JOURNAL OF LAW, ECONOMICS & POLICY

VOLUME 3 WINTER 2006 NUMBER 1

CONTENTS

PUTTING HUMPTY DUMPTY BACK TOGETHER: EXPERIMENTAL EVIDENCE OF ANTICOMMONS TRAGEDIES

Ben Depoorter & Sven Vanneste[*]

ABSTRACT

This Article conducts an experimental investigation of anticommons dilemmas. The results confirm that anticommons deadweight losses increase with the degree of complementarity and the degree of fragmentation of property. Our study further provides three novel insights into the problem of fragmentation. First, the data illustrates that individual right holders ignore the expected value of bundling and instead focus on the maximum profit he or she could realize by bundling. Second, the experiments suggest that uncertainty amplifies the anticommons pricing effect. Finally, cooperation is higher in cases where the value of bundling is more uncertain as opposed to scenarios where there is relative certainty of creating surplus but there is a (modest) chance of losses from bundling.

1. INTRODUCTION

An anticommons is a property regime in which multiple owners hold effective rights of exclusion in a scarce resource.[1] Economic theory has illustrated how the coexistence of multiple exclusion rights may lead to sub-optimal uses of resources held in common.[2] If a common resource is

---

[*] Ben Depoorter is associate professor at University of Miami School of Law and Ghent University Law School. Sven Vanneste is research fellow at Ghent University, Faculty of Psychology, Dept. of Developmental, Personality & Social Psychology and visiting scholar Ghent University School of Law, Center for Advanced Studies in Law and Economics.

[1] This definition of the anticommons, employed by Heller, provides a powerful tool for property theory. Heller first revitalized the concept in an article on the transition to market institutions in contemporary Russia. He discusses the intriguing prevalence of empty storefronts in Moscow. Storefronts in Moscow are subject to under use because there are too many owners (local, regional and federal government agencies, mafia, etc.) holding the right to exclude. *See* Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621 (1998). The title of this paper refers to the fairy tale of Humpty Dumpty to illustrate the anticommons. When Humpty Dumpty is shattered into pieces all of the king's horses and all of the king's men cannot re-assemble him, which stands in contrast to the ease with which he broke into pieces in the first place. *See* Michael A. Heller, *The Boundaries of Private Property*, 108 YALE L. J. 1163 (1999).

[2] Anticommons theory relies on Cournot's model of duopoly: a single monopolist producing a composite good will charge a price lower than the sum of the prices that would be charged by two

subject to multiple exclusion rights held by two or more individuals, each co-owner has incentives to withhold resources from other users to an inefficient level. As a result, exclusion rights will be exercised even when the use of the common resource by one party could yield net social benefits, a problem known as the "Tragedy of the Anticommons."[3] Take the example of medical innovation. It is generally understood that awarding private property rights to discoveries promotes innovation and the commercial development of new technologies.[4] In light of the anticommons, intellectual property rights to research may actually retard life-saving developments of medical products when each owner of various stages of research block each other from the use of his research in creating these products.[5] The tragedy of the anticommons may occur because the multiple holders of exclusion rights do not fully internalize the cost created by the enforcement of their right to exclude others.[6]

The intuition underlying the anticommons is that it is often harder to regenerate separated bundles than it is to fragmentize.[7] Economic models assume the costs to rebundle independently-owned property fragments are higher than the costs involved in the initial fragmentation. Such "stickiness" of fragmentation is problematic when the costs of bundling prevent value-maximizing uses of the resource. When a value-enhancing opportunity arises which requires the unification of each fragmented right, the ex-ante rational decision to fragment may turn out to be ex-post sub-

---

complementary duopolists selling the single component parts. AUGUST COURNOT, RESEARCHES INTO THE MATHEMATICAL PRINCIPLES OF THE THEORY OF WEALTH (1838).

[3] The pioneering articles include: Heller, *The Tragedy of the Anticommons*, *supra* note 1; Heller, *The Boundaries of Private Property*, *supra* note 1; Michael E. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. MAG. 698 (1998), *excerpted as Upstream Patents = Downstream Bottlenecks*, 41 Law Quad. Notes, 93-7 (1998). The concept of the tragedy of the anticommons was formalized in James Buchanan & Yong J. Yoon, *Symmetric Tragedies: Commons and Anticommons Property*, 43 J. L. & ECON. 1, 1-13 (2000); Norbert Schulz et al., *Fragmentation in Property: Towards a General Model*, 158 J. INSTITUTIONAL & THEORETICAL ECON. 594 (2002); Francesco Parisi, et al., *Duality in Property: Commons and Anticommons*, 25 INT'L REV. L. & ECON. 578 (2005).

[4] *See* Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977).

[5] Heller & Eisenberg, *supra* note 3, at 698: "more intellectual property rights may lead paradoxically to fewer useful products for improving human health." For an empirical investigation, *see* John P. Walsh et al., *Working Through the Patent Problem*, 299 SCI. MAG. 1021 (2003) (providing survey evidence of the anticommons problem and the creativity of innovators in solving the problem).

[6] *But see, e.g.*, Richard A. Epstein & Bruce Kuhlik, *Is there a Biomedical Anticommons?*, 27 REG. 54 (2004) (arguing that researchers have ample incentives to solve anticommons problems in the biomedical field).

[7] In the words of Heller: "Once an anticommons emerges, collecting rights into usable private property may prove to be brutal and slow." Michael A. Heller, *Three Faces of Private Property*, 79 OR. L. REV. 417, 418, 424 (2000).

optimal, given the greater costs of reunification.[8] Prior theoretical research on anticommons fragmentation further suggests that the severity of the deadweight losses from concurrent possession of a complementary right increases monotonically with the number of independent holders.[9]

Despite these theoretical underpinnings, the literature to date has omitted analysis of the precise factors that lead reunification efforts to fail. Are negotiations unsuccessful because of transaction costs and strategic behavior, or is the bargaining process troubled by cognitive error? What social and cognitive processes lie at the root of the anticommons problem? In what way does non-cooperation in anticommons dilemmas differ from the well-known tragedy of the commons? Because empirical evidence on anticommons tragedies is hard to obtain directly,[10] analysis of the processes that create anticommons tragedies is especially important.

This Article sets out to deepen our understanding of the anticommons problem by conducting a number of social dilemma experiments in a laboratory setting. We measure the impact of various aspects of property fragmentation and provide an interpretation of the social and cognitive processes that might cause problems of reunification. We examine a number of alleged attributing factors of anticommons tragedies that have been highlighted in the theoretical literature.[11] These factors include the complementarity of fragmented parts, the number of fragmented parts, and the degree of uncertainty in obtaining value from rebundling fragmented ownership.

The results confirm the theoretical proposition that anticommons deadweight losses increase with the degree of complementarity between individual parts and with the degree of fragmentation. Our study further provides three novel insights into the problem of fragmentation. First, the data illustrates that each individual right holder ignores the expected value of the purchaser's project, and instead focuses on the maximum profit he could possibly realize by bundling. Second, the experiments suggest uncertainty amplifies the anticommons pricing effect. Finally, cooperation is higher in cases when the value of bundling is more uncertain, as opposed

---

[8] Schulz et al., *Fragmentation in Property: Towards a General Model*, *supra* note 3.

[9] Ben Depoorter & Francesco Parisi, *Fair Use and Copyright Protection: A Price Theory Explanation*, 21 INT'L REV. L. & ECON. 453, 460-61 (2002): "The greater the number of individuals who can independently price an essential input, the higher the equilibrium price that each of these individuals will demand for his own license. At the margin, as the number of [right] holders approaches very large numbers (or infinity), complete abandonment of valuable resources will result."

[10] Underuse and missed opportunities are not as easily observed as are, for instance, visually apparent commons tragedies of overuse. Evidence of anticommons tragedies, for instance, of research avenues forsaken due to licensing bottlenecks, must be observed indirectly through survey evidence. For an example of empirical research on anticommons tragedies in the biomedical field, *see* Cohen et al., *supra* note 5.

[11] Heller, *supra* note 7, at 4. *See also* Parisi et al., *Duality in Property: Commons and Anticommons*, *supra* note 3.

to scenarios where there is a relative certainty of creating surplus but a (modest) chance of loss from bundling.

Our experiment demonstrates the burden of negotiation that rests with a buyer who seeks to rebundle independently-owned property fragments. The results indicate the price concessions a prospective seller will need to obtain to bring the price of bundling within the limits of the net expected value of bundling.

Section 2 describes the structure of the experiment. Section 3 presents the results of our experiment. Section 4 provides a discussion of the results. Section 5 summarizes our conclusions.

2. DESCRIPTION OF THE STUDY

An anticommons is characterized by a conflict between private incentives of the various right holders and their common interest. Although an individual right holder should take into account the cross-effects of his pricing decision in order to safeguard successful reunification, he has a conflicting incentive to try and obtain as much as possible from the surplus that results from the process of bundling the individual parts. This divergence between private and public interests creates the "social dilemma" that lies at the heart of anticommons regimes.[12]

In our experiment, each participant[13] was informed in a script that he or she was one of five partial right-holders (owners) to a unitary resource.[14] In each of the scenarios, participants were informed that a third-party was looking to purchase a number of parts under the terms described in the particulars of the sub-experiment. Each participant further learned that he possessed an individual piece of land worth fifty chips (each chip being the equivalent of .05 euros), which he could cash in at the end of the experiment in return for his right. If he sold his individual right in return for chips from the bundler-purchaser, he would be able to return these chips at the end of the experiment for the money equivalent. After explaining the scenario, some comprehension questions were asked to verify that participants understood the situation. All participants answered these questions correctly.

---

[12] In a social dilemma: (1) a non-cooperative choice is always more profitable to the individual than a cooperative choice, regardless of the cooperativeness of others; (2) a non-cooperative choice is always harmful to others compared to a cooperative choice; and (3) the aggregate amount of harm done to others by a non-cooperative choice is greater than the individual's profit. *See* SHIRLI KOPELMAN ET AL. *Factors Influencing Cooperation In Commons Dilemmas: A Review Of Experimental Psychological Research*, *in* THE DRAMA OF THE COMMONS, 113–156 (Elinor Ostrom et al. eds., Nat'l Acad. Press 2002).

[13] We surveyed 300 first-year undergraduate students at Ghent University in Belgium.

[14] No significant effect was found for age or gender.

Using different scenarios for each participant, as described below, we explore how respondents, as individual right-holders, demand different prices when the following independent variables vary: 1) the degree of complementarity among fragmented parts; 2) the number of other rights holders with complementary rights to the resource; 3) the synergy resulting from fragmentation; and 4) the degree of uncertainty of the surplus obtained by bundling each of the individual rights.

The questions in the script were ordered randomly to avoid learning experiences.[15] The experiments were conducted in various different rooms to prevent participants from communicating or learning each other's reservation prices. The experiment was designed to measure the statistical data on a parametrical level. To this purpose, we used a multivariate repeated measure ANOVA.[16] Two sub-experiments deviate from this statistical method and were replaced by a one-way ANOVA to comply with the between-subject measurement format.[17]

3. RESULTS

3.1. *Surveys A and B: Complementarity*

Each participant[18] was informed that he or she was one of five partial right holders (owners) to a unitary resource. The participants were informed that a third-party was looking to purchase a number of parts. In the various parts of the test, the number of individual parts ranged between 2 and 5. Students were further informed that each individual part, by itself, had a market value of 50 euros.[19] The aggregate value of the unified bundle

---

15 When all subjects receive the script with questions in the same order, the first trial could influence their opinion in the second trial and so on. The learning effect is nullified when subjects receive the scripts in random order.

16 This involves the application of the analysis of variance to data in which a single dependent variable is measured on more than one occasion on the same subject. In the case of an orthogonal factorial design, the method essentially combines, in a linear fashion, the information of the several response variables in such away as to detect any existing treatment effects. *See* RICHARD A. JOHNSON, & DEAN W. WICHERN, APPLIED MULTIVARIATE STATISTICAL ANALYSIS (5th ed. 1998).

17 Various groups of participants were assigned to the different variables (2, 3, 4, or 5 parts) and every group had to decide on the price of the part assigned to them.

18 This study's population consists of a random group of first-year students of the departments of law, political science, and economics at Ghent University. Each student was randomly assigned to one of the experiments.

19 We operate from the stylized assumption that there is no difference between the market price of each individual part and the subjective value to each of the owners. In other words, we control for any idiosyncratic qualities of the parts or cognitive attachments to the parts, such as negative endowment effects. The cognitive effects involved in the decision-making process of rebundling are explored further on in this study.

was 250 euros. No further information on the incentives of the third-party (such as profitability and synergies resulting from bundling) was disclosed at this point. While this approach reduces the control of the experiment, relative to the other treatments in the surveys described below, this scenario is useful because it has more external validity by aligning more closely with real life anticommons situations where a third party purchaser (such as an oil company) tries to hide its identity or project in order to prevent inflation of prices. In the first hypothetical scenario, each student was informed that the purchaser sought to obtain 2 out of 5 parts that were divided among five participants. In a subsequent condition, other participants were informed that the purchaser needed to obtain 3, 4, or all 5 parts. In each of these scenarios, each participant listed his reservation price[20] while attempting to maximize his personal gain. The survey thus measures the differences in reservation prices arising in situations involving varying degrees of complementarity. Where the third party only looked to purchase two parts, this represents a relative low degree of complementarity, or, conversely, a case of relatively high substitutability.[21] By contrast, when the hypothetical scenario indicates the third party needed to purchase all five parts, this is a situation of perfect complementarity.

Parts	*Mean*	*Standard Deviation*	*N*
2	64.6	18.65	20
3	69.5	15.27	20
4	76.3	35.57	20
5	100.1	48.34	20

Table 1: Descriptive statistic, between subjects

Parts	*Mean*	*Standard Deviation*
2	67.4	19.57
3	72.6	26.38
4	80,2	36.12
5	107.1	57,99

Table 2: Descriptive statistic, within subjects (N = 20)

[20] We employ the term 'reservation price' to denote the initial selling price, as stated by the individual right holder. Strictly speaking this price is not necessarily a reservation price in that this stated price is the lowest outcome a negotiator is willing to accept. However, because there are no negotiations, we assume that initial right holders, in effect, will not accept an agreement that is below the initial selling price.

[21] *See* Depoorter & Parisi, *supra* note 9, at 460-61.

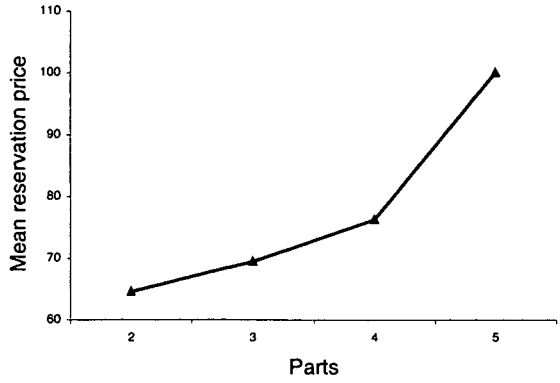


Figure 1: The mean demand price for the different parts measured between subjects (ANOVA, $F(3,76) = 4.73, p < .01$)

Table 1 and accompanying *Figure 1* map the variation between mean reservation prices. The mean demand price in the case of low complementarity was sixty-seven euros. The aggregate mean price was thus 134 euros; a total of thirty-four price units above the objective value of two combined parts. In the case of perfect complementarity, the mean demand price was 100 euros, totaling a mean demand price of 500 euros for the combined purchase of all individual parts. While reservation prices for "2 out of 5 complementarity" totaled 34% over the objective value, a case of perfect complementarity averaged a combined demand price that was 100% above the objective value. These simple findings confirm the theoretical finding that reservation prices correlate with the strength of the veto right into the successful bundling of the individual parts.

We repeated the same experiment, but measured repeatedly with the same subjects in each of the different conditions (2, 3, 4, and 5 parts) (between subjects). This measurement enabled us to verify whether subjects reasoned differently when they were asked to list a price in just one of the above scenarios, compared to situations where each individual subject was asked to formulate prices for all of the scenarios (within subjects).[22] The results—see *Table 2*—significantly correspond with the previously shown within-subject findings (Repeated Measure ANOVA, $F(3,17) = 5.42, p < .01$).

---

[22] When an experiment is conducted "within subjects," every participant is assigned to all treatments in a randomly selected order. In such experiment there is a risk that participants' selling prices differs according to the initial scenario (assembly of 2, 3, 4, or 5 required parts) first assigned to them. Such bias could be attributed to the initial scenarios working as a reference point in the mind of the participants. In such a case participants might not fully focus on the amount of parts the third party seeks to gather (degree of complementarity).

3.2. *Survey C: Opportunity Costs*

In Experiment *C*, we attached various degrees of profitability to the effort of rebundling by the third party. We measured the impact on the reservation prices of the individual right holders. As before, each participant (N = 84) was one of five partial right holders (owners) to a unitary resource. They were informed that a third-party was looking to purchase *all five parts* held by the individual owners. Again, each individual part had an objective value of fifty euros and the aggregate value of the unified bundle was 250 euros. By explicitly assigning the value of each right, we attempted to eliminate the "attribution effect" whereby each person systematically overvalues the role of his right in the overall project.[23]

Contrary to surveys *A* and *B*, we disclosed the opportunity costs of the third party upfront. Each participant was requested to state his demand price in each of five hypothetical scenarios *with varying profits to be obtained* by the third-party purchaser from bundling all five parts. In five different scenarios, each participant was informed that bundling would create a surplus for the third party of 100, 300, 500, 1000, or 10,000 euros. These scenarios each represent different values resulting from reunification. In the last hypothetical, the "sum is worth more than its parts" by 9,750 euros (10,000 – 250). In such a scenario, unsuccessful rebundling imposes considerable deadweight losses—as higher valued uses are not consummated. This situation represents a more significant anticommons tragedy relative to the first hypothetical, where a modest 100 euros was at stake in the effort to rebundle. *Figure 2*, plots the reservation prices in all five instances of surplus profitability. The vertical axis marks the asking price, expressed in relative amounts of the profits, or synergies of bundling. The horizontal axis indicates the cases of a third party profit of 100, 300, 500, 1000, 10,000 euros respectively. As *Figure 2* below indicates, there was no significant difference ($F(3,81) = 1.28, p = .168$) between reservation prices in the profit range between 300 and 10,000 euros; the average price stated by each right holder was approximately 26% of the total value of the surplus attained by bundling. In the case of a surplus of 10,000 euros, the purchaser was faced with an aggregate mean asking price of 12,300 euros. This price is 24.6% above the price that he or she can offer so that the project remains profitable. Similarly, when the profit from bundling was a

---

[23] The attribution bias holds that individuals systematically overvalue their assets and disparage the claims of their co-right holders. *See* LEE ROSS & CRAIG A. ANDERSON, *Shortcomings in the attribution process: On the Origins and Maintenance of Erroneous Social Assessments*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, 129-52 (Daniel Kahneman et al. eds., Cambridge Univ. Press 1982). Heller and Eisenberg suggest that this particular cognitive bias explains bargaining breakdowns in the biotechnology industry, where scientists tend to overvalue the importance of their discoveries for the development of follow-up, aggregate inventions. *See* Heller & Eisenberg, *supra* note 3, at 701.

more modest 300 euros (plot 2 on graph 2, a median asking price was 26.6% or 79.8 euros per part), the combined reservation price was 399. Thus, the difference between reservation prices in the surplus range of 300 and 10,000 is non-significant.

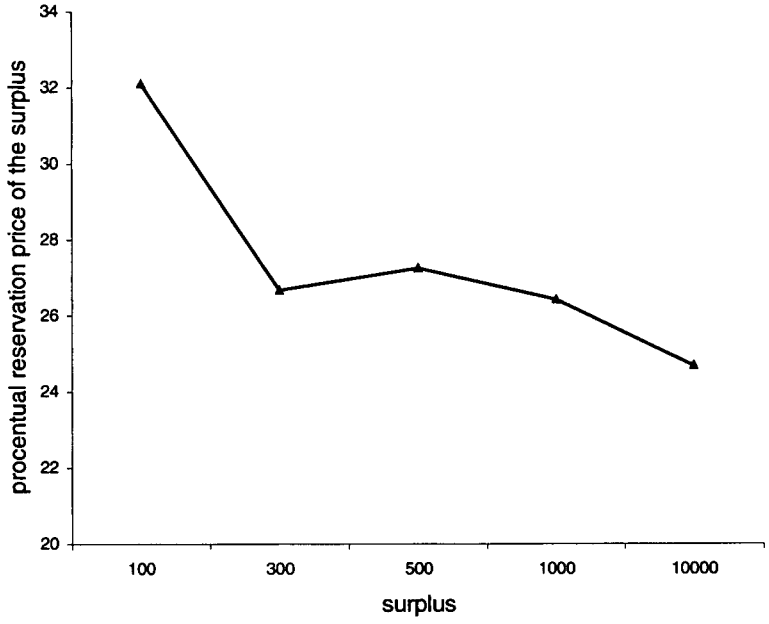


Figure 2: The degree of profitability from bundling of fragmented property entitlements on the prices charged by individual right holders ($F(4, 80) = 5.391$, $p < .001$)

3.3. *Survey D & E: Uncertainty*

Experiment *D* measures the effect of uncertainty regarding the expected benefits of the bundling of fragmented property entitlements. Again, each participant (N = 40) was informed that she was one of five partial-right holders (owners) to a unitary resource. A third-party was looking to purchase *all five parts* held by the individual owners. Each student was informed that each individual part had an objective value of fifty euros. The aggregate value of the unified bundle was 250 euros. As in Section 3, we disclosed the opportunity costs of the third party. This time, however, the subjects were informed that the purchaser faces considerable uncertainty as to the profitability of the project. Each participant was requested to state her reservation price in each of five hypothetical scenarios *with varying profits to be obtained by* the third-party purchaser

from bundling all five parts. Additional information was provided as to the uncertainty of the project's profitability. In four different scenarios, participants were informed that bundling would create a surplus for the third party of 100, 500, 1000, or 10,000 euros, each with a probability of 10%. In each of the scenarios, there would be a 90% chance that bundling did not create any surplus. The expected value of each of these projects was respectively 10, 50, 100, and 1000 euros. Are the subjects responsive to the lower expected value generated by the high degree of uncertainty? Again, the results give rise to pessimism. The results show that subjects consistently demanded a proportional share of 10% of the maximum profit that could possibly be realized by bundling. The mean reservation price, set by one individual right holder, was 14.25% of the surplus (see *Fig. 3*). Put differently, the aggregate reservation price was seven times above the expected value of the project ($F(3,37) = 20.31$, $p < .001$).[24] Given the expected benefit of the project (market value of the parts), the gap between purchaser's willingness to pay and individual owner's willingness to accept is non-negligible.

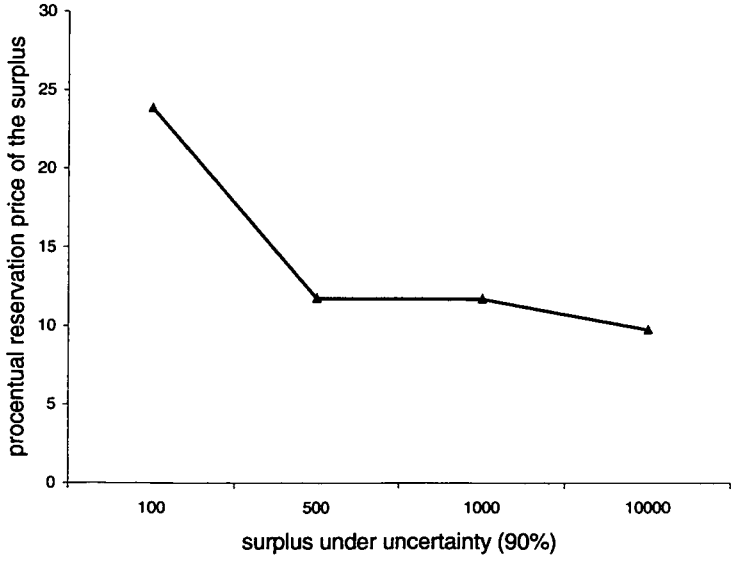


Figure 3: The expected profit of bundling the fragmented property under a 90% uncertainty for the individual holders. ($F(3,37) = 4.43$, $p < .01$)

[24] When there is certainty of 10% of surplus from bundling, every individual holders' maximum price is 2% of surplus. When individual right holders ask 14.25%, the aggregate price totals seven times the expected value of the projects. The statistical difference between the 2% case and the observed reservation prices ($F(3,37) = 20.31$, $p < .001$) is significant.

These results were confirmed in a second similar test (see *Figure 4*) when a higher degree of uncertainty is imposed: there is a 99% chance that bundling does not create any surplus. The expected value of each of these projects is respectively 1, 5, 10, and 100. Again, the subjects were unresponsive to the lower expected value generated by the high degree of uncertainty. From the results, subjects consistently demanded a proportional share of 11.44% of the maximum profit. The median price, set by one individual right holder, was fifty-seven times above the expected value of the project. The aggregate of the individual right holders' willingness to accept was fifty-seven times beyond the willingness to pay price of the purchaser, given the expected benefit of the project.[25]

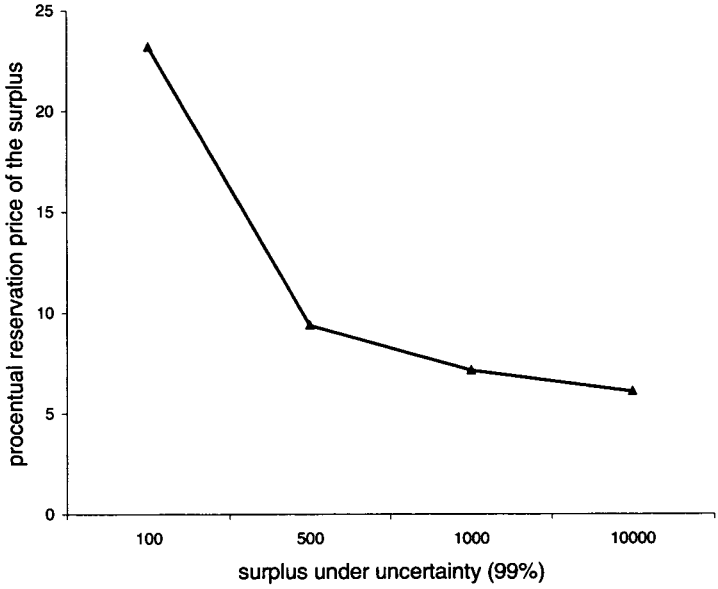


Figure 4: The expected profit of bundling the fragmented property under a 99% uncertainty for the individual holders. ($F(3,37) = 2.40$, $p < 0.05$)

Again, each subject (N = 78) was informed that he is one of the partial-right holders to a unitary source and that a third-party was interested in purchasing all five parts. Each individual part had a value of 50 euros and when the third-party bundles the five parts, this would generate a

[25] A similar deduction can be made as in footnote 21. When there is certainty of 1% of surplus from bundling, this means that every individual holders' maximum price is 0.2% of surplus. When individual right holders ask 11.44%, the aggregate price totals fifty-seven times the expected value of the projects.

surplus of 125 euros with a probability of 80% and a 20% probability of a loss of 450 euros. In two different scenarios, students were asked the same questions, but with a surplus of 1,250 or 12,500 and a loss of 4,500 or 45,000. The expected values of each of these projects were respectively, 10,100, and 1,000 euros.[26]

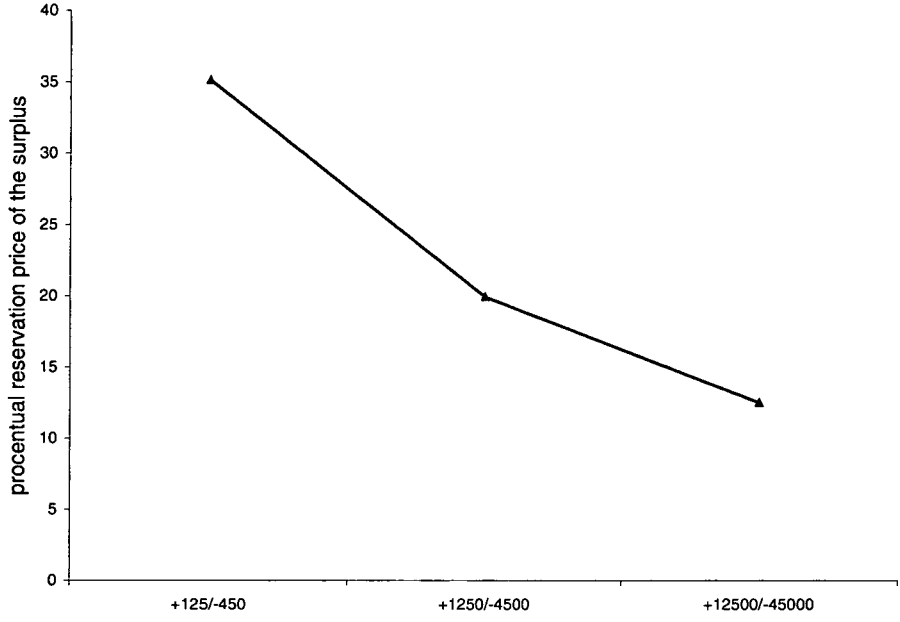


Figure 5: The expected profit of bundling the fragmented property for the holders under a 80% certainty of a surplus vs. 20% uncertainty of losing an amount of money for the purchaser. ($F(2,76) = 15.19$, $p < .001$)

Figure 5 confirms the findings of the other experiments. When stakes were minor, the individual right holders set disproportionately high reservation prices—35% in the case of a project with expected value of 100 euros (this totals a combined reservation price of 175% of the expected value of bundling). When stakes were higher, the average reservation price remained relatively stable at 14-19% of the expected surplus.

Next, we compared the reservation prices for two types of scenarios. Although the expected value of bundling was identical in both scenarios, one scenario promised high returns from bundling, but with great uncertainty, while the other scenarios promised only a more modest payoff

[26] A probability of 80% to win a surplus of 125 euros gives an expected value of 100 euros, while the chance of loss is 450 euros with a probability of 20%, giving us 90 euros. 100 euros minus 90 euros gives us an expected benefit of 10 euros.

to the third-party purchaser, but with higher certainty. In the case where bundling lead to a 100 dollar surplus with 10% probability, the mean reservation price was 24% of the expected value, compared to 35% of the expected value of the *low risk-low payoff* variant of experiment *E* ($F(2,75) = 9.44$, $p < .001$). In the case of a 1,000 dollar surplus with 10% probability in *D* (*high risk-high profit*), the mean reservation price was 12%, versus 19% in the *low risk-low profit* variant of *E* ($F(2,75) = 3.29$, $p < .05$). In the case of a 10,000 dollar surplus with 10% probability in *D* (high risk high payoff), the mean reservation price was 10%, versus 13% in the *low risk-low payoff* variant of *E (80% chance of +12,500 and 20% chance of -45,000).* Although the expected value in each of these scenarios was identical, reservation prices seem to be consistently lower (and cooperation higher) where there was considerable uncertainty regarding high returns than when there was relative certainty, but with a chance of losses for the third-party purchaser (See *Figure 6*, $F(2,75) = 4.92$, $p < .01$). Upon further examination, we find analogous results for instances where the surplus was 100 euros and 1,000 euros (see *Figure 7*) under high risk levels vs. low risk levels ($F(2,75) = 10.43$, $p < .001$).

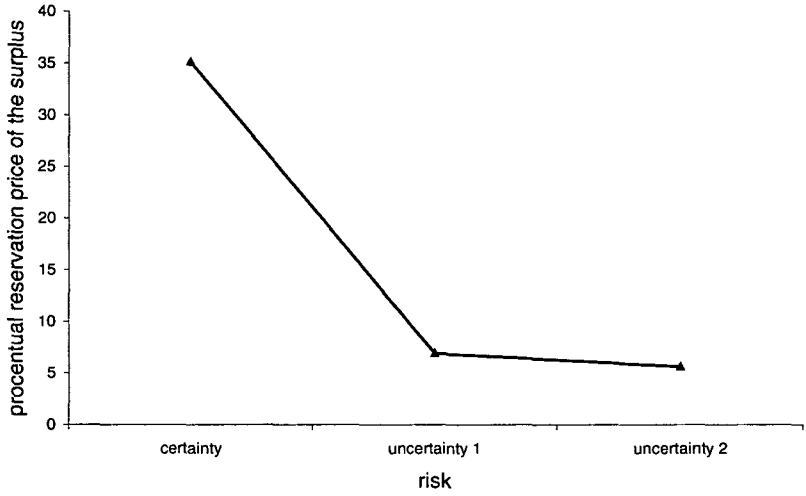


Figure 6: Different results for under certainty and uncertainty under identical expected surplus (10 euros) ($F(2,75)=4.92$, $p < .01$)

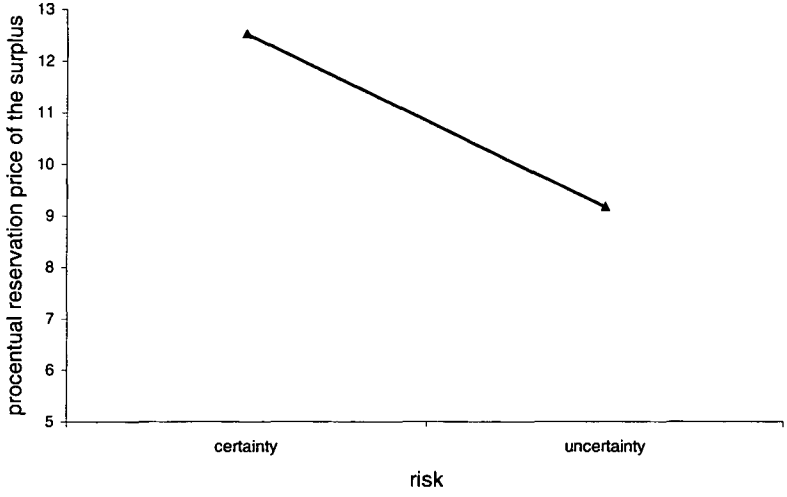


Figure 7: Different results when a same surplus (1,000 euros) given under certainty and uncertainty $F(1,76)= 4.13, p < .05$

3.4. *Survey F*

Experiment *F* was constructed along the lines of the previous surveys. Again, each subject (N = 62) was informed that she was one of five partial-right holders of a unitary resource and that a third party was looking to purchase all five parts. Every individual part was valued at 50 euros. If the purchaser was successful at rebundling the five parts, he would obtain a surplus in a range between a minimum and a maximum expected value. In a random order, the six trials indicated an expected surplus between respectively 100–500, 1000–5000 and 10,000–50,000 in the different trials. This experiment differs from section 3.3. because the exact probability and payoff from bundling remains unknown. The knowledge of subjects was restricted to the range within which the profits were situated. This experiment is more realistic because, as in real-life situations, precise probabilities remain unknown. For instance, when a real estate developer seeks to purchase five adjacent tracts, it is more likely that the land owners base their initial reservation prices on a rough, highly subjective estimate of the value to the entrepreneur, rather than probability and profit estimates of the individual provided to subjects in surveys *D* and *E*.

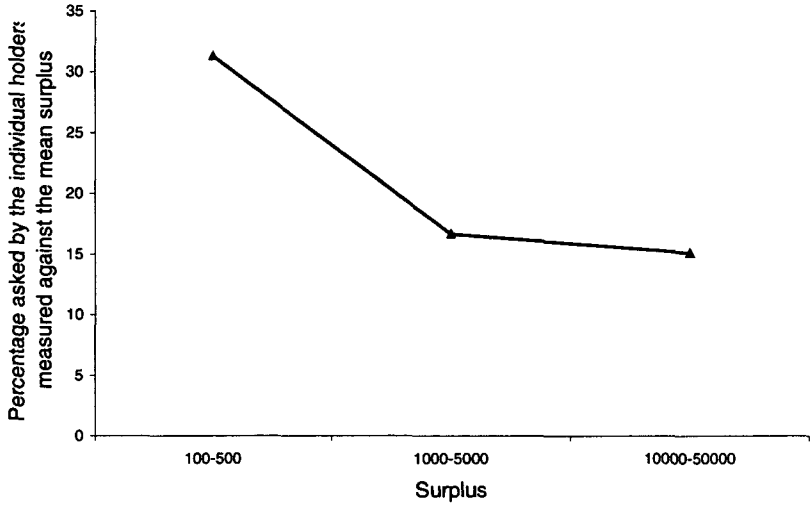


Figure 8: The expected profit of bundling the fragmented property for the holder's purchaser under the uncertainty about the amount of the surplus. ($F(2,60) = 4.15$, $p < .05$)

When the surplus was situated in the 100-500 dollar range, the individual owners demanded 32%, or eighty-three euros, of the average surplus of 300. The average reservation price was 415 euros. With regard to the higher profit ranges, the average reservation price was 17.5% of the average surpluses of 3000 and 30,000. Again, two observations appear. First, participants employed an all or nothing strategy, demanding relatively high prices, when stakes were minor.[27] Secondly, when stakes were high, subjects' reservation prices were based on a proportion of the expected profit of the buyer, irrespective of objective market value of an individual part.

