on valuations from financial markets for their data. Such valuations reflect the beliefs of market participants. While market participants are likely to be better informed than the general populace, their beliefs may not reflect the underlying reality, particularly where there is limited information as is the case with corruption.

---

97 Andrei Shleifer, *Does Competition Destroy Ethical Behavior?*, 94 AM. ECON. REV. 414 (2004).

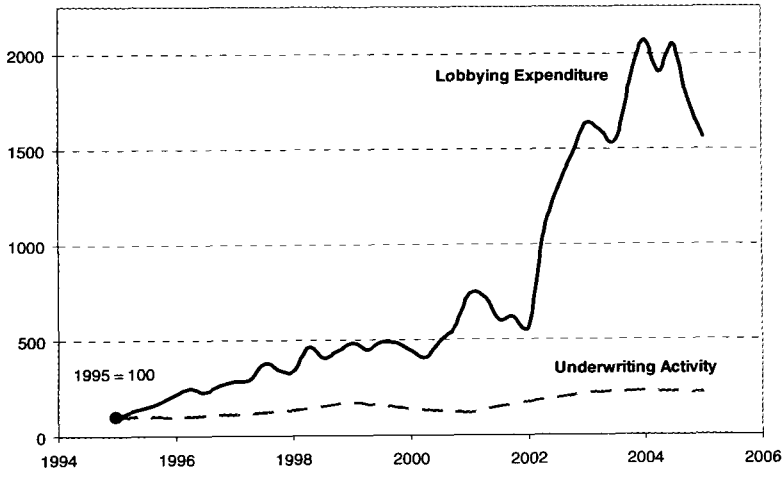
98 Rafael Gely & Asghar Zardkoohi, *Measuring the Effects of Post-Government Employment Restrictions*, 3 AM. L. AND ECON. REV. 288 (2001).

99 Brian E. Roberts, *A Dead Senator Tells No Lies: Seniority and the Distribution of Federal Benefits*, 34 AM. J. POL. SCI. 31 (1990).

100 Ray Fisman, *Estimating the Value of Political Connections*, 91 AM. ECON. REV. 1095 (2001).

The use of lobbyists in the public finance industry has grown dramatically since the enactment of G-37 in 1994, from less than $1 million annually to a high of $17 million in 2003. Lobbying expenditure has retreated slightly since 2003, partly because the underwriting firm that had previously been the biggest spender on lobbying found itself embroiled in a lobbying scandal and abruptly ceased all lobbying activity,[101] and partly because of a small decrease in underwriting volumes. While municipal underwriting activity has also increased, roughly doubling during the period from $160 billion to $350 billion, the growth in lobbying expenditure has been outstripped by a ratio of 7 to 1. This rapid growth has led some to believe that lobbyists are being used to replace campaign contributions in improperly influencing decision makers.

FIGURE 2 - GROWTH IN LOBBYING EXPENDITURE BY TOP TEN UNDERWRITERS




A few scandals have emerged from the use of lobbyists.[102] In response to these scandals, the MSRB in June 2004 proposed that lobbyists could only be hired as "affiliated persons" and would be subject to the same rules of conduct as the underwriters who hire them. This proposal would make

101 Evan Halper, *Insiders Are Cashing In on State's Bond Market*, LOS ANGELES TIMES, Sept. 27, 2004, at A1.

102 *E.g.*, Mysak, *supra* note 84.

underwriters responsible for supervising their lobbyists, and would subject the lobbyist to a myriad of rules including the G-37 ban on campaign contributions. Response to the proposed ban was overwhelmingly negative from the investment banking community.[103] A primary complaint was that the ban was not based on any substantial proof that lobbyists were generally acting improperly and that the few proven cases of corruption were not representative of the industry.[104] At the time of writing, the proposal was before the SEC awaiting approval.[105] The underwriting industry has threatened to challenge the proposal, arguing it is an unconstitutional limitation of their First Amendment free speech rights, should the proposal be approved.[106]

To test whether corruption is in fact wide-spread in these lobbying activities, this paper looks at differences in patterns of lobbying expenditures. It exploits the geographic heterogeneity in the supply of corrupt decision makers, comparing the use of lobbyists in states known to have corrupt government to the use of lobbyists in states known to have honest government. If lobbying works, as feared, through corruption, we would expect it to be most effective in states with corrupt government as these states would be most receptive to lobbyists' overtures and consequently more would be spent on lobbying in these corrupt states. Conversely, if lobbying worked according to its ideal by spreading information and improving communications, we would expect more uniform levels of lobbying expenditures.

### III.2. *A Model of Corruption Through Lobbyists*

To formalize the above intuition, the following model of the market for lobbyists is introduced.

Consider a single winner takes all contest for some government decision $b$, in state $s$, yielding profit $P_{bs}$ for the party that has the decision rendered in their favor.[107] Assume all parties incur equal costs in competing and normalize the model so that all losing parties face a loss of zero.

