The Use of Economics for Understanding Law:
An Economist’s View of the Cathedral

Thomas J. Miceli
University of Connecticut

Working Paper 2011-25
December 2011

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

This working paper is indexed on RePEc, http://repec.org/
The Use of Economics for Understanding Law:
An Economist’s View of the Cathedral

by

Thomas J. Miceli*

Abstract: This essay offers some observations, from the perspective of an economist, on the usefulness of economics for understanding law. Economic analysis provides a coherent theoretical framework for unifying different areas of law based on the pursuit of efficiency. It does this by recognizing common problems across different areas, which give rise to solutions that, while outwardly different, have the same underlying form. In this way, economics provides a theory of law. But economists can also learn a lot about how the economy functions by thinking more carefully about the role of law in facilitating economic activity. The success of law and economics ultimately resides in the recognition of this fundamental interrelationship between the two disciplines.

Key words: Law and economics
JEL codes: K00

December 2011

*Professor, Department of Economics, University of Connecticut, Storrs, CT 06269; Ph: (860) 486-5810; Fax: (860) 486-4463; e-mail: Thomas.miceli@uconn.edu.
The Use of Economics for Understanding Law: 
An Economist’s View of the Cathedral

1. Introduction

The use of economic analysis to understand law is part of the general expansion of economic reasoning to explain behavior outside of the traditional market setting.¹ This trend reflects a recognition that economics is not defined by the subject matter that it is supposed to study (e.g., markets), but that it represents a methodology that is useful in understanding rational decision-making wherever it occurs. Law is an ideal subject for economics in this regard because legal disputes provide a wealth of material for evaluating theories of rational behavior. Indeed, the most creative scholar could not dream up the variety of situations that even a casual examination of such disputes reveals.² In this sense, the case law provides economists “data” with which to test their theories.

Another reason that economics is compatible with law is that both disciplines are concerned with incentives. According to the axiom of rationality, economic agents act to maximize their self-interests (however that is measured) subject to the various constraints that they face. In market settings, these constraints consist of prices and income; in legal settings they consist of laws backed by sanctions. The application of economic analysis to law presumes that individuals view the threat of these sanctions—whether in the form of damages, fines, or prison—as implicit prices for certain kinds of behavior, and that these prices can be set to guide behavior in socially desirable directions. According to this view of law, notions of “legality” and “illegality” are stripped of their moral or ethical connotations and are instead interpreted in the context of efficient and inefficient behavior.

¹ Other notable examples include economic analysis of political behavior (Buchanan and Tulluck, 1962; Downs, 1967) and the family (Becker, 1991).
² The best example of this that I have come across is the famous case of Palsgraf v. Long Island R.R. (248 N.Y. 339, 162 N.E. 99, 1928), which illustrates the principle of proximate cause in tort law.
The great American jurist Oliver Wendell Holmes, Jr., in his classic article “The Path of the Law” (Holmes, 1897) described a theory of law that reflects (one might say, anticipates) this perspective. The key figure in Holmes’s theory is the so-called “bad man,” who has much in common with the rational decision-maker in economics. The bad man is not bad in the sense of being immoral; rather, he is a rational calculator who will stretch the law to its limit and will violate it without remorse if the perceived gains exceed the costs. The bad man thus has a strong interest in knowing what the law is and what the consequences are of breaking it. “If you want to know the law and nothing else,” Holmes argues, “you must look at is as a bad man, who cares only for the material consequences which such knowledge enables him to predict” (Holmes, 1897: 459). This “prediction theory” of law is law seen through the eyes of an economist.3

This interpretation of law may strike many as crass; critics of the economic analysis of law object, for example, to the materialization that it necessarily entails (such as the need to attach a dollar value to personal injury or other harms), or the implication that all people are basically amoral. To be sure, many people behave in socially desirable ways out of a sense of moral obligation and would do so even in the absence of a formal law,4 but laws are not designed to restrain angels.5 Some people will only behave like angels for fear of violating the law, and others will knowingly violate it. To examine the incentive effects of law—that is, to examine how it affects behavior at the margin—economists must focus on these people, those for whom it is a binding constraint.

3 Posner (1995: 1-29) therefore likens the law and economics movement to the “pragmatic” approach to law as epitomized by Holmes, which similarly adopts a forward looking, instrumentalist view of the law.
4 In fact, Shavell (2002) explicitly recognizes the substitutability of morals and laws in restraining behavior, and he goes on to discuss the relative costs and benefits of these two approaches to social control.
5 As Madison (2008 [1787-1788]: 257) wrote in Federalist 51, “If men were angels, no government would be necessary.”
Other critics of law and economics argue that it is inappropriate to evaluate law based on economic efficiency. The goal of law should instead be to pursue justice or fairness. Richard Posner’s response to this criticism has been twofold. First, he argues that justice in the sense of distributional equity is a value that most economists also embrace, along with efficiency, as relevant in judging the performance of markets. And although economists may not have any special insights into what the optimal degree of distributional justice is, they can contribute much to its measurement, as well as to how a just distribution can be achieved with the least sacrifice of resources (for example by evaluating the efficiency of various proposed redistributive policies). Second, and more pointedly, Posner argues that an important meaning of justice may in fact be efficiency because “in a world of scarce resources waste should be regarded as immoral” (Posner, 2003: 27).