4. DISCUSSION

4.1. *Survey A & B: Complementarity*

In survey *A* and B, we measured the magnifying effect of complementarity of fragmented property entitlements on the occurrence of anticommons losses.

[27] The wider variance within this cell suggests that this finding possibly is a confound resulting from the low values.

Prior theoretical research on anticommons fragmentation claims that the severity of the deadweight losses from concurrent possession of complementary rights increases monotonically with the number of independent holders: "The greater the number of individuals who can independently price an essential input, the higher the equilibrium price that each of these individuals will demand for his own license. At the margin, as the number of [right] holders approaches very large numbers (or infinity), complete abandonment of valuable resources will result."[28]

While reservation prices for "2 out of 5" complementarity totaled 34% over the objective value, a case of strict "5 out of 5" complementarity averaged a combined demand price that was 100% above the objective value. These simple findings confirm the theoretical findings that reservation prices correlate with the strength of veto-right into the successful bundling of the individual parts.

These basic results of surveys *A* and *B* are not surprising. Selling prices are higher when a seller has more individual bargaining power.

4.2. *Survey C: Reservation Prices and the Size of the Pie*

Experiment *C* examined the influence of higher degrees of profitability on the reservation prices of the individual right holders. Contrary to surveys A and B, we now disclosed the gains from bundling in order to measure the effect on reservation prices. Furthermore, we contrasted situations where reunification of fragmented parts resulted in very substantial profits with situations where reunification created very modest gains. The results give little reason to believe that, from the perspective of uncoordinated selling prices, the problem is less pronounced when opportunity costs are higher, i.e. when the costs of idleness or under use are more pronounced. The results indicate that, with regard to initial reservation prices, respondents do not discriminate between projects of rebundling that are very profitable and cases that generate more modest payoffs. As illustrated in *Figure 2* above, there was no significant difference of reservation prices in the profit range between 300 and 10,000 euros: the average price stated by each of the right holders was approximately 26% of the total value of the surplus attained by bundling. In the case of a surplus of 10,000 euros, the purchaser was faced with an aggregate mean asking price of 12,300 euros. This price was 24.6% above the price that he could offer so that the project remained profitable. Similarly, when the profit from bundling was only 300 euros (plot 2 on graph 2, a median asking price was 26.6% or 79.8 euros per part), the combined reservation price was 399 euros. Thus, the difference in reservation prices between a surplus of 300 and 10,000 is non-significant.

---

[28] Depoorter & Parisi, *supra* note 9, at 460-61.

The implication is that, in attempting to rebundle subdivided parts, a third party purchaser faces reservation prices that significantly outweigh the expected profitability of the attempted reunification, regardless of the size of the interest at stake. All else being equal, a third party with a highly profitable project or with a more modest project, faces prices that are, more or less to the same extent, beyond the expected value of the project. An oil company seeking to acquire four adjacent parcels of land for the purpose of optimal drilling, with a potential for efficiency savings of 2 million euros, faces a negotiation problem comparable to an editor trying to assemble the copyrights from four different authors for an anthology on American writing (with profitability of 1000 euros). This confirms the findings of Libecap and Wiggins that unitization of oil fields, involving multiple right holders, might fail despite the tremendous gains that can be reaped by uniting oil fields.[29]

This survey indicates that subjects hold a certain amount (approximately 25%) of the profit as a focal point as to what they deem to be the price at which they are willing to sell their individual part. Regardless of any endogenous motivation for this proportion (evaluations of fairness, etc.), five people are each asking a combined price that exceeds the expected benefits of bundling by 25%.

4.3. *Survey D & E: The Role of Uncertainty*

Next, we measured the effect of uncertainty regarding the expected benefits of the bundling of fragmented property entitlements. Surveys *D* and *E* compare conditions: 1) where there was considerable uncertainty regarding high returns; and 2) where there was relative certainty, but with a chance of losses for the third-party purchaser. The expected value was identical in both conditions.

4.3.1. High Degrees of Uncertainty with Large Upside

From the results, it follows that subjects consistently demanded a proportional share of 10% of the maximum profit that could possibly be realized by bundling. The mean reservation price, set by one individual right holder, was 14.25% of the surplus (see *Figure 3* above). In our results, aggregate reservation prices were seven times above the expected value of the project.

These results indicate that subjects ignored the expected value of the purchaser's project, and instead focused on the most optimistic outcome of

---

[29] *See* Gary D. Libecap & Steven N. Wiggins, *Contractual Responses to the Common Pool: Prorationing of Crude Oil Production*, 74 AM. ECON. REV. 87 (1984).

the scenario. This expectation leads to higher demands from right holders than in the previous surveys. Pricing decisions seem to be anchored on the maximum payoff that the third party purchaser might obtain by bundling, rather than the expected value of the project. Subjects seem to take the most positive outcome of bundling as a focal point for the division of surplus with the purchaser. From the manipulation check, it seems that respondents were making a conscious choice rather than being confused about the expected value of bundling.

In the aggregate, however, this is a gloomy outcome. The focus of right holders on the optimistic outcome of the scenario imposes a heavy burden on the third party acquirer. The third party will need to negotiate in order to drive the initial reservation prices down to a price level that is below 50% of the initial stated price. Prior experimental research has demonstrated that initial selling prices are sticky, i.e. they influence the outcome of negotiations.[30] In the advent of these bargaining costs, projects with uncertainty have a higher chance of failing, by placing such considerable burden of negotiation on those engaged in high risk projects. The prospect of such high demands by complementary right holders may lead projects that involve higher degrees of uncertainty to be forsaken, despite positive expected values.

These findings are particularly relevant for the domain of patent law. Intrinsically, the development of medical products from broad inventions involves a high risk of uncertainty—history has demonstrated that the path of innovation is unpredictable.[31] In this area, substantial investments in research and development provide no guarantees. When the risk of research and development is high and is not accounted for in the licensing prices of upstream patents, medical research may be biased towards low-risk enterprises.

On a general level, the profits obtained by bundling the individual parts can be conceptualized as a commons. As individual right holders, each has a veto right to the third party's project of bundling these resources.

---

[30] Anthony N. Doob et al., *Effect Of Initial Selling Price On The Subsequent Sales*, 11 J. PERSONALITY & SOC. PSYCHOL., 345-50 (1969). A number of field experiments investigated the effect of an initial selling price on subsequent sales of common household products. The results are consistent with dissonance theory in that subsequent sales prices track initial prices.

[31] A major historical example of the difficulty of getting an accurate estimation of the expected value of inventions is IBM's underestimation of the future market of home computers. *See* Robert P. Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75, footnote 41 (1994) (*citing* NATHAN ROSENBERG, EXPLORING THE BLACK BOX: TECHNOLOGY, ECONOMICS AND HISTORY 220 (1994)): 'The computer was regarded by its inventors as a purely scientific device . . .' (*quoting* Barbara G. Katz & Almarin Phillips, *The Computer Industry, in* GOVERMENT AND TECHNOLOGICAL PROGRESS 162, 171 (Richard R. Nelson ed.,1980)). *See also* JON ELSTER, *EXPLAINING TECHNICAL CHANGE* 111 (1983); JOEL MOKYR, THE LEVER OF RICHES: TECHNOLOGICAL CREATIVITY AND ECONOMIC PROGRESS 154 (1990); CHRISTOPHER FREEMAN, THE ECONOMICS OF INDUSTRIAL INNOVATION 75 (2nd ed. 1982).

As with over-harvesting of common resources, uncertainty about the size of a commons leads to lower levels of cooperation. In our anticommons findings, higher degrees of uncertainty regarding the profitability of the project lead to higher demands by the stakeholders. In the face of these increasing demands, projects with higher uncertainty (even if they have identical expected values) are more likely to be forsaken as right holders demand more on an individual basis, while expecting that others will demand more.[32] This result aligns with research on common resource dilemmas where levels of cooperation are reduced when there is more uncertainty as to the size of the common pool.[33]

4.3.2. Low Degrees of Uncertainty with Large Downside

Experiment *E* measures prices under situations where the purchaser/entrepreneur faced a high probability of modest gains, but there was also a modest risk of a more substantial loss (*low risk-low profit model*).

Although the expected values of each of the several scenarios were identical, reservation prices were consistently lower in cases with a large uncertainty regarding the size of the (strictly) positive outcome than in cases with relative certainty but with a modest chance of a negative outcome (See *Figure 6* above). A possible explanation for this result is that subjects emphasized the relative low probability of success in *D* over the possibility of a negative outcome in *E*.

According to the well-known framing effect,[34] it is assumed that individuals adopt different reference points as decision outcomes are framed differently. Similarly, our results illustrate the influence of the

---

[32] *See* David V. Budescu & Amnon Rapoport, *Generation of Random Binary Series in Strictly Competitive Games*, 29 BULL. PSYCHONOMIC SOC'Y 530 (1985). David Budescu et al., *Simultaneous vs. Sequential Request in Resource Dilemmas with Incomplete Information*, 80 ACTA PSYCHOL. 297 (1992); David Budescu et al., *Common Pool Resource Dilemmas Under Uncertainty: Qualitative Tests of Equilibrium Solutions*, 10 GAMES & ECON. BEHAV. 171 (1995).

[33] Kopelman et al., *supra* note 12, at 125-27.

[34] The prototype of a framing task is the Asian disease problem. Participants are told about an epidemic of Asian flu, which is expected to kill 600 people in the USA. They then have to choose between two options: option A saves 200 people with certainty; option B saves all 600 people with probability $p = 1/3$ or nobody. Options A and B are framed as gains. Options C and D introduce a negative framing. By implementing option C, 400 people will die for sure, and by implementing option D all 600 people will die with probability $p = 2/3$ or nobody will die. Although each of the options have an identical expected value (in terms of lives saved), it is attributed to the framing effect that participants prefer option A (the sure option) over B (the risky option) in the positive framing condition, and prefer option D (the risky option) over C (the sure option) in the negative framing condition. *See* Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979); Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCI. MAG. 453 (1981).

communicated frame by the bundler. Although the expected value from bundling in survey *D* and *E* were identical, reservation prices were lower when the expected value was denoted solely in terms of gains. There are a number of possible interpretations of this outcome. The results parallel the findings of de Dreu, et al., that individual right holders are less likely to make concessions when the payoffs of the third party are conceptualized from a loss perspective.[35] In our experiment, subjects seem to be more mindful of uncertainty with regard to gains than losses.[36] Put differently, in considering the price at which they would sell their rights, sellers disregarded potential losses of the purchaser; however, they seemed more willing to lower the price to take into account potential profits. The tendency of right holders to decrease reservation prices when the reference-outcome was strictly positive, suggests a higher willingness of individual right holders to cooperate when a project is termed solely in terms of positive payoffs. Alternatively, the added complexity in the aggregate calculation of expected values involving positive and negative outcomes might lead to more exaggerated demands because of the stronger non-calculative nature of collective decision making in those instances.[37]

5. CONCLUSION

Over the past three decades, economists, psychologists, philosophers, and political scientists have conducted intensive research on social dilemmas. Such research has demonstrated that social dilemmas, such as public good and prisoner's dilemmas, are very context specific.[38]

---

[35] *See* Carsten K.W. de Dreu et al., *Effects of Gain-Loss Frames in Negotiation: Loss Aversion, Mismatching, and Frame Adoption*, 60 ORGANIZATIONAL BEHAV. & HUMAN DECISION PROC. 90 (1994).

[36] The adoption of a positive or negative frame has empirically been found to affect the outcome of dyadic negotiations. Such frames may influence the outcome of further negotiations. For example, negative framing induces greater risk seeking so that negotiators with a negative frame make fewer concessions and more often fail to reach agreement than negotiators with a positive frame. Max H. Bazerman et al., *Integrative Bargaining in a Competitive Market*, 345 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 294-313 (1985); William P. Bottom & Amy Studt, *Framing Effects and the Distributive Aspects of Integrative Bargaining*, 56 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 459 (1993); Margaret A. Neale & Max H. Bazerman, *The Effects of Framing and Negotiation Overconfidence on Bargaining Behaviors and Outcomes*, 28 ACAD. MGMT. J. 34 (1985).

[37] *See* Peter Colett, *The Rules of Conduct, in* SOCIAL RULES AND SOCIAL BEHAVIOUR (Peter Collett ed., Oxford: Basil Blackwell 1977) (Individuals often seek fast and satisfactory solutions rather than rational consideration of all choices.).

[38] Peter Kollock, *Social Dilemmas: The Anatomy of Cooperation*, 24 ANN. REV. SOC. 183, 185 (1998).

In this contribution, we addressed the specific elements of anticommons dilemmas, while deferring the interesting research task of contrasting commons and anticommons dilemmas.[39]

The "Tragedy of the Anticommons" is a social dilemma where veto rights are exercised even when the use of the common resource by one party could yield net benefits for all parties involved. This experiment explores how, when a common resource is subject to multiple exclusion rights held by two or more individuals, these co-owners may withhold these rights from other users to an inefficient level.

A number of conclusions can be drawn from the experiment:

1. Our results confirm the theoretical proposition that anticommons deadweight losses increase with the degree of complementarity between individual parts, and with the degree of fragmentation. This paper illustrates the pricing effect of the anticommons. The results in experiments *A* and *B* clearly show a positive correlation between the amount of the surplus demanded by the individual property right holders and (i) the degree of complementarity of individual parts into the buyer's project (*A*); and (ii) the number of individual right-holders (*B*).

2. Individual right holders base their reservation price on a proportion of the expected surplus of the bundler-purchaser. They disregard the objective value of the good altogether. In one instance (experiment *C*), the purchaser faced *five* sellers each of who each demanded *25%* of the expected value of his project.

3. In cases of uncertainty, the anticommons dilemma becomes more pronounced. In experiments *D* and *E*, pricing decisions seem to be anchored on the maximum payoff the third party purchaser might obtain by bundling, while disregarding the expected value of the project. Subjects seem to take the best possible result of bundling as a focal point for the division of surplus with the purchaser. In Experiment *D*, this focal point led to a total reservation price that was seven times beyond the expected value of the project. The extremely high reservation price created a serious gap between what individual right holders were asking, on the one hand, and what a third party entrepreneur could reasonably offer.

Another, more subtle response to uncertainty emerges from the comparison of experiments *D* and *E*. When deciding the price at which they will sell their rights, sellers seem to disregard potential losses of the purchaser, while they were more willing to take into account uncertainty with regard to profits. The tendency of the right holders to decrease reservation prices when the reference-outcome is strictly positive, suggests

---

[39] Elsewhere, we have investigated the empirical and theoretical question on the symmetry between commons and anticommons dilemmas. *See* Sven Vanneste et al., *From 'Tragedy' to 'Disaster': Welfare Effects of Commons and Anticommons Dilemmas*, 26 INT'L. REV. L. & ECON. 104 (2006) (Finding that anticommons situations generate greater opportunistic behavior and a greater risk of under use compared to equivalent commons dilemmas.).

a higher willingness of individual right-holders to cooperate with projects termed solely in terms of gains (see *Figure 6*). Subjects seem to emphasize the relative low probability of success in *D* over the possibility of losses in the survey *E*.

4. When stakes are minor, the individual right holders state disproportionately high reservation prices—35% in the case of a project with expected value of 100 euros. Where stakes are higher, the average reservation price remains relatively stable at 14-19% of the expected surplus. This all-or-nothing strategy surfaces throughout the various experiments.

* * *

To summarize, our experiment indicates the pricing effect in settings where complementary units are fragmented over individual right-holders. Absent price coordination among these right holders, the independent pricing decisions place a high negotiation burden on a third-party purchaser.

Our experiment leaves the dynamics of negotiations among fragmented owners to further research.[40] However, the results provide a proxy for the burden of negotiation placed upon the shoulders of a buyer who seeks to rebundle independently-owned property fragments. The results also provide an indication of the extent of the price concessions that a prospective seller will need to obtain to bring the price of bundling within the limits of the net expected value of bundling. If we assume initial selling prices are sticky,[41] the prospective costs of negotiations might lead to abandonment of value maximizing projects, leading to the tragic outcome of under use or idleness.

In this regard, our results reinforce the normative hypothesis of the anticommons: property right systems should be careful in allowing the

---

[40] *See*, Robyn M. Dawes et al., *Cooperation for the Benefit of us—Not me, or my conscience, in* BEYOND SELF-INTEREST (Jane J. Mansbridge ed., 1990) (discussing the impact of discussion and interaction in enhancing cooperation in social dilemmas).

[41] When the height of reservation prices is due to the attribution effect, it is likely that price concessions will be hard to obtain. Cognitive psychology documents peoples' inclination to discount new evidence that conflict with their prior beliefs (belief perseverance). According to confirmatory bias, people tend to misconstrue or misinterpret information, so that it becomes additional information that supports the initial hypothesis. The initial experiments include John M. Darley & Paget H. Gross, *A Hypothesis—Confirming Bias in Labeling Effects*, 44 J. PERSONALITY & SOC. PSYCHOL.. 20 (1983) (establishing that identical additional information is interpreted differently because of prior beliefs or backgrounds); Matthew Rabin & Joel L. Schrag, *First Impressions Matter: A Model of Confirmatory Bias*, 114 Q. J. ECON. 37 (1999) (providing a formal model demonstrating how confirmatory bias may induce overconfidence).

liberal creation of new property rights and fragmentation of existing property rights.[42]

42 *See* Francesco Parisi, *Entropy in Property*, 50 AM. J. COMP. L. 595 (2002).

THE ROLE OF STATUS QUO BIAS AND BAYESIAN LEARNING IN THE CREATION OF NEW LEGAL RIGHTS

Robert L. Scharff[*] *& Francesco Parisi*[**]

ABSTRACT

We consider the role of status quo bias and Bayesian learning on the creation of new legal rights utilizing a model of legal evolution in which judges have limited discretion to create new precedent based on personal values, but, in the long-run, are constrained by efficiency criteria. Our model demonstrates that status quo bias may effectively transform an ex ante inefficient rule into an ex post efficient rule. Because these legal rights are internalized over time through a process of Bayesian learning, new precedent is vulnerable to reversal until the new right has been sufficiently recognized and accepted.

INTRODUCTION[#]

Common law is built on the evolution of legal rights through the accretion of precedent through *stare decisis*. Over time, this system has greatly expanded the legal rights and remedies available to prospective plaintiffs. A number of scholars have suggested that the law resulting from this decentralized process is relatively efficient (See Erlich and Posner (1974), Rubin (1977), Priest (1977), Priest and Klein (1984), Landes and Posner (1985), and Cooter and Rubinfeld (1989)). The efficiency of the common law hypothesis posits that evolutionary forces will act to eliminate inefficient law. Nevertheless, the robustness of this conclusion has been challenged by analyses of case selection effects and judicial bias. Most recently, Fon and Parisi (2003) argued that adverse selection by plaintiffs likely results in a disproportionate number of cases being heard by judges with an ideological predisposition towards the expansion of plaintiffs' rights. While this may explain the expansion of rights in pro-plaintiff courts, it does not explain the eventual wholesale adoption of these new

* Corresponding Author. JD, PhD. Assistant Professor, Department of Consumer Sciences, Campbell Hall, Room 265E, The Ohio State University, 1787 Neil Ave., Columbus, OH 43210. Phone: 614-292-4549. Fax: 614-688-8133. E-mail: Scharff.8@osu.edu.

** Professor of Law & Director, Law and Economics Program, George Mason University School of Law. Distinguished Professor of Law and Economics, University of Milan.

We would like to thank the participants of the Levy Workshop in Law and Economics at George Mason University School of Law for valuable comments and suggestions.

rules by pro-defendant courts not bound by the precedent created in parallel courts.

In this paper, we examine how the interaction between status quo bias and Bayesian learning may play a role in explaining this phenomenon. In particular, we find that, under certain conditions, status quo bias acts to transform ex ante inefficient rules into ex post efficient rules.[1] Status quo bias also acts to stabilize existing law by making it less subject to repeal by future legislative or judicial interference. We also investigate the effects of Bayesian learning on the adoption of new legal rules. A process of Bayesian learning suggests that new precedent is most vulnerable to being overturned immediately after being instituted. Furthermore, Bayesian learning can help to explain the gradual adoption of liberal rules in more conservative districts.

The first section of this paper examines the theoretical and empirical support for status quo bias. We find that the phenomenon is well supported in the literature and is empirically verified by numerous experimental studies. In section two, we analyze the effects of status quo bias and Bayesian learning on individual utility. We explain how the initial allocation of rights affects ex post efficiency. In section three, a model of judicial decision making in the presence of status quo bias and Bayesian learning is constructed. We show how judicial uncertainty or bias may lead an ex ante inefficient rule to be embraced by the judiciary when the rule is ex post efficient. Finally, in section four, we examine some of the implications of the model.

I. THE EXISTENCE OF STATUS QUO BIAS

Status quo bias exists when decision makers choose to remain with the status quo more often than traditional economic theory would suggest. (Samuelson and Zeckhauser 1988). Thus, if the revealed preferences of an individual, adjusted for transactions costs, suggest that the individual make a change in consumption habits, the individual may still choose to do nothing. This behavior represents, in effect, a cognitive asymmetry in the estimation of value. Empirical evidence of this cognitive asymmetry is present in a large body of literature examining the difference between the willingness to pay (WTP) for the acquisition of a property right and the

---

[1] "Efficiency" in this paper is defined as any allocation that meets the Kaldor-Hicks compensation test. Under the Kaldor-Hicks test, an allocation is efficient if no alternative allocation can be made in which all persons would be at least as well off as they were before the reallocation, given optimal side-payments. For example, if A owns X and values it at $5 and B values X at $10, the current allocation is not efficient. A reallocation of X from A to B is efficient because B could adequately compensate A for the loss. This is an efficient outcome regardless of whether B does, in fact, compensate A.

willingness to accept compensation (WTA) for the loss of an equivalent property right.

A. *Theoretical Basis for Status Quo Bias*

Economic theory has traditionally recognized a difference between values for WTP and WTA. However, under neoclassical microeconomic theory, cognitive biases are not the basis for this difference. Instead, WTA is recognized as being equal to WTP plus a wealth (or income) effect. The wealth effect is defined as being equal to the difference in consumer's surplus under Hicksian demand functions for different wealth levels. (See Willig 1976; Freeman 1993). Because ownership of a good effectively increases an individual's wealth, the owner of a particular good will hold a higher value for that good than they would have had for the good if their wealth were reduced by the value of the good. This is consistent with a diminishing marginal utility of wealth.

Hanemann (1991) found an alternative explanation for this disparity in the substitution effect. The Hanemann explanation submits that the relatively large difference between measured values of WTP and WTA is the result of the relative nonsubstitutabilty of the goods in question. Specifically, using a neoclassical framework, Hanemann showed how a public good with no close private substitutes may have a WTA value that exceeds its WTP value by a greater amount than the income/wealth effect alone would predict.

The neoclassical view of value was challenged by the publication of Kahneman and Tversky's (1979) seminal paper on prospect theory. Kahneman and Tversky showed that individuals' perceptions and valuations of risk were subject to psychological phenomena such as framing effects and reference effects. This was a controversial position because it implied that individuals do not act in a fully rational manner. Of particular significance for the present paper, Kahneman and Tversky defined value as being a function of deviations from a reference point. The resultant value function is concave for gains and convex for losses. Furthermore, the reference effect suggests that there is an asymmetry in the marginal utilities of losses and gains. A steeper function for losses implies that the WTA value for losses exceeds the WTP value for gains. This has become known as loss aversion or regret avoidance.

Thaler (1980) followed with a critique of mainstream consumer choice theory. Thaler argued that individuals typically under-weight the opportunity cost for current consumption relative to out-of-pockets costs for alternative choices. This leads to an exaggerated preference for current consumption over other consumption possibilities. As a result, a loss of current consumption goods is valued more than the gain from the purchase of equivalent items. This 'endowment effect' leads to a value for WTA that exceeds the corresponding value for WTP.

While the theories advanced by Thaler (1980) and Kahneman and Tversky (1979) imply a degree of irrationality in decision making, the work done by Smith (2003) suggests a rational explanation for these cognitive asymmetries. Smith argues that a scarcity of cognitive resources results in an allocation of decision making between the mind (the active thinking part of cognition) and the brain (the automatic pilot). Because the mind's resources are costly to use, most decisions are relegated to the brain. Choices made by the brain follow established patterns of decision-making and are, therefore, highly context dependent. However, when an individual is faced with a new situation (such as the existence of a new legal right), continued reliance on the set patterns used by the brain may be inefficiently costly. At this point, either costly mental resources must be used to recalibrate the brain or (more slowly, but less costly) the brain will recalibrate itself in accord with the new reality. The rational desire to avoid these adjustment costs can result in an exaggerated preference for the status quo.

Each of the preceding theories falls under the rubric of status quo bias. This phenomenon may be built on both the psychological biases underpinning loss aversion, regret avoidance, and the endowment effect, as well as behavior consistent with rational choice theory. As described by Zeckhauser and Samuelson (1988), status quo bias simply represents a preference for the current state of affairs that biases the decision maker against both selling a good that is part of their current endowment and buying a new good. In fact, Zeckhauser and Samuelson explain this effect as both the result of rational decision making in the presence of transitions costs and uncertainty, and as the result of cognitive misperceptions and psychological commitment stemming from misperceived sunk costs, regret avoidance, or a drive for consistency.

In this paper, we adopt this broad definition of status quo bias. Whether the phenomenon being measured is due to income and substitution effects, mental adaptation costs, or cognitive biases, the result is the same. In each of these cases, there is a valid theoretical basis for a significant disparity between WTP and WTA measures. This disparity, in effect, leads to a situation in which preferences are context dependent and there is independent positive value placed on the current state of affairs.

B. *Empirical Support for Status Quo Bias*

The argument in favor of adopting a behavioral approach to law and economics is bolstered by empirical support for the phenomenon. In particular, evidence of WTA values that exceed WTP values by more than would be explained by income effects alone permeates the literature. In one early study by Viscusi (1987), the amount individuals would require in compensation for a risk increase of 1/10,000 from insecticide poisoning was a full order of magnitude greater than the amount the individual would

be willing to pay for a similar decrease in risk. Market simulating studies under controlled conditions have replicated this result, though not with the same degree of disparity. Kahnemen, Knetsch, and Thaler (1990) set up a market for pens and mugs in a classroom experiment and, after carefully correcting for the possibility of income effects, found a mean WTA value for the commodities that was double the mean WTP value. Samuelson and Zeckhauser (1988) found evidence for status quo bias both in an experimental study of investment decision-making and in empirical studies of choices of health plans and retirement fund options. Finally, Boyce, Brown, McClelland, and Schultze (1992) used the derivation of disparate WTA and WTP measures for the preservation of environmental amenities to show that the assignment of property rights is important for the acceptance of moral responsibility. These studies represent only a fraction of the studies that indicate the existence of a disparity between WTP and WTA measures.

C. *The Use of Status Quo Bias in the Analysis of Legal Rules*

Recently, the law and economics literature has recognized the importance of status quo bias. Korobkin (1998) cast doubt on the efficiency of contract default rules by demonstrating how the status quo bias inhibits efficient contracting around default rules. Prentice and Koehler (2003) view status quo bias as a psychological foundation for a normality bias that is found in lawmaking. Posner (2003) has suggested that endowment effect may explain why state laws conferring a right to maternity benefits have led to an increase in the value of those benefits.[2] More generally, Korobkin and Ulen (2000) argue for the wholesale adoption of a behavioral approach towards law and economics, including the recognition of status quo bias.

In a review of the literature Korobkin (2003) examines the circumstances affecting the presence and magnitude of the endowment effect. He finds that the effect exists for both tangible and intangible items; the effect tends to be stronger when the right was acquired through skill or performance rather than through chance; the disparity between WTA and WTP is larger when the items being traded are difficult to compare; the disparity is larger for entitlements with no close substitutes; the effect is larger when the item/entitlement is evaluated for its use value than when it is assessed for its exchange value; and agents without a stake in the outcome of a decision are less likely to exhibit the endowment effect than are principals. Because legal rights can take many forms, the role that

[2] The endowment effect recognized by Posner (2003) is defined as being due to a combination of rational adaptive preference, wealth effects, and the nonsubstitutability of the good in question. This definition is consistent with rational choice theory and is at odds with Thaler's endowment effect.

status quo bias is likely to play for a given right depends on the nature of that right. For example, the right to be free from sexual harassment is more likely to be subject to status quo bias than is the right to keep a dollar bill that you have found on your sidewalk. This is the case because a right to be free from sexual harassment has no close substitutes,[3] it is an entitlement that is likely to be evaluated for its use value—not for its exchange value, and there are significant transitions costs associated with the loss of the right.[4] Alternatively, the right to keep found money is likely to exhibit less of a status quo bias because it has many substitutes, is evaluated for its exchange value, and is not likely to be subject to large transitions costs. This distinction between types of legal rights is an important one that plays a role in how the evolution of a particular right is likely to be affected by status quo bias.

D. *Prescriptive Limitations*

It is important to keep in mind that this paper does not attempt to make normative prescriptions about how the law should be changed to account for status quo bias. Although some have argued that behavioral economics should be fully incorporated into a normative theory of law and economics,[5] we are hesitant to do so at this juncture. While it is clear that status quo bias does exist, it is far from clear why it exists. Furthermore, even if one could determine the root cause of the bias, accurately measuring the magnitude of the bias would be a formidable, if not impossible, task. Consequently, we have tried to minimize the normative content of the paper.

II. EFFECTS OF STATUS QUO BIAS AND BAYESIAN LEARNING ON THE VALUATION OF LEGAL RIGHTS

Most studies that examine status quo bias focus on disparities in the valuation of property rights. In this paper, we extend the analysis to legal rights in a more general sense. Although property rights have some unique aspects, the studies confirming a status quo bias are generally applicable to

---

[3] Rights that are protected by tort law typically would not be expected to have close substitutes because they deal with unique features (i.e. personal integrity) that cannot be replaced or protected by contract.

[4] A vulnerable person who suddenly loses the right to be protected from sexual harassment would likely have to expend significant cognitive and physical resources to reestablish an equilibrium mode of behavior in a world in which only self-protective measures were available.

[5] *See* Jolls, Sunstein, and Thaler (2000) for an argument in favor of using behavioral law and economics to make specific recommendations for improving the legal system. *See also* Rachlinski (2003) for arguments in favor of and against the use of behavioral social science to justify paternalism.

rights other than property rights (Korobkin and Ulen 2000). As a result, our examination includes rights that are less tangible than property rights. These less tangible rights, though certain to exist, are unlikely to be in the information set of all persons in society.[6] In addition, court cases not directly on point may create general uncertainty about legal rights by suggesting the existence of a legal right without actually conferring one. Nevertheless, ignorance is not static. Through media reports and interactions with others, individuals are constantly learning about and updating their beliefs regarding the legal rights they possess. We expect that this knowledge about the existence of legal rights will be incorporated into individuals' utility functions over time in a Bayesian manner.

A. *Valuation of Rights*

Assume that there are two parties, A and B, who both value the acquisition of an opposing right. This right could be a right to property, or it could be an intangible right, such as the right to obtain an equitable remedy under tort law. In the absence of ownership of the right, each party has a value for their willingness to pay to acquire the right (WTP_A, WTP_B). Alternatively, once a party has acquired ownership of the right, that party has a distinct new value for their willingness to accept compensation for the loss of the right (WTA_A, WTA_B). The presence of status quo bias leads each party's value for a currently owned right to exceed their value for a desired right that they do not own, or $WTA_A > WTP_A$ and $WTA_B > WTP_B$.

Now, assume that the right in question is the right to a tort remedy that has been allocated to A. Ex ante, if A and B have perfect information about the allocation of the right, the parties have values for obtaining the right equal to WTA_A and WTP_B respectively. Assume further that a benevolent dictator (i.e. a judge) who wants to maximize efficiency has the ability to reallocate the right. Ex ante, efficiency is determined by comparing A's WTA measure with B's WTP measure.[7] If $WTA_A > WTP_B$, it is more

---

[6] In fact, so many rights have been conferred in our society that it is impossible for one person to fully understand all of the rights they possess. We may realize that we have certain basic rights such as the right to privacy, or to speak freely, but most people probably do not know what the boundaries are for those rights. In addition to these intangible rights, the modern welfare state is constantly creating new rights to the use of tangible resources, such as the right to use the local community center, the right to surplus cheese at no cost, and the right to camp in a national forest. Only a small subset of rights will be known to a particular individual at a given point in time.

[7] This analysis assumes that status quo bias reflects legitimate preferences. If WTA is not a true preference, but is only a distorted version of WTP, WTP would be the correct measure for the benevolent dictator to use in all cases. Because WTP and WTA are, as we use them, measures of revealed preference, we take them as true preferences. In any case, the ultimate conclusions of the paper are not affected by this assumption because we are merely attempting to explain the phenomenon of legal evolution. We are not attempting to make any subjective evaluations.

efficient to let A keep the right. What happens, however, if our dictator mistakenly gives the right to B? Ex post, after B has received the right and all adjustments have occurred, the new values for the right to A and B are WTP_A and WTA_B. Status quo bias has resulted in a fall in A's value and a rise in B's value for the right. In a future determination of efficient rights allocations our dictator should now only return the right to A if $WTP_A > WTA_B$. Given a wide enough disparity between WTA_B and WTP_B, it is possible that $WTA_B > WTP_A > WTP_B$. If this is the case, our benevolent dictator should not reallocate the resource to A. Consequently, if the status quo bias is large enough, an allocation that is inefficient ex ante may become efficient ex post.

B. *The Role of Bayesian Learning*

The disparity between WTP and WTA measures is typically predicated upon the assumption that the individual who has a legal right knows that she has the right. The real world is not so accommodating. Often times an individual will be unaware of a court decision that has conferred a right upon her. Other times, an individual will have heard of the right but will not fully understand the implications of the decision. Still other times, the individual will rationally deduce that a new right has a nonzero probability of being overturned either by a higher court, or by the legislature. In each of these examples, the full value of the status quo bias will not be internalized.[8] To take this into account, we model the legal right as being internalized over time through Bayesian learning using a model suggested by Viscusi (1992).[9] At a given point in time (t), a typical individual's realized value of a right that has been conferred is:

$$WTA^*_t = \alpha_t \times WTP + \gamma_t \times WTA, \quad (1)$$

where WTP is the individual's willingness to pay for the right in the absence of status quo bias, WTA is the amount the individual would have to be paid for the right once the status quo bias has been fully internalized, α_t is the weight placed on the belief that the individual does not possess the right in question, and γ_t is the weight placed on the belief that the individual

---

[8] Similarly, an individual may believe they have a right when no such right exists.