---

103 Letter from the Bond Market Association, *Comment to Proposed Rule G-38 Amendment*, to the Municipal Securities Rule Making Board (Jun. 4, 2004); Letter from the Bond Market Association, *Comment to Second Draft of Proposed Rule G-38 Amendment*, to the Municipal Securities Rule Making Board (Dec. 15, 2004 ).

104 Municipal Securities Rulemaking Board Notice 2005-17 (Mar. 17, 2005). *Amended by* Municipal Securities Rulemaking Board Notice 2005-41 (Aug. 9, 2005).

105 *See id.*

106 *See id.*

107 The assumption that the winning contestant will receive profit $P_{bs}$ conceals two underlying assumptions. First is assumes that there is in fact a real profit and that it will not be dissipated by the contestants competing based on price. In the public finance context this is reasonable given that it is

For each contest assume that there are only two lobbyists available for competing parties to hire:[108] one lobbyist who improves information sharing and communication with the decision-maker (hereafter information lobbyist) and who extracts fees $F_{Ibs}$ from the party that hires them, and a second lobbyist who uses corruption (hereafter corrupt lobbyist) to influence the decision maker and extracts fees $F_{Cbs}$ from the party that hires them.[109]

Now, if for each contest to provide services, there is a $C_{bs}\%$ chance that the decision-maker will act corruptly, selecting any party that hires the corrupt lobbyist they are affiliated with, and will otherwise choose the best candidate. And assuming both competing parties and lobbyists are rational profit maximizing actors. Then, in a competitive market for lobbyists, underwriters will compete to hire the corrupt lobbyist, until at equilibrium the benefit of hiring the corrupt lobbyist equals the cost.[110] Hence:

$$(3) \qquad F_{Cbs} = C_{bs} \times P_{bs} + \in$$

Where $\varepsilon$ is an error term.

Now assuming for each contest a competing underwriter can also improve their chance of winning by an additional $I_{bs}\%$ by hiring the information lobbyist whom extracts fees $F_{Ibs}$,[111] then, in a competitive market, the information lobbyist would be paid fees:

---

known to be a lucrative field and that the contestants (large investment banks) are known to be reluctant to compete on price. Anand, *supra* note 52.

Secondly, it assumes the model assumes profit $P_{bs}$ will be the same for each winning bidder. Again, this is reasonable the in following application, where we restrict our sample to the largest ten underwriters whom each do many public finance underwritings each year, which are relatively routine and are thus likely to have similar costs.

[108] This assumption simplifies calculations. Allowing multiple lobbyists significantly complicates the analysis without adding significant predictive power.

[109] This assumption of a single corruption lobbyist and a single communications lobbyist simplifies calculations as we do not have to consider the effects of competition between lobbyists. Relaxing the assumption adds complexity without adding predictive power.

[110] Note that it is not essential that the underwriter knows how effective a lobbyist will be in advance. (i.e. the underwriter does not need to know $C_{bs}$ or $I_{bs}$) Indeed the underwriter need not even know whether the lobbyist they hire uses the corruption or communication mechanism (or both). By use of contingency fees or bonus payments, the hiring underwriter can utilize ex-post information (i.e. whether the underwriter won or lost) to determine how effective the lobbyist was and consequently the appropriate fee. Alternatively if the same underwriter hires the same lobbyist on multiple occasions they can use past performance to estimate what would be an appropriate fee for the future.
Note that the majority of lobbyists receive bonuses or contingency fees.

[111] *See* note 112.

$$(4) \qquad F_{Ibs} = P_{bs} \times I_{bs} + \in$$

Summing equations 3 & 4 gives:

$$(5) \qquad F_{Tbs} = P_{bs} \times I_{bs} + C_{bs} \times P_{bs} + \in$$

Where $F_{Tbs}$ is the total amount paid in lobbying fees.[112]

Now, assuming that $C_{bs}$ is randomly distributed within each state, that $I_{bs}$ is randomly distributed both within and between states, and that both $C_{bs}$ and $I_{bs}$ are independent variables,[113] it can be shown that for any given time period that:

$$(6) \qquad \sum_{b} F_{bs} \approx I_{\bar{b}} \times \sum_{b} p_{bs} + C_{\bar{bs}} \times \sum_{b} p_{bs} + \in$$

Where $C_{\bar{bs}}$ is the average chance that a decision maker in state s will act corruptly and select the party that hires a corrupt lobbyist. And $I_{\bar{b}}$ is the average chance of wining due to hiring an information lobbyist.

III.3. *Data and Summary Statistics*

III.3.1. Corruption Data

The rate of corruption ($C_{\bar{bs}}$) among government officials for the selection of underwriters in each state is unknown since it is likely many acts of corruption go undetected. As a proxy for this variable, an assumption is made that the rate of corruption in each state is proportional to the number of federal public corruption convictions made in the that state in the preceding decade, per capita. More formally:

---

112 Since by definition there are only two lobbyists, $F_{bs} = C_{bs} + I_{bs}$.

113 This assumption could be loosened to assuming that $I_{bs}$ is independent of $C_{bs}$.

(7) $C_{\bar{bs}} = k_1 \times R_s$

Where $k_1$ is some constant and $R_s$ is the rate of Federal Public Corruption Convictions in the preceding decade per million persons in state $s$.