Kaplow and Shavell (2002) have argued that social welfare, by which they mean the aggregation of some index of the well-being of all members of society, should be the sole basis for evaluating legal policy. Based on this view, notions of fairness should matter, but only in so far as they affect people’s well-being (meaning that people attach some value to fairness). By this logic, narrow concepts of efficiency, like pure wealth maximization (or Kaldor-Hicks efficiency) are inappropriate precisely because they exclude factors like fairness that people value.

It is nevertheless the case that most law and economic analysis focuses on wealth maximization as the metric of efficiency. Although changes in legal rules generally affect the distribution of income, in many cases it will be quite difficult to ascertain these effects. And even when the effects can be traced, tinkering with legal rules will not generally be the best means of achieving a more equitable distribution of wealth, or will entail a trade-off between
fairness and efficiency. For example, the assignment of liability for product-related damages will likely affect both the distribution of wealth between businesses and consumers, but also their incentives to avoid accidents. Because there are alternative mechanisms for redistributing wealth that have no relation to products liability (such as the income tax and welfare systems), it makes sense to use the tort system solely for creating incentives to minimize accident costs.⁶

In my own research on law and economics, I have generally followed this wealth-maximization approach, which I believe represents mainstream of law and economic analysis as reflected by the majority of articles published by the major journals in the field.⁷ The fact that I followed this path was not inevitable, however, because my introduction to the field came in an undergraduate law and economics course that was taught in a decidedly non-mainstream way.⁸ Ironically, the primary effect of this course was to deepen my understanding of economics because for the first time I saw economic reasoning applied to non-market behavior, thus offering me a glimpse of its power for explaining social phenomena outside of the conventional setting. This inspired me to pursue graduate training in economics (which was in urban economics and did not include any law and economics courses), after which I returned to law and economics as my primary field of research. At that time, I viewed law as just another subject area to which I could fruitfully apply my newly minted neo-classical economic tools. And in the late 1980s there were still lots of interesting questions that an applied theorist could tackle with little or no formal legal training. All that was required was some ad hoc research.

---

⁶ See the discussion in Shavell (2004: Chapter 28).
⁷ For a good comparative discussion on the different schools of thought in law and economics, see Mercuro and Medema (1997).
⁸ I took this course several years before the first edition of Cooter and Ulen (1988) was published, but I know that had the book been available, my professor, Richard Adelstein, would not have used it. (I know this because he still teaches the course and does not assign any textbook, opting instead for selected readings.)
In the remainder of this essay, I will offer my view, based on more than two decades of research in this area, on the usefulness of economics for understanding law. In doing so, I will divide the discussion into two parts, the first of which I will title “economic analysis of law” and the second “law and economics.” Readers will note that these represent the two labels that have been most commonly used to describe the field, but they also aptly reflect the evolution of my own thinking, as a trained economist interested in law, about the nature of the relationship between the two disciplines. Throughout the discussion, I will offer several examples to illustrate the type of analysis that I believe exemplifies each approach.

2. Economic Analysis of Law

By economic analysis of law I mean the use of economic tools and methods to evaluate legal rules. This is essentially Posner’s program as first outlined in his classic treatise, although as he acknowledges, the roots of economic analysis of law go back at least to Bentham (Posner, 1995: 437). This type of analysis comes in two varieties: positive analysis and normative analysis.

Positive analysis uses economic methods to examine the law as it is. It thus addresses questions about how the law affects behavior, but it often goes further to argue that the various doctrines of the common law reflect an underlying economic logic. In other words, it maintains that the goal of efficiency has somehow become embedded in the law. This is not a claim that judges and juries consciously pursue efficiency, but rather that the common law has market-like tendencies that cause it to evolve in that direction.

---

Normative analysis, in contrast, asks how the law can be improved to better achieve the goal of efficiency. Of course, this type of analysis relies on the assumption that efficiency is a norm that the law should pursue, and that rules should be changed when it is demonstrated that more efficient ones exist. As suggested above, many critics of law and economics focus on the insufficiency of efficiency as a norm for evaluating law (though most would concede that it is not irrelevant).

Mainstream economic analysis of law encompasses both positive and normative analysis, as I will illustrate below. I will first describe the positive analysis as it has been applied to tort law, but then extend the logic derived from study of that area to illustrate the ability of economic theory to unify disparate areas of law. I will then use the economic theory of criminal law to illustrate normative analysis.