[9] Other studies suggest alternative mechanisms by which a legal right is likely to grow over time. Strahilevitz and Loewenstein (1998) have demonstrated that the history and length of ownership of a legal right have a significantly positive effect on the size of the endowment effect measured. Additionally, Loewenstein and Issacharoff (1994) have shown that a person may initially discount the value of a right they have because the right is tainted by negative associations. If these negative associations dissipate over time due to growing general acceptance of the right, the value of the right will increase over time.

does possess the right. The decision weights α_t and γ_t are normalized by setting $\alpha_t + \gamma_t = 1$. Over time, as information about the new right pervades the public's consciousness, and as people begin to have confidence in the permanence of the new right, the average weight placed on WTA increases ($\partial\gamma/\partial t > 0$). Because the decision weights are normalized, the average weight placed on WTP declines symmetrically, or $\partial\alpha/\partial t = -\partial\gamma/\partial t$. Thus, in the absence of new negative precedent, the value to a typical individual of a new right increases over time at a rate of:

$$\frac{\partial WTA^{*}_{t}}{\partial t} = \frac{\partial \gamma}{\partial t} \cdot (WTA - WTP) \qquad (2)$$

As a result, if status quo bias exists for a right (WTA > WTP), the value of a newly acquired right will increase over time, all other factors being held constant.[10]

The exact dynamics of the change in γ are uncertain in a given case. However, it is not implausible to assume that the function might be s-shaped if initial doubts about the future existence of the right were followed by broad acceptance of the rule once it became clear that the initial precedent would not be overturned ($\partial^2\gamma/\partial t > 0$). This case is illustrated in figure 1. Prior to the creation of the right ($t < 0$), the average individual who is to benefit from the reallocation of the right has a value for the right equal to WTP. At time $t = 0$, a court ruling confers the right. Initially, few people know about the right and the future survival of the right is in question. Thus, while the value of the right would be WTA if it were fully internalized, incomplete information leads to a value of WTA^{*}_{0} at $t = 0$. By $t = a$, broad acceptance of the precedent begins to take hold, peaking at $t = b$. After time $t = b$, knowledge of the accepted right is slowly disseminated to the remainder of the population and the average value of WTA^{*}_{t} for the population approaches WTA asymptotically.

C. *Effect of Precedent on the Internalization of Rights*

One favorable ruling in an individual's judicial district of residence generally is not sufficient to cement that person's belief in a new legal right or remedy. Typically, new precedents are tested by other litigants, which

[10] Posner (2003) analogizes the stock of precedent to capital stock and views the precedential value of a given case as depreciating over time if new precedent does not support it. Posner's assumption is based on the fact that the value of the precedent is diminished because it is ignored or distinguished in similar cases. This does not apply to our analysis, however, because in our stylized example, we assume that either the precedent has been upheld by other similar cases or no similar cases have come before the court.

process determines the robustness of the original ruling. As precedent accumulates, the original precedent is affirmed, overturned, or distinguished.[11] Knowing this, individuals will not fully internalize a right until it is either affirmed by a higher court, or a sufficient body of precedent has accumulated to provide the individual with confidence in the rule. Therefore, γ increases with the accumulation of direct precedent J_D, or $\partial\gamma/\partial J_D > 0$.

Similarly, the absence of favorable precedent in a given jurisdiction does not negate the possibility of some internalization of a right. Individuals who observe a right being conferred in a neighboring jurisdiction may rationally update their beliefs to presume that the right in question is, or may soon be, recognized in their jurisdiction. This phenomenon will be enhanced where there is a great deal of favorable media reporting on the case in question. Consequently, the persuasive precedent (J_P) of neighboring jurisdictions will also have a positive effect on γ, or $\partial\gamma/\partial J_P > 0$.

Because γ directly modifies WTA, the net result is an expected increase in WTA^*_t values with the accumulation of precedent ($\partial WTA^*_t/\partial J_D > 0$, $\partial WTA^*_t/\partial J_P > 0$). The net effect of a change in precedent at time t is illustrated in equation 3.

$$\frac{\partial WTA^*_t}{\partial J_t} = \frac{\partial WTA^*_t}{\partial J_D} + \frac{\partial WTA^*_t}{\partial J_P} > 0 \qquad (3)$$

As we have discussed, allocative efficiency is a function of the values people place on rights. Therefore, through its impact on WTA^*_t, precedent changes the efficient allocation of rights.

III. JUDICIAL DECISION MAKING IN THE PRESENCE OF LEGAL PRECEDENT

The efficiency of the common law hypothesis minimizes the discretionary role of judges in the evolution or preservation of legal precedent. Instead, under this view, the evolution of the common law will tend towards efficiency because inefficient rules will be litigated more often than efficient rules, leading to a greater probability that these inefficient rules will be overturned (Rubin 1977; Priest 1977). While we do not take issue with the basic premise of common law efficiency, we do argue that judges are not effectively automatons who act as bureaucrats at the local DMV might. Because, as Posner (1993) has argued, judges are induced by

[11] *See* Fon, Parisi, and Depoorter (2002) for an examination of the dynamics that result in the expansion and contraction of precedent.

the "conditions of judicial employment . . . to vote their personal convictions and policy preferences—or, in a word their values," judges are likely to have tremendous influence over the allocation of legal rights.[12] Posner sees the ability of federal judges to be free from outside influences as ultimately supporting the efficiency hypothesis. This may be true; however, where there are multiple locally efficient allocations of rights (i.e., when status quo bias is sufficiently large), the efficiency hypothesis only guarantees that law will evolve towards one of these allocations. Which locally efficient allocation is ultimately chosen is indeterminate. In this section, we examine the effect of judicial bias and uncertainty on the development of legal rights.

A. *A Model of Jurisprudence*

We begin by positing a model of judicial decision-making. The model assumes that a given judge, i, will rule in favor of the recognition of a right or remedy at time (t) with probability P_{it}. In making her ruling, the judge will take into account both the merits of the present case M, and the existing stock of binding and persuasive precedent (J_t) in cases like the one at hand. The judge will also bring to the table her own biases based on her belief system B_{it}. A drive towards efficiency E is a part of every judge's belief system, but, for judges predisposed to equity, the weight attached to E may be small. As described above, the relative values of each individual's WTA^*_t is affected by precedent. Consequently, E is affected by precedent, or $E(J_t)$. One's belief system is also affected by other values (O) such as distributional equity and fairness more generally. Equation 4 illustrates the resulting probability of ruling in favor of the recognition of a right.

$$P_{it} = f(M, J_t, B(E(J_t),O_i)) \quad (4)$$

This differs from the models suggested by Rubin (1977), Priest (1977), and Fon, Parisi, and Depoorter (2005) in which judges' beliefs are exogenously determined. In our model, beliefs are endogenous. As equation 5 demonstrates, the probability of recognizing a right or remedy increases with positive precedent.

---

[12] Posner's focus in this article was on the incentives facing federal appellate judges and Justices of the Supreme Court. Our analysis is as relevant to state courts as it is to federal courts. While these judges do not face incentives that are as conducive to the freedom of voicing personal values, we believe that personal values still play a significant role in the judicial decision making process.

$$\frac{\partial P_{it}}{\partial J_t} = \frac{\partial f}{\partial J_t} + \frac{\partial f}{\partial B_{it}} \cdot \frac{\partial B_{it}}{\partial E} \cdot \frac{\partial E}{\partial J_t} \geq 0 \qquad (5)$$

Equation 5 can be divided into two components. The first component, $\partial f/\partial J_t$ deals with the direct effect of prior precedent on current decision making through *stare decisis* and the use of persuasive precedent as a tool to minimize information costs. The second component, $\partial f/\partial B_{it} * \partial B_{it}/\partial E * \partial E/\partial J_t$, deals with the indirect effect of precedent on decision making through its effect on efficiency.

The direct effect of binding precedent on decision making through *stare decisis* and the use of persuasive precedent to minimize information costs has been addressed in the law and economics literature (See, e.g., Talley 1999; Posner 2003; Fon, Parisi and Depoorter 2005). A single precedent in a given district does not automatically bind all similar decisions in that district. Rather, *stare decisis* evolves from the gradual accumulation of reinforcing precedent. Nonetheless, the probability of winning a case is much greater when there is one precedent on your side, than when there is none. Conversely, when there are thirty cases on-point that support your argument, the value of a thirty-first case is likely to be negligible. In other words, there are diminishing marginal returns to new precedent in the creation of *stare decisis*, or $\partial f/\partial J_t > 0$ and $\partial^2 f/\partial J_t < 0$.

Persuasive precedent from parallel districts has a similar, but smaller, effect. However, because persuasive precedent does not directly represent the will of the court, persuasive precedent only has value as a mechanism to avoid potentially large information costs (Talley 1999). It may be more efficient for a court to rely on the reasoning of a parallel court rather than engage in a potentially expensive and uncertain analysis of its own. This option has the further benefit of minimizing the chance that the court's decision will be overturned, especially if the first court's decision has been upheld on appeal.

The second part of equation 5 deals with the indirect efficiency-based effect of precedent on judicial decision making. As equation 3 demonstrates, an increase in precedent, J_t, results in an increase in WTA^*_t for the class of persons benefited by the precedent. This is due to the increased confidence in the future existence of the right or remedy that is inspired by each new precedent. As the value of WTA^*_t increases for persons given, or expecting to be given, the right in question, the net social gain from allocating the right to these people increases. This, in turn, increases the probability that this allocation is efficient ($\partial E/\partial J_t > 0$). As the subjective efficiency (E) for a rights allocation increases, so does the judge's belief in that right ($\partial B_{it}/\partial E > 0$). Finally, a greater belief (B) in the utility of awarding the right leads to an increase in the probability that judge i will actually rule in favor of the right's existence, or $\partial f/\partial B_{it} > 0$. Together,

these dynamics imply that the efficiency enhancing effects of precedent will strengthen path dependence.

B. *Efficiency of the Common Law Versus Precedential Cascades*

The above analysis may appear to support a theory of rampant inefficient precedential cascades.[13] This is not likely to be the case. Talley (1999) lists a number of criteria that must hold for precedential cascades to be sustained. First, judges must be rule-bound. Where *stare decisis* is the rule, path dependence is the inevitable result. However, as Posner (2003) and Talley (1999) suggest, *stare decisis* is not rigidly adhered to. Second, information about past decisions (with the exception of the holding) must be costly or unavailable. Although there is a cost involved in reading cases, detailed written opinions are widely available on-line. Third, there must be judicial homogeneity. Posner (1993) and Fon and Parisi (2003) argue that such homogeneity does not exist. In this paper, by allowing idiosyncratic ideology to affect judges' decision making ($\partial f/\partial O_i \neq 0$) we explicitly foreclose this possibility. Fourth, there must be short judicial tenures. This typically is not the case, especially for federal judges who have lifetime tenure. Fifth, the hierarchy of the court system must be relatively flat. The existence of a number of higher courts discounts this factor. Finally, there must be population stationarity. The population characteristics of most states, and of the country as a whole, have changed dramatically over the last half-century. In sum, none of the factors necessary for sustained precedential cascades are present in our model.

Nevertheless, in our model, limited ex post efficiency-producing precedential cascades may occur. Consider the following scenarios, as illustrated in figures 2 and 3, in which there is a right initially held by type B persons, which is incorrectly reallocated to type A persons.

Scenario 1. Ex ante, as illustrated in figure 2, group A has a willingness to pay (WTP_A) for a right or remedy that is less than the willingness to accept value for group B1 (WTA_{B1}).[14] Suppose a judge makes the inefficient ex ante decision to give the right or remedy to group A due to ideology, misunderstanding of previous precedent, or miscalculation of efficiency criteria. If other judges follow suit, a cascade may begin. Seeing the accumulation of precedent, the members of group A begin to internalize the right. Consequently, the value of the right to group A increases to WTA^*_{At}, which increases over time. Concurrently, B types

---

[13] A precedential cascade is a phenomenon that occurs when ostensibly independent decisions are so heavily influenced by previous decisions that a path of inefficient decisions is created that is immune to the pressures at work in the efficiency of the common law hypothesis.

[14] The value of the right for each group is assumed to be the sum of the values of the members of the groups.

begin to believe that they may not have the right in question. If B types respond rationally, WTA^*_{B1t} will begin to fall. At any time before t = a, the cascade is vulnerable and is subject to the external pressures implicated by the efficiency of the common law hypothesis. Prior to t = a, members of group B1 have the incentive to try the rule in districts with efficiency-minded judges so that it will be overturned. However, if the right is one that has been given to plaintiffs, defendants may not be able to have their cases heard. This would occur because, until t > a, plaintiffs only have the incentive to try those cases in districts with judges predisposed to their clients. Once t > a, plaintiffs will be willing to file anywhere because it is now (ex post) efficient to hold for A. Eventually, the values of group A and group B1 will approach WTA_A and WTP_{B1}, and a new, locally-efficient rights allocation will be established.

Scenario 2. Ex ante, group A has a willingness to pay (WTP_A) for a right or remedy that is less than the value to group B2 (WTA^*_{B2}). As in scenario 1, a limited precedential cascade will lead to a reduced value for the right among B types and an increased value for the right among A types. Nevertheless, as figure 3 demonstrates, given that group B2's minimum value for the right (WTP_{B2}) is greater than group A's maximum value for the right (WTA_A), there is no amount of accumulated time or precedent that will cause the allocation of the right or remedy to A to be even locally efficient. In this case, the cascade will always be vulnerable to the pressures underlying the efficiency hypothesis of the common law.

The above scenarios demonstrate that, while precedential cascades may occur, it is important not to overstate the probability of occurrence or long-term survival. The efficiency of the common law hypothesis suggests that precedential cascades are only likely to be sustainable where there are multiple locally efficient rights allocations ($WTA_A > WTP_B$ and $WTP_A < WTA_B$). If the status quo bias creates a WTA value that is an order of magnitude greater than the corresponding WTP measure, precedential cascades may be widespread. If however, the status quo bias is more modest, as we suspect, fewer precedential cascades will occur and large deviations from globally efficient legal rules will be rare.

IV. IMPLICATIONS

A number of interesting implications flow from the model outlined above. First, given a presumption of valid context dependent preferences and status quo bias, multiple allocations may be (ex-post) Kaldor-Hicks efficient. Second, a combination of media effects and adverse selection of disputes for litigation will have the effect of equalizing rights across jurisdictions over time. Third, while plaintiffs' rights may increase over time, in the absence of external shocks or broad cultural shifts, we would not expect to see systematic contraction of these rights over time. Finally,

the model of Bayesian learning suggests that precedent is most vulnerable to attack soon after it is made.

A. *Multiple Efficient Allocations*

The first implication of the model is that, for a given set of rights, there may be a number of allocatively efficient outcomes. Although there may be only one path that maximizes efficiency ex ante, there are likely to be a number of paths that prove to be ex post efficient. This does not mean that every ex post allocation will be efficient. Differences in ex ante and ex post efficiency are most likely to occur where opposing parties have values for a right which are close, or where the disparity between WTP and WTA measures are large.

Both theoretical and empirical studies show that the disparity between WTP and WTA is greatest when the right is not substitutable. As a result, the status quo bias is least likely to play a role when rights to fungible commodities are involved, and is most likely to be determinative when unique rights are at issue. Therefore, we expect more equilibria to develop around rights such as the right to privacy and the right to bodily integrity, while property rights dealing with substitutable chattels may have only one efficient outcome.

Alternatively, the Boyce, Brown, McClelland, and Schultze (1992) conclusion that the assignment of property rights is important for the acceptance of moral responsibility suggests that the disparity in value measures will be greatest where the assignment of the right has moral implications. Thus, in cases where we are dealing with risk decisions affecting children (as in the Viscusi 1987 study) or where environmental stewardship is at stake (as in Boyce et al. 1992), the disparity is likely to be greatest and it is more likely that there will be multiple efficient equilibria.

In any event, the experimental studies involving tangible and largely fungible items, such as coffee cups and pens, demonstrate that the status quo bias exists; even in the absence of substitution effects and assignment of moral responsibility (see Kahnemen, Knetsch, and Thaler 1990). Thus, multiple ex post efficient equilibria are still possible, if not likely, for rules affecting the allocation of these types of goods.

B. *Rights Will Tend to Equalize Across Jurisdictions Over Time*

Another phenomenon that one would expect to observe in a free society with mass media is an evolutionary drive towards equalization of rights across jurisdictions over time. The adoption of a rule in one district is likely to have effects on both neighboring districts and, if the allocation of the right is sufficiently newsworthy, on distant locals.

As individuals outside the precedent-producing jurisdiction hear about a new precedent through media reports and conversations with acquaintances, those individuals will begin to believe that such a rule is possible in their jurisdictions at some point in the future. This will manifest itself as an increase in the weight attached to their WTA values (γ). Because the effects of media coverage and personal contacts are likely to be largely confined to the region in which the initial precedent is made, we expect any expansion of a right to occur in neighboring jurisdictions initially. However, where the allocation of a right is sufficiently controversial or otherwise newsworthy, national coverage may affect values of γ far from the source of the decision. Where this occurs, adoption patterns are less likely to be regional and more likely to be ideological.

Within a given region there are likely to be both efficiency-minded judges and equity minded judges. Assume that a potential plaintiff in a given district wants to litigate a claim that would expand their rights. If the expansion of the right would not serve ex ante efficiency goals, but would serve equity goals, the party will only litigate the claim if they draw an equity judge for their trial. Thus, selection effects will lead early cases that are ex ante inefficient to be tried in front of judges motivated more by equity. As precedent accumulates, communication of this fact to the population increases the value of the right by increasing γ. Eventually, the right may be of great enough value that even efficiency judges will recognize the right. At this point, plaintiffs will litigate suits even if they draw an efficiency judge and the new right will gain broad acceptance. This process may explain why it is not unusual for conservative jurisdictions to eventually adopt liberal rules.

The spread of information about court decisions and the process of adverse selection in litigating disputes both explain how initially isolated decisions to expand rights may come to be generally accepted.[15] Nevertheless, where there are entrenched special interests, regional cultural differences exist, or where there are large demographic differences between populations, the spillover effect may not occur and differential rules may persist indefinitely.

C. *Plaintiffs' Rights are Unlikely to Contract in the Long Run*

It is important to note that the mechanisms described above typically advance plaintiffs' interests. Defendants' rights are unlikely to be advanced

---

[15] It is also possible that the media will have the opposite effect. Litigators who keep current with legal news from across the nation may incorrectly view the reallocation of a right in a distant jurisdiction as a sign that the time is right to try such a case locally. If the litigator optimistically tries such a case before the public has internalized the right, he may only succeed in establishing precedent against the reallocation of the right.

through a similar mechanism because defendants typically do not have the same ability that plaintiffs have to force a ruling in a friendly jurisdiction. Because defendants typically do not initiate suits, plaintiffs often determine the venue in which a case is to be tried. Furthermore, if the plaintiff is unsuccessful in securing a friendly judge for trial, they always have the option to drop the suit, thereby avoiding negative precedent. While defendants may be able to take the initiative and successfully argue for declaratory judgments in some cases, the opportunities and incentives to do so are minimal. For example, the asymmetric application of collateral estoppel in federal cases gives defendants an exaggerated incentive to avoid litigation altogether. In most cases, the nature of the case will make declaratory judgment infeasible, if not impossible. Therefore, defendants generally are forced to pursue their rights through other means, such as legislation.

The foregoing suggests that plaintiffs' rights will generally advance over time.[16] However, defendants' rights may be advanced by the judicial system when there are large shocks to the system, where broad cultural shifts have taken place, or where a media campaign has successfully shifted γ in the absence of favorable precedent. Nevertheless, holding these factors constant, the effect of status quo bias coupled with adverse selection of disputes for litigation tends to favor plaintiffs over defendants.

D. *Precedent is Most Vulnerable to Attack Soon After it is Made*

The final implication of our model is that the gradual process of Bayesian learning over time leads precedent to be most vulnerable immediately after it is made. Soon after a court issues new precedent, γ is still small and, hence, the value for the right conferred by precedent is relatively low. As time passes, γ increases and the odds of a reversal of precedent diminish. Therefore, a defendant might be expected to fight precedent most vociferously, either through a media campaign or through legislation, soon after it is first recognized. Of course, if few people know about the precedent and discussion of the debate only acts to inform more potential plaintiffs of the new right, it is not necessarily in the potential defendant's interest to raise the issue. In an atmosphere of rampant

---

[16] This does not mean that plaintiffs' rights will be advancing at all times. Ex post inefficient precedential cascades will still be vulnerable to the pressures suggested by the efficiency of the common law hypothesis. A wide ranging ideological movement in favor of plaintiffs' rights will produce new rules that are both ex post efficient and ex post inefficient. An ideological shift away from plaintiff's rights within the judiciary would, in the long run, result in the retraction of those plaintiffs' rights that are ex post inefficient. Nevertheless, in the long run, there is a net gain for plaintiff's rights over defendant's right.

ignorance, noisily fighting precedent might only act to increase γ, thereby increasing the chance that the precedent will stick.

CONCLUSION

The evolution of the common law has been extensively studied. However, neither accounts based on the efficiency of the common law hypothesis nor those based on judicial bias have satisfactorily addressed the trend towards the increasing recognition of plaintiff rights. In this paper we have shown that judicial decision making in the presence of status quo bias and Bayesian learning, when coupled with adverse selection of disputes for litigation, leads to an accumulation of precedent favoring plaintiffs. In those cases where the resulting reallocation of rights is ultimately ex post efficient, the efficiency of the common law hypothesis does not act to constrain the growth of plaintiff rights.

FIGURE 1

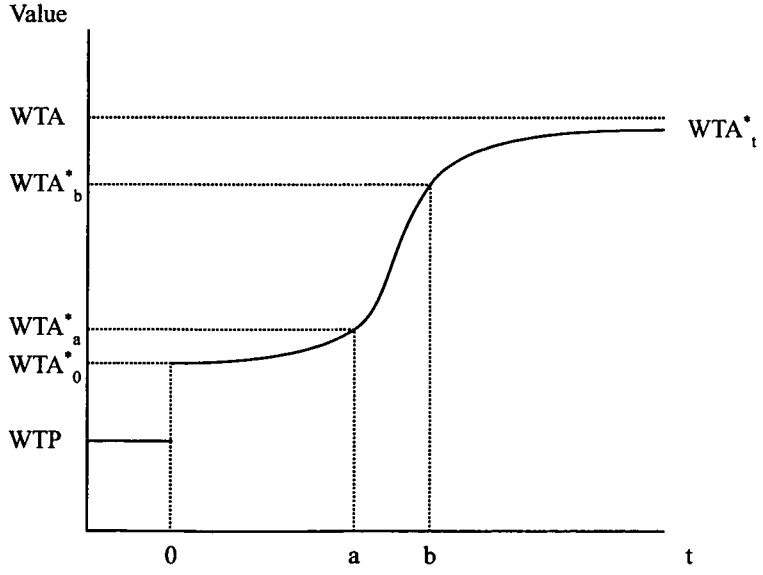


FIGURE 2

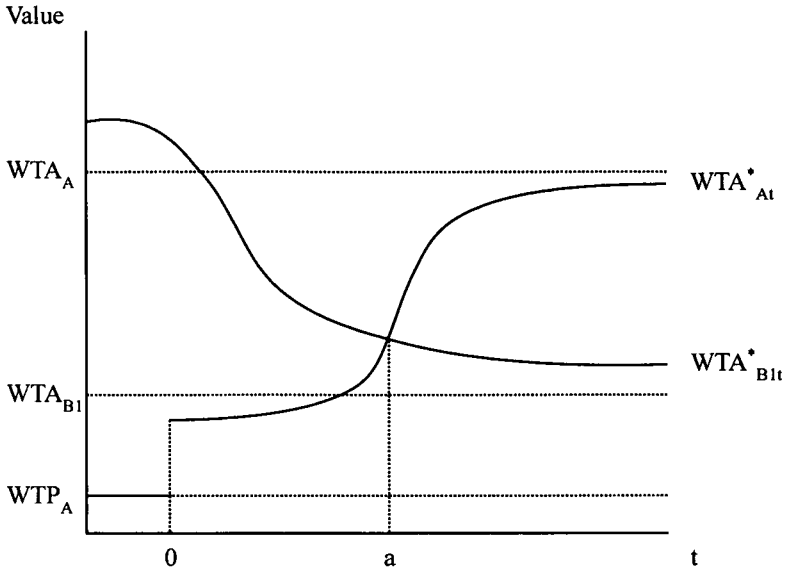
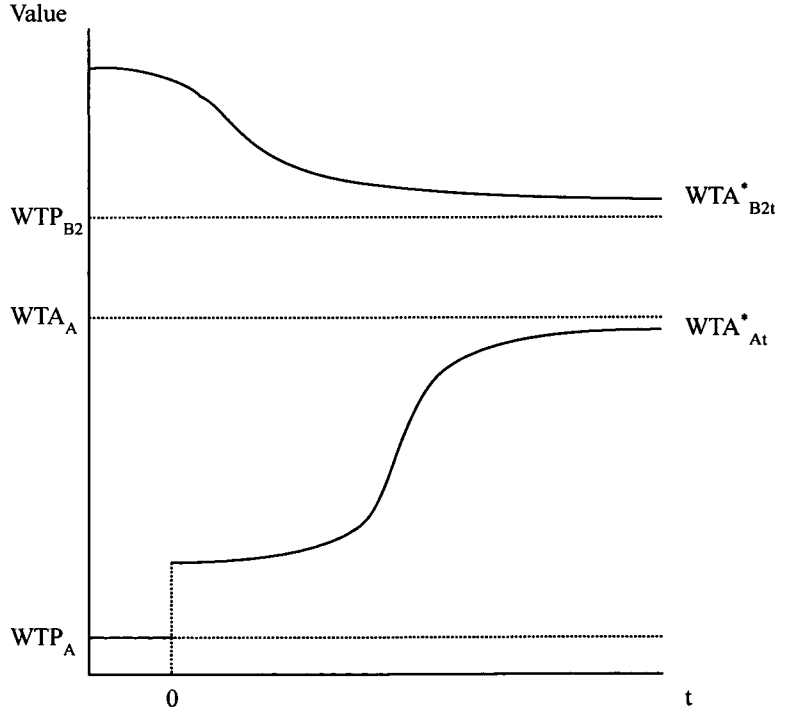


FIGURE 3



REFERENCES

Boyce, Rebecca R., Thomas C. Brown, Gary H. McClelland, George L. Peterson, and William D. Schulze. 1992. An experimental examination of intrinsic values as a source of the WTA-WTP disparity. *American Economic Review* 82: 1366-73.

Coase, Ronald H. 1960. The problem of social cost. *Journal of Law & Economics* 3: 1-44.

Cooter, Robert D., and Daniel L. Rubinfeld. 1989. Economic analysis of legal disputes and their resolution. *Journal of Economic Literature* 27: 1067-97.

Erlich, Isaac, and Richard A. Posner. 1974. An economic analysis of legal rulemaking. *Journal of Legal Studies* 3: 257-86.

Fon, Vincy, Francesco Parisi, and Ben Depoorter. 2005. Litigation, judicial path dependence, and legal change. *European Journal of Law & Economics* 20: 43-56.

Fon, Vincy, and Francesco Parisi. 2003. Litigation and the evolution of legal remedies: a dynamic model. *Public Choice* 116: 419-33.

Freeman, A. Myrick. 1993. *The measurement of environmental and resource values: theory and methods.* Washington, D.C.: Resources for the Future.

Hanemann, W. Michael. 1991. Willingness to pay and willingness to accept: how much can they differ? *American Economic Review* 81: 635-47.

Jolls, Christine, Cass R. Sunstein, and Richard H. Thaler. 2000. *A behavioral approach to law and economics.* In *Behavioral Law & Economics.* Cambridge, UK: Cambridge University Press.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economy* 98: 1325-48.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: an analysis of decisions under risk. *Econometrica* 47: 263-91.

Korobkin, Russell B. 1998. The status quo bias and contract default rules. *Cornell Law Review* 83: 608-87.

Korobkin, Russell B., and Thomas S. Ulen. 2000. Law and behavioral science: removing the rationality assumption from law and economics. *California Law Review* 88: 1051-1144.

Korobkin, Russell B. 2003. The endowment effect and legal analysis. *Northwestern University Law Review* 97: 1227-91.

Landes, William M. 1971. An economic analysis of the courts. *Journal of Law & Economics* 14: 61-107.

Landes, William M., and Richard A. Posner. 1985. A positive analysis of products liability. *Journal of Legal Studies* 14: 535-67.

Loewenstein, George, and Samuel Issacharoff. 1994. Source dependence in the valuation of objects. *Journal of Behavioral Decision Making* 7: 157-68.

Posner, Richard A. 1993. What do judges and justices maximize? (The same thing everybody else does). *Supreme Court Economic Review* 3: 1-41.

Posner, Richard A. 2003. *Economic Analysis of Law.* 6th edition. New York, NY: Aspen Law & Business Publishers.

Prentice, Robert A., and Jonathan J. Koehler. 2003. A normality bias in legal decision making. *Cornell Law Review* 88: 583-650.

Priest, George L. 1977. The common law process and the selection of efficient rules. *Journal of Legal Studies* 6: 65-82.

Priest, George L., and Benjamin Klein. 1984. The selection of disputes for litigation. *Journal of Legal Studies* 13: 1-55.

Rachlinski, Jeffrey J. 2003. The uncertain psychological case for paternalism. *Northwestern University Law Review* 97: 1165-1225.

Rubin, Paul H. 1977. Why is the common law efficient? *Journal of Legal Studies* 6: 51-63.

Samuelson, William, and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1: 7-59.

Smith, Vernon L. 2003. Constructivist and ecological rationality in economics. *American Economic Review* 93: 465-508.

Strahilevitz, Michal A., and George Loewenstein. 1998. The effects of ownership history on the valuation of objects. *Journal of Consumer Research* 25: 276-89.

Talley, Eric. 1999. Precedential cascades: an appraisal. *Southern California Law Review* 73: 87-137.

Thaler, Richard H. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1: 39-60.

Viscusi, W. Kip, Wesley A. Magat, and Joel Huber. 1987. An investigation of the rationality of consumer valuations of multiple health risks. *Rand Journal of Economics* 18: 465-79.

Viscusi, W. Kip. *Fatal Tradeoffs: Public and Private Responsibilities Towards Risk.* New York, NY: Oxford University Press, 1992.

Willig, Robert D. 1976. Consumer's surplus without apology. *American Economic Review* 66: 589-97.

EXPLORING THE SHAVELLIAN BOUNDARY: VIOLATIONS FROM JUDGMENT-PROOFING, MINORITY RIGHTS, AND SIGNALING

Nicholas L. Georgakopoulos[*]

I. INTRODUCTION

Economic analysis of law used to be subject to arrest by redistributive objections. Despite the efficiency that an economically optimal rule had, the objection could be raised that it also had undesirable consequences for the distribution of wealth. A thesis by Professor Shavell, later repeated with Kaplow, overcame this objection by showing that using an optimal tax for redistribution is superior to altering a non-tax, or substantive, rule to achieve the same redistribution.[1] When either an optimal substantive rule or an optimal tax is altered, the change induces a reduction of work in favor of leisure (the "chilling effect"). However, when a substantive rule is altered it also distorts the market for the activity that is the subject of the rule. This conclusion was instrumental for the success of economic analysis of law because it enabled economic analysis, as a normative tool, to proceed without being hampered by objections about the distributional effects of its proposals. Shavell's thesis can also be interpreted as identifying a boundary of proper objectives for substantive rules. Policy makers are cautioned not to use substantive rules to achieve redistributive

---

* Harold R. Woodard Professor of Law, Indiana University Law School—Indianapolis. I wish to thank comments of Phil Curry, Francesco Parisi, Antony Page, Rob Katz, Dan Cole, the audience at the 2006 annual meeting of the Canadian Law and Economics Association and two anonymous referees.

[1] Steven Shavell, *A Note on Efficiency vs. Distributional Equity in Legal Rulemaking: Should Distributional Equity Matter Given Optimal Income Taxation?* 71 AM. ECON. REV 414 (1981). *See also* Louis Kaplow & Steven Shavell, *Why the Legal System is Less Efficient than the Income Tax in Redistributing Income*, 23 J. LEG. STUDIES 667 (1994) [hereinafter *Redistributing*]; and Louis Kaplow & Steven Shavell, *Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income*, 29 J. LEG. STUDIES 821 (2000) [hereinafter *Clarifying*].

goals. The thesis therefore establishes a taboo, an option that rule designers must ignore, a boundary that they must not cross.[2]

Several law and economics scholars have explicitly taken the opposite position, and have defended the pursuit of redistributive goals with non-tax rules.[3] The literature is rich and varied. Some argue, for example, that redistribution beyond that which maximizes aggregate welfare is desirable.[4] Others argue that utility can be compared between persons, and that this allows efficient distributional outcomes.[5] Still others argue the more general point that distributional goals can be pursued more effectively with substantive legal rules rather than with tax rules.[6] This paper follows this last path.

Shavell and advocates of his theory acknowledge that substantive rules may, on rare occasions, be superior to an optimal tax, but they insist that this does not undermine their thesis that tax rules are the generally preferable method of redistribution.[7] This article explores the Shavellian boundary and describes three types of substantive rules that are exceptions to it (one might refer to rules that violate the Shavellian boundary as exo-Shavellian[8]). The exceptional rules are substantive rules that are motivated

---

[2] That distribution must not drive the design of rules does not mean that rule design must avoid distributional consequences. Take the goal, for example, of minimizing the cost of accidents by designing tort rules. Suppose, also, that the rule that minimizes the cost of accidents also has desirable distributional consequences, while a second rule induces slightly larger costs of accidents and has no distributional consequences. The Shavell boundary does not imply that the second rule should be preferred.

[3] *See, e.g.*, Mathew D. Adler & Eric A. Posner, *Re-thinking Cost-Benefit Analysis*, 109 YALE L.J. 165, 204-09 (1999) (arguing that utility can be compared between persons, which leads to a validity of distributional concerns); Christine Jolls, *Behavioral Economic Analysis of Redistributive Legal Rules*, 51 VAND. L. REV. 1653, 1656 (1998) (arguing that cognitive errors may indicate that distributional goals can be pursued more effectively with non-tax legal rules rather than with tax rules). *See also* Nicholas L. Georgakopoulos, *Solutions to the Intractability of Distributional Concerns*, 33 RUTGERS L. J. 279 (2002) (offering several candidates of rules that violate the Shavellian boundary). Some in the law and economics community also argue that redistribution beyond that which maximizes aggregate welfare is desirable, *see* Chris W. Sanchirico, *Deconstructing the New Efficiency Rationale*, 86 CORNELL L. REV. 1003 (2001) (arguing that infinitesimal adjustments in favor of redistribution do not violate the Shavell thesis); *see also* Ronen Avraham & Kyle D. Logue, *Redistributing Optimally: Of Tax Rules, Legal Rule and Insurance*, 56 Tax L. Rev. 157 (2003). In his plenary address to the American Association of Law Schools, Judge Guido Calabresi addressed the issue of altruism as a utility enhancing strategy for inducing people to like the altruist and expanded to social altruism through (possibly sub-optimal) redistributive schemes. *See* Guido Calabresi, *The Lawyer As Institutional Empiricist: The Case of Law and Economics*, AALS Annual Meetings Proceedings (2006).

[4] *See, e.g.*, Calabresi, *supra* note 3.

[5] *See, e.g.*, Adler & Posner, *supra* note 3.

[6] *See,. e.g.*, Jolls and Avraham & Logue, *supra* note 3.

[7] *Redistributing*, *supra* note 1, at 680-81.

[8] The "exo" prefix originates from the Greek ἔξω and means "out," so that exo-Shavellian rules are those that are outside the Shavellian boundary. Compare exogamy (marrying outside a group),

by redistributive goals and are superior to a substitute based on optimal taxation. The redistribution that the change of the rule provides could be provided by the tax system with an incremental, optimal tax. The "optimal tax substitute" of a substantive rule is the incremental, optimal tax that funds the same redistribution, paired, if necessary, with a substantive rule that induces the same substantive outcome.

The distinction between tax rules and substantive rules may appear simple on the surface, but it hides a complication. To a large degree, the legal system can be replicated using tax rules alone. Taxes that turn on "substantive" events, such as injuries or breaches of contracts, can be designed that produce incentives identical to those of all substantive rules about such matters as torts or contracts. Therefore, we must establish a definition of "tax rules" that distinguishes them from rules that have non-tax effects despite being ostensibly taxes. Yet, the definition must be broad enough to include not only taxes paid to the fiscus, but also benefits received from it, such as subsidies. Also, even though this analysis rests on optimal income taxation, the definition should include taxes on the basis of transactions, such as sales taxes.