Data for Federal Corruption Convictions is based upon surveys conducted by the Department of Justice's Public Integrity Section that record the number of successful public corruption convictions secured in each US attorneys office[114] between 1990 and 2003.[115] Federal data is used rather than state data, as there is presumably less inter-state variation in the type, detection, prosecution, and recording of corruption between Federal US Attorney's offices than between State Prosecutor's offices. This is important because interstate variations in the type, detection, prosecution, and recording of corruption will lead to a violation of our assumption that the rate of corruption in each state is proportional to the number of recorded convictions.

A small fraction of Federal U.S. Attorney's offices, comprising 3% of office year observations, do not report data in some years. Where this is the case, these observations are replaced by the average number of corruption convictions made by that office in the remaining years of the sample. This yields a data set of just under 12,000 federal public corruption convictions.

Approximately half of these convictions are of Federal officials, a quarter are of state and municipal officials, and a quarter are of private citizens. The vast majority of convictions are not related to public finance. However, it is assumed that states with a greater prevalence of general corruption also tend to have a greater prevalence of municipal bond corruption.

The population of each state was determined using figures from the 1990 and 2000 Census.[116] For the years 1991-1999 and 2001-2003 where there is no census data, a geometric rate of population growth is assumed.

The average number of federal public corruption convictions for the United States is approximately 3.2 federal public corruption convictions per million persons, per year. States range from just over 7 convictions per million persons (North Dakota and Mississippi) to less than 1 conviction per million persons (Nebraska and Oregon). (See Figure 3). By visual inspection, the rate of corruption does not appear to exhibit any geographic patterns.

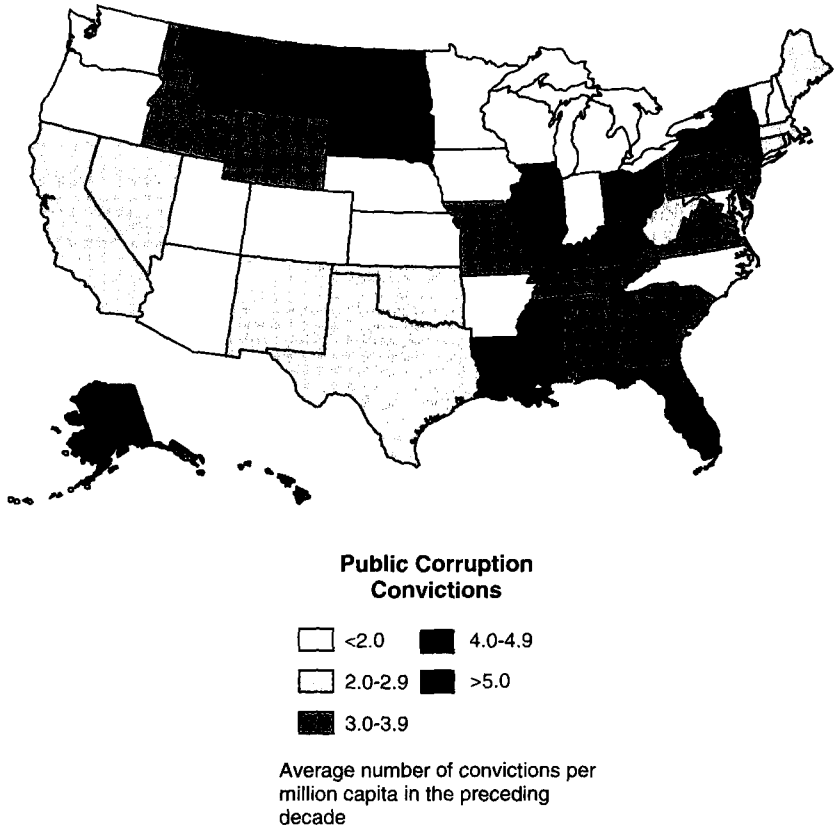
---

[114] US attorney's offices are divided by state. Some larger states have multiple offices. For example California is divided into four offices.

[115] DEPARTMENT OF JUSTICE—CRIMINAL DIVISION, REPORT TO CONGRESS ON THE ACTIVITIES AND OPERATIONS OF THE PUBLIC INTEGRITY SECTION (1999-2003).

[116] http://www.census.gov

FIGURE 3 - RATE OF FEDERAL CORRUPTION CONVICTIONS (1990 - 2003)




Data on lobbying fees ($F_{Tbs}$) was collected from disclosures made to the Municipal Securities Rulemaking Board (MSRB). Since 1994, underwriters have been required to disclose all payments made to lobbyists under MSRB General Rule 38 (hereafter G-38). All disclosures made by the ten largest underwriters for the years from 2000 to 2004 were hand coded for the time and amount of payments, the area of operation, the form of compensation, size and location of campaign contributions made by the lobbyists, the amount of business that was attributed to the lobbyist, and location of the lobbyist.