2.1. The economic analysis of tort law

The general proposition that tort liability can serve as a means of internalizing harmful externalities is, by itself, not especially insightful in the sense that it is just a version of Pigovian taxation that is administered privately through the liability system rather than publicly by a regulatory authority. The real insights, I would argue, arose from the efforts of Posner (1972) and Brown (1973) to develop an economic theory of negligence law.

Although negligence is the principal standard of civil liability for accidents in the U.S., it only emerged as a separate component of tort law during the early to mid-nineteenth century. Prior to that time, claims for negligence were limited to certain “public” callings, such as common carriers, innkeepers, blacksmiths, or surgeons, who were contractually obligated to provide proper service, the breach of which subjected them to liability; or to sheriffs, who were held negligent when they allowed debtors to escape custody. In none of these areas, however,

---

did negligence carry the modern meaning of a failure to take due care—rather, it connoted “nonfeasance;” that is, failure to perform a pre-existing, usually contractual, duty. The broader realm of accident law—so-called accidents between “strangers”—was a relatively unimportant field of law, with few accidental injuries coming before courts. And in those that did, there was “considerable uncertainty” over whether they were governed by strict liability or some form of negligence (Landes and Posner, 1987: 2-3). The rise of the modern concept of negligence, meaning the failure to take due care (i.e., to be “at fault”), coincided with the onset of the industrial revolution and the consequent explosion of accidents caused by machinery and railroads.

Posner (1972) was the first to propose an explicitly economic theory of negligence law based on the ordinary care standard famously advanced by Judge Learned Hand in the case of U.S. v. Carroll Towing Co.\(^\text{11}\) According to the so-called “Hand test,” the determination of negligence is based on three factors: the probability that an accident will occur \(P\), the damages, or loss, in the event of an accident \(L\), and the burden of precautions necessary to prevent the accident \(B\). If the burden is less than the expected damages, or if \(B<PL\), then a defendant who failed to take the necessary precautions should be held negligent, whereas if the reverse is true, or if \(B>PL\), then the defendant should not be held negligent. As Posner argued, this standard, properly interpreted, creates exactly the right incentives for injurers to invest in efficient accident avoidance. In particular, the \(PL\) must be interpreted as the expected marginal reduction in accident costs if the precaution in question is taken, while \(B\) is the marginal cost of care.

To illustrate, suppose than an injurer can avoid an accident by spending $100 on some safety measure (for example, regular brake inspections), whereas if he does not invest, there is a one in four chance of an accident that would cause damages of $500 to the victim. Is it efficient

\(^{11}\) 159 F.2d 169 (2d. Cir. 1947).
for the injurer to invest in the safety measure? The answer is yes because, by spending $100, the
injurer can prevent an expected cost of $125 (\(=.25\times$500\)). If, in contrast, the damage from an
accident is only, say, $300, but everything else is the same, then it would not be efficient for the
injurer to invest the $100 because by doing so the injurer would only save $75 (\(=.25\times$300\)) in
expected costs.

To see how the Hand test induces efficient behavior in this setting, suppose that
precaution is efficient in the above sense. Since \(B<PL\) in this case, the court will find an injurer
negligent if he failed to invest in the safety measure and an accident occurs. The expected
liability for the injurer at the time he must make his decision is therefore $125. The injurer
knows, however, that by spending the $100 on accident prevention, he can avoid any future
liability. Thus, he will clearly invest in precaution because by spending $100 he avoids an
expected cost of $125. Suppose alternatively that accident prevention is not efficient (i.e., the
loss from an accident is only $300). Since \(B>PL\) in this case, the court will not find the injurer
negligent if he failed to take care and an accident occurs. He therefore has no incentive to invest
the $100 in accident prevention because he gets no benefit from doing so. Again, he makes the
efficient choice. Brown (1973) used a game-theoretic model to show that these conclusions
generalize to settings involving continuous care.

It is important to emphasize that a negligence rule creates efficient incentives for injurers
regarding accident prevention because it creates a threshold of behavior—namely, due care—
that shields injurers from liability if they meet that standard. Injurers thus have a strong
incentive to do so. Strict liability also induces injurers to invest in efficient accident prevention,
but it does so by imposing on them the full liability for any accident losses that they cause. (This
is the sense in which it resembles a Pigovian tax.) Injurers therefore cannot escape liability by
taking efficient (ordinary) care, but they still have an incentive to be careful so as to minimize their overall exposure to accident costs (prevention plus liability). The importance of this distinction in the way negligence and strict liability work becomes apparent when we extend the simple accident model to allow victims as well as injurers to take precaution (i.e., to make it a “bilateral” rather than “unilateral” care model).