A *tax rule* is a rule that produces a monetary obligation to (or monetary entitlement from) the government as a result of income or transactions and has no direct non-monetary effect. All rules that involve direct non-monetary obligations or entitlements and all rules that, despite involving only monetary obligations or entitlements, also have direct non-monetary effects are *substantive rules*. This definition means that some rules within the tax code may be considered substantive rules—particularly if they involve, beyond the payment of a tax or receipt of a subsidy, a physical obligation or entitlement or require the creation of additional information, such as the publication of the taxpayer's identity.

Part II of this article describes the components of Shavell's theory and explains the Shavellian boundary. Parts III through V provide examples of exo-Shavellian substantive rules: rules that breach the Shavellian boundary by being superior to their optimal tax substitutes. Part III explains how a tax on an activity that is complimentary with leisure improves even an optimal income tax. Part IV demonstrates how anti-majoritarian protections may have a redistributive nature that makes them superior to the rule that the majority would impose. Part V shows that substantive rules may produce signaling equilibria that have redistributive effects that cannot be obtained by even optimal tax rules. Part VI concludes.

---

exocrine glands (that do not secrete to internal organs), exosphere (outside the atmosphere), exoskeleton (a skeleton outside the body).

II. THE SHAVELLIAN BOUNDARY

Shavell's thesis rests on two fundamental economic principles. The first is the realization that taxes on specific activities or goods tend to be undesirable because they distort the market for those activities, as well as for their supplements and complements.[9] The second, more direct foundation applies the former observation, but considers not just activities that produce income, but also distortions in the market for activities that produce enjoyment directly: leisure. The first principle favors an income tax over sales taxes on various activities. The second reveals that no income tax is perfect, because it creates a bias in favor of leisure, particularly among the most skilled individuals.[10]

Shavell brought these principles to bear on the desire to produce some degree of income equality. The thesis requires the reader to suppose that the ideal shape of the legal rule in question has been established. The issue becomes how to evaluate changes for the purpose of redistribution. The question's very definition leads to the answer. Since it is given that the rule is ideal, any proposed changes to the rule could not improve its substance. Because the changes have redistributive purposes, Shavell compares them to an incremental amount of taxation that produces the same amount of redistribution.

That the redistribution achieved by the rule change is equal to the redistribution achieved by an incremental tax is a central component of Shavell's argument. It is also an assumption. While, as an assumption, it is proper, it also limits the thesis to those redistributive changes that have tax-based substitutes. The effect of this assumption is central for the present article. A second central assumption of Shavell's thesis is that the equivalent tax is optimal.

The components of Shavell's thesis are in place. First, a rule has the optimal form. Second, the change to be evaluated is motivated by redistribution. Third, an incremental tax that produces the same redistribution as the proposed rule change is feasible. Finally, that incremental tax is optimal. The conclusion is unavoidable. The optimal rule accompanied by an optimal tax is necessarily superior to changing the rule, because any

---

[9] The analysis of optimal taxation is more nuanced but the small size of its deviation and the complexity of its administration argues in favor of equal taxation. The intuition that all taxes on commodities and activities should be equal was actually rebutted by Ramsey who showed that taxes should be "such as to diminish the production of all commodities in the same proportion." F.P. Ramsey, *A Contribution to the Theory of Taxation*, 37 ECON. J. 47, 54 (1927).

[10] Again, the analysis of optimal taxation is more nuanced. A closer reading indicates that in a society with no taxation, a significant fraction of society would forego work; taxation restores their incentive to work. *See* J.A. Mirrlees, *An Exploration of the Theory of Optimal Taxation*, REV. ECON. STUDIES 175, 201 (1971) ("only men for whom $n > x^0$ [i.e., having skill above some level,] actually work").

change would lead to a substantively inferior rule and would impose a sub-optimal burden as a quasi tax. In other words, changing the rule is likely to cause two distortions compared to the ideal of both an optimal rule and optimal tax. The first distortion is that the rule will no longer be optimal, and the second distortion is that its effect as a tax will not be optimal. Characteristically, Shavell's argument is also called the double-distortion argument.

The conclusion of Shavell's theory allows some arguments for legal change to be considered improper or false. Arguments motivated by redistribution are improper because redistribution is achieved better by taxes. This conclusion can be considered akin to setting a boundary in legal argumentation. Arguments for substantive rule changes are "out of bounds" if they are motivated by redistribution.

This article explains three likely exceptions, offered as evidence that the Shavellian boundary is permeable. Redistribution alone justifies (i) excess liability on accidents from a leisure activity; (ii) rights to conduct distasteful to political majorities, such as abortion[11] or gun carrying in some jurisdictions; and (iii) rules that produce signals of skill, such as a physical housing subsidy. Parts III through V below argue that the redistribution these types of substantive rule changes produce is better than what would be possible using taxation.

III. JUDGMENT-PROOF LIABILITY BURDENING LEISURE

The theory of optimal taxation considers as a standard result that tariffs or use taxes are suboptimal because they distort the choice of conduct.[12] A tariff or use tax reduces demand for specific commodities or services, whereas optimality can be achieved only if the supply of, and demand for, commodities and services is determined by the market without distortions. For example, a tariff on boating would reduce demand for boating and related goods, and it would increase demand for some substitute activities, which would not have occurred if boating were available without the tariff.

In exceptional cases, however, tariffs may improve an income tax system. Because the enjoyment from leisure is not subject to an income tax, the unavoidable drawback of any income tax system is that it induces leisure. Thus, the optimal income tax has the sub-optimal effect of reducing work and increasing leisure. Consequently, a tariff on activities that correlate with leisure can increase the efficiency of an optimal income

---

[11] Roe v. Wade, 410 U.S. 113 (1973); *modified in* Planned Parenthood of Southeastern Pennsylvania v. Casey, 505 U.S. 833 (1992).

[12] *See generally* David E. Wildasin, *Distributional Neutrality and Optimal Commodity Taxation*, 67 AM. ECON. REV. 889 (1977).

tax by reducing the inducement toward leisure that is created by the tax. The same result can also be achieved by subsidizing a non-discretionary activity, such as housing.[13] Shavell and Kaplow concede that this is an exception to their theory.[14]

The liability system can reduce the distortions caused by the optimal tax even more effectively than a tariff on activities complementary with leisure. Whereas a tariff would distort demand generally, and may even have counter-redistributive effects,[15] the fear of liability does not fall on those who are judgment proof. Judgment-proof individuals can avoid their unpaid liability through bankruptcy's fresh start. As a result, a substantive rule change creating excess liability for leisure activities offers an improvement over a tariff on the same activities. Poor individuals can engage in the activity because they do not have to pay a tariff, and they are not deterred from engaging in the activity by the excess liability because they are judgment-proof.

An example illustrates. Assume that a society is comprised of individuals with various levels of skill and corresponding incomes. Without any income tax, individuals in aggregate would devote some fraction of their time to labor and the remaining to leisure. An optimal income tax burdens all individuals and has its chilling effect, leading individuals to reduce labor in favor of leisure.

Suppose that an activity exists that correlates perfectly with leisure and that can be subject to either a tariff or excess liability. In an exchange on this topic, Sanchirico and Kaplow & Shavell use as an example of a leisure activity "boating"[16] and, for the sake of consistency, I will follow this precedent despite the fact that some caveats are necessary: For example, the tariff on boating or the excess liability for boating accidents must

---

[13] *See, e,g.*, Helmuth Cemer & Firouz Gahvari, *Uncertainty, Optimal Taxation, and the Direct versus Indirect Tax Controversy*, 105 ECON. J. 1165 (1995).

[14] In a brief comment on optimal tax, Kaplow and Shavell concede that superior alternatives to the optimal income tax exist:

> [T]axes or subsidies on particular commodities might have indirect effects that reduce the distortion of an income tax. In particular, by taxing complements of leisure and by subsidizing substitutes, one can reduce the labor-leisure distortion and thereby improve welfare by more than the inefficiency that results from distorted purchases of the taxed or subsidized commodities. ... Thus, although a complete and sophisticated analysis does not demonstrate that it could never be efficient to change legal rules from what narrowly seem to be the most efficient ones, there is no general argument for adjustments of a conventionally redistributive type.

Redistributing, *supra* note 1, at 680-81.

[15] A fixed tariff, or a tariff comprised of a fraction of the activity's expenses, would have counter-redistributive effects by deterring those who could least afford the tariff from engaging in the activity. On the other hand, a tariff comprised of a percentage of the participant's income (so as to more accurately off-set the effects of the income tax) would not have the counter-redistributive effect.

[16] Chris William Sanchirico, *Taxes Versus Legal Rules as Instruments for Equity: A More Equitable Approach*, 29 J. LEGAL STUD. 797 (2000); *Clarifying*, *supra* note 1 (responding to Sanchirico).

exclude boating for vocational fishing or transportation; the example would be inapt in Venice or the Greek archipelagos but apt, perhaps, in a landlocked state with boating on lakes or rivers that is mostly recreational. Also, the intuition is clearer if boating is the only possible leisure activity, so that individuals who are deterred from boating by the tariff or the extra liability cannot turn to other, substitute leisure activities.[17]

According to optimal tax theory, a tariff on boating may improve the optimal income tax.[18] By deterring leisure, it increases labor, partly offsetting the chilling effect of the income tax. Even though individuals react to the tariff by reducing leisure, making receipts from the tariff lower, their increased labor compensates for that effect by causing individuals who reduce leisure to earn more income and pay higher income tax revenues. The revenues from the income tax plus those of the tariff therefore exceed the revenues of the optimal income tax alone.

Compare this optimal tariff to the imposition of excessive liability on the accidents from boating. Because all individuals have some probability of causing accidents if they engage in boating, the probability-adjusted excess liability is akin to a tariff. If debts could not be discharged in bankruptcy and no judgment-proof debtors avoided payment of their liability, the result would be identical to that of the tariff. In actuality, however, the less wealthy segment of society can avoid this liability through bankruptcy or merely by being judgment proof. The result is that the danger of liability produces its disincentive unequally. The wealthy (and productive) segment of society faces a deterrent from boating (and leisure) that is greater than that faced by the less wealthy segment of society. This difference makes excess liability on a leisure activity a likely superior addition to a regime of optimal income taxation.[19]

IV. MINORITY-PROTECTIVE MAJORITY-DISTASTEFUL RIGHTS

Some activities may be distasteful to the majority while strongly desirable to a minority. In certain regions of the country, the possession of firearms or the availability of abortion may be apt examples. There is some evidence that the strongest benefit from the possession of firearms accrues to the physically weakest segment of society.[20] For the majority, however,

---

[17] If substitute leisure activities exist the results are weaker but still hold. The results are weaker to the extent that those who are deterred from boating will engage in the substitute leisure activities. The result still holds, however, if not all of the deterred boating time is spent on substitute leisure activities but some is spent on labor.

[18] *See supra* note 14.

[19] Furthermore, the actual tariff may have counter-redistributive effects, deterring the consumption of leisure by the less wealthy disproportionately more than that by the wealthier.

[20] Defending oneself from crime without a firearm is a dominated strategy for either sex, but some evidence indicates that men experience only a small change in their rates of injury if they use a firearm

any protective benefit may be outweighed by distaste for firearms. Similarly, there is reason to believe that the strongest benefit from access to abortion accrues to the poorest, sexually active, teenage girls rather than to a majority of society.[21] For a majority (in certain jurisdictions) with a strong distaste for abortion, its availability would be harmful when benefits to the minority are outweighed by costs to the majority. As no market exists for the minority to compensate the majority for either "harmful" practice (firearm possession or abortion), the optimal rule under standard economic calculation appears to be a prohibition of the activity. Nevertheless, a rule of greater authority, such as a Constitutional provision or interpretation that prevents the nominally optimal rule from materializing, is desirable from a redistributional perspective. An example illustrates.

Assume that a society consists of 100 individuals. All prefer being educated to being illiterate. They divide into two sub-groups, one whose members achieve high utility with education (H)—i.e., a large net welfare gain from greater capacity for earning and leisure after educational costs in money, effort, and opportunities—and another whose members achieve less high utility with education (L). There are 35 high types, H, and each achieves a utility of 30 from education. There are 65 low types, L, and each achieves a utility of 20 from education. Receiving no education—i.e., being illiterate—corresponds to a utility of 10 for both H and L types.

This society is plagued by a type of accident, teen pregnancy, which prevents individuals from receiving education. Suppose that teen

---

to defend themselves against crime (from 4.85% to 3.51%). Women, on the other hand, experience a more significant drop in their rate of injury when using a firearm to defend against crime (from 3.08% to 1.25%). In this particular comparison, the data reach statistical significance, although if defending without a firearm is included, then it does not. *See* Lawrence Southwick, Jr., *Self-Defense with Guns: The Consequences*, 28 J. CRIM. JUSTICE 351-370 table 6 (2000). The example would be supported more vividly if the evidence distinguished between average men and women and weak or elderly men and women, and showed that the latter receive a greater benefit from firearm possession, but for the purposes of this example, I assume these conclusions to be valid.

[21] Evidence supports the hypothesis of the text that the availability of abortion increases the opportunities for schooling in minorities (read "in the disadvantaged" according to the main text) and little effect in the majority (read "those having large wage gains from education" according to the hypothesis of the text), *see generally* Joshua D. Angrist & William N. Evans, *Schooling and Labor Market Consequences of the 1970 State Abortion Reforms*, NBER Working Paper W5406, *available at* http:// ssrn.com published in 18 RESEARCH IN LABOR ECONOMICS 75 (S. Polachek, ed., 1999); similar results are reported by Adam Ashcraft, *Identifying the Consequences of Teenage Childbearing*, *available at* http://ssrn.com. The choice to not have a potentially desirable (in-wedlock) pregnancy in the face of increased difficulty of obtaining an abortion is documented, along with a thorough review of the literature, by Thomas Kane & Douglas Steiger, *Teen Motherhood and Abortion Access*, 111 Q. J. ECON. 467 (1996). The choice to travel to states with liberal abortion laws before the universal legalization, is used to determine the benefit due to the availability of abortion, *see generally* Timothy A. Deyak & V. Kerry Smith, *The Economic Value of Statute Reform: The Case of Liberalized Abortion*, 84 J.POL. ECON. 83 (1976).

pregnancy occurs with a probability of 20%. Accordingly, 7 of the high type individuals and 13 of the low type individuals would become pregnant and would not become educated unless they underwent an abortion. The aggregate utility cost from not educating any of these individuals would consist of the foregone utility of the 7 H individuals who each miss a gain of 20 (for a cost of 140), plus the foregone utility of the 13 L individuals who each miss a gain of 10 (for a loss of 130). The loss of both types sums to 270.

Further assume that abortion causes social displeasure of 280, or 2.8 per person. Thus, a ban of abortion costs 270 in foregone gains from education but saves harm of 280, and appears to be desirable. This calculation, however, ignores the counter-redistributive effect of the ban. Without the ban, individuals attain two levels of welfare. With the ban, some individuals attain a third, lower level, causing an increase in wealth inequality that itself tends to cause social displeasure. The average welfare appears to be higher with the ban only by ignoring any displeasure that arises from the greater wealth inequality caused by the abortion ban. If the increased inequality increases social discomfort by more than 10, or 0.1 per person, then the ban actually decreases welfare.

To state the model in the abstract, consider that the society consists of two types of individuals, high and low, symbolized by H and L. The individuals differ on how much their welfare increases by education. Welfare or utility is symbolized by u and depends upon each individual's type (high or low) and whether or not each receives education. Both type and education level are denoted by subscripts, with u_i corresponding to the welfare of uneducated or illiterate individuals. Educating a high-type individual (H) enables her to reach welfare u_H, rather than remain at u_i. Educating a low-type individual (L) results in a smaller gain, but her welfare reaches u_L. Thus, the welfare of educated low-type individuals is less than that of the educated high-type individuals, i.e., $u_i < u_L < u_H$. The probability of teen pregnancy is p. The fraction, or density, of high-type individuals is d_H, and $1\text{-}d_H$ is the density of the low type individuals. The displeasure, per person, of the use of abortion is c_A. And, the incremental displeasure from the counter-redistributive effect of its ban is c_B. The society seeks to maximize average welfare. The average welfare loss under the ban is symbolized by O_B, and is:

$$O_B = pd_H(u_H\text{-}u_i) + p(1\text{-}d_H)(u_L\text{-}u_i) + c_B\,. \qquad (1)$$

According to our example, the first two terms in the equation above amount to an average loss of 2.7 per person. The average welfare loss without the ban is symbolized by O_w and is:

$$O_w = c_A \ . \qquad (2)$$

According to our example, that was an average loss of 2.8 per person. All individuals would be educated and the baseline redistribution would take place. The ban appears desirable provided its immediate costs are smaller than the displeasure from abortions, i.e., if:

$$pd_H(u_H-u_i) + p(1-d_H)(u_L-u_i) < c_A \ . \qquad (3)$$

Despite the apparent desirability of the ban, however, it is undesirable provided that the disutility from abortions is smaller than the total disutility from the ban, including the displeasure that results from the incremental counter-redistributive effect, i.e., if:

$$c_A < pd_H(u_H-u_i) + p(1-d_H)(u_L-u_i) + c_B \ . \qquad (4)$$

It is important to note that this example is one of a desirable rule—the ban against abortion—that becomes undesirable when its redistributive effect is taken into account. In the United States, several individual states banned abortions until the U.S. Constitution was interpreted to prohibit such bans.[22]

As our exploration of the Shavellian boundary continues, the prohibition of the ban must be compared with its optimal tax substitute. In this setting, though, a tax can never be equal to the prohibition of an abortion ban. The remedy is costless, and no other cure is available. Neither could the increased welfare be taxed. First, only part of it is monetary. Second, even if it were all monetary, and that entire gain could be captured by a tax, giving that amount to those offended by the remedy would be insufficient. By definition, their harm exceeds the monetary gains. Since no tax substitute is available, the prohibition of abortion bans is an exo-Shavellian rule in the circumstances described above.

The relaxation of other simplifying assumptions strengthens the conclusions of the model. If the decision to ban abortion is made by the vote of an electorate that is mostly already educated, they are likely to ignore the next generation's distributional consequences.

A further aggravation of the ban's distributive effect appears if we make the plausible assumption that skill may influence the probability of pregnancies. Then, the more skilled teen girls avoid pregnancies anyway and benefit little from the ban. If pregnancies occur predominantly to the less skilled teens, they bear disproportionately the burden of the ban. The

---

22 *See supra* note 11.

effect, then, of the ban is not that some skilled and some less skilled remain uneducated but only that some of the less skilled remain uneducated, producing a greater disparity of wealth.

Under the specifications of the model, no substitute remedy existed. But, in fact, governmental policies may be able to restore educational capacity without abortion. A rule could require every educational institution to provide full daycare services, for example, which the government would fund with an optimal tax. In either case, even if the consequence is a complete recapture of the redistributive gain of *Roe*, an additional cost exists, the chilling effect of the additional tax to fund the daycare centers.

Available empirical evidence supports the model's conclusions. One study compares the effect of *Roe v. Wade*'s prohibition of abortion bans on disadvantaged groups: groups disproportionately comprised of low-type individuals. The prohibition of abortion bans improved their educational attainment more than it influenced that of the general population.[23]

V. SIGNALING ARRANGEMENTS

The actions of individuals may send signals about their capacities. One of the great drawbacks of any income taxation system is that it does not have the ability to tax capacity and is limited to taxing only realized income, letting able individuals avoid taxation by consuming leisure. An optimal income tax system would be improved if it extracted information about individuals' abilities. Then, abilities would command greater taxation and lack of ability would lead to lower tax or to subsidies. However, the tax system does not have the capacity to verify individuals' representations about their own ability. If the tax system were to ask individuals about their ability, able individuals could plead inability and ask for low tax rates or subsidies. However, economic theory suggests that circumstances may arise that induce the transmission of truthful signals. Their analysis is the object of *signaling theory* and its holy grail is *separating signaling equilibria*, i.e., settings that induce actors to separate according to their attributes by choosing different conducts.[24]

A legal regime may establish an environment that induces individuals to signal their ability—that is, produce a separating signaling equilibrium. If the result of this signaling equilibrium is desirable from a redistributive perspective, it is superior to an optimal tax substitute.

---

[23] *See supra* note 21.

[24] The signaling literature falls under the general topic of the economics of information and uncertainty. Despite the plethora of research papers on the topic, few introductory works exist. *See, e.g.*, INES MACHO-STADLER & J. DAVID PEREZ-CASTRILLO, AN INTRODUCTION TO THE ECONOMICS OF INFORMATION: INCENTIVES AND CONTRACTS (Richard Watt trans., Oxford University Press 1997).

An example illustrates. Assume that the employment of individuals is divided into periods: one "just hired" period and one or more "veteran" periods. Individuals divide into two types: high and low skill. Low-skill individuals are one half of the population, and high-skill individuals are the other half. Although skill determines productivity, an individual's skill cannot be communicated credibly to prospective employers. Employers find out an employee's skill only at the end of the first, just hired, period.

In an attempt to further redistribution, the government awards the visible benefit of housing to individuals with low income.[25] Moreover, only individuals with low skill find the benefit attractive.

Employers may then treat the benefit as a revelation of skill and award the high wages to individuals without the benefit and low wages to individuals who take the benefit. While this would tempt individuals to decline the benefit so as to obtain the high wage, this is not an appealing strategy because the employer will soon recognize the low skill and terminate the employment relation.

The subsidy achieves a redistributive goal directly. Moreover, it does not chill productivity, because the subsidy creates a separating equilibrium that reveals skill. An example illustrates.

Each period, high-skill individuals earn wages of 100 and low-skill individuals earn wages of 50. The housing subsidy is 20, and it is financed by a tax of 20 on the earnings of the high-skilled, which then drop to 80. In order to obtain the housing subsidy for any period, each individual would have to accept it before the beginning of the first employment period. A worker who incorrectly signaled high skill before the first period would be disqualified from receiving the subsidy in the second. Both high-skill and low-skill individuals are assumed to work the same amount of time during each employment period.

High-skill workers do not have an incentive to send a false signal. If skilled workers were to ask for the subsidy before starting to work, as soon as their high skill would be revealed and their wage would rise to 100, they would no longer qualify for the subsidy. Thus, they would enjoy one period of low wages and subsidy, or earnings of 50+20=70. The second period, they would be taxed and enjoy earnings of 80. By contrast, if they refused the subsidy and signaled their high skill, they would enjoy two periods of 80. (For simplicity, the discount rate is set at 0.)

Workers of low-skill also prefer not to send a false signal. If low-skill workers were to decline the subsidy, they would only receive the high wage for one period, receiving, net of taxes, 80. At the end of the first period, the employer would re-classify the worker as low-skill, and the worker would receive the low wage in the second period. Because qualification for the housing subsidy must occur before the first employment period, however,

---

[25] For reasons I will discuss, only the provision of subsidized physical housing, as opposed to a cash payment, will satisfy the requirements of a substantive rule that breaches the Shavellian boundary.

this individual would not qualify for the subsidy. The worker therefore earns only 50 in the second period, and the two-period total would be just 130. Instead, a low-skill worker would accept the subsidy before the beginning of the first employment period and enjoy earnings of 70 in each of two periods, for a total of 140.

Modeling the example allows its necessary conditions to be specified and clarified. Again, the model has two periods. The employer pays wages in each period, but in the second period the skill of employees has been revealed. Individuals divide into high and low types (H or L), and would obtain wages w_H or w_L, respectively, under perfect information. The subsidy has size s and has as its conditions that the recipient's income is low (w_L) and that the individual signaled truthfully in the first period. A high wage is burdened by a tax that finances the subsidy and depends on the proportion of the population, r, that has low skill. The resulting tax that burdens the high wage is:

$$x = r \cdot s/(1 - r) .^{26} \qquad (5)$$

Equilibrium requires that low-type individuals have an incentive to be truthful. Individuals of low type might have the incentive to pretend to have high skill to employers so as to obtain one period of high wage ($w_H - x$). If they did so, they would receive the high wage in period one, but they would receive the low wage in period two and forfeit the subsidy. Their total earnings would be $w_H - x + w_L$. If they acted truthfully in the first period, their total earnings would be $2 (w_L + s)$. To have the incentive to be truthful, individuals of low skill must prefer the latter income. For deception not to be attractive, the following inequality must hold:

$$2 (w_L + s) > w_H - x + w_L . \qquad (6)$$

Solving the above for s, reveals that individuals of low type would be truthful if:

$$s > (w_H - w_L - x)/2 . \qquad (7)$$

In other words, the subsidy must be greater than half the excess after tax earnings of high-skill types compared to low skill types. By

---

[26] Its derivation is easiest by assuming that the population is known to be n. Since a fraction r receives s, their aggregate subsidy is $n \cdot r \cdot s$. The remaining population, n (1-r), must raise the amount that solves for x the equation $n \cdot r \cdot s = n (1 - r) x$.

substituting the definition of the tax x from equation (5), we obtain a more definitive figure:

$$s > \frac{(r - 1)(w_H - w_L)}{r - 2} \qquad (8)$$

In the simplified case where the population segments are equal, and, therefore, the subsidy is equal to the tax and $r = .5$, then the subsidy must be greater than one-third of the difference between the two wages for low types to have the incentive to be truthful.

For the separating equilibrium to be maintained, the high-type individuals must also not prefer to pretend to be of low type so as to receive the subsidy. Truthful individuals of type *H* are assured of two periods with high wages, a total income of $2 (w_H - x)$. Taking the subsidy and the low wage in the first period will not prevent the employer and the government from finding they are high types, leaving $w_H - x$ as their sole income in the second period. Consequently, total income in both periods is $w_L + s + w_H - x$. To have the incentive to be truthful, individuals of high type must prefer the former income. For the deception not to be attractive, the following inequality must hold:

$$w_L + s + w_H - x < 2 (w_H - x) \ . \qquad (9)$$

Solving for x, after substituting s by its expression as a function of the tax, which is $s = x(1 - r)/r$, reveals that individuals of high type will be truthful if:

$$x < (w_H - w_L) r . \qquad (10)$$

In other words, for the individuals of high skill to have the incentive to be truthful, the subsidy must be smaller than the low-type's fraction of the difference between the two wages. A greater subsidy would not make sense, as it would allow the low skilled workers to enjoy greater total earnings than those of high skill.

Notice that the conditions for truthfulness of both groups can be satisfied. The example used a subsidy that was greater than a third of the difference between the wages and, while the low types fraction was half, a tax that was smaller than half the difference between wages. It was large enough to induce truthful signaling by individuals of low type and small enough to prevent individuals of high type from attempting to get the subsidy.

The signaling model shows, however, that only the visibility and verifiability of a non-tax rule can render it superior to a tax rule. If

employers cannot verify which candidates waive the subsidy, then the employers cannot use that information to distinguish the skill of employees and the separation unravels. Both types of employees would be offered the average wage, 75, by the employer in the first period. Moreover, the subsidy cannot be based on wage, since all employees have the same income in the first period.

A physical housing subsidy would satisfy the verifiability criterion of the substantive subsidy rule, as verifiability would be possible by checking the physical address of the employee. A monetary subsidy, on the other hand, which would result from a pure tax rule, would be easy to conceal, leading to the collapse of the separating equilibrium. Unlike abortion, tax law can emulate the outcome of substantive law by announcing the recipients of subsidies. Unless tax law does this, however, the substantive rule is superior.

VI. CONCLUSION

By identifying three potential violations of the Shavellian boundary, this article calls for empirical research to determine the exact contours of each violation in practice. To the extent that these violations are confirmed, the normative implications are manifest. Excessive tort liability should be brought to bear on activities that are synergistic with leisure, and jury instructions in the determination of punitive damages should include a relation of the injury to leisure as an aggravating circumstance. The interpretation of the constitutional protections of the Second Amendment and *Roe* should become more cognizant of the minority groups that receive their greatest benefits. Firearm purchase waiting periods and lock requirements should perhaps have exceptions for elderly purchasers and parental notification requirements for abortions should be reconsidered as being inimical to young women, who are the minority group that benefits most from *Roe*. Finally, a physical housing subsidy may be preferable over a voucher subsidy.

By identifying three sources of Shavellian boundary violations, this Article also demonstrates its permeability and lack of complete generality. While these three exceptions (and the collection of others referenced in footnote 3) may not appear to be exceptions of a phenomenon with generality, they are results of a limitation of a boundary that heretofore has been viewed with generality. From the perspective of economic theory, these examples exploit an assumption that restricts Shavell's thesis to redistributive rules that operate on monetary thresholds. From the perspective of legal theory, this does reveal one general exception to the Shavellian boundary. All rules with a non-monetary redistributive effect are potentially superior to the combination of the otherwise optimal rule and an optimal tax. This Article offered the examples of rules that improve optimal taxation, create welfare in minorities that the majority would

choose to eliminate, and allow signaling arrangements that a pure tax system could not. Others have offered rules that remedy tendencies for errors,[27] or the satisfaction of preferences by altruistic means,[28] and so on. This should be an unsurprising conclusion. Taxation operates on a small number of variables, essentially income and transaction price. Substantive rules operate on an array of conditions and caveats. Therefore, taxation is an instrument that is not likely to emulate the sophistication of substantive rules.

---

27 Jolls, *supra* note 3.

28 Calabresi, *supra* note 3.

FCC LICENSE AUCTION DESIGN: A 12-YEAR EXPERIMENT

*David Porter & Vernon Smith**

I. INTRODUCTION

Recent policy discussions regarding broadband Internet access have revived debates about various methods for allocating electromagnetic spectrum rights and the appropriateness of spectrum auctions (Telecomm. Rep. 2007). Debates over the assignment of spectrum rights via auction are hardly new (e.g. Herzel 1951; Coase 1959). Economists have long argued that auctions would promote efficiency in various ways, including the reduction of rent seeking and the avoidance of transaction costs used to reassign licenses in secondary markets (Kwerel & Felker 1985). Still, auctions have attracted vigorous opposition from defenders of traditional "public interest" licensing, who have argued that competitive bidding was not even feasible,[1] and that it would undermine the government's ability to regulate broadcasters. When this market-oriented approach was actually implemented in the United States in the early 1990s, these debates ceased being purely theoretical. Twelve years of actual experience with spectrum auctions now allow us to look back and assess what lessons can be learned from the adoption of this approach.

This short article provides a glimpse at the auction process's evolution, from its initial design to the rules governing the Advanced Wireless Services (AWS) auction held from August 9, 2006 to September 18, 2006. It also provides observations about the strengths and weaknesses of various auction designs, and it proposes ways to improve future auctions. Section II briefly describes the political environment that lead to the first electromagnetic spectrum auction in 1994 and the nature of the rights that are auctioned. Section III explains the basic auction design that the Federal Communications Commission (FCC) adopted, while sections IV and V discuss the rules that were implemented to accommodate that initial design and the bidding strategies that arose to take advantage of those rules. Sections VI and VII describe the rules FCC adopted for later auctions in order to put boundaries around that strategic behavior. Section VIII

* David Porter is a professor in the Interdisciplinary Center for Economic Science at George Mason University. Vernon L. Smith, is a professor of economics and law at George Mason University and a research scholar at the Interdisciplinary Center for Economic Science.

[1] Among those who argued that it was not feasible, is a former FCC Chief Economist, Dallas Smythe (1952).

concludes with observations about the shortcomings of FCC's chosen auction design, and it proposes an alternative design that should lead to more efficient use of the electromagnetic spectrum.

II. BACKGROUND

From the 1927 Radio Act until the mid-1980s, "comparative hearings" (distributing transmission rights by political fiat) constituted the sole license assignment method (Hazlett 1998). In 1981, federal legislation authorized the FCC to use lotteries for spectrum allocation—a compromise that stopped short of auctions. However, the lotteries, which the FCC used to assign hundreds of cellular licenses starting in 1984, made visible what had previously been hidden: failure to employ competitive bidding left billions of dollars of potential revenue on the table. Consequently, economists, who promoted auctions for efficiency reasons, had their advice bolstered by political demands to capture additional revenues.

Political realignment in the 1992 elections gave the Democratic Party control of the executive branch and both houses of Congress. This change in presidential administrations was accompanied by a fresh outlook, and the shared party affiliation allowed new accommodations to be reached between Congress and the White House that enabled the FCC to begin selling licenses to high bidders. The actual reform was included within the federal Omnibus Budget Reconciliation Act passed in the summer of 1993. The Commission was given one year to initiate auctions, and auctions did indeed begin, as mandated, in July 1994.

Before discussing auction design issues, however, a word should be said about the rights that are to be auctioned. The licenses awarded in spectrum auctions do not grant full property rights to a certain spectrum frequency. Rather, they give purchasers only those use or access rights that are defined in the license. For example, the 1993 legislation specifically authorized the award of bandwidth licenses for use by Personal Communications Service (PCS) networks—which include such items as mobile telephones, personal digital assistants, and similar devices. Thus, because the auctions were for licenses, and not broader rights to the spectrum frequency, they defined one, and only one, use to which the specified bandwidth could be put. However, the best use, in terms of highest value, for any spectrum frequency will always be subject to changing technology and economics. In a dynamic world, it is unwise to build in constraints on how any resource may be used; otherwise, today's efficient allocation may quickly become obsolete.

Consider another example. Digitalizing analog frequencies can greatly increase information transmittal capacity, but early allocations of certain spectrum bands to defined analog transmission—such as television—have locked that bandwidth into a comparatively low value use. Thus, television stations (channels 50 to 100) that were originally awarded rights in the 900

MHZ band have an asset that would be more valuable if digitalized and assigned to other uses. This is an argument for allowing previous winners at auction to enter new FCC auctions as sellers in a two-sided exchange. Doing so would allow any earlier property rights to be reassigned to higher value uses when technological innovations expand the ways in which the bandwidth can be used.

III. BASIC AUCTION DESIGN

Use of auctions was authorized by the federal budget statute enacted in 1993. However, the statute did not specify what form the auctions should take, so scholars and policymakers quickly began an investigation to resolve this issue. In the summer of 1993, a conference was held at the California Institute of Technology at which the merits of various proposals were discussed and demonstrated. Most of the main contributors to the academic auction literature attended the conference, and participants discussed three potential auction forms.

Participants first considered an initial proposal for a sequential auction in which licenses would be auctioned one-by-one, using either an English auction[2] format or a first-price sealed-bid[3] process.[4] This structure appeared easy to conduct and implement, but it had important shortcomings. Most of the license areas that would be auctioned were small relative to the area needed for efficient scale of operations, exactly the situation that obtained in mobile phone services. It is typical in PCS license auctions to have hundreds of non-overlapping franchise areas for blocks of spectrum with potential bidders having strong economies (in both consumption and

---

[2] In an English auction, bids are taken in real time, each successively higher, until only one bidder remains. The price paid is the last price bid.

[3] In a first-price sealed-bid auction, bidders submit the price they are willing to pay. The highest bidder wins, paying the price offered.

[4] There was also a discussion on the use of Vickrey, or second price, auctions. However, it was noted that these auctions have poor political appeal because they leave the impression that money has been left on the table. Thus if the highest sealed bid is $2 million and the second highest is $1 million the high bidder wins the item and pays $1 million leaving the superficial appearance that a million dollars was foregone by the seller. If, however, these were indeed the maximum amounts that the two highest bidders were willing to pay, then in the English progressive auction the bidding would have stopped the moment the most eager buyer raised the bid to $1million because the second most eager bidder would have declined to raise the bid. It follows that no one would have known that the winner of the auction was willing to go as high as $2 million. The second price rule was actually used in the New Zealand spectrum auctions in part because it was supported—in fact recommended—by auction theory and theorists as the favored rule in sealed bid auctions. Hence, this first meeting of theorists at Caltech was informed by earlier experience based on what had been thought to be a straight forward application of (private values) auction theory. This, in microcosm, is the first in the long sequence of reappraisals of the state of understanding of auction theory, based on experience, as it has been applied to the auctioning of spectrum rights.

supply) generated by regional or national networks. Sequential auctioning would not allow for the assembly of many interrelated (complementary or substitute) area licenses that could make most efficient use of the spectrum. It would be akin to trying to solve a general equilibrium pricing problem one market at a time.

A second auction format proposal suggested that large groups of licenses be auctioned simultaneously, so that bidders could see all of the prices forming and move their bids accordingly. This would at least acknowledge the area assembly problem posed by a sequential auction format, but it too fails to maximize auctioning's full potential. Because participants must bid on licenses without knowing whether they will also be able to acquire complementary licenses, they are likely to enter lower bids for any given license than they would be willing to enter if they could be assured of assembling a larger, contiguous license area.