Data from 2000 to 2004 was used because disclosures prior to 2000 do not uniformly contain some critical information making them less useful. The MSRB strictly enforced compliance with the G-38 lobbyist reporting regarding lobbyist fees in these latter years, punishing non-compliance with

draconian sanctions.[117] Thus, we should expect the data in this date range to be highly reliable.

Data for the ten largest underwriters, all divisions of large Wall Street investment banks, are used since smaller underwriters are substantially different in their operations. Smaller underwriters, typically regional banks, do most of their business in a limited area.[118] They have an advantage over the larger national underwriters in these areas in that they typically have pre-existing personal relationships with local government officials and are accorded affirmative discrimination.[119] However, they suffer the disadvantage of not being regarded as highly as the national firms and are not as well-compensated.[120] Also, smaller firms may hire consultants for technical assistance, which would require disclosure under G-38, while it is unlikely that the ten largest underwriters, who have plenty of in-house technical capabilities, would hire such consultants. Consequently, the benefit the smaller underwriters receive from hiring a lobbyist, either for information or corruption purposes, is substantially different from that received by the large underwriters violating the model assumptions. Thus the smaller underwriters are excluded from the sample. The ten large underwriters who are used in the sample account for over of 70% of municipal bond underwriting by market value.[121]

Fees are allocated to the quarter in which the fees were actually paid. Expenses were included as lobbying fees, but excluding them from the analysis does not yield substantially different results.

Lobbying fees are allocated first to the state in which the disclosure stated the lobbyist was targeted. To break ties where a disclosure reports more than one target state, fees are allocated to the state in which the lobbyist reported their business address, because, presumably, that state is where the lobbyist wields the most power. The very small fraction of fees that still could not be allocated to a state using this procedure is excluded from the sample.

The territories, District of Columbia, Puerto Rico, Guam, and the Virgin Islands, were all excluded from the sample. The first is excluded since as the political capital of the United States, it may be home to many lobbyists who conduct business that does not relate to the territory. The later three are excluded because these territories are very different markets

---

117 Jon B. Jordan, *The Regulation of "Pay-to-Play" and the Influence of Campaign contributions in the Municipal Securities Industry*, 1999 COLUM. BUS. L. REV. 489, 553-60 (1999) (giving examples of trivial and unintentional violations that have resulted in costly bans); Lynn Stevens Hume, MSRB Asked to Allow Waiver for Inadvertent First G-37 Offense, BOND BUYER, Mar. 6, 1997, 5.

118 Alexander W. Butler, Distance Still Matters: Evidence from Municipal Bond Underwriting 2-3 (2004) (unpublished manuscript, available at http://ssrn.com/abstract=334080).

119 *Id.*

120 *Id.* at 3, 9.

121 *See, e.g.*, THOMSON FINANCIAL, PUBLIC FINANCE 2004 (2005) (note in allocating issues, full value is allocated to the lead underwriter).

than the fifty states and are expected to have different characteristics both for bond issues and corruption.

This yields a data set of over 2100 lobbyist quarters with a total of over $37 million of fees. The average fee paid to a lobbyist during a quarter was $20,000 and the median fee $10,000. These fees were paid in a variety of ways. About 70% of lobbyists received a monthly retainer, averaging $6,000 per month with a similar median. Around 50% were eligible for expenses, with the average expense claim being $2400 and the median $800. Around 30% were eligible for discretionary bonuses. And about 20% were paid a fixed share of the revenue obtained by the lobbyist with the average and median share being about 15% of net revenue. A small number of lobbyists were paid an hourly rate or were given one-time payments.[122]

Lobbyists reported winning business for their clients on average once every 2.5 quarters. However, note that the quality of information supplied with regard to winning business was poor. Some firms did not complete this section of the disclosure form, and others used a stock sentence denying that any business was obtained by consultants.[123] It is thus likely that this figure understates the amount of business won by consultants.

Just under 10% of lobbyists report making campaign contributions in any given quarter, donating an average of $6,000. Again, the quality of information regarding lobbyists campaign contributions appeared poor, likely understating the extent of contributions.

### III.3.2. Underwriter's Fees Data

The profitability of each bond transaction ($P_{bs}$) is private information, but as a reasonable proxy we can assume that profit is some fixed fraction of the bond issue size.[124] Formally it is assumed that:

---

122 Note that the total exceeds 100% since a lobbyist may be compensated using more than one mode of payment. For example a lobbyist might receive both a monthly retainer and a discretionary bonus.

123 *See, e.g.*, MERRILL LYNCH, PIERCE, FENNER & SMITH, INC., Municipal Securities Rulemaking Board Form G-37/G-38, Document ID A0976033 through A0976053 (2001-05).