It should be obvious that in many accident settings, victim care is at least as important as injurer care in reducing accident risk. For example, drivers should stop and look both ways at railroad crossings, and consumers of dangerous products should only use them for their intended purposes. In these bilateral care contexts, imposing all liability on injurers is not generally efficient because then victims would have little or no incentive to be careful. In the language of economics, strict liability creates a moral hazard problem for victims. In contrast, the negligence rule potentially creates incentives for both injurers and victims to invest in efficient accident prevention.\(^\text{12}\)

The reason for this striking result can be illustrated as follows. As demonstrated above, the due care standard creates a strong incentive for injurers to meet the standard in order to avoid liability, which implies that any accident losses will remain on the victim. And since victims rationally anticipate that this will be the case, they too have an incentive to invest in efficient accident prevention so as to minimize their expected losses. The negligence rule thus induces efficient prevention by both injurers and victims because it combines the two methods for creating efficient incentives—namely, it sets a threshold level of care so that the injurer can

\(^{12}\) Brown (1973) was apparently the first to demonstrate this result. Also see Shavell (2004: 182-193).
avoid liability by meeting the threshold (as shown above), and it simultaneously imposes full damages on the victim, thus eliminating moral hazard.\textsuperscript{13}

The preceding argument illustrates the superiority of negligence over strict liability in bilateral accident settings, but the logic of the argument has implications beyond tort law. In particular, it reveals the general usefulness of what I will call “threshold rules” for creating bilateral incentives. As noted, the problem with a rule of strict liability was that the awarding of compensation to victims eliminates any incentives for them to avoid the accident because they (theoretically) expect to be made whole.\textsuperscript{14} This problem is a consequence of the dual function of liability to deter dangerous activities as well as to compensate victims of those activities. Deterrence and compensation are perfectly compatible when only injurers can take precaution, but when victims can also take precaution, the compensatory and incentive functions come in conflict. Cooter (1985) has called this the “paradox of compensation.” The problem is the legal requirement that any amount of money that the injurer is required to pay in liability must be passed on to the victim as compensation.\textsuperscript{15} The brilliance of the negligence rule is that it resolves this paradox by establishing a standard of behavior that the injurer can meet to avoid liability, which then leaves the victim to bear her own damages, thereby simultaneously giving her the

\textsuperscript{13} Negligence law often supplements the due standard for injurers with a corresponding standard for victims, referred to as contributory negligence. (See Davies v. Mann, 11 East 60 (K.B. 1809). Under contributory negligence, failure of victims to meet the due standard of care prevents them from recovering damages regardless of the injurer’s care level. The logic of this rule is the same as that for the Hand test, and it similarly creates a powerful incentive for victims to invest in efficient accident prevention. In terms of efficiency, however, creation of a due standard for victims is redundant since, as we have just seen, the “simple” negligence rule which only establishes a due standard for injurers is adequate to create efficient bilateral incentives. Landes and Posner (1987: 76) offer an efficiency argument for contributory negligence.

\textsuperscript{14} A symmetric argument applies to a rule of “no liability,” which creates incentives for victims to take care but not injurers. The bilateral care model thus reveals the symmetry between strict and no liability, as epitomized by the “least cost avoider” approach to liability, as well as the superiority of negligence.

\textsuperscript{15} This is why some economists have advocated “decoupled” liability rules (see, for example, Polinsky and Che (1991)).
correct incentives to invest in precaution.\textsuperscript{16} Beginning with Brown (1973),\textsuperscript{17} economists have formally shown that this approach works perfectly to create efficient bilateral incentives if the “due standard” is set at the injurer’s cost-minimizing level of care, as described in the example above.

Grady (1988: 16) identifies the key feature of negligence law that allows it to coordinate the behavior of injurers and victims in this way: “if each party is placed in a position where its conduct is judged on the assumption that the other side has taken appropriate precautions, then it would act in the socially appropriate manner.” Brown and subsequent analysts used game theory to formalize this idea to show that efficient care by both injurers and victims, each acting in ignorance of the other’s behavior, emerges as the Nash equilibrium under a properly designed negligence rule. Specifically, efficient care is the “best response” of each party under the presumption that the other party is also taking efficient care. What’s more, the law of negligence actually requires parties who find themselves in risky situations to hold such presumptions; that is, to assume that other parties with whom they may come in contact are taking ordinary care (Keeton, et al., 1984: 240). In this way the law facilitates the attainment of the efficient equilibrium by compelling parties to hold Nash conjectures about what others are doing.\textsuperscript{18}

2.2. Extending the logic to other areas of law: the model of precaution

Cooter (1985) revealed the power of economics to provide a unifying framework for law when he used Brown’s accident model—what he called the “model of precaution”—to demonstrate an underlying unity in torts, contracts, and property law. The key insight was