Finally, there was a call to accomplish area assembly more directly in the auction process itself, by allowing a more integrated auction form in which bids could be placed in packages with various restrictions—what are known as combinatorial auctions.[5] The need for this type of auction arose because bidders argued that grouping licenses together into packages (to create regional networks) would be worth more than the sum of licenses bid a la carte.[6] When such possibilities arise, auctions that do not take into account these complementaries can result in financial losses to bidders (see Cull et al. 2000; Banks et al. 2002). However, combinatorial auctions have been thought to face difficult computational issues, sometimes referred to as the 2^N bogyman or NP-completeness.[7] Critics have also argued that

---

[5] Combinatorial auctions were first invented by Rassenti, see Stephen Rassenti, 0-1 Problems with Multiple Resource Constraints: Algorithms and Applications (1981) (PhD. thesis, on file with Univ. of Arizona), and subsequently published in Rassenti et al., *A Combinational Auction Mechanism for Airport Time Slot Allocation*, 13 BELL J. OF ECON. 402 (1982). For more information on these types of auction see CRAMTON ET AL., COMBINATORIAL AUCTIONS (2006).

[6] For example, bidder's valuation for one license in a market might depend on "who" will be the winner of the licenses in neighboring markets. A bidder might care what type of service its neighbor provides if roaming agreements are required. For PCS, there are three competing technologies: Code Division Multiple Access (CDMA), Time Division Multiple Access (TDMA), and Global System of Mobile Communications (GSM). Thus, bids conditional on which technology would be implemented in adjacent areas would be an important means by which all such information could be incorporated into the set of submitted bids, and taken into account in the awards. Failure to do this carried the potential of leaving lots of unrealized gains from exchange on the table, and creating financial uncertainty for the bidders.

[7] If there are bids for combinations of items and there are N items, then to completely enumerate all of the possible combinations (if there were bids submitted for all possible packages) then the computations grow exponentially but computational resources do not. Specifically, computationally easy problems can be solved by computer algorithms that run in polynomial time; *i.e.*, for a problem of size N, the time or number of steps needed to find the solution is a polynomial function of N. Algorithms for solving such difficult problems require times that are exponential functions (Non-Polynomial (NP) time) of size N.

combinatorial auctions would make it more difficult for new entrants to compete in auctions with incumbents, which we address in Section VIII below.

Erring on the side of conservatism, and without benefit of any laboratory testing in this early phase, the FCC decided to implement the second proposal, a simultaneous auction informed by what economists had learned from auction theory applied to simpler environments. The auction form FCC used is typically called the Simultaneous Multi-Round (SMR) Auction or the Simultaneous Ascending Auction (SAA) (see Milgrom 2000). However, one needs more than a name to implement an auction; specific rules are required.

While various types of wireless licenses would be auctioned, the initial interest was dominated by the assignment of PCS licenses.[8] These would enable additional competition in the mobile phone market, then structured as a duopoly with each U.S. market having two licenses allocated 25 MHz of bandwidth (50 MHz total). The FCC had allocated another 120 MHz for use by mobile carriers, with the bandwidth spread across six new license types (three allotted 30 MHz, three allotted 10 MHz). The country was divided into 51 non-overlapping license areas for two of the 30 MHz license types (PCS A and PCS B), and into 493 markets for the other four license types (C, D, E, and F). Hence, there were some 2,074 total licenses ([2*51] + [4*493]) to assign by auction.[9] This highly disaggregated licensing scheme stands in stark contrast to global markets, where the great majority of countries award nationwide licenses for mobile telephony. It adds complexity to the bidding process, as mobile operators attempt to construct regional or national networks by winning multiple licenses.[10]

IV. SPECIFIC AUCTION RULES

In order to implement its chosen auction scheme, the FCC had to make many decisions about rules that would govern the auctioning itself. The general structure of the SAA auction is that participants submit a series of

---

[8] While the FCC was required to begin holding auctions by the end of July 1994—and met that deadline with its auction for Interactive Video Data Service (IVDS) licenses—the first PCS auctions (PCS A and PCS B) did not begin until December 1994, concluding March 1995.

[9] The FCC actually pulled three licenses out of the PCS A and B license auction, awarding them to companies the Commission determined had made notable contributions to advancing PCS technology. This was under the "pioneers' preference" policy, later discontinued. The three awardees were charged license fees that were based on a formula using the winning bids paid for other licenses and that incorporated a discount for the pioneers' preference.

[10] Of course, networks can be (and are) pieced together via roaming agreements, such that it is not necessary for a single entity to own each license. Yet roaming agreements are not perfect substitutes for ownership, as seen in U.S. wireless networks where operators have made pointed efforts to aggregate licenses.

single-item, sealed bids for desired licenses. Following the submission of such bids, the high bids for each license are posted. These high bids then become the standing bids for the next round of bidding. Still, this basic design left many questions unresolved. For example, how long should a round last? Since the auctions were allocating highly valuable assets, some suggested that each round should be an entire day long, so that bidders would have plenty of time to digest the information from the auction and make intelligent bids.[11] Whatever the length, rounds should be long enough for firms and their consultants to peruse the data from each round in order to make informed bids.[12]

On the other hand, some observers believed that a set of rules to help speed-up the auction and allow for flexibility of the bidders was required. In particular, the FCC and its advisors were worried that participants would not bid on the cadre of licenses in which they were interested, but would hold back and wait to see what others were bidding on to gain an informational advantage. To discourage this type of strategy, a set of rules was designed to force the pace of the auction. First, for a bid to be *acceptable* in any round, it had to be greater, by a pre-specified *increment*, than the standing bid for that license. The FCC had to determine the increment size (this was listed in percentage amounts over the standing bid) and had the right to change the size of the increment during the auction for any license. Obviously, changing the amount of the increment can affect the speed of the auction and its allocative efficiency.

Second, the FCC introduced an *eligibility* requirement for bidding after the first round. Each license was assigned a numerical value in terms of *activity units*. This number is typically derived from the MHz and population (referred as MHz-Pops) associated with the license territory. For example, if a license consisted of 20 MHz of spectrum and within the boundary of that license there were one million people, the license would be assigned 20 million activity units. At the beginning of a bidding round, a participant would be eligible to bid only on a number of activity units that was related to the number of activity units on which he bid in prior rounds. The exact amount of a participant's *free eligibility* was equal to (1) the sum of the activity units of licenses for which he submitted acceptable bids in the previous round and for which that participant did not have the standing bid, plus (2) the sum of the activity units of licenses for which the participant had the standing bid two rounds previous, but no longer has the

---

11 The FCC did, however, use an open out-cry auction for the IVDS auction in 1994.

12 It should be noted that many of the participants in the design of the FCC auction rules were also at some point consultants to bidders in the actual auctions. The FCC auctions have created a cottage industry for consulting firms in assisting bidders through the myriad rules and strategies in the SAA. Critics have remarked that this is akin to a new "Military-Industrial complex"—call it the "FCC-Consultant complex." Many of the consultants argue, on the other hand, that much of their billing time was spent familiarizing the bidders with the complexities of the rules.

standing bid. Given this total eligibility budget, a participant is then constrained to bid on licenses for which he has the standing bid and additional licenses whose sum of activity units is less than his free eligibility.[13] Thus, if a bidder is interested in obtaining a license or set of licenses totaling 20 million activity units, he must actively bid on licenses totaling that level of activity or he will not be able to obtain that license set at the end of the auction. This rule obviously forces bidders not to lay in wait.

Given that the activity rules push the pace of the auction, the FCC created several rules to provide flexibility to bidders. First, eligibility for all participants would be adjusted within each round by the FCC's choice of a numerical activity rule factor to apply in each auction *stage*. The auction begins in stage one, and a factor $0 < \alpha_1 < 1$ is selected, so that if a bidder has X amount of current eligibility he is required to bid on licenses totaling only $X \cdot \alpha_1$ of activity units to maintain X. Thus, if the stage factor is .8 and a participant bids on licenses totaling 16 million activity units, then his eligibility for the next round can be as high as 20 million units. As the auction proceeds, the FCC has the right to move the auction to different stages with $0 < \alpha_1 < \alpha_2 < \ldots < 1$.

The FCC also created a limited number of *waivers* for each bidder, each of which allowed a bidder to advance to the next bidding round at his current level of bidding eligibility without bidding in the current round. This rule was implemented to provide bidders flexibility to respond to either a computer hardware or software problem or an unexpected need to consult on bidding or financing matters with senior management or paid advisors. For example, a bidder could use one of his waivers to sit out a round while he sought additional budget authority.

Because a bidder may value a combination of licenses more than the sum value of the individual licenses alone, there is a possibility of a "failed aggregation." To provide flexibility to bidders for this possibility, bids on provisionally winning licenses at the beginning of a round could be *withdrawn*. After a withdrawal, the FCC becomes the standing bidder for the withdrawn license and replaces the bid with one that is less than or equal to the withdrawn bid (typically the previous high bid for the license). An individual who withdraws a bid pays a *penalty* equal to the greater of zero or the difference between the amount of the bid he withdrew and the highest bid submitted by a participant other than the FCC after his withdrawal.[14] Thus, a bidder desiring both license X and Y may decide to withdraw a standing bid on X because he no longer wants to stay in the bidding for Y (and X), and vice versa. Of course, in a combinatorial

---

[13] This sounds complicated, and it is. Consulting firms and the FCC have created special software programs to assist bidders in tracking and managing their eligibility.

[14] Because a standing bid on a license may be withdrawn multiple times, the highest bid after a withdrawal need not be the final bid on a license.

auction, the bidder will only send a bid message for X and Y together, and he will not bid on X or Y separately unless he is also willing to buy X or Y without its complement. Unfortunately, the FCC rejected the combinatorial auction design.

An auction would come to a close when all bidders have no free eligibility remaining—that is, when the amount of their eligibility is equal to the amount of their standing bids. Licenses are awarded to the participants with the standing bids, and any withdrawal penalties are computed and paid at that time. However, to ensure that participants are committed to paying for the licenses on which they bid—or, as the FCC puts it, to "help deter frivolous or insincere bidding"—an upfront payment deposit must be placed with the FCC in order to be qualified to bid in the auction. The size of the upfront payment determines a bidder's initial activity unit budget. The larger the upfront payment made, the larger the initial eligibility of the bidder. At the end of the auction, the upfront deposit is returned if the bidder does not win any licenses. If a winner of a license fails to pay for any of the licenses on which he placed the standing bids at the end of the auction, he forfeits his upfront deposit.

In addition to the rules listed above, the FCC provides all information (including bidder identities, bids, eligibility amounts, etc.) to all participants during every round. The full information rule was justified for two reasons. First, the FCC believed full transparency was necessary because it was allocating rights to what is considered a public good. Second, because spectrum licenses have important common and affiliated value, knowing what others have bid can increase auction revenue. The hypothesis here is that, if some bidders are better informed about values, an open auction can allow this information to be revealed and reflected in the bidding; other bidders would update their value assessments and, thus, what they would bid.[15]

Every rule in the above list was implemented based on intuitions from the theory of auctions for allocating a single unit, however. At no time were these rules tested in a scientific manner, and the problem was far too complex to admit of formal modeling. Nevertheless, armed with this set of rules, the FCC began its experimental journey.

---

[15] Since the full information characteristics of the auction force revelation of value by the better-informed bidders, this undermines the original incentive to invest in acquiring information. To the extent, then, that there is under-investment in information that is not firm specific, this feature diminishes the importance of common values as an assumed feature of the auction environment, i.e., the form of the auction provides disincentives to invest in acquiring the information whose postulated existence is what justifies the form of the auction.

V. STRATEGIC USE OF THE RULES

After several auctions, a variety of bidding strategies started to arise (see Cramton and Schwartz (2000) for a detailed list of some of these strategies). New terminology was created to describe these emergent behaviors.

1. *Jump Bidding*: Bidding above the prescribed minimum increment to stay active. The use of jump bidding has been associated with a strategy to signal strength to "scare" away bidders from a license. It is also associated with an attempt to secure a license for which the bidder may not have the highest value, but, because of the size of the bid increment, the highest valued bidder would lose if he tried to bid after the jump.[16]

2. *Up Yourself*: Increasing one's bid despite being the standing bidder on a license. Ordinarily, this is viewed as a patently irrational action in auction theory, but, in FCC license auctions, it has been associated with the same signaling strategies as those associated with jump bidding.

3. *Retaliatory Bids*: Placing bids on the licenses bid upon by rivals to force them not to bid on the licenses the bidder desires. For example, if a bidder is interested in license A and another bidder is interested in licenses A and B, the first bidder can drive up the price of B, signaling that the second bidder should cease bidding on A.

4. *Parking*: Bidding on a license one does not want, but which is in high demand (many bids on the license), in order to stay active without revealing interest in the licenses the bidder does want. This allows a bidder to not drive up the price on an item he wants and still maintain eligibility, a tactic intended to mislead other bidders.

5. *Eligibility Management*: Bidding on a license that has a higher eligibility point total to have the option value to return to it if bids change later in the auction.[17]

6. *Lateral Hand-Off*: Bidding on a license, then withdrawing and bidding at a lower level, in order to signal that a bidder is not interested in the license, but will punish bidders on other licenses in which the bidder is interested.

7. *Bid and Waive*: Combining jump bidding on a license with the use of a valuable waiver in order to signal a strong desire for a license and ensure that other bidders "get the message." In each round of the auction, bidders tend to examine four important statistics on each bidder because of their costliness: reduced eligibility, jump bids, withdrawals, and waivers.

---

[16] Jump bidding is also used to ensure that one's bid is not tied, in which case, one of the tied bids is randomly selected to be the provisionally winning bid, and the others are discarded.

[17] Banks et al., *Theory, Experiment and the Federal Communications Commission Spectrum Auctions*, 51 J. OF ECON. BEHAVIOR & ORG. 303 (2001), shows that the asymmetric eligibility points on licenses can have a significant effect on revenue obtained in the auction.

8. *Trailing Digits*: Attaching market numbers in the last three digits of a bid to tell another bidder where it would be punished if it continued its bidding on a certain license, or on which license the rival should back off if it wants to avoid further punishment.[18]

9. *Budget Bluffing*: Bidding above one's budget to fool rivals into believing the budget is larger than it is. Each bidding team is typically provided a budget from its corporate management to use in the auction. Bidders normally track the *bid exposure* for other bidders in every round (the sum of the other bidders' previous round provisionally winning bids plus new round provisionally winning bids plus non-winning new bids in a round), because the maximum of any participant's bid exposure during the auction provides some insight into that competitor's potential budget. With this in mind, a bidder could bid above his budget, knowing that he will likely be outbid on some licenses, thus sending a false signal about his budget.

The behaviors observed in these auctions may have been surprising to the FCC and some economists, but they had been observed many times in laboratory experiments. McCabe et al (1988) found jump bidding behavior in their attempts to test Vickrey's proposal to use English auctions for multiple units. They found that allowing bidders to announce bid prices from the floor is not a good design feature in multiple unit auction environments—a problem that does not arise in simple, single-unit auctions. This may explain why one does not typically observe the simultaneous auctioning of multiple units in other fields. Rather, what typically occurs is a market, such as the Australian wool markets, in which multiple units are auctioned sequentially, one lot at a time.

Using clocks that move price based on bidder demands eliminates jump bidding and is more effective in yielding efficient outcomes.[19] Porter (1999) found that the ability to withdraw bids leads to worse outcomes due to a false security from "getting out" and to retaliatory bidding. Experimentalists have long known that, as one adds more information to the auction results, outcomes can, counter intuitively, get worse because the ability to signal becomes greater. Sometimes less information yields more efficient awards. In addition, adding more rules to fix or fine-tune a

---

[18] Market numbers are two or three digits, and bids are typically six figures or more. So, a bid could contain, at negligible cost, the market number as its last few digits, prefaced by leading zeroes to make the trailing digits stand out.

[19] In a single-unit English auction with strictly private (or strictly common) value, jump bidding cannot lead to an inefficient allocation since the award must be to the highest value bidder. If the high bidder jumps past the value of the second highest value bidder it just means that he pays more than he needed to. When multiple identical units, say $Q > 1$, are auctioned simultaneously, jump bids can raise the price so much that only $q < Q$ units are sold. Similarly, in the FCC auctions, if the efficient allocation is to a bidder who wants both license A and license B, the aggregation attempt may fail if someone jumps the bid for A so that the combined price of A and B is out of reach of the bidder trying to efficiently combine A and B.

process tends to create a new series of unanticipated problems. As noted in the introduction by Smith in Cramton, Shoham and Steinberg (2006):

> "The ideal incentive mechanism design should lead managers to a two-step procedure: (1) an estimation of the value of the auctioned item(s), followed by (2) a readiness to reveal this value in the form of a bid, if necessary, such action being a fair approximation to that which serves the interest of the bidder. Market design should focus on how to facilitate this procedure. Very complex market allocation problems for runway rights, gas in pipeline networks, energy on a high voltage grid, and so on, can be made simple for the participants. Humans make the value judgments, and smart markets handle the complexity. Participants are not required to be experts in anything except their own business uses of the auctioned items, and must apply what they know to determine the private values of those items. That must be their specialty and their focus, and strategizing should not be plainly required of them. Privacy is essential: public information on who is bidding for what, how much, and when, fosters manipulation, gaming, collusion, and inefficiency . . . ."

VI. NEW RULES TO REDUCE STRATEGIC BIDDING

Many of the strategies summarized above are used in an attempt to reduce the amount a bidder pays and potentially reduce the efficiency and/or revenue of the auction. The FCC imposed new, untested rules to overcome some of these practices. In 1997, the FCC imposed the following rules:

1. *Click-box Bidding*: This form of bidding only allows the bidder to increase his bid in integer multiples of the identified increment. Thus, if the increment amount were 10 percent for a particular license, any bid submitted for that license was restricted to be equal to the Standing Bid times $(1 + .10\pi)$, where π is a positive integer greater than or equal to 1. This was used to eliminate trailing digits and typos after learning that trailing digits were used to signal intentions. Note, however, that any given rule potentially maps into many behavioral effects, not just the effect the rule was designed to control. There are no assurances that 'Undesirable Behavior X is controlled if and only if rule R is imposed.' This is why rule changes should be tested in the laboratory to determine if there are unintended consequences whose costs may exceed the presumed benefits. For example, the Vickrey rules for multiple unit English auctions, although logically unassailable, in practice led to jump bidding and, in some cases, inefficient rewards—an unintended consequence that may have been foreseen if they had been tested prior to adoption.

2. *Limit Withdrawals*: Bidders were limited to two withdrawals in an auction. This rule was implemented to reduce the lateral hand-off problem, but it obviously could interfere with assembling complementary items and may reduce efficiency and/or revenue.

3. *Increment Smoothing*: The FCC now changes the percentage bid increment from round to round based on the number of new bids received on a particular license. This was adopted in order to speed-up the auction.

In addition to the observations that spurred these rule changes, there were other lessons learned from conducting the auctions. In particular, the C Block auction lasted approximately six months.[20] While this was a boon for consultants, it created high transaction costs for participants—costs not part of the design criteria. In the early FCC auctions, the system averaged two rounds each day. Recently, this has been increased to an average of nine rounds each day in later rounds. Finally, the Justice Department has shown a willingness to pursue bidders they believe are fostering collusive bidding, such as retaliatory bidding or bid signaling. Thus, bidders are becoming more sensitive to the type of bids they will submit.

VII. NEW RULES ON THE HORIZON AND THE AWS ACTION

There have been two specific proposals to increase the efficiency and revenue from the FCC auctions: a change in the auction form and a change in the informational structure of the auction.

A. *Combinatorial Auctions*

An important concern for bidders in FCC auctions is the inability to bid for packages of licenses. That is, bidders are not able to make bids such as "I want License B AND C together or neither." Not allowing "AND" bids handicaps bidders who have regional or national business plans. Combinatorial auctions allow for such bidding possibilities. In addition to AND bids, combinatorial auctions allow for "OR," "ONLY IF" and other logical bid constraints.

In 2000, the FCC held its first combinatorial auction conference to begin looking at ways to implement a combinatorial auction for future licenses. Software was designed and redesigned to be used for combinatorial bidding, and two additional conferences were held in 2001

---

[20] The C block auction, completed in May 1996, extended bidding credits to Designated Entities (DEs), small businesses or rural telephone companies determined by the FCC to be handicapped in accessing credit markets. DE bidders winning licenses were extended long-term (10-year) credit on extremely favorable terms (U.S. Treasury debt interest rates), paying for licenses via installments. The two largest bidders defaulted on their payments, leading to court battles resolved by the U.S. Supreme Court in 2003. F.C.C. v. NextWave Personal Commc'ns Inc., 537 U.S. 293, 123 S.Ct. 832 (2003) (Bankruptcy Code prohibits FCC from revoking licenses held by a debtor in bankruptcy upon the debtor's failure to make timely payments owed for purchase of the licenses). Eventually, however, the licenses were re-auctioned. *US FCC announces NextWave settlement agreement*, INT'L. TELECOMMUNICATIONS INTELL., Apr. 21, 2004 (Pg. Unavail. Online) 2004 WLNR 6953743. For more details on the use of bidding credits and their effects see Thomas Hazlett & Robert Munoz, *What Really Matters in Spectrum Allocation Design* (AEI-Brookings Working Paper 04-16), available at http://www.aei-brookings.org/publications/abstract.php?pid=821.

and 2003.[21] To date, no such auction has been used by the FCC. Such auctions, however, have been used effectively in various other public and private venues (see Ishikida et al. (2001), Ledyard et al. (2002) and Porter et al. (2006)). A new, improved design has been proposed in Porter et al. (2003). It remains to be seen if such auctions will be adopted by the FCC.

For the AWS auction held from August 9 to September 18, 2006, the FCC sought comment on "the feasibility and desirability of allocating the AWS-1 licenses among two auctions, run concurrently, with one of the auctions using the standard SMR format and the other using the FCC's package-bidding format ("SMR-PB"). Under the SMR-PB format, bidders can place bids on groups of licenses they wish to win in combination, with the result that they win either all of the licenses in a group or none of them, in contrast to the license-by-license bidding in the FCC's SMR format. In the SMR-PB auction format, each bidder can have, at most, a single winning bid, so that, in order to win any particular license combination, the bidder must have placed a package bid on that license or specific group of licenses."[22] Given the complexity of participating in two concurrent auctions and the overwhelming negative comments, the FCC opted to not try any combinatorial auction designs for the AWS auction.

B. *Information Provided to Bidders*

The bidding strategies outlined in section V require that bidders have access to full information on bidder identities, bids submitted, and other information. Thus, the FCC also sought comments on potential rules concerning limitations on the specificity of information provided to bidders about the identities and actions of other bidders during the AWS auction.[23] Specifically, The FCC initially proposed "not to reveal until the close of the auction: (1) bidders' license selections on their short form applications and the amount of their upfront payments; (2) the amounts of non-provisionally winning bids and the identities of bidders placing those bids; and (3) the identities of bidders making provisionally winning bids." After each bidding round, the FCC would reveal the number of bidders who placed bids for each license and the amount of the current highest bid.

Both the Federal Trade Commission and the Department of Justice submitted *ex parte* comments supporting the no information condition and

---

[21] A potential combinatorial design, based on a similar rule structure as the SMR was tested, and though it performed better than the SMR, it took many more rounds to complete.

[22] Notice and Filing Requirements, Minimum Opening Bids, Upfront Payments and Other Procedures for Auction No. 66 AU Docket No. 06-30.

[23] Federal Communications Commission, Auction of Advanced Wireless Services Licenses Scheduled for June 29, 2006 Comment Sought on Reserve Prices or Minimum Opening Bids and Other Procedures, AU Docket No. 06-30, 71 Fed. Reg. 6486 (Feb. 8, 2006) ("Public Notice").

noting that providing full information would "likely result in a loss of competition with lower government auction revenues and less efficient allocations of markets among bidders." However, the FCC opted for a modified rule that examined the "competitiveness" of the auction. Competitiveness was defined by the upfront payments of the participants, so that, if the amount of eligibility obtained from the upfront deposit was three times the aggregate level of the license offered in the auction, then all information would be provided; otherwise the rule identified above would be used.[24] FTC provided no explanation why the competition defined by three-times the aggregate level is enough to not worry about collusion, and it remains to be seen whether this is true. Note, though, that each of theses rules is testable, and they can be compared with alternative variations.

VIII. AWS AUCTION RESULTS AND CONCLUSION

Based on the upfront payments of the bidders, the modified eligibility ratio in the August-September 2006 AWS auction was 3.04, just barely above the require 3.0 amount to make the auction open. Interestingly, though, the amount of eligibility shed after round one of bidding pushed the modified eligibility ratio well below 3.0. The auction lasted 29 days, with 161 rounds of bidding. There were 168 qualified bidders, and 104 bidders ended up winning at least one license (35 licenses went unsold). The auction raised $13.7 billion in revenue, which translates into a price of 54.34 cents per MHzPop. This is significantly below the amounts paid in previous international Third Generation (3G)[25] auctions and below the prices in the PCS C and F block auctions ($4.00 MHzPop) or recent private transactions of spectrum ($1.70 MHzPop).[26]

One interesting insight from the AWS auction arose from the observation that there were at least two clearly identified bidders who would be new nationwide entrants, a consortium of the Cable companies (SpectrumCo) and a partnership of the satellite TV companies (Wireless DBS). The incumbent national wireless operators (Verizon, Cingular, and T-Mobile) were also participants. There was a fear that the incumbents would bid to keep the new entrants out, or at least raise their entry costs. Given the structure of Simultaneous Multi-Round auctions, an incumbent could keep a national entrant out by bidding on just a handful of licenses to

---

[24] Specifically, the eligibility ratio is defined as the sum of the bidding units from the upfront payments divided by the total number of bidding units on licenses in the auction. Each bidder's total eligibility was capped at 50 percent of the total bidding units in the auction. This capped amount was used in calculating the "modified eligibility ratio."

[25] The most advanced PCS devices are considered to be "Third Generation" technologies, which combine high-speed mobile access with Internet protocol based services.

[26] In the appendix, a history of the revenue and rules for the FCC auctions is provided.

cripple any attempt to assemble a national aggregation of licenses. If, instead, a combinatorial auction were to be used, an incumbent would have to keep out a new entrant by potentially outbidding a nationwide package of licenses, which would be costly if the incumbent actually won the licenses. This is precisely the opposite argument from the one that suggests a combinatorial auction hinders entry.

In view of this experience, we think a fresh reexamination of the FCC auction design protocols is overdue. An important objective of redesign should be to find auction procedures that reduce the participation and transactions costs of the bidders, and that make it easier for bidders who desire packages of the elemental rights to assemble those packages by expressing their willingness-to-pay, with minimum incentive for strategic behavior. We think the likely features of an improved auction design include: English clock procedures for advancing the price on each offered item; publicity of those items on each round that are or are not still actively bid; bid privacy; and efficiency (not revenue) optimization algorithm support, as needed, after bidding on the last item(s) becomes inactive. Whatever the proposed redesign, it should be thoroughly tested first in the laboratory. After shakeout, that testing should be further refined with industry and government professionals, using environments that are thought to be both relevant and challenging to the design. We think it is also time to consider two-sided auctions, in which incumbent rights holders can offer their rights for sale in combination with new rights being offered.

APPENDIX

The Figures below chart the revenue from each FCC Spectrum auction by year. The first figure tracks the revenue and major results and design decisions for the auctions from 1993 to 2000. The second figure does the same for 2001 to 2006. However, the vertical scale is decreased by a factor of 10 since the revenues are much lower in these later auctions.[27]

[27] The figures show the reported revenues from each auction. The aggregate winning bids per auction as reported by the FCC total $45.118 billion (1994-2006). However, approximately $8 billion has proven to be uncollectible in Auction 5 (the PCS C Bloc). In addition, the approximately $16.9 billion for Auction 35 (Jan. 2001), has resulted in virtually zero actual revenues when the federal courts overturned the FCC's ownership of the licenses in the Nextwave case. (The uncertainty concerns licenses other than Nextwave's that may have been legally sold in the auction.). Roughly, this means that $20 billion of the reported $45 billion has gone uncollected, for a grand total of about $25 billion in actual revenues over 12 years, a little more than $2 billion per year.

FIGURE 1

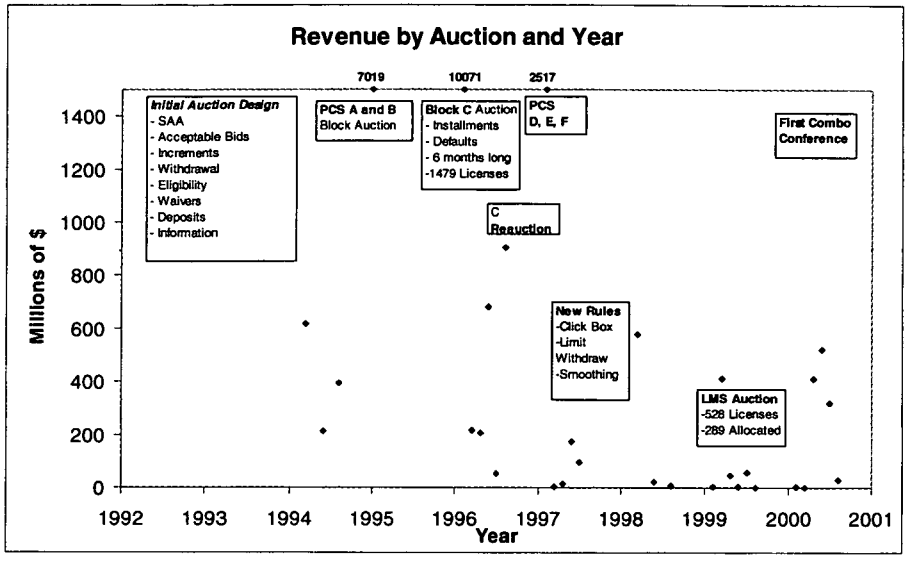
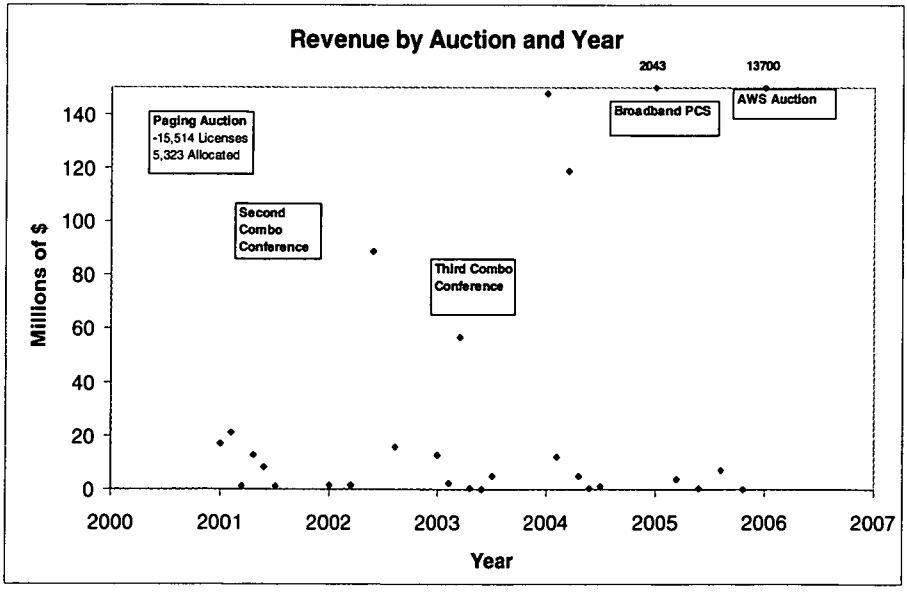


FIGURE 2



REFERENCES

Banks, Jeffrey, Mark Olson, David Porter, Steve J. Rassenti, and Vernon L. Smith. 2002. Theory, experiment and the Federal Communications Commission spectrum auctions. *Journal of Economic Behavior and Organization* 51: 303-50.

Coase, Ronald H. 1959. The Federal Communications Commission. *Journal of Law and Economics* 2: 1-40.

Cramton, Peter, Y. Shoham, and R. Steinberg, eds. 2006. *Combinatorial auctions*. MIT Press.

Cramton, Peter, and Jesse Schwartz. 2000. Collusive bidding: lessons from the FCC spectrum auctions. *Journal of Regulatory Economics* 17: 229-52.

Cull, Robert J., Mark M. Bykowsky, and John O. Ledyard. 2000. Mutually destructive bidding: the FCC auction design problem. *Journal of Regulatory Economics* 17: 205-28.

Hazlett, Thomas. 1998. Assigning property rights to radio spectrum users: why did FCC license auctions take 67 years? *Journal of Law & Economics* 41: 529-75.

Hazlett, Thomas, and Robert Munoz. 2004. What really matters in spectrum allocation design. *AEI-Brookings Working Paper* 04-16.

Herzel, Leo. 1951. Public interest and the market in color television regulation. *University of Chicago Law Review* 18: 802-16.

Isikida, Takashi, John Ledyard, Mark Olson, and David Porter. 2001. The design of a pollution trading system for southern California's RECLAIM emission trading program. *Research in Experimental Economics* 8: 185.

Kwerel, Evan, and Alex D. Felker. 1985. *Using auctions to select FCC licensees*. Federal Communications Commission, OPP Working Paper Series #16.

Ledyard, John, Mark Olson, David Porter, and Joseph Swanson. 2002. The design of an auction for logistics services. *Informs*.

McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith. 1988. Testing Vickrey's and other simultaneous multiple unit versions of the English auction, rev. 1989. In *Research in Experimental Economics* 4: 45-79, 1991. R. M. Isaac, ed. JAI Press, Greenwich, C.T.

Milgrom, Paul. 2000. Putting auction theory to work: the simultaneous ascending auction. *Journal of Political Economy* 108: 245-72.

Porter, David. 1999. An experimental examination of bid withdrawal in a multi-object auction. *Review of Economic Design* 4: 73.

Porter, David, Steve J. Rassenti, Anil Roopnarine, and Vernon L. Smith. 2003. Combinatorial auction design. *Proceedings of the National Academy of Sciences* 100: 11, 153-57.

Porter, David, Stephen J. Rassenti, William Shobe, Vernon L. Smith, and Abel Winn. 2006. The design, testing and implementation of Virginia's NOx allowance auction. *ICES Working Paper.*

Rassenti, Stephen J. 1981. *0-1 Decision problems with multiple resource constraints: algorithms and applications.* Unpublished PhD. thesis, Univ. of Arizona.

Rassenti, Stephen J., Vernon L. Smith, and Robert L. Bulfin. 1982. A combinatorial auction mechanism for airport time slot allocation. *Bell Journal of Economics* 13: 402-17.

Smythe, Dallas. 1952. Facing facts about the broadcast business. *University of Chicago Law Review* 20: 96-107.

Telecommunications Reports. 2007. *Copps: Better analysis needed to spur broadband deployment.* Telecommunications Reports (February 1, 2007) 2007 WLNR 1485120.

JUSTICE AND THE EVOLUTION OF THE COMMON LAW

Richard O. Zerbe, Jr.[*]

> "The first and chief design of every system of government is to maintain justice; to prevent the members of a society from encroaching on one another's property, or seizing what is not their own. The design here is to give each one the secure and peaceable possession of his own property.—When this end, which we may call internal peace . . . is secured, the government will next be desirous of promoting the opulence of the state."[1]
>
> —Adam Smith

I. INTRODUCTION

Empirical evidence shows, and theory suggests, that the common law tends toward economic efficiency.[2] While various theories attempt to explain this phenomenon, no single one is well accepted. This article provides a simple explanation.[3] It suggests that efficiency arises as a matter of justice-seeking as efficiency and justice overlap. Judges discover just and efficient common law not only through the adoption of norms, but also through intellectual knowledge. When neither is available, justice is more difficult to determine. Judges seek justice because justice-seeking is a social norm with its own sanctioning force. When justice is sought, efficiency is achieved because the two principles overlap substantially.