124 Generally larger issues carry lower per bond issued. However this is at least partly counterbalanced by the lower cost of underwriting larger issues per bond that are incurred by the underwriter through economies of scale. Further, there is some conservatism built into this assumption as we would expect that corrupt issuers pay higher fees than legitimate ones. Thus assuming a linear rate of profit will tend to understate the level of corruption.

$$(8) \qquad P_{bs} = k_2 \times B_{bs}$$

Where $k_2$ is some constant. And $B$ is the value of the bond issue $b$ in state $s$.

Aggregate data for the total value of long term bond issues for each state in the sample years of 2000 to 2004 was sourced from The Bond Buyer, the industry newspaper which compiles figures from Thompson Financial. This gives a data set of 250 state, year observations comprising $1.5 trillion worth of debt.

Substituting equations 7 & 8 into equation 6 and rearranging gives:

$$(9) \qquad \sum_b F_{Tbs} = k_2 \times I_{\bar{b}} \times \sum_b B_{bs} + k_1 \times k_2 \times R_s \times \sum_b B_{bs} + \in$$

Rearranging this gives:

$$(10) \qquad \sum_b F_{Tbs} = k_{INF} \times \sum_b B_{bs} + k_{COR} \times R_s \times \sum_b B_{bs} + \in$$

Where $k_{INF}$ and $k_{COR}$ are constants.[125] In interpreting the results the reader will note that:

- $k_{INF}$ represents lobbying based on communication of information. $k_{INF}$ is equal to the number of dollars spent on this type of lobbyist per dollar of bonds issued, and
- $k_{COR}$ represents lobbying based on corruption. $k_{COR}$ is equal to the number of dollars spent on this type of lobbyist per dollar of bonds issued, per federal public corruption conviction per million capita per year.

### III.4. *Results*

Since the sample data contains both large and small states it exhibits mild heteroscedasticity (White's Test is positive at the $p>0.15$ level). Hence, the regression analysis is conducted using the Weighted Least Squares (WLS) method, weighting by state population. Results are also

[125] Note, $k_{INF} = k_2 \times I_{\bar{b}}$ and $k_{COR} = k_1 \times k_2$.

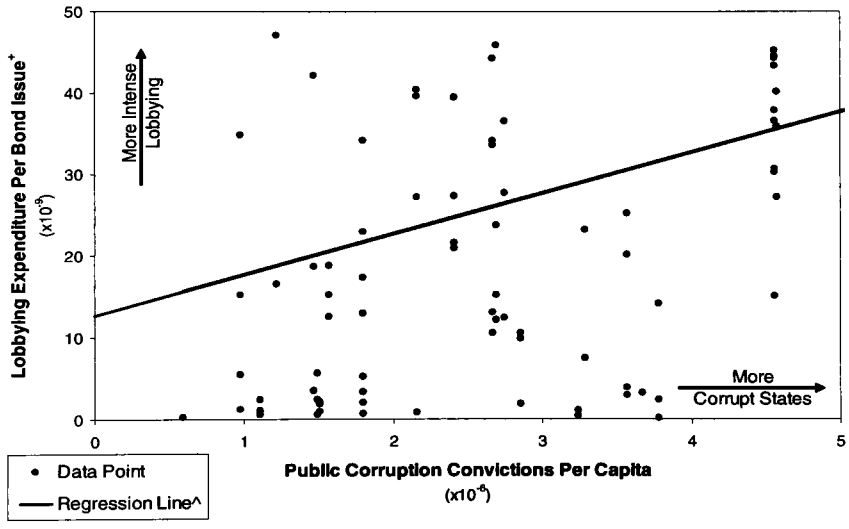
shown using the more widely used ordinary least square (OLS) method for reference. Hereafter, all references will be to the WLS results as these are expected to the most reliable.[126] Noted that the OLS method yields similar and statistically significant results. (See Table 2 below).

The regression constant is fixed at zero; not fixing the constant produces a constant that is not statistically significant. The results are shown below.

Results are also reported for the subset of state years where a state issued more than $3 billion worth of bonds in a calendar year, since there may be concern that low issuing states do not attract the attention of the large underwriters that are the focus of this investigation. This subset is about half the size of the full sample.

A visual representation of this subset of the data is also included in Figure 4.

FIGURE 4 - LOBBYIST CORRUPTION REGRESSION PLOT*




*Data includes all states years in which more than $3 billion worth of bonds was issued

^The regression line is calculated using the Ordinary Least Squares method

+A small positive scatter is added to the y-axis to increase clarity

126 W. H. GREENE, ECONOMETRIC ANALYSIS, 225-27 (5th ed. 2002).

TABLE 2 - LOBBYIST CORRUPTION REGRESSION RESULTS

| | All State Years | | State Year Issuing > $3b/yr | |
|---|---|---|---|---|
| | **WLS** | **OLS** | **WLS** | **OLS** |
| $K_{INF}$ ( x $10^{-6}$) | 17.87 *** (4.35 ) | 12.73 ** (4.15 ) | 17.86 ** (5.99 ) | 12.78 * (5.77 ) |
| $K_{COR}$ ( x $10^{-6}$) | 3.93 ** (1.26 ) | 5.00 *** (1.15 ) | 3.94 * (1.73 ) | 5.02 ** (1.59 ) |
| **Observations** | 250 | 250 | 133 | 133 |
| **Adjusted $R^2$** | 0.75 | 0.70 | 0.75 | 0.71 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parenthesis)

As seen in Table 2 above, both the corruption coefficient $k_{COR}$ and the information coefficient $k_{INF}$ display a high degree of statistical significance, implying that both communication and corruption are significant reasons for hiring a lobbyist.