\textsuperscript{16} A rule of strict liability coupled with a contributory negligence defense maintains bilateral incentives but also results in compensation of victims, which shows that deterrence and compensation are not necessarily incompatible.
\textsuperscript{17} Also see Landes and Posner (1987) and Shavell (1987).
\textsuperscript{18} Grady (1988) goes on to discuss what happens when parties can observe what others are doing before choosing their own actions. In the case of these “sequential accidents,” strategic behavior becomes possible (Shavell, 1983; Wittman, 1981). Still, Grady argues that the law has developed fairly sophisticated methods for dealing with this problem.
Cooter’s recognition that most legal conflicts, regardless of the subject matter, can be conceptualized as “bilateral care” problems in the sense that they involve a dispute between two parties (a plaintiff and defendant), each possessing the ability to take some action to avoid or mitigate the resulting “damages.” Cooter illustrated this approach by showing that the basic accident model that Brown had used to evaluate the efficiency of various liability rules in tort law can also be used to examine rules applicable to breach of contract cases and property disputes.

In contract law, for example, Cooter modeled the breach of a contractual promise as an “accident,” where the “injurer” is the breaching party (the promisor), and the “victim” is the party to whom the promise was made (the promisee). The promisor’s “care” in this setting is the decision of whether (or when) to breach the contract, and the promisee’s “care” is how much to invest in reliance on performance, which determines the amount of her loss in the event of breach. In this context, one can show that awarding victims full damages for breach (expectation damages) is like strict liability in torts in the sense that it creates efficient incentives for promisors to avoid breach, but inadequate incentives for promisees to mitigate damages (i.e., they overrely on performance). Cooter showed, however, that a rule that limits damages to the amount that promisors could have reasonably foreseen will, like negligence, create efficient incentives for both parties. Although such a rule does not literally establish a threshold for the promisor, it employs a similar mathematical device to create efficient incentives; namely, limiting the promisor’s liability to the level of damages that would result from efficient promisee reliance.19

---

Cooter’s application of the model of precaution to property law concerns government takings of private property, or eminent domain. In this context, the injurer’s care corresponds to the government’s taking decision, while the victim’s care represents the landowner’s initial investment in improvements to the land, which determines the value of the lost property in the event of a taking. Miceli and Segerson (1994, 1996) applied this model to the law of regulatory takings by showing that efficient decisions by both regulators and landowners can be achieved by a conditional (threshold) compensation rule that requires the government to pay compensation only when it inefficiently regulates a piece of land. They then argued that such a rule goes a long way toward explaining actual practice in an area of law that has been regarded by many commentators as muddled or “incoherent” (Rubenfeld, 1993: 1080).

The preceding examples illustrate the usefulness of economic theory for providing a unifying framework for understanding disparate areas of law. In each of the examples, the problem was to design a rule for coordinating the behavior of parties engaged in an activity that would potentially entangle them in a legal dispute. Although the actual rules that have arisen in these different areas vary in their apparent form, Cooter’s analysis shows that they often embody the same inherent structure. In this sense, the law has converged on similar solutions (though often in outwardly different forms) to a common set of economic problems.20

This conclusion naturally raises the question of where this underlying unity in the law came from. This is a central question in the positive economic theory of law, which asserts that the common law displays an inherent economic logic. Originally, Posner attributed this underlying logic to the decision-making of judges, who he claimed actively promote efficiency

---

20 In addition to Cooter (1985), see Posner (2003), Shavell (2004) and Miceli (2009) for textbook treatments of law and economics that emphasize these unifying principles. Interestingly, there is a corresponding concept in evolutionary biology, referred to as “convergent evolution,” which describes the independent emergence of certain adaptive forms or traits (Gould, 1985, : 411-412).
because they “cannot do much … to alter the slices of the pie that various groups in society receive, [so] they might as well concentrate on increasing its size” (Posner, 2003, p. 252). More recent theories, however, have attempted to show that efficiency can emerge without the conscious help of judges based on the argument that the common law, like the competitive market, is driven by invisible-hand type forces that propel it in the direction of efficiency. A large literature has arisen to evaluate the merits of this argument.\textsuperscript{21}

2.3. \textit{Normative analysis: the economics of crime}

I will conclude this section by providing an illustration of normative analysis based on the economic theory of law enforcement as first formalized by Becker (1968) and elaborated on by Polinsky and Shavell (2000, 2007), among others.\textsuperscript{22} (I will therefore refer to this as the BPS model.) Although many of the prescriptions of that model are in fact compatible with actual law enforcement practices (representing further examples of positive analysis), I will highlight here a couple of divergences.