---

* Thanks to Louis Wolcher and Stephen Vineyard for useful comments, and to Stephen Vineyard for research help and Elizabeth Tutmarc for copy editing. I would also like to thank members of the Political Economy workshop organized by Aseem Prakash at the University of Washington for comments. I note a useful comment by Katherine Stovel of the UW Sociology Department.

1 *See* ADAM SMITH, LECTURES ON JURISPRUDENCE 1 (R.L. Meek et al. eds., Oxford University Press 1978).

2 *See, e.g.*, Paul Rubin, *Why is the Common Law Efficient*, 6 J. LEGAL STUD. 51 (1977); ROBERT D. COOTER & THOMAS S. ULEN, LAW & ECONOMICS (2nd ed. 1997); RICHARD POSNER, ECONOMIC ANALYSIS OF THE LAW (2d ed. 1992); George Priest, *The Common Law Process and the Selection of Efficient Rules*, 6 J. LEGAL STUD. 65 (1977); John Goodman, *An American Theory of the Evolution of the Common Law*, 7 J. LEGAL STUD. 393 (1978); Robert D. Cooter & Lewis Kornhauser, *Can Litigation Improve the Law without the Help of Judges?*, 9 J. LEGAL STUD. 139 (1990). *But see* Mark Kelman, *On Democracy-Bashing: A Skeptical Look at the Theoretical and 'Empirical' Practice of the Public Choice Movement*, 74 VA. L. REV. 199 (1988) (describing the proposition as ideologically based).

3 *Compare* Robert D. Cooter, *Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant*, 144 U. PA. L. REV. 1643 (1996) (expressing the analogous view that social norms explain common law efficiency). *See also* RICHARD O. ZERBE, JR., ECONOMIC EFFICIENCY IN LAW AND ECONOMICS (Edward Elgar ed., 2001) (arriving independently at the same view).

Inefficiency is created by change. The faster the pace of change, the less likely norms and knowledge are in determining whether efficiency is available. Thus, efficient and just common law is least likely to be created in periods of rapid change. Several examples exemplify this model of common law efficiency.[4]

II. ECONOMIC EFFICIENCY IS ENHANCED WHEN ITS DEFINITION IS EXPANDED TO INCLUDE MORAL SENTIMENTS

A problem for the congruence between justice and efficiency is the limited definition of efficiency. Traditional efficiency fails to consider equity and, more generally, moral sentiments. Mainstream efficiency is defined by the Kaldor-Hicks (KH) criterion which, by definition, eschews issues of equity and moral sentiments, making it unlikely that it will directly correspond with justice. Thus, it is not difficult to find common law examples that are not KH efficient.[5]

Efficiency becomes more powerful, useful, and accepted when the concept is expanded by including equity and moral sentiments. I, and others, explore this idea elsewhere.[6] I call this expanded criterion KHM, where the M represents moral sentiments.[7] KHM represents a version of KH that includes, *inter alia*, moral sentiments.[8] That is, all goods,

---

[4] We do not consider here examples of statutory law and its interrelation with common law. Although it is understood and generally accepted that statue law deriving from legislative power has preference over judge made law in the absence of constitutional restraints, statue law is often sufficiently vague that common law develops around it. This topic, while interesting, requires another paper.

[5] *See* ZERBE, *supra* note 3. *See also* Robert H. Lande & Richard O. Zerbe, Jr., *Anticonsumer Effects of Union Mergers: An Antitrust Solution*, 46 DUKE L. J. 197 (1996) (describing the development of labor law as an example of this).

[6] *See generally* ZERBE, *supra* note 3 (using the term "KHZ" to represent Kaldor-Hicks-Zerbe). Subsequently, I changed this to the more appropriate term KHM, which stands for Kaldor-Hicks-Moral. *See, e.g.*, Richard O. Zerbe, Jr., Yoram Bauman, & Aaron Finkle, *An Aggregate Measure for Benefit-Cost Analysis*, 58 ECOLOGICAL ECON. 449 (2006).

[7] Additional characteristics are: (1) the use of the willingness to pay (WTP) for gains and the willingness to accept (WTA) for losses; (2) the use of WTP and WTA from a legal status quo; (3) the exclusion of gains or losses that are legally illegitimate, as with goods held by the thief, or that violate well-accepted moral principles (benefit-cost rationale is provided for this); (4) a recognition and inclusion of non-pecuniary effects; (5) an efficiency test that is passed when and only when the aggregate benefits exceed aggregate losses so that (6) there is no use of the potential compensation test; (7) an assumption of equal marginal utility of income so that each person is treated the same; (8) the absence of reliance on market failure or externalities to justify the use of benefit-cost analysis; (9) the inclusion of transactions costs of operating a project; and (10) an understanding that the role of benefit-cost analysis is to provide information to the decision process and not to provide the answer. This list of characteristics is explored more fully elsewhere. *See generally* Zerbe et al, *supra* note 6.

[8] KHM efficiency differs from KH by its clearer grounding in legal rights, by its inclusion of all sentiments for which there is a willingness to pay, by its abandonment of the potential compensation

including moral sentiments, are to be treated as economic goods as long as there is willingness to pay for them or reluctance to give them up.[9] As this definition is more inclusive of sentiments generally, it will better correspond with the requirements of justice and, thus, is more likely to be consistent with the common law.

III. JUSTICE AND EFFICIENCY

The following example is meant to illustrate the way in which KHM can amend the divergence between KH efficiency and justice.[10] I shall say that justice arises from meeting reasonable expectations. Consider the location of a NIMBY, say a waste incinerator, in one of two neighborhoods; one neighborhood is rich, the other poor. The incinerator produces negative environmental effects and no corresponding benefits, making it undesirable for both neighborhoods. It is, of course, efficient to locate the incinerator in the poorer neighborhood. The land is cheaper, and the willingness to pay to avoid it is greater in the richer neighborhood. We can ask, however, if the poorer neighborhood should be compensated, monetarily or with, say, the provision of a new park. Is compensation just? Certainly it can be seen as just when society in general feels and expects that compensation is the right thing to do. Traditional efficiency (KH) has nothing to say about the issue. If moral sentiments are included, as required by KHM, then compensation is also efficient. Sentiments about compensation are a part of justice and are a part also of KHM, but they are not a part of a traditional efficiency

---

criterion for one of net benefits, by its reliance on transaction costs rather than market failure to determine where to apply benefit-cost analysis, by its inclusion of transactions costs of operating a project though not of the costs of institutional change, and by its view of efficiency as a technique to provide information relevant to the answer, not to provide the answer. This view is essentially identical to the view that has been presented elsewhere as the KHZ view. *See* Zerbe et al, *supra* note 6.

[9] KHM differs from tautological efficiency. *See, e.g.,* Yoram Barzel: *Transactions Costs: Are They Just Costs?* 141 J. INSTITUTIONAL & THEORETICAL ECON. 4-16 (1985). Barzel explains tautological efficiency as a state in which "individuals must spend resources to discover inefficiencies and arrange to take advantage of their profit potential. Suppose that after taking account of these costs, some of these activities are still found profitable but some are not. The former will be eliminated whereas the latter will be allowed to stand. The latter ones, however, are not worth eliminating . . . . It is tautological that . . . . given profit maximization efficiency will prevail." Logically, efficiency should further require that spending on discovery itself must be at the efficient level.

[10] Generally it is thought that KH-efficiency does not include an evaluation of moral sentiments. This view is an extension of Kaldor's rejection of the consideration of distributional effects for benefit-cost analysis because of the mistaken notion that value judgments would be avoided by such exclusion. Some have further claimed that to include moral sentiments can lead to a failure to pass the potential compensation test and to double counting. These claims, however, are either incorrect or not a legitimate justification for ignoring moral sentiments. Rather, moral sentiments are fundamental to justice and likewise should be for efficiency. Indeed, one can show that a Pareto improvement can be rejected by KH because it does not include moral sentiments. *See* Zerbe et al., *supra* note 6.

analysis. A project with compensation is thus different from the same project without compensation under KHM.

Table One illustrates, from a benefit-cost perspective, how the KH and KHM models differ when applied to the hypothetical in which the incinerator is placed in the poorer rather than the richer neighborhood. Table One assumes that the incinerator will be located in the poorer neighborhood. Compensation to the poorer neighborhood and mitigation of environmental effects are possible in three of the four scenarios, but are alternatives (such that only one can be chosen in each scenario).

TABLE 1.

BENEFITS AND COSTS (thousands of dollars) (PV)	[1] No Compensation or Mitigation Occurs	[2] Compensation Occurs	[3] Mitigation Occurs	[4] Neither Compensation Nor Mitigation Are Feasible
Benefits to Society	100	100	100	100
Ordinary Costs	60	60	60	60
Harm to Neighborhood	20	20	0	20
Mitigation Costs			21	--
Moral Harm to Non-Residents	50	0	0	50
Transfer Cost of Compensation	-	20	-	--
Administrative Costs of Compensation	-	2	-	
KH NPV	20	18	19	20
KHM NPV	-30	18	19	-30
Conclusion	Neither compensation nor mitigation appear worthwhile under KH, as moral harm is ignored	Compensation eliminates moral harm, which is relevant only under KHM	Mitigation eliminates moral harm, which is relevant only under KHM	Moral harm renders project undesirable under KHM but not under KH

* Note that not all figures are relevant to KH and that mitigation and compensation are substitutes so that one or the other, but not both, are included in the AM calculation.

In column one, the net present value (NPV) is a positive $20,000 under KH because KH fails to count the costs arising from moral sentiments. The NPV of KHM is a negative $30,000. In column two, compensation is made to the poorer neighborhood for bearing the environmental costs of the incinerator. Moral harm is eliminated so that KH and KHM both give the same NPV of a positive $18,000. In column three, mitigation is made that eliminates the harm so that again the NPV under KH and KHM are the same ($19,000). Under KH, however, economic efficiency will not lead to either mitigation or compensation, as the NPV is greatest when neither mitigation nor compensation occurs (column one). KHM, however, leads to a choice of mitigation as this is the highest NPV including all sentiments. If neither compensation nor mitigation is feasible, KHM suggests the project be abandoned as the NPV is a negative $30,000, while the KH analysis does not change. Table One shows that KH fails to convey relevant information. KH suggests putting the incinerator into the poorer neighborhood without compensation or mitigation.

Moral sentiments regarding the allocation of rights and goods must be included as a matter of justice if we are to have confidence that actions passing a benefit-cost test are welfare-enhancing and just. Without such inclusion, projects will be undertaken that do not increase net gains while counting the value of moral sentiments, and projects that will increase net gains while excluding the value of moral sentiments will be undertaken.[11] The simple moral rationale for benefit-cost analysis (BCA) is that when it is applied broadly, everyone tends to gain. Those that lose from one project gain from another so that when applied broadly and over time, the use of the efficiency criterion in policy can tend to make social policy Pareto-superior. This welfare result is, however, considerably more likely when distributional effects, including a sense of fairness and due process, are considered goods in an efficiency analysis. Without the use of such goods in the analysis, efficiency tends to diverge from justice and becomes less reliable. In the incinerator example, using the traditional KH efficiency criteria will lead to locating the incinerator in the poorer neighborhood without compensation. This will result in the further degradation of that neighborhood. When the city is faced with a future decision based on willingness to pay or willingness to accept, the poorer neighborhood will be in an even weaker position to protect itself. This situation can lead to a downward spiral for those less fortunate and vitiate the moral standing of efficiency.

Justice means meeting reasonable expectations. In a liberal society, reasonable expectations arise from existing rights. Given existing rights, it will be efficient to incorporate changes in rights that increase aggregate

---

[11] *See* ZERBE et al., *supra* note 6.

well-being. This is also a description of KHM efficiency. It rests on existing rights and, when applied broadly, tends to increase the welfare of all when welfare includes those moral sentiments that are also an integral part of well-being.[12] The simple moral argument for the use of efficiency to allocate resources is that its use has a reasonable chance to increase net wealth for the most people, particularly if applied with due regard for others' rights and moral sentiments. Losers from a project today will be winners tomorrow. The BCA results in an increase in wealth across all projects that meet the benefit-cost standard. As net wealth increases, there is a clear potential for all to be winners; that is, the systematic application of a net-benefits approach has potential to satisfy a Pareto test.

IV. THE VALUE OF NORMS FOR JUSTICE AND EFFICIENCY

Reasonable expectations, and thus justice, arise largely from the extant set of rights and the perceived fairness of those rights. Increasing rights or ownership increases justice as well as efficiency. Ownership lowers transaction costs. First, rights tend to be efficient as they reduce costly conflict. Second, rights rearrange valuations so that a change from them is less likely to be efficient. Ownership establishes a reference point from which losses are to be calculated by the willingness to accept (WTA) and gains by the willingness to pay (WTP). Benefits and costs are measured, respectively, by the WTP and by the WTA under KHM as well as under KH.[13] The WTP is simply the amount one will pay for a gain, which is the measure of the willingness to buy. It represents the amount that someone who does not own a good would be willing to pay to buy it. In other words, it is the maximum amount of money one would give up to buy some good or service, or would pay to avoid harm.[14] The WTP is always finite, because the resources of the payer are limited. The accompanying measure, the WTA, is the minimum amount one would accept to give up a right or a good. That is, it represents a willingness to sell. Unlike the WTP, the WTA can be infinite. There may be a case, for example, when one would not sell at any price. For example, the WTP to save one's life is finite but

---

[12] *See generally* Ernest Fehr & Klaus Schmidt, *Theories of Fairness and Reciprocity-Evidence and Economic Applications* (Inst. for Empirical Research in Econ., Univ. of Zurich, Working Paper No. 75, 2001), *available at* http://www.iew.unizh.ch/home/fehr (summarizing fairness experiments).

[13] RICHARD O. ZERBE, JR. & DWIGHT D. DIVELY, BENEFIT COST ANALYSIS IN THEORY AND PRACTICE (1994).

[14] These are non-technical definitions and, as such, are not wholly accurate. The compensating variation (CV) is the sum of money that can be taken away or given to leave one as well off as one was *before* the economic change. The equivalent variation is money taken or given that leaves one as well off as *after* the economic change. *See generally id.* (deriving these concepts in terms of indifference curves).

the WTA to sell one's life may be infinite.[15] The WTP and WTA are part of one class of exact utility indicators developed by John Hicks. In welfare economics, this measure is called the compensating variation or the KH measure.[16] Under this measure, the loss from a change in ownership is properly measured by the WTA and the gains from a change by the WTP. For the same individual, the WTA of a good will equal or exceed the WTP.[17] Thus, the measure of loss from a change of ownership will tend to exceed the measure of gain, *ceteris paribus*.

The compensating or KH measure is used generally as these representations of value apparently correspond with the associated psychological states of valuation. From a legal perspective, the use of the WTA to measure losses and the WTP to measure gains rests on a normative decision to recognize ownership. Ownership establishes a psychological state that efficiency must recognize because it is relevant for the calculation of gains and losses. Thus, efficiency corresponds with the psychological states associated with ownership—that is with a set of reasonable expectations.[18]

One's sense of ownership will usually conform to one's knowledge of legal ownership. Most people feel that they have a moral right to possess what they legally own and do not feel that they have the moral right to possess something they do not own. For most cases, the law will determine whether the WTP or WTA will be used even if the economic standard is psychological ownership. The common assumption is that a choice based on assigned legal entitlements will usually be correct. However, it is correct because of the correspondence between the legal and psychological states. That is, it is not correct as a matter of principle, and it is actually incorrect in important cases.[19]

---

[15] The benefits from a project may be either gains (WTP) or losses restored (WTA). The costs of a project may be either a loss (WTA) or a gain forgone (WTP).

[16] The other measure is the equivalent variation which uses the WTA for gains and the WTP for losses.

[17] Except in rare circumstances in which inferior goods are involved. *See* ZERBE & DIVELY, *supra* note 13; ZERBE, *supra* note 3; Zerbe et al., *supra* note 6.

[18] Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Experimental Tests of the Endowment Effect and the Coase Theorem*, J. POL. ECON. 1325 (1990); Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSP. 193 (1991); Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Uncertainty*, 47 ECONOMETRICA 263 (1979); Daniel Kahneman & Carol Varey, *Notes on the Psychology of Utility*, *in* INTERPERSONAL COMPARISONS OF WELL-BEING (Jon Elster & John Roemer eds., 1991).

[19] Daniel Levy & David Friedman, *The Revenge of the Redwoods? Reconsidering Property Rights and the Economic Allocation of Natural Resources*, 61 U. CHI. L. REV. 493, 509 (1994). Levy and Friedman incorrectly assert "the determination of the conceptually appropriate form of CV query is a matter of property rights, not economics or psychology." This implies that the law ought to govern in the event of a conflict between rights given by law and those recognized as a psychological reference point. The authors use the term "CV query" in reference to questionnaire studies. "CV" here stands for

Both efficiency and justice recognize legitimate ownership, not only in the proper allocation of rights, but in determining the calculation of gains and losses.[20] For some time it has been recognized that the policy and welfare implications of any substantive economic analysis depend upon the legitimacy of the property rights that underlie the relevant supply and demand functions.[21] Heyne notes that, "Because this legitimacy depends on existing law . . . the foundations of economics may be said to rest in the law."[22] The legitimacy of ownership will in turn depend on moral sentiments about what is justly owned and rewarded. If the legitimacy of ownership allocations is then to conform to those suggested by efficiency, efficiency itself must recognize these moral sentiments.

In a sense, this concept has long been noted. Atiyah pointed out that David Hume and Adam Smith both said that expectations arising out of rights of property deserved greater protection than expectations in regard to something that had never been possessed. To deprive somebody of something that he merely expects to receive is a wrong, deserving of less protection, than to deprive somebody of the expectation of continuing to hold something that he already possesses.[23]

The law has long recognized that it is more intrusive to stop an owner from conducting an ongoing activity than to prohibit the owner from

---

"contingent valuation," not compensating variation. This result is contrary to economic efficiency. Economic efficiency in the KHM form would recognize the psychological status quo as primary and change ownership to conform to it. The psychological reference point is, however, not just that of the individual but of society generally, so that in so far as the law embodies the general understanding, Levy and Friedman are correct in that the law should govern. Because the underlying basis is the general psychological reference point, however, where this differs from the law, it furnishes a guide for further development as indeed it has done with the development of common law.

20 *See* ZERBE, *supra* note 3. This approach makes clear the irrelevancy of the critical legal studies objection to benefit-cost analysis as Heyne has shown. The KHM approach shows the failure of the critical legal studies argument that the measurement of benefits and costs is incoherent. Put briefly, the critical legal studies argument is that one cannot use the concept of efficiency without endorsing some concept of property rights, from which it is seen to follow that the concept of efficiency cannot be used to resolve disputes over property rights without begging the question. Benefit-cost analysis takes, as does the law, the existing structure of rights as extant. But there are disputes that reflect uncertainty about some small portions of these rights. Benefit-cost analysis merely furnishes information relevant to the legal decision about the allocation of such a right. Take a simple case: A change in technology makes valuable rights to the radio wave spectrum that has hitherto been unowned. No party has a superior claim. The assignment of the right to a particular party will be a gain. Gains in economic analysis are to be measured by the WTP. The WTP will in turn be partly determined by the pattern of wealth that rests on the existing system of rights. Economic analysis suggests auctioning off the right. The right in general should go to that party who would pay the most for it if transactions costs were zero. Cases where conflicting prior claims exist raise more difficult questions, but these are answerable and elsewhere I have provided answers.

21 *Id.* at 53–71.

22 Paul Heyne, *The Foundations of Law and Economics: Can the Blind Lead the Blind?*, 11 RESEARCH IN L. AND ECON. 53 (1988).

23 *See* PATRICK S. ATIYAH, THE RISE AND FALL OF FREEDOM OF CONTRACT 5 (1979).

undertaking the same activity if he has not yet begun it. The currently fashionable expression of this may be found in Justice Brennan's phrase in *Penn. Central Trans. Co. v. City of New York*,[24] that a restriction is more likely to cause a taking if it destroys "investment backed expectations." Norms confer a type of ownership status. Thus, they tend to confer the same benefits of efficiency and justice as legal rights. Therefore, legal rights that build on norms will also conform to both expectations and a sense of what is deserved and will a fortiori promote efficiency and justice. When norms are well established, they are more likely to be efficient because the rights they establish will be more certain and accepted. Those rights that are long standing are more likely to be efficient than when those rights are contentious. Justice involves people receiving what they deserve and in meeting their reasonable expectations. Thus, as a principle of efficiency, justice suggests that those who contribute more should receive more. Norms will cause less contention when they are seen as just and efficient.

V. JUDGES ADOPT NORMS AS COMMON LAW BECAUSE THEY SEEK JUSTICE

Judges act according to a norm of justice.[25] According to Glick, empirical evidence suggests that judges seek to do justice in deciding cases.[26] As John Chipman Gray points out, "The essence of a judge's office is that he shall be impartial . . ."[27] A similar sentiment was expressed by the commission of four bishops, two earls and six other barons who were appointed after the triumph of Henry III over the baronial faction: "Furthermore, we ask the same lord king . . . that, for doing and rendering justice, he will nominate such men as, seeking not their own interests, but of those of God and the right, shall justly settle the affairs of subjects according to the praise-worthy laws and customs of the kingdom."[28]

Posner's view of judges is not apparently at variance with the one expressed here. Posner notes that Holmes's *The Common Law* is an extended paean to judges' skill in adapting common law doctrines to

---

[24] Penn. Central Trans. Co. v. City of New York, 438 U.S. 104 (1978). The author would like to thank William B. Stoebuck for this reference.

[25] This view is not universal. *See, e.g.*, George Everson, *The Human Element of Justice* 10 J. CRIM. L. & CRIMINOLOGY 90 (1919); David Blanck, *The Appearance of Justice Revisited* 86 J. CRIM. L. & CRIMINOLOGY 887-927 (1996).

[26] HENRY R. GLICK, COURTS IN AMERICAN POLITICS (McGraw Hill 1990).

[27] JOHN CHIPMAN GRAY, THE NATURE AND SOURCES OF THE LAW 114 (Columbia University Press 1909).

[28] ARTHUR HOGUE, ORIGINS OF THE COMMON LAW 67 (Liberty Press 1966).

durable public opinion.[29] Durable public opinion, of course, is what we mean when we speak of norms. This opinion then helps to define efficiency so that the efficiency of the common law, far from being unusual, should be expected. Judges act according to a norm—a norm that expects them to dispense justice;[30] they use the language of justice. I will cite one example among almost endless possibilities because, first, it uses the language of justice; second, it illustrates the regard for others; third, it shows concern for the income distribution; and fourth, it keeps with my custom here of using historical references.

In *Gilmore v. M'Kelvey*, a case arising out of the Irish land law of 1881, the court writes:

> With respect to the question of value, the court is perfectly unanimous. One cannot help having a certain feeling with respect to a gentleman who having in 1878 voluntarily and without coercion taken a couple of fields outside the town from a lady, not very wealthy, at a rent of £30 a year, comes in the year 1882, and seeks to get a perpetuity in that land as against her at a rent of £12 15s. I have no doubt Mr. Gilmore reconciled himself to the transaction, *but there are many people who would not.*[31]

A more powerful and more modern example is found in the history of racial covenants. In *Shelley v. Kraemer*, 344 U.S. 1, 5 (1948), state enforcement of restrictive covenants were found to be an illegal state action. This ruling overturned a very long history of holding such covenants and their enforcement private action. What allowed this change? I note that strong norms tend to become law. "When nudged along by judicial recognition, norms become law, in the formal as well as the informal sense."[32] It is not that norms of racial discrimination were not widely prevented but that in enforcing covenants, the courts looked as well to the sentiments of future generations and gave standing to those previously without it. In discussing restricted racial covenants tied to land ownership or occupancy, Carol Rose notes:

---

[29] RICHARD POSNER, THE PROBLEMS OF JURISPRUDENCE 17 (Harvard University Press 1990) (referencing OLIVER WENDELL HOLMES, THE COMMON LAW (Brown Little ed., 1881).

[30] The trickier question, of course, is what it means to say that a judge should "do justice." Under the jurisprudential doctrine of positivism, a judge does justice (especially in a democratic country) by following the "plain meaning" of statutes. *See* Anita Allen, *Autonomy's Magic Wand: Abortion and Constitutional Interpretation* 72 B. U. L. REV. 683, 692 (1992). Under the natural law and legal realist theories of law, a judge does justice by recognizing either transcendent moral values (natural law) or public policy and common sense (legal realism). *Id.*; Kenneth Pennington, *The Spirit of Legal History* 64 U. CHI. L. REV. 1097 (1997).

[31] *See* E.O. MACDEVITT, LAND CASES, BEING A COLLECTION OF REPORTS OF DECISIONS UNDER THE IRISH LAND ACTS SINCE THE PASSING OF THE LAND LAW (IRELAND) ACT, 1881, WITH OTHER CASES NOW REPORTED FOR THE FIRST TIME 10 (Alex. Thom & Co. 1884) (emphasis added).

[32] Carol Rose, The Story of Shelly v. Kraemer, *in* PROPERTY STORIES at 198 (Gerald Korngold and Andrew Morriss eds., 2004).

> [W]hen a court holds that a covenant "runs" to subsequent purchasers as a part of the property, the court implicitly makes assumptions about what landowners in general would be likely to expect and to value in owning the land in question. These subsequent parties have not explicitly stated what they want: the question rather is whether they are bound by things that prior owners wanted. According to the limiting doctrines of covenant law, when a court imposes such an earlier arranged covenant on subsequent landholders, the court must explicitly or implicitly suppose that some substantial number of persons in their situation would find the covenant restrictions valuable.[33]

To limit such an assumption to only white purchases, that is to give standing only to white purchases, would have been to enforce norms that were not unexceptional or widespread. Thus we have a powerful example of the role of common law in reaching justice.

Attempts such as those of Landes and Posner to explain judicial behavior from an interest-group perspective are "simply unconvincing," as North and Buchanan have pointed out.[34] As Hogue states, "When judges in medieval England failed to maintain the high standards of learning and disinterested action expected of them, English feudal barons, churchmen, and merchants insisted on reform."[35]

Efficiency itself is such an important norm that we should not be surprised when impartial judges advance changes in rules that are efficient. Gray argued that judges and jurists approached the law from the side of public welfare, and sought to adapt it to the common good.[36] Holmes in *The Common Law* asserted that when revenge was a prevailing sentiment, the law provided a remedy for a wrong that approximated what would have been considered necessary to give victims their vengeance.[37] Later, when revenge became less important relative to the values of deterrence and compensation, the old doctrines ingeniously adapted to the new sentiments.[38]

The relationship between the British king and the judiciary may explain the norm in part. Efficient norms that promote the wealth of a nation are likely to increase the sovereign's wealth as well, and reduce dissention, and will tend to then be left undisturbed.[39] Thus, norms such as

---

33 *Id.* at 197.

34 William Landes and Richard Posner, *The Independent Judiciary in an Interest Group Perspective*, 18 J.L. & ECON. 875 (1975); DOUGLAS NORTH, STRUCTURE AND CHANGE IN ECONOMIC HISTORY 57 (W.W. Norton & Co. 1981); James Buchanan, *Comments on the Independent Judiciary in an Interest Group Perspective*, 18 J.L. & ECON. 903 (1975).

35 Hogue, *supra* note 28, at 253.

36 *See* GRAY, *supra* note 27.

37 *See* HOLMES, *supra* note 29.

38 *See* POSNER, *supra* note 29.

39 A full explanation of the origin of the norm that judges should dispense justice would require a treatise on English history, which I do not provide. I will, however, note that a straightforward extension of the work of Weingast and of Calvert suggests that a norm of justice arises as an equilibrium condition in a "game" that produces stable democracies. The equilibrium condition requires that citizens agree on the boundary of the state and that those boundaries be self enforcing. Barry R.

those requiring payment of debts or enforcement of contracts among citizens are wealth-increasing, and will not be disturbed by the sovereign. The sovereign's judges have, then, an incentive to see that those norms become legally enforceable.[40]

VI. ADOPTION OF UNCONTENTIOUS NORMS IS EFFICIENT

A norm is a set of rights or ownership established by custom. A norm contributes to efficiency by setting or clarifying rights.[41] When norms are well established they are more likely to be efficient because the rights they establish will be more certain. By definition, uncontentious norms are widespread, long-established, and without controversy.[42] Norms lacking one of these elements are contentious. An uncontentious norm is more apt to be efficient than a contentious norm. Being without controversy such norms can reasonably be regarded as fair. Property rights established under an uncontentious norm are likely to be settled, accepted, better known, and clearer than those established under a contentious norm. When uncontentious norms are in effect, it is less likely that an efficient rule change will exist.[43] If a rule is uncontentious, it is settled and enforced through social pressure.

A court that adopts uncontentious norms into the common law establishes legal property rights where they did not exist before and ensures that these legal rights correspond with established economic ownership. Suppose, for example, the norm in a community is that group A has the right to collect driftwood along a certain beach to the exclusion of group B. Psychological ownership in driftwood among group A has been established. In the absence of knowledge that rights have been mis-specified by custom, the assignment of rights to A is more likely to be efficient. A's loss—were the right to be assigned to B—would be measured by its WTA, but B's gain

---

Weingast, *The Political Foundations of Democracy and the Rule of Law* 91 AMERICAN POLITICAL SCIENCE REVIEW 245 (1997); Randall L. Calvert, *A Rational Choice Theory of Social Institutions: Cooperation, Coordination, and Communication*, MODERN POLITICAL ECONOMY (Jeffrey Banks and Eric Hanushek eds., New York: Cambridge University Press) (1995). In a game-theory setting, this occurs when constitutional or other provisions are held in sufficiently high esteem that citizens are willing to defend them. I note also that disturbing a norm is costly because it fuels opposition.

40 Myers notes the puzzling fact that England produced a more stable democracy earlier than other European countries but also had a stronger monarch earlier. This puzzle may be resolved by considering the role of norms. A.R. MYERS, ENGLAND IN THE LATE MIDDLE AGES 1 (Penguin Books 1971). A stronger monarch produces a more uniform system of norms. This more uniform system will make it more difficult for the monarch to play off some groups against others.

41 Richard O. Zerbe, Jr. and C. Leigh Anderson, *Culture and Fairness in the Development of Institutions in the California Gold Fields*, 61(1) J. ECON. HIST. 114-43 (2001).

42 WILLIAM BLACKSTONE, COMMENTARIES (19th ed. 1900).

43 There is likely to be an exception to this statement when a rule change can harness the enforcement power of government and in this way improve on a norm.

would be measured by its WTP. Even if group B were the least-cost collector, the WTP may be less than the WTA of A if the right has more than purely commercial value. Assigning the right to A is likely to be the efficient assignment even if B is the least-cost collector. First, to assign the right to B, without compensation to A when it has psychologically belonged to A, will be seen as unfair by others. Society's moral sentiment that the rule change was unfair would be measured as an efficiency loss resulting from the assignment to B. Second, it will be expensive to determine that B is the least-cost collector. The main role for the courts in this sort of situation is simply to specify property rights and thus lower the transactions costs of A's selling the right to B so that the right is more likely to transfer. When the existing psychological ownership is accommodated, any transfer from the group with psychological ownership will be compensated and thus will be more likely to be seen as fair.

A court's decision to reallocate a right away from party A, who holds psychological ownership, to party B would involve greater risk. To determine that it is more efficient to assign a right to party B, the court would need to determine that the right was worth more to B than to A, and this determination would be both expensive and error-prone. The court knows, moreover, that if it allocates rights according to existing social norms, parties will likely reallocate rights to resolve any inefficiency if transaction costs are low enough.[44] The court recognizes that any loss caused by allocating the right to collect driftwood to party A instead of party B is limited to the transaction costs of transferring the property right. These transaction costs are likely to be less than the costs of having the courts attempt to determine whether B is the efficient holder of the right. Thus, granting the right to A rather than B and following the norm is more likely to result in efficiency.

VII. THE COMMON LAW INCORPORATES LONG-STANDING CUSTOM

To establish rights that correspond to economic ownership when conditions are unchanging is, by definition, efficient. Norms that are uncontentious and long-standing involve the establishment of economic rights. When law adopts a norm, legal ownership corresponds to and codifies economic ownership. When conditions change, the common law seeks an efficient adoption of norms and may or may not find one, depending on the pace of change and the corresponding difficulty of determining the efficient rule.[45] A wide variety of common law was developed by judges in response to such scenarios.[46]

---

[44] Transaction costs may be prohibitively high in the case of public goods.

[45] Underlying this proof is a notion that preferences are to be taken as given. It is true that in benefit-cost analysis, preferences are usually taken to be given. However, efforts to change the law to

In his treatise, Blackstone provided a list of criteria judges should consider before codifying norms into the common law.[47] Blackstone contended that norms must be long-established and uncontentious before being incorporated into the common law. According to Plucknett, the civilian jurist Azo, held in high esteem by Bracton, noted that "a custom can be called long if it was introduced within ten or twenty years, very long if it dates from thirty years, and ancient if it dates from forty years."[48] This requirement helped ensure that the public's willingness to accept the changes was greater than their willingness to maintain the status quo.[49] Blackstone suggested that prior to codifying a norm, a social sanction for failure to obey the norm should already exist in order to guarantee that only important customs became enshrined into law. Thus, Blackstone's criterion ensured the incorporation of only true norms, which are efficient norms.[50]

Judges have historically sought out custom to incorporate into common law. One of the attempts to codify social norms was made by Lord Mansfield, who acted as chief justice of the court of King's Bench in England from 1756 to 1788. During his tenure, Mansfield adeptly incorporated the merchant law into the common law, thus fashioning what had been a body of special customary law into general rules within the common law.[51] Hogue notes:

> When a case touched commercial law, [Mansfield] saw to it that reputable merchants of the city of London formed the jury. Thus he secured in his court the participation of jurors who presumably understood every detail of material evidence. Outside court, on social occasions, he cultivated the acquaintance of merchants to acquire for himself a precise knowledge of their ways of doing business.[52]

VIII. THE SOURCES OF INEFFICIENCY

A. *Inefficiency Arises From a Change in Conditions*

A law will be more just and efficient when it is more complete and fit. A law is complete if it specifies all relevant rights. It is fit to the extent that

---

accord with preferences may themselves be KH-efficient. Underlying this proof is the notion that changing preferences to be in accord with the law cannot be described as efficient.

46 David Cohen and J.L. Knetsch, *Judicial Choice and Disparities between Measures of Economic Values*, 30 OSGOODE HALL L. J. 737 (1992).

47 BLACKSTONE, *supra* note 42, at 56-57.

48 THEODORE PLUCKNETT, A CONCISE HISTORY OF THE COMMON LAW 308 (5th ed. 1956).

49 BLACKSTONE, *supra* note 42.

50 Again, Blackstone's (1800, pp. 56-57) norms represent the ideal along a continuum, not a rigid requirement. *Id.* at 56-57.

51 *See* Hogue, *supra* note 28.

52 *Id.* at 248-49. *See also* PLUCKNETT, *supra* note 48, at 350, 664.

the particular specification of rights is efficient. Consider a static society in which all rights are specified. A static society is one in which there is no change in technology, sentiments, or knowledge. This society will be efficient to the extent that transactions costs allow. Insofar as transactions costs can be lowered through private innovation or through government intervention, profit seekers will lower them to the extent that transactions costs of political action allow. Thus, such a society is efficient in that it can be not improved upon.

An increase in inefficiency in such a static society with a norm of justice arises from, and only from, changing conditions. Again I refer to changes in sentiments, technology, or knowledge.[53] The deterioration rate of a law or norm in terms of completeness and fitness will be a positive function of the rate of change. Thus, in a static society there is no increase in inefficiency and the base level of inefficiency is limited by experience. Similarly, a rule that adopts uncontentious norms is almost certainly efficient where conditions are not changing.[54]

For example, suppose residents in a hot climate have no right to run noisy air conditioners without approval of their neighbors. This law is complete. Suppose it is not fit since the aggregate WTP to run noisy air conditioners is significantly greater than the WTA payment to allow them to be run. Overheated residents are able to purchase or obtain approval from their neighbors but transactions costs are higher and the net social surpluses are lower than they would be if all residents had the right to run air conditioners. In considering a challenge, the court will note that the custom is for users of air conditioners to purchase the right to run them in many cases. The court, relying on custom, will then find a right to use air conditioners as a matter of common law so that both justice and efficiency will be served.