The results imply that for every billion worth of bonds sold, $18,000 is spent on lobbyists who use communication and information to help their clients win business. And in a state with an average rate of corruption,[127] for every billion dollars worth of bonds issued about $12,000 will be spent on a corrupt lobbyist, which is about 40% of the total amount spent on lobbying. The amount spent on corrupt lobbyists increases by about $3,000 dollars per billion dollars of bond issued per federal corruption conviction per million persons per annum. Thus, we estimate that in the most corrupt states $21,000 is spent on corrupt lobbyist per billion dollars of bond issued as compared to $3,000 in the least corrupt states.

If we add a time term to the regression as follows we can analyze the changes in information lobbying and corrupt lobbying over the sample period:

[127] Recall that the average rate of corruption was 3.2 convictions per million capita per year.

$$(11)\ \sum_b F_{Tbs} = \left(k_{INF} + Y \times k_{INFg}\right) \times \sum_b B_{bs} + \left(k_{COR} + Y \times k_{CORg}\right) \times R_s \times \sum_b B_{bs} + \epsilon$$

Where Y is the year and 2000 is set as Year 0. Where $k_{INFg}$ represents growth in information lobbying each year and is equal to the number of extra dollars spent on information lobbyists per dollar of bonds issued per year. And where $k_{CORg}$ represents growth in corruption lobbying and is equal to the number of extra dollars spent per year on this type of corrupt lobbying per dollar of bonds issued, per federal public corruption conviction per million capita per year. The results are shown below.

TABLE 3 - LOBBYING REGRESSION RESULTS - WITH TIME FACTOR

| | All States | | States Issuing > $3b/yr | |
|---|---|---|---|---|
| | WLS | OLS | WLS | OLS |
| $K_{INF}$ ( x $10^{-6}$) | 48.6 $^{+++}$ (9.99 ) | 21.91 $^{++}$ (9.04 ) | 48.79 $^{+++}$ (13.8 ) | 21.7 $^{+}$ (12.68 ) |
| $K_{INFg}$ ( x $10^{-6}$) | -12.6 $^{+++}$ (3.54 ) | -4.33 (3.29 ) | -12.58 $^{++}$ (4.91 ) | -4.22 (4.59 ) |
| $K_{COR}$ ( x $10^{-6}$) | -6.28 $^{++}$ (2.85 ) | -0.62 (2.50 ) | -6.31 $^{+}$ (3.94 ) | -0.473 (3.50 ) |
| $K_{CORg}$ ( x $10^{-6}$) | 4.23 $^{+++}$ (1.04 ) | 2.44 $^{+++}$ (0.93 ) | 4.24 $^{+++}$ (1.45 ) | 2.39 $^{+}$ (1.30 ) |
| Observations | 250 | 250 | 133 | 133 |
| Adjusted $R^2$ | 0.76 | 0.72 | 0.76 | 0.72 |

+/++/+++ Denotes significance at 0.10/0.05/0.01 level
(Standard Error in Parenthesis)

The results show that the amount spent on corrupt lobbyists has been growing strongly during the sample period from close to zero during 2000 to about $32,000 per billion worth of bonds for an average state in the last year of the sample.

Conversely, the amount spent on information lobbyists has fallen during the sample period from about $50,000 per billion worth of bonds issued to close to zero.

This implies that the hiring of lobbyists has become increasing related to corruption over the last five years, and increasingly unrelated to providing information and improved communication.

### III.5. *Alternative (Non-Corruption) Hypotheses*

#### III.5.1. Endogeneity

There may some concerns that there is some endogeneity in the model. For example, it may be that less populous states have higher rates of corruption due to their being fewer persons checking up on government and that less populous states also have a greater need for lobbyists due to the relative unfamiliarity of Wall Street underwriters with these states.

To test for endogeneity, a number of variables are tested in the model, including population density, state GDP per capita, government spending per capita, state debt per capita, and state credit rating. None prove to be statistically significant, nor does their inclusion significantly alter results. (See Appendix 3).

#### III.5.2. Allocation Methodology

It might also be possible that the allocation methodology caused the results. For example, there may be concerns that the method of allocating fees where there were multiple targets to the state containing the lobbyists office may have favored regional centers such as New York, California, and Illinois as lobbyists may tend to locate their offices in these centers. To allay these concerns a variety of different allocation methodologies were tested, including allocating fees evenly between all reported targets, and by completely excluding all ambiguous disclosures with multiple state targets. No substantial difference was found when using any of these alternative allocation methodologies. (See Appendix 4).