The economic approach to crime control is based on the assumption of rational offenders who decide whether or not to commit illegal acts by comparing the dollar value of the gains from the act in question to the expected cost, where the latter is computed as the magnitude of the legal sanction multiplied by the probability of apprehension and conviction. Although this “rational offender” assumption strikes many as far-fetched or inappropriate, those same doubters, when asked what the social function of criminal punishment is, will usually offer deterrence as one of the justifications (if not the primary one).\textsuperscript{23} But of course, this makes the economist’s

\textsuperscript{21} This literature began with the seminal papers by Rubin (1977) and Priest (1977). See also Gennaioli and Shleifer (2007), and Miceli (2011).
\textsuperscript{22} Polinsky and Shavell (2000: 45) note that as early as the eighteenth century, writers like Beccaria and Bentham used economic ideas to discuss optimal law enforcement. Becker, however, was the first to use formal analysis to systematically examine optimal enforcement.
\textsuperscript{23} My undergraduate law and economics students provide the basis for this observation.
point, for without rational calculation on the part of at least some would-be offenders, deterrence is an empty concept. Given the goal of deterrence, the optimal law enforcement policy prescribed by the BPS model involves choosing the probability of apprehension, and the type and magnitude of criminal sanction (fine and/or prison), to minimize the overall cost of crime.\textsuperscript{24}

The assertion of efficiency as the norm for deriving the optimal enforcement policy is what makes the analysis largely normative. Although minimizing the cost of operating the system is surely an important goal, actual policies often diverge from what the BPS model prescribes, suggesting that other goals are also at play. I will give two examples. First, the BPS model says that the optimal sanction, whether a fine or imprisonment, should be maximal. The reason this is true for a fine is clear: since it is costly to raise the probability of apprehension but not the fine,\textsuperscript{25} then given that offenders act based on the expected fine, it is cost-minimizing to raise the fine as much as possible (up to the offender’s wealth) before raising the probability. (The reason that the optimal prison term should also be maximal is less obvious, but relies on the same logic.) Actual punishment policies clearly do not implement this prescription since fines and prison terms are not maximal for most crimes.

Second, the model predicts that imprisonment should only be used after fines have been used to the maximum extent possible, and only then if additional deterrence is cost-justified. Again, the logic is clear: fines are costless to increase while prison is costly, so it is optimal to exhaust the costless form of punishment first. Actual punishment policies also fail to implement this prescription, as offenders are often sentenced to prison when they could have afforded a substantially higher fine. (Think of white collar criminals.)

\textsuperscript{24} It is a somewhat debated point in the literature as to whether the gains to offenders should count in the social calculus. Most would concede that it should in some cases (e.g., double parking) but not in others (e.g., murder or rape). Economic models generally do count it, while acknowledging the debate. See, for example, Polinsky and Shavell (2000: 48). Friedman (2000: 229-231), Lewin and Trumbull (1990), and Stigler (1970).

\textsuperscript{25} Extensions of the basic model, however, suggest that raising the fine is costly for various reasons.
The reason for both of these divergences almost certainly reflects the importance of norms besides efficiency, such as fairness or equal protection, which society deems relevant for the determination of optimal criminal punishment. For example, the idea of sentencing poor defendants to lengthier prison terms than wealthy defendants for the same crime would strike many as unacceptable (if not unconstitutional) because it would appear that the wealthy were being allowed to “buy their way out of prison” (Lott, 1987).

3. Law and Economics

I want to shift gears here and discuss a different use of economics for understanding law. This approach, rather than treating law as the subject matter, involves the use of economics to understand how law and markets interact as co-equal social institutions for coordinating human behavior. Thus, the label “law and economics” better describes this type of analysis. Ironically, I arrived at this view of the relationship between law and economics from teaching economics rather than from studying law, because I began to appreciate the idea that markets cannot function in isolation but must rely on the existence of a stable background of legal rules that protect property rights and enforce contracts. Most economists who don’t study law take this background as given in the same way that they take tastes and technology as given. But law and economics scholars have shown that important insights can be gained about the operation of markets by stepping back and examining the interrelationship between markets and law. Indeed, the nascent field of development law and economics, as reflected, for example, by De Soto’s

---

26 See, for example, Harris (1970) and Miceli (1991).
27 In an interesting historical essay, Stigler (1992) notes that Aaron Director, one of the founding editors of *The Journal of Law and Economics*, originally suggested the title *Law or Economics*, reflecting the absence of much overlap between the two disciplines at the time. In light of this suggestion, the essence of the current discussion, as will become clear, might best be described by the combined label “law and/or economics.”
(2000) study of the role of secure property rights in promoting economic growth, exemplifies this perspective.  