Similarly, suppose that air conditioners are a new invention and that the right to their use is unclear. The law is not complete. In a hot climate, the likely custom will be that people will use their conditioners and complaints about their noise will be generally disregarded. Again, the courts, relying on custom, under traditional common law procedure, would create a right of use. Thus, custom in a static society will suggest what changes are efficient so that a static society will tend toward efficiency.

---

53 It has been suggested to me that changes in the extent to which foreigners operate in the society would also affect efficiency.

54 Thus, Ellickson's discussion of the conditions under which norms are efficient, and his discussion of why norms in whaling and norms with respect to wandering cattle are more efficient than other hypothetical norms, fails to recognize that, by definition, norms are efficient if they are stable and uncontentious. Ellickson's own examples, if read from the perspective illustrated here, demonstrate just this point. His interesting discussion indicates why norms of closely knit groups may be different from those of more diverse groups, and why it might to better to belong to a closely knit group than not; but it does not explain why norms are efficient. Robert Ellickson, *A Critique of Economic and Sociological Theories of Social Control*, 16 J. LEGAL STUD. 67 (1991).

In a static society, better evidence will exist about which result among competing outcomes and legal rules is efficient. Unchanging conditions over time will provide cleaner and less ambiguous information. As society members gather knowledge about mis-specified rights, pressure arising from a sense of fairness and justice, as well as self-interest, will build to change existing rules. Knowledge will become cheaper to acquire with time. As the price to attain additional knowledge decreases, rules will change to reallocate mis-specified rights. Judges will adjust rights at common law as a matter of justice so that society becomes KHM-efficient. The level of residual inefficiency in a static common law society that adopts norms will be limited by transactions costs that prevent an efficient trade. If transactions costs are sufficiently high such that no trades take place when the right is mis-specified, there will be no norm to examine and some residual of inefficient rights might remain.

B. *The Greater The Pace Of Change, The Less Likely Common Law is Efficient*

Changes in sentiments, technology, and knowledge create a dynamic world that guarantees both the existence and continual creation of inefficiencies. With change, laws tend to become less complete and less fit. A social change may render a previously efficient rule inefficient when the change results in ambiguous ownership, as with the discovery of a new valuable resource created by new technology or knowledge. Similarly, a change in moral sentiments may make a change in ownership efficient, as when the exercise of a right that harms others is no longer seen as acceptable. Thus, it is generally efficient to change the legal precedent only when conditions change.

A change in conditions creates inefficiencies; a more rapid change will create greater inefficiencies. A change in conditions implies, as North noted, a change in relative prices.[55] As relative prices change, behavior will change in response as will the efficient equilibrium. North explains historical change on the basis of just such responses to changes in relative prices. More rapid change increases the pressure for efficient rule changes but may also increase the costs of discovering which rule changes are efficient. Hogue notes that in every generation both lawyers and laypeople seem to have been drawn toward two desirable but separate and contradictory goals. The first of these is the goal of permanence, stability,

---

[55] DOUGLAS NORTH, STRUCTURE AND CHANGE IN ECONOMIC HISTORY (W.W. Norton & Co. 1981).

and certainty in legal doctrines. The second is the goal of flexibility and adaptability, permitting adjustment of the law to social necessity.[56]

The slower the pace of change, the easier it is for changes in custom to precede law. The slower the pace of change, the easier it is for judges to accurately determine the social standards of the age and incorporate custom into law. Thus, one would expect common law to be more efficient in a quieter age.[57]

Efficiency in law is found through completeness and fitness. Inefficiency arises from a change in conditions that increases incompleteness and reduces fitness. As conditions change, the law or custom loses both completeness and fitness. The rate at which new issues are raised in a particular area of the law will be a positive function of both the level of incompleteness and lack of fitness in those areas of the law. The rate of new cases will increase as the net gains from a change toward greater efficiency increases, and the law's injustice increases. As law becomes less just, advocates that seek greater efficiency and justice will have greater chances of persuading judges who value justice. If custom has changed and had time to become settled and uncontentious, the new custom will be a guide to efficient changes in the law. Otherwise, intellectual knowledge may be a guide. By this process, judges adopting new law by adapting to either new custom or new intellectual knowledge will tend to create efficient common law precedent.[58]

Custom has changed over time, and the law has changed with it. Plucknett notes, "The Middle Ages seem to show us bodies of custom of every description, developing and adapting themselves to constantly

---

56 Hogue also notes that "[t]he result of the pull in these two directions has been an unresolved tension between factions, parties, or groups of men; not always a tug-of-war between conservatives and radicals. The dual objectives can exist in the legal thought of a single jurist." Hogue, *supra* note 28, at 8.

57 It does not follow, however, that the common law should not be used in eras of rapid social change. Just as the common law runs into trouble during eras of rapid social change, so too would any other system of law that attempted to match norms and laws. In a civil law system, judges are simply administrators, and play little or no role in creating law. In such a system, it is the legislature that would run into trouble, as it attempted to find an uncontentious norm when no uncontentious norm existed. If either a common law or civil law country attempted to ignore norms, this would create inefficiency.

58 La Porta et al. have found that law enforcement with respect to investor protection is stronger in common law systems: "there is no clear evidence that different countries favor different types of investors; the evidence rather points to a relatively stronger stance favoring all investors in common-law countries." Rafael La Porte, Florencio Lopez-de-Silanes, Andrei Schleifer & Robert W. Vishny, *Law and Finance*, 106 J. POLIT. ECON. 1113, 1151 (1998). Moreover, as La Porta et al. note, recent evidence suggest that stronger protection promotes economic growth. *Id.* at 1152. That is, in my terminology, common law countries show greater fitness with respect to investor protection. This result is what we would expect if judges seek justice in common law countries, if we might also suppose that the path to justice through legislation is more time consuming than through changes to the common law, at least in situations where the adjustment to efficiency is gradual rather than heroic. Legislatures are more likely to respond to interest group considerations than are judges that seek justice.

changing conditions." He continues, "Indeed nothing is more evident than that custom in the Middle Ages could be made and changed, bought and sold, developing rapidly because it proceeded from the people, expressed their legal thought, and regulated their civil, commercial and family life."[59]

The more rapidly conditions change, the less opportunity there is for uncontentious norms to develop and the more difficult it is for judges to determine what is in fact efficient. When conditions change more rapidly, there may be no particular custom or norm that the common law can incorporate. There may, however, be reasonable generalizations from existing particular customs that represent the reasonable expectations of rights-holders and that are thus efficient. There will also be general norms or general custom that can be applied, although it may be doubtful that a general norm will be superior if a particular norm exists. The phrase "general custom" refers to general norms that may be regarded as principles. Such norms may include an expectation that one is entitled to what one earns, that promises should be honored, or that equals should be treated equally.

When social conditions change, the analogy between past cases and the current issues may become strained, which may make it difficult for the parties to predict how the law will be applied to a current dispute. The harder it is to predict how a law will be interpreted, the higher transactions costs will be as lawyers and experts are enlisted as consultants (in the hopes of avoiding a lawsuit) or litigators (after a lawsuit begins).

To substitute for loss of uncontentious custom in a period of rapid change, judges will seek evidence of efficiency in intellectual knowledge. Such a source will, as with common law, be more reliable the more settled it is and the longer such settled knowledge has existed. Changing conditions will increase inefficiency through incompleteness and lack of fitness more rapidly than custom or intellectual knowledge can answer, thus causing an increase in legal inefficiency.

In the modern era, when the pace of changes in conditions has been more rapid, greater reliance has been placed on judicial judgment concerning what is efficient as compared with well-established custom in the development of efficient common law. As Friedman notes, "Since the First World War the tempo of social change has accelerated beyond all imagination. With it the challenge to the law has become more powerful and urgent."[60] Today, courts recognize that the law must change in response to changes in sentiments, knowledge, and custom. For example,

---

[59] PLUCKNETT, *supra* note 48, at 308.

[60] Friedman notes the law's response to some of these changes. Examples of common law adapting to change may be found in *McPherson v. Buick*, 217 N.Y. 382, 111 N.E. 1050 (1916), and *Donoghue (or McAlister) v. Stevenson, (1932) All ER Rep 1; (1932) AC 562; House of Lords (1932).* WOLFGANG FRIEDMAN, LAW IN A CHANGING SOCIETY 26f (1959). A host of similar examples are mentioned by Friedman.

the U.S. Supreme Court noted in 1997 that antitrust law must change to reflect "new circumstances and new wisdom," and that the common law cannot remain "forever fixed where it was" in a previous era.[61] The problem is that in a period of rapid change it is more difficult for a judge to determine whether sentiments, knowledge, or customs are changing, and to determine the course of such changes.[62] These propositions will be examined through the following case studies.

IX. EXAMPLES OF THE INTERACTION OF COMMON LAW AND NORMS OF JUSTICE

I have suggested that the law achieves justice by adopting uncontentious norms. I have suggested that the common law tends toward efficiency and that this tendency will be more successful in a quieter age. Logic and analysis can be imperfectly substituted for the experience of norms in the modern age where norms are less available, as seen in the example of vertical integration in antitrust law. When economists and lawyers ignore considerations of justice, as when moral sentiments are ignored, the normative analysis is flawed and the correspondence between efficiency and the common law is less apparent. This is illustrated by a consideration of slavery and of dueling in the antebellum South. When no such norms are available, the efficient law is difficult to determine. This is illustrated by the cases dealing with segregation and the law of nuisances discussed below.

A. *A Change in Knowledge: Antitrust Law*

Even though a federal statute governs antitrust law, it is generally accepted that courts supply the content of the antitrust law by creating antitrust "common law."[63] In deciding antitrust cases, courts recognize that the law should change to reflect new economic theory and data.[64]

Changes in knowledge have made it efficient to change the law of vertical restraints. In antitrust vernacular, a "vertical restraint" is an attempt by a manufacturer to control the activities of wholesalers, distributors, or retailers. There are two basic categories of vertical restraints. First, there are price restraints where the manufacturer sets either a minimum or a maximum price at which a retailer may sell its products to customers. Second, there are non-price restraints where the manufacturer limits the

---

61 *See* State Oil v. Khan, 522 U.S. 3, 21 (1997).

62 *See* FRIEDMAN, *supra* note 60.

63 *See Khan*, 522 U.S. at 20, 21.

64 *See id.*

customers to whom a retailer may sell its products. Non-price restraints usually take the form of territorial restraints in which a retailer is given an exclusive right to sell the manufacturer's product within a certain area in return for a promise not to sell the product to any customers outside of the area. Horizontal restraints, on the other hand, refer to agreements between firms at the same level—i.e., two or more manufacturers *or* two or more retailers—not to compete. Like vertical restraints, horizontal restraints usually involve either price-fixing or territorial market divisions.

Although the Sherman Antitrust Act[65] prohibits "every contract, combination, or conspiracy" from suppressing competition, the courts quickly realized that every contract suppresses competition in some sense (because an agreement to sell 100 widgets to one person is an implicit agreement not to sell those particular widgets to anybody else) and that Congress could not have intended to outlaw every business agreement or even every agreement between competitors.[66] Therefore, courts developed a "rule of reason" stating that only "unreasonable" restraints (those that harm competition more than they benefit it) are violations of the antitrust laws.[67] On the other hand, courts realized that some types of agreements, such as horizontal price-fixing, were so likely to harm competition that an in-depth analysis of each one was not justified.[68] Such agreements are unlawful "per se." If it is proved that a defendant engaged in an agreement subject to the per se rule, the defendant cannot escape liability by arguing that his agreement had pro-competitive effects and will be punished. In holding that a type of agreement is unlawful per se, courts essentially make an economic *prediction* that the probability injuring competition by an agreement of that type is so much greater than its probability of benefiting competition that it is not worth the court's time to analyze the competitive consequences of a particular agreement of that type.[69] Therefore, economic theory greatly aids judges who must decide whether to hold a type of agreement unlawful per se.[70] In characterizing an agreement as unlawful per se, the court is denying the defendant economic standing: indeed, Justice Harlan once noted in a dissent that the per se rule is a "no trial rule."[71]

The first Supreme Court case involving a vertical restraint, *White Motor Co. v. United States*, was not decided until 1963.[72] In *White Motor Co.*, a truck and auto parts manufacturer placed both price and non-price

---

65 15 U.S.C. §§ 1-7 (2005).

66 *See* Chicago Bd. of Trade v. United States, 246 U.S. 231, 238 (1918).

67 *See* United States v. Standard Oil Co. of New Jersey, 221 U.S. 1, 59-60 (1911).

68 *See* N. Pac. R.R. v. United States, 356 U.S. 1, 5-6 (1958).

69 *See State Oil*, 522 U.S. at 10 (1997).

70 *Id.*

71 Albrecht v. Herald Co., 390 U.S. 145, 159 (1968).

72 White Motor Co. v. United States, 372 U.S. 253 (1963).

restraints on distributors.[73] The government argued that both the price and non-price restraints should be subjected to the per se rule, and the lower court agreed.[74] White Motor Company did not contest the ruling that vertical price-fixing was illegal per se, but it did argue that vertical non-price restraints should be governed by the rule of reason. The Supreme Court agreed with White Motor Company. Specifically, the Court held that because the application of the per se rule is a prediction that agreements of a certain type are almost always profoundly anticompetitive, the rule should not be applied to a type of agreement where the courts did not have enough experience to make a reliable prediction.[75] In other words, the court decided that too little was known "about the economic and business stuff out of which [non-price restraints] emerge" to say with certainty that vertical non-price agreements would almost always harm competition.[76] Therefore, the court remanded the case to the district court to determine whether White Motor Company's non-price restraints could be justified under the rule of reason.[77] Specifically, the court speculated that vertical non-price restraints, unlike horizontal territorial restraints, might benefit competition by allowing small companies to break into a business, and such restraints might be necessary to save a failing manufacturing company.[78]

Four years later in *United States v. Arnold, Schwinn & Co.*, the Supreme Court imposed the per se rule on vertical non-price restraints unless the restraint was part of a consignment contract.[79] Schwinn manufactured bicycles and sold them through retailers. About 75 percent of the sales to retailers were characterized as "consignment contracts," while the other 25 percent were described as "sales contracts." Both the consignment and sales contracts with retailers placed territorial restraints on the sellers' ability to sell the bicycles.[80] The Court apparently felt that it had become familiar enough with vertical non-price restraints to make a reliable economic prediction about their competitive effect.[81] It began its analysis by interpreting *White Motor Co.* narrowly, stating that *White Motor Co.* extended the rule of reason to non-price restraints only when the

---

73 *Id.* at 257-59.

74 *Id.* at 256.

75 *Id.* at 263.

76 *Id.*

77 *Id.*

78 White Motor Co., 372 U.S. at 263.

79 *See* United States v. Arnold, Schwinn & Co., 388 U.S. 365, 379 (1963). In a consignment contract, a manufacturer delivers a product to a distributor, but retains title to the product until it is actually sold to a customer. The distributor keeps some of the sales money for itself, and sends the rest to the manufacturer at a previously agreed upon ratio.

80 *Id.*

81 *See id.* at 373-74.

manufacturer was a new, small company or a failing business, and noting that Schwinn was neither.[82]

In analyzing the competitive effect of vertical non-price restraints, the *Schwinn* court concluded that some small companies could compete with manufacturing giants only if they could offer dealers exclusive sales contracts that involved vertical non-price restraints.[83] On the other hand, the court felt that "prudence" dictates that it would be foolish to allow a company to give a dealer an exclusive contract while retaining "title" to and "dominion" over the goods.[84] Therefore, the court compromised by applying the rule of reason to non-price restraints in consignment contracts, but applied the per se rule to restraints in sales contracts.[85] The court justified this compromise by arguing that it was consistent with the "*ancient* rule against restraints on alienation."[86] The court dismissed Schwinn's argument that its exclusive dealerships enabled it to compete more effectively with larger competitors, because Schwinn was not a failing business.[87]

Justice Stewart, in a forceful dissent, argued that the majority's reliance on an "ancient" rule to resolve a difficult antitrust issue was misplaced. He argued that the "ancient" rule against restraints on alienation is of little help in predicting whether a vertical restraint will benefit competition today.[88] He noted that, in any event, the "ancient" rule against restraints on alienation outlawed only unreasonable restraints, and therefore operated much more like the rule of reason than the per se rule.[89] He agreed with the majority and Schwinn that being able to offer exclusive dealerships would be necessary if a company was to attract quality retailers and distributors. However, he felt that whether a transaction with a retailer was characterized as a consignment or a sales agreement made little practical difference in the manufacturer's ability to restrict competition and therefore should not determine whether an agreement violates antitrust laws.[90]

One year later, the Supreme Court extended the per se rule to vertical price-fixing agreements in *Albrecht v. Herald Co.*[91] *Albrecht* involved a

---

82 *See Schwinn*, 388 U.S. at 374-375; *White Motor Co.*, 372 U.S. at 263.

83 *See Schwinn*, 388 U.S. at 379. This is almost certainly a misreading of *White Motor Co.*, since there is no evidence that White Motor Company itself was a new company or a failing business. *See White Motor Co.*, 372 U.S. at 263. The references to new companies and failing businesses were intended to be *examples* of situations in which the rule of reason might be satisfied, not to constitute an *exclusive* list. *See id.*

84 *See Schwinn*, 388 U.S. at 380.

85 *See id.* at 379.

86 *See id.* (emphasis added).

87 *See id.* at 374-75.

88 *See id.* at 392.

89 *See Schwinn*, 388 U.S. at 391.

90 See *id.*

91 Albrecht v. Herald Co., 390 U.S. 145 (1968).

newspaper company that terminated a paperboy's route when he charged more than the maximum price specified in the contract.[92] Although the price the Herald set was in no way predatory, the court applied the per se rule and[93] offered three justifications. First, the court noted that part of the purpose of the antitrust law is to preserve an entrepreneur's independent business judgment, and that an entrepreneur's judgment was restricted regardless of whether he was forced to offer low prices or high prices.[94] The court insisted that a firm should not be able to substitute "the perhaps erroneous judgment of the seller for that of the competitive forces of the market."[95] Specifically, a manufacturer might set prices so low that the dealer would be unable to make a profit, or it might set prices so high that the dealer would be prevented from offering essential services to customers.[96] Second, the court argued that the Herald could not justify its maximum price-setting rule by claiming that it protected consumers from paperboys who themselves enjoyed a monopoly; it was the Herald that granted the paperboys a monopoly in the first place.[97] In other words, if the Herald argued an exclusive paper route gave a paperboy monopoly power that he could use to demand super-competitive prices, the correct solution was to refuse to give him an exclusive paper route in the first place, not to grant it and then make the additional anticompetitive act of price-fixing. Third, the court noted that a maximum price-setting agreement might actually be a minimum price-setting agreement in disguise.[98] That is, a manufacturer might characterize something as a maximum price in a contract, but the dealers might realize that the manufacturer *really* wants them to charge that price at a minimum.

Dissenting from the majority, Justices Harlan and Stewart argued that all three of the court's justifications for the per se rule were economically naive. First, Justice Stewart pointed out that the antitrust laws are not concerned with protecting the independent business judgment of an entrepreneur when the entrepreneur is exercising monopoly power.[99] In fact, the paperboy's "business judgment" is less likely to be consistent with the needs of the market than the Herald's judgment because a paperboy will complain about a reasonable maximum price only when it prevents him from charging a super-competitive price to consumers.[100] Justice Harlan argued a company could completely eliminate independent entrepreneurs by hiring its own sales employees, and, while such an act would not violate

92 *See id.* at 147-48.
93 *See id.* at 154.
94 *See id.* at 152.
95 *See id.*
96 *See Albrecht*, 390 U.S. at 152-153.
97 *See id.* at 154.
98 *See id.* at 153.
99 *See id.* at 169.
100 *See id.*

the antitrust laws in any way, it would be more destructive to competition than the Herald's modest price ceiling rule.[101] Second, Justice Stewart asserted that a paperboy's exclusive territory was most likely a natural monopoly, which was a product of the market's inability to support more than one paperboy per territory rather than a grant of monopolistic power by the Herald.[102] Third, Justice Harlan pointed out that while it might be true that some maximum price agreements are disguised minimum price agreements, many maximum price agreements are not.[103] In deciding whether to apply the per se rule to maximum price agreements, he noted that the question "is not whether dictation of maximum price is ever illegal, but whether it is always illegal."[104]

Continental T.V., Inc. v. GTE Sylvania, Inc.[105] reversed *Schwinn*[106] and held that vertical non-price restraints should be subjected to the rule of reason. The *Sylvania* court noted that *Schwinn* had been wrongly decided for a number of reasons.[107] First, *Schwinn* ignored *White Motor Co.*'s warning that a per se rule was justified only when a court had sufficient experience with a business practice to make a reliable economic prediction about its consequences.[108] The majority in *Schwinn* did not identify any new information that had not been available at the time of *White Motor Co.*, yet it changed the rule.[109] Instead, *Schwinn* attempted to resolve its difficulties by turning to "ancient" common law distinctions.[110] Second, to the extent data existed that was available to *Schwinn* but had not been available in *White Motor Co.*, that data strongly indicated that a per se rule against vertical non-price restraints would be inefficient and possibly even disastrous.[111] Third, developments in economic theory after *Schwinn* strengthened the case that a per se rule against vertical non-price agreements was a mistake and that a rule distinguishing between consignment and sales contracts was wrongheaded.[112] In fact, it was the large companies with little legitimate need for exclusive dealerships that were most likely able to characterize their transactions as consignment

---

101 *See Albrecht*, 390 U.S. at 160-61.

102 *See id.* at 169.

103 *See id.* at 165-66.

104 *See id.*

105 Cont'l T.V., Inc. v. GTE Sylvania, Inc., 433 U.S. 36 (1977).

106 *Schwinn*, 338 U.S. at 365.

107 *See Sylvania*, 433 U.S. at 59.

108 *See Sylvania*, 433 U.S. at 47-48; *White Motor Co.*, 372 U.S. at 263; *Schwinn*, 388 U.S. at 379.

109 *See Sylvania*, 433 U.S. at 47-48; *Schwinn*, 388 U.S. at 379-80; *White Motor Co.*, 372 U.S. at 263.

110 *See Sylvania*, 433 U.S. at 53; *Schwinn*, 388 U.S. at 380.

111 *See Sylvania*, 433 U.S. at 53; *Schwinn*, 388 U.S. at 382-94.

112 *See Sylvania*, 433 U.S. at 48-49; *Schwinn*, 388 U.S. at 365.

contracts, and the small companies with a strong need to offer exclusive dealerships that were least likely able to do so.[113]

In general, most economists became convinced that because interbrand competition was more important to consumer protection than intrabrand competition, a manufacturer's interests were more likely to be consistent with the public's interest than with those of a distributor or a retailer.[114] For example, vigorous interbrand competition ensures that a dealer with an exclusive territory cannot exploit his monopoly power because a consumer would turn to a different brand name rather than pay super-competitive prices.[115] Additionally, economists noted that exclusive dealerships allowed a manufacturer to eliminate "free riders" who might dissuade retailers from offering vital services and repairs or from marketing the manufacturer's products.[116] As the economic evidence began to indicate *Schwinn* had been wrongly decided and as more and more scholars advocated its reversal, it became increasingly apparent that it would be efficient to reverse the decision. Because *Sylvania* explicitly relied on the expertise of economists and recognized it is desirable to change a common law rule when new knowledge suggests the old rule is inefficient, *Sylvania* has been hailed as a turning point in antitrust legal history and the beginning of modern antitrust analysis.[117]

When the recent case of *State Oil Co. v. Khan* reversed *Albrecht*,[118] the *Khan* Court noted that none of *Albrecht*'s "dire predictions" of the consequences of legalizing vertical price maximums were founded in fact and that such predictions only created additional problems.[119] In essence, Justice Harlan's predictions about the probable effects of *Albrecht* were borne out by the Court's subsequent experience.[120] *Albrecht* actually contributed to the elimination of independent entrepreneurs because it encouraged manufacturers to replace dealers with sales employees.[121] Further, many economists concluded that *Albrecht* hurt consumers because a dealer was considerably more likely to set a super-competitive price than a manufacturer because the latter either prevented dealers from making a reasonable profit or prevented them from offering services consumers desired.[122] Also, *Albrecht*'s logical underpinning was abrogated by *Sylvania* because it was now lawful in many circumstances to give a dealer

---

113 *See Sylvania*, 433 U.S. at 56.

114 *Id.* at 52.

115 *See id.*

116 *See id.* at 55.

117 Stephen Calkins, *An Enforcement Official's Reflections of Antitrust Class Actions*, 39 ARIZ. L. REV. 143 (1997).

118 *See State Oil Co.*, 522 U.S. at 3.

119 *Id.* at 19.

120 *Id.*

121 *Id.* at 16-17.

122 *See id.* at 17-18.

an exclusive territory, so it seemed foolish to prevent the manufacturer from protecting consumers by setting a maximum price.[123] Finally, the court agreed with Justice Harlan that the possibility a maximum price was a disguised minimum price was hardly a justification for outlawing all maximum-price agreements.[124] Under the rule of reason, the court could identify any alleged maximum-price agreement that was actually a minimum-price agreement.[125]

In conclusion, a survey of the line of antitrust cases dealing with vertical restraints reveals two changes over time. First, there is a shift in the Supreme Court's attitude toward the significance of new economic knowledge.[126] *Schwinn* quite consciously chose an "ancient" rule to draw a line on a difficult economic issue, and turned its back on recent economic knowledge.[127] *Albrecht* similarly cited a "noneconomic" concern: protecting the independent business judgment of dealers.[128] *Sylvania* and *Khan*, in contrast, recognized the importance of new economic knowledge, because the decision whether to apply the per se rule is an economic prediction about an activity's impact on the marketplace.[129] Second, there was an increase in the availability and prominence of economic literature discussing antitrust law. In fairness to the *Schwinn* court, the volume of economic literature available to assist courts in deciding whether to extend the per se rule was much smaller than it was at the time of *Sylvania*. Indeed, *Schwinn* itself provoked a great deal of the economic literature *Sylvania* relied upon.[130]

---

123 *See Khan*, 522 U.S. at 14; *Sylvania*, 433 U.S. 36; *Albrecht*, 390 U.S. 145.

124 *See Khan*, 522 U.S. at 17; *Albrecht*, 390 U.S. at 165-66.

125 *See Khan*, 522 U.S. at 17.

126 It is worth noting that in the early cases there are virtually no references to the economic literature, while the later cases are peppered with them.

127 *See Schwinn*, 388 U.S. at 380.

128 *See Albrecht*, 522 U.S. at 152, 158. In fact, the existence value of independent businesspeople is an "economic" good but *Albrecht* does not discuss the value of independent business judgment in economic terms. *See id.*

129 *See Khan*, 522 U.S. at 21; *Sylvania*, 433 U.S. at 54.

130 *See Sylvania*, 433 U.S. at 47-48; *Schwinn*, 388 U.S. 365. According to Westlaw, the *Schwinn* case generated over three hundred case citations, some administrative material and a handful of law review articles before it was reversed in 1977 in *Sylvania*. The journal articles are significantly undercounted because many journals were not included in Westlaw until the 1980s or 1990s.

B. *A Change in Sentiment: the Examples of Slavery, Dueling and Segregation*

While it is easy to find examples where the common law is efficient,[131] it is almost as easy to find examples where it is not.[132] The following examples discuss the difficult formation of common law under changing conditions. The first, dealing with the law of dueling, shows the relationship between changes in sentiments and changes in law.[133] The second illustrates how courts attempt to craft an efficient rule to deal with changing sentiments about the importance of sunlight and changing technology for converting sunlight into energy. The third example, the development and demise of the "separate but equal" standard in segregation, demonstrates the relationship between changing sentiments, shifts in the regard for others, and changes in law. There is an entire class of examples in which economists' judgments about efficiency are flawed because they ignore the sentiments of those not directly affected: the sentiments of third parties. These include the efficiency of rape, abortion, dueling, and slavery. I consider the examples of slavery, dueling, and segregation below.

1. Slavery

In response to criticisms from Dworkin, Posner offered that efficiency "probably" condemns slavery as inefficient because a person could, if he chose, be more productive and produce a greater physical output as a free person than as a slave.[134] This argument is both technically incorrect and misguided. It is technically incorrect because it ignores those sentiments that must be taken into account when determining efficiency, even if we confine ourselves to the sentiments of the slave and of the owner and ignore

---

131 For example, the general common law rule that a landowner is not liable for negligently harming a trespasser is probably efficient. The exceptions where a landowner is liable to a trespasser are probably efficient as well.

132 The law of mining and significant pieces of labor law come to mind. *See generally,* Robert H. Lande & R.O. Zerbe, Jr., *Anticonsumer Effects of Union Mergers: An Antitrust Solution*, 46(2), DUKE L. J. 197 (1996).

133 An ongoing change in sentiments is happily reflected in Sen's criticism of values that have led to the phenomenon of "missing women," that is, women in developing countries whose survival has not been given proper weight. *See* AMARTYA SEN, DEVELOPMENT AS FREEDOM 20, 104-07 (Alfred Knopf ed. 1999).

134 For example, see Richard A. Posner, *Ethical and Political Basis of Efficiency*, 8 HOFSTRA L. REV. 501-502 (1980). Posner notes, "For example, if we started with a society where one person owned all the others, soon most of the others would have bought their freedom from that person because their output would be greater as free individuals than as slaves, enabling them to pay more for the right to their labor than that right was worth to the slave owner." *Id.* at 501.

the regard for others. Consider first the transactions between the slave and the owner alone. Imagine that the status quo position is one of slavery. Then the efficiency question is whether the WTP of the slave is greater than the WTA of the owner. Even if the slave could be more productive as a free person and capital markets were perfect such that the slave could borrow against future earnings, we cannot say whether the WTP of the slave would be greater than the WTA of the owner. The owner may have a taste for owning a slave and may be willing to suffer the financial loss inherent in retaining the slave because his psychological gain is greater than the financial loss. Posner's mistake arises partly from his focus on efficiency in terms of material wealth, but primarily from ignoring the regard for others.[135]

One cannot maintain that slavery, at least by the 1850s in the United States, was efficient on the basis of available evidence, professors Fogel and Engerman notwithstanding. While professors Fogel and Engerman[136] maintain that slavery was efficient, what they really show is that it was economically viable. They answer in the negative the question, "If slavery had been eliminated, would the GNP have been greater?" There is no reason to doubt their answer. However, to determine whether slavery was efficient is a different question from the one Fogel and Engerman, as well as Posner, addressed. The question should have been: "Would the WTP of those opposed to slavery have been greater than the WTA of slaveholders?"[137] To this question no answer has been provided. That slavery might not have been efficient is suggested when we consider the sentiments of others, and the growing social antipathy towards it.[138] One cannot imagine that the practice would have survived the turn of the century, even without the Civil War.

---

135 Similar criticisms can be made about Posner's discussion of rape. The regard for others is, however, recognized by Guido Calabresi and A. Douglas Melamed. They note, "[I]f Taney is allowed to sell himself into slavery, or to take undue risks of becoming penniless, or to sell a kidney, Marshall may be harmed, simply because he is a sensitive man who is made unhappy by seeing slaves, paupers or persons who die because they sold a kidney." G. Calabresi and D. Malemed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARVARD L. REV. 1089, 1112 (1972).

136 *See* ROBERT FOGEL & S. ENGERMAN, TIME ON THE CROSS: THE ECONOMICS OF AMERICAN NEGRO SLAVERY (1974).

137 If the possession of slaves is regarded as a matter of rights in dispute, the WTA of both parties will also play a role, as I showed earlier.

138 *Uncle Tom's Cabin* was the greatest fiction success of the nineteenth century. ALFRED KAZIN, Introduction to HARRIET BEECHER STOWE, UNCLE TOM'S CABIN *vi* (Bantam Books ed., 1981) (1852). It was not for nothing that President Lincoln greeted Mrs. Stowe as "the little woman who wrote the book that made this great war." *Id.* at ix. Kazin also notes that "Mrs. Stowe had brought to her indictment of slavery . . . a moral passion that in the book is the most powerful antagonist of slavery and one that so worked on people's feelings from 1852 to the end of the Civil War that no other single book can be said to have contributed so much to the end of slavery." *Id.* at viii.

2. Dueling and Economic Efficiency

The social convention of dueling in the antebellum South has been held to have been an efficient norm.[139] However, the offered proof, which consists of pointing out elements of efficiency in the practice, cannot be accepted under either wealth maximization or KHM because Schwartz's definition of efficiency does not take into account moral sentiments (what I have called the regard for others.[140] The problem that Schwartz et al. does not overcome is that dueling, even in the antebellum South, was a contentious norm.[141] Its efficiency cannot therefore be proved without a closer examination of the regard for others; the sentiments of the general population need to be determined through a detailed evaluation.

The evidence suggests the sentiments against dueling were considerable. Schwartz et al. note "one other important feature of the larger social context was that the duel was explicitly made illegal and subjected to severe penalties."[142] The laws were, in fact, carefully designed to eliminate the practice. This contention was not confined to the antebellum South. In England, for example, there was never a time when private dueling was legal, according to Bothwick.[143] From a KHM perspective, what is of interest is why the practice of dueling arose and why it waned.

Trial by combat was a part of the legal system in England only after the time of William the Conqueror.[144] It arose, in large part, in response to widespread perjury, and from reasoning that it was better to risk one's body than one's soul.[145] Possibly, in a more Christian period, it was felt that God gave victory to the right, although, somewhat ironically, the Christian church was actually attempting to abolish dueling. As Gibbon noted, "Is it not true that the event both of national wars and of private combat is directed by the judgment of God? And does not Providence award the victory to the juster cause."[146] From the Crown's point of view, dueling was narrowly efficient in that it seems to have brought more money into the treasury.[147] In England, probably from the time after the reign of Henry I, there was no battle in civil cases unless the property in dispute was worth at

---

139 *See* W. Schwartz, K. Baxter & D. Ryan, *The Duel: Can These Gentlemen be Acting Efficiently?*, 13 J. LEGAL STUD. 321 (1984).

140 *See* ZERBE, *supra* note 3.

141 *See supra* note 138.

142 Schwartz et al., *supra* note 139, at 326.

143 Bothwick notes, "That although, in times of ignorance, our ancestors had recourse to a blind method of trial by duels, yet there never was a period in the annals of Britain, when duels could be lawfully engaged in by private parties." WILLIAM BOTHWICK, REMARKS ON BRITISH ANTIQUITIES 19 (1776).

144 *See* GEORGE NEILSON, TRIAL BY COMBAT (MacMillan 1891).

145 *Id.*

146 EDWARD GIBBON, DECLINE AND FALL OF THE ROMAN EMPIRE 552 (P.F. Collier & Son 1899).

147 *See* NEILSON, *supra* note 144.

least ten shillings.[148] However, in the more "primitive" country of Scotland, parties had recourse to "cold iron" even in disputes concerning the most trivial property.[149] Bothwick noted, "In a rude age, this method of proceeding was exceedingly natural."[150]

A change in sentiments played a role in the decline of dueling.[151] The practice was never universal.[152] It was not practiced by the Greeks or the Egyptians, nor was it part of the Roman codes or the treatises of their jurists. In Europe, from its earliest days, the influence of the Christian church was directed against trial by combat, and seems to have been in the main directed against it during succeeding centuries. Clearly, by the late 18th century the practice was regarded with repugnance. Neilson notes "its roots must be sought in lands inhabited by a people not yet advanced beyond the barbarian stage."[153] There was a steady process of restriction of trial by battle to the writ and the appeal of felony. By 1219, a rigid line had formed around dueling, which it could not pass:

> In burgh after burgh it passed away . . . in the other courts in which it was competent, the judges more and more found reasons and made them, for disallowing a mode of trial in which they could have little faith, and in which the people at large by no means loved . . . . When the century ended, trial by battle was far advanced on the high road to extinction. It had become uncommon before the close of the reign of Henry VI.[154]

In the South after the Civil War, the value of honor probably declined. A similar explanation may apply to England and Scotland. Bothwick notes that "Expressions which go for nothing in the year 1776 would not have gone for nothing in the year 1400. In proportion as honesty is become rare, a sense of personal honour is become less delicate."[155] Thus, without taking into account the change in sentiments, Schwartz cannot prove that dueling was efficient.