#### III.5.3. Underwriters Mistaken Perceptions

It is also possible that underwriters believe they need to hire lobbyists to win work in corrupt states when this is not the case. Such a mistaken belief would cause underwriters to spend more on lobbying in corrupt states in the mistaken belief that it was necessary, which would lead to the observed result that underwriters in corrupt states spend more on lobbying.

But given the high cost of lobbyists, taking up to 20% of an underwriters gross fees, we would expect underwriters to have strong incentives to discover which consultants are effective and to shed those that are not. We would expect that underwriters have the ability to learn which lobbyists are effective given that: (1) underwriters receive rapid feedback, learning if they won underwriting business within weeks of submitting their bids[128] and (2) underwriters can learn from their competition—G-38 creates transparency in the market for lobbyists, an underwriter can observe the track record of a lobbyist prior to hiring them. Consequently, we should expect that even if this hypothesis were initially correct, it would correct itself over time. However, the opposite trend is apparent in the data, as lobbying spending has shifted over the last five years from being substantially uniform across states to being greater in corrupt states.

#### III.5.4. Communicating with Corrupt Governments is Harder

Another complementary hypothesis may be that corrupt governments benefit more from a communications consultant. Perhaps governments in corrupt states are less sophisticated or have more complex financing needs such that there is greater benefit in these states from hiring a communications lobbyist. Thus, underwriters would hire more communications consultants and/or be willing to pay them greater amounts, creating a greater amount spent on fees in the corrupt states.

There is little in the empirical results that would prove or disprove this hypothesis, but it seems unlikely for the following two reasons. First, most issuers hire financial consultants whose purpose is to help government understand the options open to them and to advise them as to which option they should take. Second, even if the elected official with decision making power does not have finance experience, they will almost certainly be surrounded by a staff that does have such experience.

### III.6. *Normative Implications*

#### III.6.1. Public Finance

Beyond public finance, the results should tend to make us more wary of lobbyists using improper means to influence government. Greater empirical study of the means by which lobbyists operate in other industries seems appropriate.

---

128 Note the feedback will contain some *noise*, with some false positives when an underwriter wins despite an ineffectual lobbyist and false negatives when an underwriter loses despite having a very influential lobbyist.

### III.6.2. Corruption Reemergence

Both the first and second studies when taken together provide an anecdotal example of how corruption evolves.

In the public finance industry, a temptation for corruption was created by having government officials make highly subjective decisions regarding lucrative contracts. This created incentives for competitors to attempt to influence these officials by offering them private benefits in the form of political contributions. We observe that once one corrupt practice, the making of political contributions, was prohibited, corruption reemerged. With the underlying incentives for corruption unchanged, lucrative contracts awarded by highly discretionary criteria, corruption quickly re-emerged in a non-prohibited form under the guise of lobbying.

This reemergence suggests that attempts to reduce corruption by only prohibiting specific acts rather than attending to the underlying conflict of interest will only achieve short-term results and will merely cause corruption to shift to another form.

A preferable solution would reduce the underlying incentives and opportunities for corruption. For example in the municipal bond context, forcing government to use competitive issues would provide a simple way of reducing the underlying conflict. The competitive bidding among underwriters would reduce the profitability of the winner, reducing the temptation to engage in corruption. Further, the lower discretion given to decision makers in selecting the underwriter would diminish the opportunity for corruption.

## CONCLUSION

This paper provides two examples of how public corruption can be measured on the macro scale. Looking at the public finance industry, an event study methodology shows that direct campaign contributions in the early 1990's were corrupting decision makers and that corruption fell following a ban on contributions. Utilizing the geographic heterogeneity in the supply of corrupt decision-makers provides evidence that corruption has reappeared in the form of lobbyists.

It is hoped that these methods cast some light on corruption in government, and that these findings will encourage greater use of statistical techniques in illuminating and monitoring public corruption, thus bringing better information into the debate on how this scourge should be treated.

APPENDIX ONE

TABLE 4—IMPACT OF CAMPAIGN CONTRIBUTIONS ON MODE OF ISSUE (INCLUDING 1985 & 1986)

| | Three Year Window | Five Year Window | Ten Year Window |
|---|---|---|---|
| $B_1$ | 0.70*** (0.012) | 0.70*** (0.009) | 0.74*** (0.015) |
| $B_2$ | 0.027* (0.0082) | 0.0204*** (0.0035) | 0.004 (0.0026) |
| CORRUPT | 0.43* (0.080) | 0.382*** (0.055) | 0.24* (0.11) |
| Observations | 6 | 10 | 20 |
| Adjusted $R^2$ | 0.89 | 0.83 | 0.19 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parentheses)

## APPENDIX TWO

To determine the effects of volatility in the bond market, results were calculated using the following equation:

$$(11)\quad NEGOTobs_t = B_1 + B_2 \times t + B_3 \times VOL_t + \left(1 - \left(B_1 + B_2 \times t + B_3 \times VOL_t\right)\right) \times CORRUPT \times LAW_t + \in$$

Where $B_3$ is a bond volatility constant and $VOL_t$ is bond volatility in year $t$.