3.1. The Coase Theorem

The starting point for such an examination is, of course, the Coase Theorem (Coase, 1960). Coase’s original motivation was to offer a critique of the Pigovian view of externalities, which asserted that some form of government intervention (taxes, fines, or liability) was required to internalize external harm, such as that caused by straying cattle or railroad sparks. Absent such intervention, this view maintained, the “cause” of the harm (the rancher or the railroad) would overinvest in the harmful activity; that is, the rancher would have too many cattle or the railroad would run too many trains. Coase challenged this view by noting that there is no single cause of the harm—the injurer and victim are both necessary for it to be felt (i.e., “two are needed to make an accident” (Grady, 1988: 15)). The designation of one party as the “injurer” in fact represents an implicit awarding to the other party (the “victim”) of the right to be free from any harm that arises from the incompatibility of their activities. Thus, under the Pigovian view, the farmer has the right to be free from crop damage—whether from straying cattle or spewing sparks—and so the rancher/railroad has to be compelled to pay the farmer’s cost. But what if the farmer-victim is in a better position to avoid the harm, say by moving his crops, or by not locating near the railroad or ranch in the first place? In that case, the designation of the rancher/railroad as the injurer may actually create inefficient incentives for avoidance of the harm, as discussed in the context of the bilateral care accident model above.

Coase’s point in raising the causation issue was not necessarily to go down the road of designing efficient liability rules, but rather to evaluate the conditions under which court-imposed liability was needed at all. Suppose, for example, that in the rancher-farmer dispute the

---

28 Other studies in the area include Besley (1995) and Alston, Libecap, and Schneider (1996).
court does not intervene to assign liability to the rancher. Does that necessarily mean that the rancher’s herd will expand inefficiently? The answer, of course, is no, as long as the parties can bargain, because if bargaining is possible, the farmer would be able to bribe the rancher to reduce the herd to the point where the marginal benefit and marginal cost of the last steer are equal. Note that in this case, property rights in straying cattle effectively belong to the rancher, and the farmer has to “purchase” them, which he will do up to the point where the two parties value the last steer equally, which is the point where the herd size is efficient. This is the reverse of what happened under the Pigovian solution, where the farmer is awarded the right to the straying cattle and the rancher has to purchase them by paying the court-imposed damages. In both cases, however, the outcome is efficient. This conclusion—that the initial assignment of property rights does not affect the final distribution of resources, which is efficient—is the Coase Theorem.

The importance of the Coase Theorem, however, lies not its practical relevance, which is limited, but in its highlighting of the role of bargaining costs. When the conditions for the Coase Theorem are satisfied—that is, when bargaining is possible—the assignment of liability for external harms doesn’t affect efficiency because the parties will re-arrange any initial assignment of rights as long as there are gains from trade. In this sense, the law doesn’t matter for efficiency (though it does affect the distribution of wealth).29 When bargaining is not possible, in contrast, the law does matter because the parties will not be able to re-arrange inefficient assignments of rights. In this case, the law must be designed with the goal of efficiency in mind (again, as demonstrated by the above discussion of tort rules). In this way, the Coase Theorem defines the

29 The conclusion that the efficient allocation of resources will be achieved regardless of the initial assignment of legal rights mirrors the First Fundamental Theorem of Welfare Economics, which says that market exchange will be efficient regardless of how property rights are initially assigned. The Coase Theorem thus shows that externalities need not preclude this outcome as long as bargaining costs are low.
efficient scope for legal intervention (Demsetz, 1972). The next section elaborates on this idea in the context of the choice between property rules and liability rules.

3.2. Property rules versus liability rules

The classic paper by Calabresi and Melamed (1972) on the choice between property rules and liability rules is probably second only to Coase’s article in the impact it has had in the field of law and economics. The issue Calabresi and Melamed addressed was the manner in which legal rights, or entitlements, are legally protected. They distinguished between two rules: property rules and liability rules. Under property rules, an entitlement can only be transferred if the holder of the entitlement consents. In other words, a would-be acquirer must purchase the entitlement from the holder in a voluntary exchange. Under liability rules, in contrast, a party seeking to acquire an entitlement can do so without the holder’s consent provided that he or she is willing to pay compensation for the holder’s loss. The transfer therefore occurs non-consensually at a price set by the court.

It should be apparent from these definitions that property rules form the basis for market (voluntary) exchange, while liability rules form the basis for what I will call legal (forced) exchange. Because market exchange is consensual, it ensures a mutual benefit, or the existence of gains from trade from any exchange that is actually completed. The role of the law in such transactions is subsidiary in the sense of preventing theft (unwanted seizures) of entitlements and enforcing contractual exchange of entitlements. In this way, law is complementary to markets in promoting the efficient allocation of resources. In the case of liability rules, on the other hand, the law takes the primary role of forcing an exchange of the entitlement on terms dictated by the

---

30 Also see Kaplow and Shavell (1996) and Polinsky (1980).
31 Calabresi and Melamed also discuss a third rule, called an inalienability rule, which prohibits the exchange of an entitlement under any circumstances, including consensual exchange. Examples include constitutional protections of certain fundamental rights, like speech and religion, as well as laws prohibiting the sale of organs, children, and cultural artifacts.
court. Here, the law is a substitute for market exchange in arranging a transfer of the entitlement because bargaining costs preclude voluntary transfers. The discussion in the previous section of this essay offered several examples of legal exchanges by means of liability rules, including tort cases (in which accident victims seek compensation for non-consensual takings of their rights to be free from damages), breach of contract cases, and government takings or regulation of property under eminent domain.