---

148 *See id.*

149 BOTHWICK, *supra* note 143.

150 *Id.* at 8 n. 157.

151 A similar experience may be seen in the history of flogging in the British Navy. "Flogging around the fleet disappeared by the mid-1800s, and by 1870 a captain's right to order flogging was severely restricted. In 1879 it was abolished," according to Massie. ROBERT MASSIE, DREADNOUGHT: BRITAIN, GERMANY AND THE COMING OF THE GREAT WAR (1991).

152 *See* NEILSON, *supra* note 144.

153 *Id.* at 3.

154 *Id.* at 72.

155 BOTHWICK, *supra* note 143, at 8.

3. *Plessy v. Ferguson*: A Misapplication of the Common Law Tradition and the Difficulty of Determining Efficiency in the Absence of an Uncontentious Norm

The common law tradition of using social norms to create law is not invariably efficient, and it is particularly likely to be inefficient if the norm is contentious. As Blackstone noted, a norm is an efficient tool in creating law only when the norm is uncontentious. An example is found in the famous case of *Plessy v. Ferguson*.[156] *Plessy* bungled the common law tradition in three ways: (1) it adopted a norm that lacked sufficient support; (2) it adopted a norm when a competing norm existed; and (3) it adopted a norm that ultimately lost out to a competing norm.[157] The *Plessy* Court faced a situation where no uncontentious norm existed.

Plessy upheld a Louisiana statute that provided for "separate but equal" accommodations for white and African-American train passengers and provided for fines and imprisonment of passengers and train employees who refused to comply with the rules.[158] Contrary to popular belief, *Plessy* did *not* require that the facilities for whites and African Americans be equal; it held only that a racially discriminatory law is constitutional if it is "reasonable" in light of the "established usages, customs, and traditions of the people."[159] Because the statute was consistent with Louisiana's "social conventions," the statute was held constitutional.[160] Clearly the *Plessy* decision was norm-seeking.

Justice Harlan argued in dissent that the "reasonableness" of the statute in light of Louisiana's "social conventions" was irrelevant.[161] At first glance, this appears to be a rejection of the common law tradition. If this is so, his dissent would be of little use in determining the efficiency of *Plessy*.[162] However, Justice Harlan's opinion makes it clear that it is not Louisiana's social conventions that are relevant, but those of the United States.[163] Thus, the Fourteenth Amendment renders Louisiana's policies

---

156 Plessy v. Fergusen, 163 U.S. 537 (1896).

157 *Id.*

158 *See id.* at 541, 550-551.

159 *See id.* at 550-551.

160 *See id.* Justice Harlan points out, in dissent, that racial *segregation* was not Louisiana's social convention in any event, because it prevented an African-American servant from waiting on a white patron during the ride, something that Louisiana's social conventions not only allowed but demanded of African Americans. *See id.* at 553. The point is not that a norm of servitude is morally superior to a norm of segregation, but merely that the alleged norm of segregation was not even historically accurate.

161 *See Plessy*, 163 U.S. at 550-51, 557.

162 *Id.*

163 *See id.* at 554. Justice Harlan notes, "[T]he Constitution of the United States does not, I think, permit any public authority to know the race of those entitled to be protected in the enjoyment of [constitutional] rights." *Id.* Saying that a public authority may not "know" a certain fact when making a decision is an apt description of what it means to deny economic standing.

unconstitutional.[164] Using the language of KHM, Justice Harlan argued that Louisiana's custom of segregation should not be considered because the United States had made a reasonable social judgment that the costs of governmental racial discrimination outweigh any benefits the citizens of the state would receive from it.[165] Just as a thief lacks standing to argue that his WTP for stolen goods is higher than his victim's WTA, Louisiana lacked standing to argue that its statute was efficient because of its consistency with Louisiana's norms.[166]

The majority in *Plessy* at least partially recognized the legitimacy of Justice Harlan's argument in that it attempted to formulate a norm that both justified Louisiana's statute and was consistent with the spirit of the Fourteenth Amendment.[167] The majority argued racial integration is only appropriate when it is "voluntary" and a product of "a mutual appreciation of each other's merits."[168] This argument was incoherent, however, because Louisiana's statute provided for fines and imprisonment if a white and an African-American passenger decided to sit together because they had a "mutual appreciation of each other's merits."[169] The majority's incoherence was inevitable, because there *was* no norm justifying Louisiana's statute that was consistent with the Fourteenth Amendment.[170]

Plessy improperly applied a norm that lacked the uncontentious quality required by Blackstone. At the time of *Plessy*, there were competing norms of racial integration and racial segregation and neither norm was sufficiently uncontentious to guarantee efficiency.[171] Furthermore, the norm *Plessy* attempted to establish did not become uncontentious over time.[172] In fact, support for *Plessy*'s norm evaporated, and the Supreme Court to backed away from its holding.[173] In *Missouri Ex Rel Gaines v. Canada*, the Supreme Court held it was unconstitutional for Missouri to provide for a legal education for African Americans by subsidizing their tuition to attend law school in an adjacent state.[174] In *Gaines*, the majority demanded that the privilege of education be extended to all races on an "equal" basis, while the dissent insisted that the question was merely whether the state had made a "reasonable" effort to provide "specialized education" to African

---

164 *See id.*

165 *See id.* at 555.

166 *See* ZERBE, *supra* note 3; COOTER & ULEN, *supra* note 2.

167 *See Plessy*, 163 U.S. at 550-51.

168 *See id.* at 550-51.

169 *See id.* at 551.

170 *See id.* at 551, 557.

171 At a minimum, Harlan's eloquent dissent provides an example of one competing norm. *See id.* at 552-64.

172 *Id.*

173 *See* McLaurin v. Okla. State Regents for Higher Educ., 339 U.S. 637 (1950); Sweatt v. Painter, 339 U.S. 629 (1950); Missouri *ex rel.* Gaines v. Canada, 305 U.S. 337 (1938).

174 *See Gaines*, 305 U.S. at 342, 349-52.

Americans.[175] The dissent's approach was probably more consistent with *Plessy*'s "reasonableness" standard than was the majority's approach, but after *Gaines*, "reasonableness" was not enough.[176]

Sweatt v. Painter involved Texas' attempt to maintain the all-white status of the University of Texas Law School by creating a smaller, adjacent law school for African Americans with many of the same faculty and textbooks.[177] The majority of the Supreme Court held this "separate" school was not "equal."[178] While the *Sweatt* Court could have relied on the tangible inferiority of the African-American law school, it instead focused on the "intangible" factors such as "reputation of the faculty, experience of the administration, position and influence of the alumni, standing in the community, traditions and prestige."[179] While *Sweatt* recognized the theoretical possibility of a separate law school equal to the white one, it is hard to imagine how any non-white school, which in Texas would necessarily be a *new* law school, could have alumni of equal "influence" or comparable community standing and "prestige."[180]

In *McLaurin v. Oklahoma State Regents for Higher Education*, the Supreme Court held that an integrated graduate school violated the Fourteenth Amendment even though it admitted an African American into its department, because it forced him to sit in designated seats in the cafeteria, classrooms, and library.[181] McLaurin's education would have been as "equal" to that received by the white students as any separate education could have been, considering that he would have heard the same lectures from the same professors and studied the same books in the same library.[182] However, the Court recognized that interaction with other students is an essential aspect of education and that McLaurin would be unfairly and unconstitutionally denied this interaction.[183] Any theoretical possibility of a segregated school's passing constitutional muster left open by *Sweatt* was closed by *McLaurin*.[184] If interaction with other students is an essential part of education such that denying an equal opportunity to interact means denying an equal education, then segregation is inherently unconstitutional. The effect and the *purpose* of segregation are to prevent interaction between students of different races. When we consider *Sweatt's* recognition that the "position and influence" of a school's alumni is an essential element of its quality, it becomes clear segregated schools

175 *See id.*, at 349-50, 353.

176 *Compare Plessy*, 163 U.S. at 550-51, *with Gaines*, 305 U.S. at 349, 353.

177 *Sweatt*, 339 U.S. at 632-36.

178 *Id.*

179 *Id.* at 633-34.

180 *Id.*

181 *McLaurin*, 339 U.S. at 638-42.

182 *See id.* at 641.

183 *See id.*

184 *See McLaurin*, 339 U.S. 637; *Sweatt*, 339 U.S. 629.

disadvantaged African-American students.[185] After *McLaurin*, there was only a short conceptual step to *Brown v. The Board of Education*.[186]

Brown formally declared that segregated schools are inherently unequal.[187] *Brown* justified its departure from *Plessy* on the grounds that social conditions had changed.[188] First, *Brown* noted public education was far more important in 1954 than it had been at the time of the Fourteenth Amendment's passage (1868) or even at the time of *Plessy* (1896).[189] Compulsory education, which dramatically increased the importance of high-quality public education, was not adopted by every state until 1918.[190] Second, *Brown* cites a series of psychological studies arguing that segregation harmed the self-esteem of African-American students.[191] Although the validity of those studies has been vigorously attacked,[192] what is more important for my purpose is the implicit recognition that society's willingness to tolerate attacks on the self-esteem of African Americans had changed: in other words, the regard for others had changed. As the regard for others shifted, *Plessy*, which had probably never been efficient, became ever more palpably inefficient. The *Plessy* court responded to the argument that segregation was intended to degrade African Americans with a callous statement that it was only insulting "if the colored race chooses that construction," implicitly stating, "That's your problem: deal with it."[193] *Brown* recognized that regard for others had shifted. That African-American students' self-esteem might suffer from segregation was a loss for both the students and to others, and implied a right to not be exposed to such a loss.

C. *A Change in Technology: With Rapid Change, Custom and Intellectual Knowledge May Be Insufficient to Determine Efficiency*

In this section I offer an additional example in which courts were unable to reach correct outcomes because there was neither custom nor sufficiently developed intellectual knowledge on the issue at bar. The example comes from a Wisconsin case involving restriction to sunlight and

---

185 *See Sweatt*, 339 U.S. at 634.

186 Brown v. Bd. of Educ., 347 U.S. 483 (1954).

187 *Id.* at 493.

188 *See id.* at 492-93.

189 *See id.*

190 *See id.* at 490.

191 *Id.* at 494-95.

192 Indeed, many of the authors of these studies retracted their findings.

193 *See Plessy*, 163 U.S. at 551. Justice Harlan, in contrast, recognized that "everyone knows" that the purpose of the Louisiana statute was to degrade African Americans. *Id.* at 556-57.

shows how difficult it can be for courts to determine what is efficient under changing conditions.[194]

X. SUNLIGHT AND THE LAW OF NUISANCE

Prah v. Maretti involved a dispute about whether the defendant committed a nuisance when he obstructed the plaintiff's access to sunlight.[195] The plaintiff, Prah, had built a system using solar collectors to provide his house with heat and hot water.[196] The defendant, Maretti, then purchased property adjacent to Prah's and began to build a home.[197] Prah argued that Maretti's construction would prevent him from receiving enough sunlight to get adequate use from his solar collectors and that by doing so, Maretti was committing a nuisance. Maretti argued that under Wisconsin law one could not commit a nuisance simply by obstructing his neighbor's access to sunlight.[198] That is, in essence, Maretti argued that Wisconsin denied standing to plaintiffs seeking access to sunlight.[199] Although the Wisconsin Supreme Court recognized that courts had consistently denied standing to such plaintiffs, it decided to overrule that line of cases and granted standing to Prah.[200]

In an era of rapid social change, it is also possible for judges to *overreact* to social change and to change the law even though new social values do not justify a change in the law. Judges may *overestimate* the extent or importance of the social change. *Prah* is an example of a case in which the majority appears to have overreacted to a social change.[201]

Under the law of nuisance, a person cannot *unreasonably* interfere with another person's ability to enjoy his or her property.[202] In a typical nuisance action, the defendant uses his or her land in a way that is

---

194 Another example may be found in Richard Posner's decision in *Lorenzen v. Employees Retirement Plan of the Sperry & Hutchison Co.*, 699 F. Supp. 1367 (E.D. Wis. 1988), *rev'd* 896 F.2d 228 (7th Cir. 1990). *See* ZERBE, *supra* note 3, at 69.

195 *See* Prah v. Maretti, 321 N.W.2d 182, 184 (Wis. 1982).

196 *Id.*

197 *Id.* at 185 (noting that Maretti had received permission to build his home from both the subdivision and the city).

198 *Id.* at 188-189. Before *Prah*, the only way to acquire a protectable interest in sunlight under the common law was to convince one's neighbor to give him an express easement for sunlight. *See id.* An easement is essentially a contract between two landowners which either grants one person the right to use another person's land in some limited way or which prevents a person from using his own land in a particular way. *See id.* In this particular case, Prah and Maretti attempted to negotiate an agreement but were unable to reach a compromise. *See id.* at 185.

199 *Id.* at 184.

200 *Id.* at 189.

201 *Prah*, 321 N.W.2d at 194 (dissenting opinion).

202 *Id.* at 187.

inconvenient or annoying to the plaintiff.[203] The plaintiff must first show he or she is not a "hypersensitive" landowner.[204] If the defendant's conduct is only disruptive to the plaintiff because the plaintiff is unusually sensitive or vulnerable, the defendant's conduct is not deemed a nuisance even if the plaintiff is suffering extreme economic losses as a result.[205]

If a court concludes that the plaintiff is not hypersensitive, it goes on to compare the utility of the defendant's conduct with the gravity of the harm to the plaintiff.[206] Under the Restatement (Second) of Torts § 827, the factors to be considered in measuring the gravity of the plaintiff's harm include: (a) the extent of the harm involved; (b) the character of the harm involved;[207] (c) the social value the law attaches to the type of use or enjoyment invaded; (d) the suitability of the particular use or enjoyment invaded to the character of the locality; and (e) the burden on the person harmed of avoiding the harm.[208] The factors considered in measuring the utility of the defendant's conduct include: (a) the social value the law attaches to the primary purpose of the conduct; (b) the suitability of the conduct to the character of the locality; and (c) the impracticability of preventing or avoiding the invasion.[209]

The Restatement's approach to nuisance law is consistent with KHM's approach to benefit-cost analysis. To determine the plaintiff's WTA or WTP, one needs to know both the seriousness of the harm caused by the defendant and also the plaintiff's opportunity cost. Similarly, to determine the defendant's WTA or WTP, one needs to know both the value of his conduct to him and his opportunity cost. The Restatement asks us to consider both the value of the defendant's conduct to the two parties and their opportunity costs allowing one to determine their WTP and WTA.

Furthermore, the Restatement's consideration of whether the plaintiff's use or the defendant's use is more consistent with the neighborhood helps determine the opportunity costs of each, especially

---

203 *Id.*

204 *Id.* at 197 (dissenting opinion).

205 *Id.*

206 *See id.* at 187.

207 By "character of the harm" the Restatement means that nuisances that cause physical damage to a structure on real property are in general more serious than personal discomfort or annoyance. The Restatement's justification for treating physical damage as more important than those that cause personal discomfort is that the former is much easier to prove than the latter. If the Restatement is simply making a prediction that plaintiffs will be successful more often when they produce evidence of physical damage to their property, because physical damage is easier to prove than discomfort, this distinction is sensible. If the Restatement is making a prescription that preventing or reimbursing people for personal discomfort be treated as less important than preventing or reimbursing physical damage, even when the economic injury (WTA or WTP) is the same, that distinction is not sensible. Whether an injury has been proved is a separate issue from the weight that should be given to an injury once it has been proved.

208 RESTATEMENT (SECOND) OF TORTS § 827 (1990).

209 *Id.* at § 828.

when the best way to avoid the injury is for one of the parties to move. It is probably more efficient to make a person change his or her lifestyle to fit the needs of the neighborhood, either by moving or adopting some measure that reduces the harm of the invasion, than to make the whole neighborhood change to fit the needs of one party.

The Restatement also incorporates moral sentiments by considering the social value the law attaches to the plaintiff's use and the defendant's use. If society attaches value to ensuring that the plaintiff wins and the defendant loses, then the regard for others favors giving the right to the plaintiff. Similarly, if society is willing to pay or willing to accept payment to ensure the defendant wins, the regard for others is in favor of the defendant.

If a nuisance is established, the plaintiff is typically entitled to damages for past interference and an injunction against future interference.[210] Injunctions are available to plaintiffs who have suffered irreparable injuries: that is, injuries for which mere damages would be inadequate.[211] Because the law views each parcel of land as unique, an injury to land is often deemed irreparable.[212]

Even if the plaintiff cannot convince the court that the defendant's use is a nuisance, the plaintiff can prevent the harm by convincing the defendant to grant a restrictive easement or a covenant.[213] On the other hand, if something is held a nuisance, the defendant can purchase from the plaintiff the right to continue committing the nuisance. Therefore, the consequence of a court declaring that something is not a nuisance is that the invasion will continue unless the plaintiff's WTP is higher than the defendant's WTA, after transactions costs. On the other hand, the consequence of a court declaring that something is a nuisance is that the invasion will not continue unless the defendant's WTP is higher than the plaintiff's WTA, after transactions costs. When an individual's WTA is higher than the WTP for normal goods, declaring something to be a nuisance increases the odds that the activity will be stopped, but it does not guarantee it.[214]

Interestingly, plaintiffs seeking access to sunlight had been historically denied standing because the law of nuisance recognized three broad social

---

210 *See* Prah v. Maretti, 321 N.W.2d 182, 184 (Wis. 1982).

211 *See* Stokes Cty. Soil Conservation Dist. v. Shelton, 67 N.C. App. 728 (1984).

212 *See* United Church of the Med. Ctr. v. Med. Ctr. Comm'n, 689 F.2d 693, 701 (7th Cir. 1982).

213 *See Prah*, 321 N.W.2d at 188. Easements or covenants are special types of agreements. They are essentially contracts in which a party either agrees to allow another person to use his property, or agrees not to use his own property in a certain manner. The law of property sometimes creates an easement or covenant even when there is no express agreement (such as an easement by prescription).

214 The finding of a nuisance also has a distributive effect, because the plaintiff's wealth is increased at the expense of the defendant. The plaintiff's wealth will increase because either a harmful activity will be prevented, or the plaintiff will receive a sum of money that is at least as valuable to him as preventing the activity.

policies widely accepted in the nineteenth and early twentieth centuries. Satisfying these policies would justify the denial of standing.[215] First, there was a widespread belief that a landowner should be able to put his land to any use he wished so long as he did not cause physical damage to a neighbor.[216] If a plaintiff could stop the defendant from developing his property simply to ensure the plaintiff's access to sunlight, society would feel the defendant was being treated unfairly, and this sense of unfairness would cause society to experience a loss due to the regard for others. Second, sunlight was valued only for its aesthetic qualities to its owner, and it was thought that the owner could acquire equivalent illumination through artificial devices.[217] In other words, society believed that seeking sunlight had a low WTA and WTP, because sunlight was of relatively little value and the opportunity costs of purchasing artificial light was relatively low. Third, society had a significant interest in encouraging property development.[218] The United States was in the middle of a growth period that people almost universally viewed as necessary to its future. That is, economic growth or development as defined by market goods was highly valued relative to non-market amenities. Most believed that American society experienced a significant gain whenever U.S. land was developed and that judges inflicted a loss on society if they recognized a right to sunlight and allowed plaintiffs to prevent development.

The *Prah* court concluded a series of social changes had occurred in the late twentieth century that undermined the three social goals outlined above.[219] First, sunlight had become something more than just an aesthetic luxury; it had become an energy source[220] increasing the value of sunlight to the plaintiff. Furthermore, the opportunity cost of losing sunlight increased, because one could not generate solar energy artificially. Artificial devices can provide illumination, but they cannot generate electrical energy. Second, the value of non-market amenities had grown relative to traditional market growth since the nineteenth century. Today, society is less willing to encourage traditional market growth at the expense of environmental and other amenities.[221] Third, the United States is rapidly depleting its supply of fossil fuels, causing policy makers to focus on developing alternative energy sources.[222] Therefore, allowing the plaintiff to develop solar energy would likely result in a *direct* benefit to American society, because it lessens the burden on fossil fuels. Fourth, American attitudes toward

---

215 *See Prah*, 321 N.W.2d at 189-90.

216 *See id.*

217 *See id.*

218 *See id.*

219 *Id.*

220 *Id.*

221 *Prah*, 321 N.W.2d at 189-90.

222 *Id.*

property owners have changed. Few Americans still believe landowners should have a completely unrestricted right to develop their property.[223] Today, Americans regard a relatively large degree of regulation of land use as reasonable, even when one has not physically injured his neighbor's property. Therefore, regard for others is less likely to favor allowing a defendant to use his property in a way that obstructs sunlight, even if the defendant is not causing physical harm to the plaintiff's property.

In light of those four changes, the court concluded it was no longer reasonable to *assume* it was in America's best interests to deny standing to plaintiffs seeking access to sunlight.[224] Therefore, it is efficient to grant standing to people seeking access to sunlight so a court can determine, on a case-by-case basis, whether a particular plaintiff's need for sunlight is greater than a particular defendant's need for development. The Wisconsin Supreme Court remanded the case, directing the trial court to consider the factors specified in the Restatement to determine whether Maretti's construction actually constituted a nuisance.[225]

Judge Callow argued in dissent that the court should continue to deny standing to plaintiffs seeking access to sunlight.[226] He made three arguments to support his conclusion. First, he argued that the state legislature was in a better position than the courts to determine whether there was a change in the regard for others.[227] He reasoned that courts should continue to deny standing to plaintiffs seeking access to sunlight until the legislature granted this right to plaintiffs. In fact, he noted that the legislature had actually drafted a statute that governed one's right to sunlight and argued that the court was wrong to ignore that statute.[228] Under the statute, one who builds a solar collector can prevent a neighbor from blocking his access to sunlight only if the plaintiff received a solar access permit from the state before the defendant received a permit to build his house from the local subdivision and the city.[229] In this case, Prah apparently never received a permit from Wisconsin, and Prah did not notify Maretti that he built a solar panel until after Maretti had received a building permit from the local subdivision and the city.[230] Therefore, under the recently adopted Wisconsin statute, Prah had no right to prevent Maretti's construction.[231]

Callow's first argument, that the judiciary should let the legislature decide what is in the public interest and what the regard for others favors, is

---

223 *See id.*

224 *Id.*

225 *See id.* at 192.

226 *See id.* at 193-99.

227 *See Prah*, 321 N.W.2d at 195.

228 *See id.* at 195-96.

229 *See id.*

230 *See id.* at 184-85.

231 *See id.* at 197.

a common argument, but it is not self-evidently correct. In this case, however, the legislature enacted a statute that struck a balance between the interests of solar power users and people who wished to develop their property. The legislature's "first in time" approach to the issue of conflicts between solar power users and other landowners is a reasonable one and the majority probably should have at least considered it.[232]

Second, Callow argued that courts should only recognize regard for others in an action alleging a public nuisance, not a private nuisance.[233] Callow's argument is not persuasive, at least under KHM analysis. Regard for others would be *more* opposed to a defendant who committed a public nuisance than it would be to a defendant who committed a private nuisance, but there is no reason to ignore regard for others in private nuisance actions. If regard for others is ignored in private nuisance actions, courts would frequently reach inefficient decisions, under KHM's definition of efficiency, by ignoring a class of sentiments that represent a WTP for regard for others. Such a conclusion would then be not only inefficient but unjust.

Third, Callow argued even if solar power is of greater value today than it was in the nineteenth century, and even if the regard for others favors plaintiffs in Prah's position, it is *still* true that Prah is a *hypersensitive* plaintiff, and thus it is still efficient to deny standing to Prah.[234] Callow notes that whether a plaintiff is hypersensitive is largely a question of relative numbers.[235] A hypersensitive plaintiff is a plaintiff who is bothered by something that most people would not find bothersome.[236] It may be true that Prah is engaging in a socially useful activity by experimenting in solar energy, but the fact remains that most people would not have been bothered by Maretti's construction.[237] Judge Callow analogized the history of the law's treatment to horses and cars.[238] When the car was first invented, it was frequently held to be a nuisance to horses.[239] Many more people owned horses than owned cars, and thus it made sense to require car owners to

---

232 In addition, the legislature's rule might have reduced the transactions costs of both those seeking access to sunlight and those seeking the right to build a home on their property. Whether a person holds a solar access permit is publicly available information. Under the legislature's approach, it would be relatively easy for solar power users to express their need for sunlight, by requesting a permit. Similarly, it would be relatively easy for people who wish to build a home to find out whether any of their neighbors had a permit. By thwarting the legislature's approach, the court makes it more difficult for people like Maretti to find out whether any of their neighbors has a particularly intense need for sunlight.

233 *See Prah*, 321 N.W.2d at 194-95.

234 *See id.* at 196-97.

235 *See id.* at 195.

236 *See id.* at 197.

237 *See id.*

238 *See Prah*, 321 N.W.2d at 195.

239 *See id.*

restructure their lives to reduce the impact on horses rather than requiring horse owners to restructure their lives to suit car owners.[240] Later, when cars became commonplace, the *horse* was held to be a nuisance.[241] Callow suggests solar energy is still in such an early stage of development that solar energy users like Prah are hypersensitive while home-builders like Maretti are behaving reasonably.[242] Callow's third argument is highly persuasive. The majority is almost certainly correct in arguing that society's attitude toward the value of sunlight has changed,[243] but Callow is almost certainly correct in responding that society has not changed enough to justify a new legal rule.[244] Despite the social changes the majority discusses, the vast majority of homeowners in Wisconsin would not be bothered by Maretti's construction because they do not rely on solar power for heat and hot water. Therefore, solar power users like Prah are hypersensitive. While society has changed, society has not changed *enough* to make solar energy use more widespread. Just as the car eventually succeeded the horse, it may be that solar energy will eventually become so valuable that one who desires access to sunlight would not be considered hypersensitive. However, it seems unlikely that Wisconsin had reached that point in 1980 and thus, the majority's decision was premature.[245]

XI. CONCLUSION

The argument presented here suggests that the considerable and long-standing debate about both the predictive and normative roles of economic efficiency on the one hand, and justice on the other, is misplaced. Economic efficiency, properly presented as KHM, is just where justice is seen as legitimate expectations arising from rights and norms. Thus, this article suggests a reason for the efficiency of common law rooted in the search for justice.[246] Efficiency and justice correspond and judges seek justice as a consequence of social norms. For both economists and judges, if their arguments seem incorrect when reformulated with the language of either justice or efficiency, they were originally formulated incorrectly.

---

240 *See id.*

241 *See id.*

242 *See id.*

243 *See id.* at 189.

244 *See Prah*, 321 N.W.2d at 197.

245 *Id.*

246 This view will allow judges and other scholars who wish to rely on economic efficiency to understand that they are also relying on principles of justice. It will allow economists to realize more fully the inherently normative nature of efficiency, and that efficiency needs not be seen as attached to selfish behavior; moral sentiments are also a consistent part of efficiency. Judges and others that wish to use the language of justice can reformulate their arguments to relate to economic efficiency.

Judges find justice and efficiency in custom and intellectual knowledge. As technology, moral sentiments and knowledge change, the law can become incomplete, unfit, and thus, inefficient. Judges seek to improve the efficiency of the law by incorporating custom and intellectual knowledge. As conditions change rapidly, however, these may be in short supply and the common law will be less efficient.

My thesis is developed through several examples. One shows efficient judge-made changes in antitrust law through the incorporation of intellectual knowledge. The examples of slavery and dueling show that KHM is a better approach to understanding change than KH. The examples of access to sunlight and racial equality show how difficult it can be to determine either efficiency or justice when there is neither sufficient custom nor knowledge. As Simpson notes, justice and efficiency are a matter of recourse to the conscience of the community.[247] One way to look at the evolution of common law is to note that common law tends toward fairness and justice: efficiency is merely a by-product.[248]

---

247 A.W. BRIAN SIMPSON, CANNIBALISM AND THE COMMON LAW: THE STORY OF THE TRAGIC LAST VOYAGE OF THE MIGNONETTE AND THE STRANGE LEGAL PROCEEDINGS TO WHICH IT GAVE RISE (University of Chicago Press 1984). Simpson's discussion of the Mignonette shipwreck gives us an example where the court would have preferred the destruction of four lives through starvation to the destruction of one life through cannibalism. This is clearly "inefficient" in the sense of the destruction of greater human capital, but it is not necessarily inefficient in a KHM sense. Whether this outcome is inefficient in a KHM sense depends on where rights and sentiments lie. In other words, if the regard for others finds cannibalism sufficiently appalling, it might be more efficient to sacrifice four lives to starvation than to sacrifice one life to cannibalism.

248 *See* Douglas Easterling, *Fair Rules for Siting a High-Level Nuclear Waste Repository*, 11 J. OF POL'Y ANALYSIS & MGMT. 442 (1992) (providing an analysis in which efficiency contributes to fairness).

BOOK REVIEW

Kenneth Dam, *The Law-Growth Nexus: The Rule of Law and Economic Development* (2006).

The industrial world has regularly and recently renewed its commitment to supporting economic development. The United States has identified development as a key pillar of its National Security Strategy.[1] Under the auspices of the United Nations, there has been a push to achieve the Millennium Development Goals by 2015.[2] Major power summits have been devoted to eliciting substantial commitments of aid.[3]

Uncomfortably, though, there is little consensus on how one should achieve this development. In the post-World War II era, the discussion in developmental economics has shifted—sometimes dramatically—on the appropriate role of governments and on the essential measures to foster economic growth. Initially, the key points of contention were economic: what should government's role be in supporting industry? Should nascent industries be sheltered from international competition? What role do regulations on labor, goods, or financial markets play? Would a sound physical infrastructure suffice?[4]

A central point of the exercise has been to develop prescriptions for countries that are ailing. Prescriptions did in fact emerge from the decades of discussion. Perhaps the most notable set has been the "Washington Consensus" that described an assortment of free market policies that countries ought to adopt. However, these prescriptions have not been universally embraced. This stems in part from the experience of countries that adopted them partially or in their entirety and did not find them a

---

[1] White House National Security Council Report, *The National Security Strategy of the United States* (September 17, 2002), Section VII "*Expand the Circle of Development by Opening Societies and Building the Infrastructure of Democracy*," *available at* http://www.whitehouse.gov/nsc/nss.pdf.

[2] Official website of United Nations Millennium Development Goals *available at* http://www.un.org/millenniumgoals/.

[3] Alan Beattie, *Campaigners Divided on Aid Promises for Africa: Anti-Poverty Initiative*, The Financial Times, London Edition (July 9, 2005); *see also* the official website of G8 Gleneagles Summit 2005 *available at* http://www.g8.gov.uk/.

[4] Anne O. Krueger, *Trade Policy and Economic Development: How We Learn*, 87 AM. ECON. REV. 1 (1997).

panacea.[5] Aid programs that sought to provide the raw materials for development met with decidedly mixed results.[6]

That prompted further searching into the factors that explained countries' development success, or lack thereof. One candidate explanation that leapt forward was that countries with better institutions tended to be prosperous. This posed a couple of tricky problems. First, there was the question of endogeneity: did good institutions lead to prosperity or did prosperity let a country afford good institutions? Second, there was the problem that economic researchers were generally less well-versed in the functioning of institutions than in matters of markets and incentives.

Setting the second problem aside, a prominent line of economic research over the last decade found an inspired way to address the first problem: What if one looked not at the present, malleable state of institutional development, but instead considered the origin of countries' legal systems? The researchers found that in matters of finance, a common law background was significantly superior to a civil law background. This much-celebrated research seemed to offer insight into both the importance of institutions as a prerequisite for growth and perhaps the policies one might adopt to strengthen them.[7]

In the present work, *The Law-Growth Nexus*,[8] Kenneth Dam dissents from this view. Dam, an eminent legal scholar and policymaker, finds the "legal origins" scholarship unpersuasive and misleading. While he works assiduously throughout the book to discredit the legal origins findings, one could consider this a mere pretext for an enormously valuable exploration into institutions and economic development. It is a rare example of bridging the interdisciplinary divide between law and economics. Nor is it hard to see why this is so rare; it requires a substantial amount of erudition in a wide array of subjects. Dam brings to bear law, sociology, political science, and history.

---

[5] John Williamson, *What Should the World Bank Think About the Washington Consensus* (August 2000) *available at http://www.worldbank.org/research/journals/wbro/obsaug00/pdf/(6)Williamson.pdf.* It is not obvious that the "tests" to which the prescriptions have been put have been valid ones. Nevertheless, even if one believes the "Washington Consensus" policies to be necessary to achieve prosperity, they are unlikely to prove entirely sufficient in themselves. *See also*, T.N. Srinivasan, *The Washington Consensus a Decade Later: Ideology and the Art and Science of Policy Advice* (August 2000) *available at http://www.worldbank.org/research/journals/wbro/obsaug00/pdf/(7)Srinivasan.pdf.*

[6] *See* WILLIAM EASTERLY, THE ELUSIVE QUEST FOR GROWTH: ECONOMISTS' ADVENTURES AND MISADVENTURES IN THE TROPICS (2000).

[7] The focus is in Raphael La Porta, Florencio López-de-Silanes, Andrei Shleifer, and Robert Vishny, *Law and Finance*, 106 J. OF POLITICAL ECON. 1113 (1998) (though this was one in a series of articles). The literature is surveyed in Ross Levine, *Finance and Growth: Theory and Evidence*, Working Paper 10766, National Bureau of Economic Research (September 2004).

[8] KENNETH W. DAM, THE LAW-GROWTH NEXUS: THE RULE OF LAW AND ECONOMIC DEVELOPMENT (2006).

Perhaps most importantly, though, he accompanies this with a solid understanding of economics. Dam objects not to the simplifications of econometric analysis *per se*, but rather to the way this strand of the literature has done it. He considers whether it is meaningful to think about families of legal origin in general and in the particular areas of law that the "legal origins" work focused upon, and finds that it is meaningful only in certain branches of the law—and not the ones that are relevant for the "legal origins" research.

Dam looks at the way that line of research coded for institutional strength, e.g., how does one rate the strength of countries' protection of private property? Should one look at facets of bankruptcy law? Or would corporate and property law be more appropriate? The "legal origins" authors opt for the former; Dam argues persuasively for the latter.

Then there is the question of whether we have been drawing inferences from a limited and misleading sample of countries. Dam cites evidence that we have.

Lest this review make the book sound like an extended referee report, it is very well-presented and at each juncture it delves into central and challenging questions about the role of institutions in development and the law in general. Which matters more—well-written laws or strong enforcement? How did strong institutions develop in the West? What are the differences between common and civil law? Which branches of the law are most relevant for allowing the sorts of transactions that will support emerging economies? This scholarship will prove essential for researchers who hope to tread a bit more carefully than their predecessors.

The downside of caution, as Professor Dam knows well from his days as a leading government official in the Treasury and State Departments, is that policymakers want to make policies and they tend to be impatient. The book does not offer any easy recipes for pursuing institutional development the right way. This may be part of the point—institutions have been hard to transplant historically and the norms and practices of the adopting country are vitally important.

The closest the book comes to demonstrating its practical implications is a review of developments in the People's Republic of China. Amidst warnings about the remaining challenges, Professor Dam lauds China's gradualist approach to reform and the way in which those reforms paid attention to incentives and existing institutions. He draws a parallel with the centuries it took for the West to develop the legal protections that undergird its economic system.

The difficulty is that growth, communication, and commerce have all accelerated tremendously in the last century. This means that there is increased international public exposure to the scenes of disease and deprivation that accompany underdevelopment, thereby increasing public pressure for action. It also means that the potential repercussions for

maintaining flawed policies are significantly greater, thus making gradualism perilous.

Anyone who hopes to craft swifter remedies, however, will find this book invaluable.

Philip I. Levy[*]

[*] Philip Levy studies international trade and development at the American Enterprise Institute. Prior to joining AEI, he was a senior economist for trade on the President's Council of Economic Advisers and a member of the Policy Planning Staff at the State Department. An economist by training, Mr. Levy has experience in many international trade and development policy issues, including free trade agreements, trade with China, antidumping policy, welfare effects of globalization, and U.S. foreign assistance and economic development policy. Mr. Levy has served as Academic Director of the Yale Center for the Study of Globalization and was a faculty member in the Yale Department of Economics. Mr. Levy holds a Ph.D. in Economics from Stanford University.