TABLE 4—IMPACT OF CAMPAIGN CONTRIBUTIONS ON MODE OF ISSUE (INCLUDING A BOND VOLATILITY)

| | Three Year Window | Five Year Window | Ten Year Window |
|---|---|---|---|
| $B_1$ | 0.83* (0.093) | 0.79*** (0.046) | 0.74*** (0.026) |
| $B_2$ | 0.024 (0.0082) | 0.015** (0.0038) | 0.0078*** (0.0019) |
| $B_3$ | -0.013 (0.0010) | -0.0093 (0.0048) | -0.0017 (0.0023) |
| CORRUPT | 0.43 (0.080) | 0.33** (0.066) | 0.30*** (0.068) |
| Observations | 6 | 10 | 20 |
| Adjusted $R^2$ | 0.90 | 0.87 | 0.48 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parentheses)

APPENDIX THREE

TABLE 5 - LOBBYIST CORRUPTION REGRESSION RESULTS

| | All State Years | | State Year Issuing > $3b/yr | |
|---|---|---|---|---|
| | **WLS** | **OLS** | **WLS** | **OLS** |
| **$K_{INF}$** ( x $10^{-6}$) | 18.0 ** (5.6 ) | 16.4 ** (4.69) | 17.8 (10.1) | 17.6 * (8.72) |
| **$K_{COR}$** ( x $10^{-6}$) | 5.87 *** (1.54 ) | 4.87 *** (1.23) | 6.48 ** (2.45) | 5.55 ** (1.97) |
| **Population Density**[129] (x $10^{-6}$) | 29.7 (135) | -9.51 (66.4) | -0.108 (0.740) | 0.22 (0.62) |
| **State GDP per capita**[130] (x $10^{-6}$) | 2.24 (8.61 ) | -0.822 (3.34) | 8.10 (1.24) | 4.10 (6.57) |
| **State & local spending per capita**[131] (x $10^{-6}$) | -50.1 (33.0 ) | 9.49 (16.5) | -82.7 (53.5) | -49.6 (37.0) |
| **State & local dept per capita**[132] (x $10^{-6}$) | -34.8 (26.4) | -18.5 (14.5) | -38.5 (40.0) | -23.3 (26.4) |
| **State bond rating**[133] ( x $10^{-9}$) | 42.8 (25.5) | 2.53 (11.7) | 45.7 (38.7) | 27.1 (25.3) |
| **Observations** | 240 | 240 | 133 | 133 |
| **Adjusted $R^2$** | 0.76 | 0.71 | 0.76 | 0.72 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parenthesis)

129 *Supra* note 116. Note, units are persons per square mile.

130 *Id.*

131 *Id.* Note figures are for 2002.

132 *Id.* Note figures are for 2002.

133 2004 Standard & Poor State Bond Rating, where 9=AAA & 1=C; note no rating information available for Nebraska or South Dakota.

APPENDIX FOUR

*Using Alternative Allocation Methods*

TABLE 6 - LOBBYIST CORRUPTION REGRESSION RESULTS
EQUAL ALLOCATION METHOD

| | All State Years | | State Year Issuing > $3b/yr | |
|---|---|---|---|---|
| | **WLS** | **OLS** | **WLS** | **OLS** |
| $K_{INF}$ ( x $10^{-6}$) | 20.63 *** (4.34) | 14.58 *** (4.19) | 20.61 *** (5.97) | 14.43 * (5.77) |
| $K_{COR}$ ( x $10^{-6}$) | 3.28 ** (1.26) | 4.63 *** (1.16) | 3.29 (1.73) | 4.69 ** (1.59) |
| **Observations** | 250 | 250 | 133 | 133 |
| **Adjusted $R^2$** | 0.76 | 0.71 | 0.76 | 0.71 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parenthesis)

TABLE 7 - LOBBYIST CORRUPTION REGRESSION RESULTS
EXCLUSION METHOD

| | All State Years | | State Year Issuing > $3b/yr | |
|---|---|---|---|---|
| | **WLS** | **OLS** | **WLS** | **OLS** |
| $K_{INF}$ ( x $10^{-6}$) | 18.70 *** (4.44) | 12.00 ** (4.18) | 18.87 ** (6.11) | 12.01 * (5.77) |
| $K_{COR}$ ( x $10^{-6}$) | 2.87 ** (1.28) | 4.38 *** (1.15) | 2.88 * (1.77) | 4.41 ** (1.59) |
| **Observations** | 250 | 250 | 133 | 133 |
| **Adjusted $R^2$** | 0.70 | 0.65 | 0.70 | 0.66 |

*/**/*** Denotes significance at 0.05/0.01/0.001 level.
(Standard Error in Parenthesis)