The distinction between market and legal exchanges raises the question of when one type of transactions is preferred over the other. The assurance of mutual benefit under property rules (market exchange) might suggest that it is the more efficient structure. The problem, of course, is that the need for consent under property rules can sometimes prevent otherwise efficient exchanges from occurring if transaction costs are high. To return to one of Coase’s examples, suppose that farmers situated along a railroad track have the legal right to be free from crop damages caused by sparks, and that right is protected by a property rule. Railroad companies would then have to secure the agreement of all farmers in order to run trains along a given route, a prospect that would likely prevent any trains from ever leaving the station. Under the assumption that running the trains is efficient, this would represent an inefficient outcome.

If the farmers’ rights were instead protected by a liability rule that only required the railroad to compensate farmers for any damages the trains caused, but did not allow the farmers to prevent trains from running, the railroad would internalize the harm and would run the efficient number of trains. The cost of using liability rules, however, is that it places a heavy burden on the court to measure the damages suffered by victims accurately. If it under-estimates the damages, the railroad will run too many trains, and if it over-estimates damages, the railroad
will run too few. (Liability rules also involve the administrative cost of using the legal system, which of course can be very high.)

The choice between property rules and liability rules as just described delineates the scope of markets versus law in achieving an efficient allocation of resources. The choice comes down to a trade-off between the possibility of high transaction costs under property rules, which can prevent some efficient exchange from occurring, and the risk of legal error under liability rules, which can result in either too many or too few exchanges. In low-transaction-cost settings where property rules are expected to function well, they are clearly preferred because they will be most conducive to efficient exchange; this is the realm where markets predominate. The role of the law here is primarily supportive—that is, to protect property rights and enforce contractual exchange. In high-transaction-cost settings where markets are likely to fail, however, legal exchange by liability rules is preferred. And although there is always a risk of error in the court’s determination of damages, the expected gains from forcing exchanges in these circumstances will generally outweigh the cost. In these settings, the law substitutes for markets in promoting the efficient allocation of resources.

3.3. The General Transaction Structure

The preceding discussion, building on the insights of Coase and Calabresi and Melamed, describes what has been called a “General Transaction Structure” for organizing exchange.\(^{32}\) Figure 1, which is adapted from Coleman (1988: 31), neatly summarizes the framework in the context of the rancher-farmer dispute. The rows represent the party to whom the entitlement is initially assigned (the “winner” of the legal dispute), and the columns represent the rule used to protect, or enforce that assignment. Cells I and II represent cases won by the farmer, and cells III and IV represent cases won by the rancher. In the cases won by the farmer, protection by a

\(^{32}\) See, for example, Klevorick (1985) and Coleman (1988: Chapter 2).
property rule (cell I) means that the rancher can only infringe on the farmer’s right to be free from straying cattle by bargaining. The farmer can therefore prevent any infringement by obtaining an injunction. In contrast, when the farmer’s right is protected by a liability rule (cell II), the farmer cannot prevent straying cattle but can only seek compensation for the resulting damages. Note that this latter outcome coincides with the Pigovian view of externalities first critiqued by Coase. As the General Transaction Structure reveals, however, this is only one out of four possible resolutions of the dispute, and it is only desirable when transaction costs preclude bargaining between the rancher and the farmer.

[Figure 1 here]

In cases won by the rancher, protection by a property rule (cell III) means that the rancher is free to let his cattle stray without penalty. As the above discussion revealed, this does not necessarily imply too much crop damage because, as Coase recognized, if the value of lost crops to the farmer exceeds the value of additional steers to the rancher, the farmer can bribe the rancher to cut back his herd, provided that transaction costs are low enough. When the rancher’s right is protected by a liability rule, in contrast (cell IV), the farmer can go to court to obtain a reduction in straying cattle, but must be willing to pay the resulting loss in profits to the rancher. As with cell II, no bargaining between the parties is required here, but the court is charged with calculating the value of additional steers to the rancher.

Notice finally that the four cells in Figure 1 provide courts with discretion over both the distributional effects of legal rules, and the manner in which they can be transferred. The choice of which party receives the initial right (the winner) has distributional implications because it determines the assignment of a valuable entitlement. Thus, regardless of the ultimate ownership of the entitlement, each party would like to be the initial recipient, either because it prevents him
from having to purchase the entitlement if he is the higher valuer, or because he can profit from selling the entitlement if the other party values it more. Either way, the initial recipient is enriched by the court’s assignment decision. As to the enforcement rule, that has efficiency implications because it determines the manner in which any transfers away from the initial assignment can occur, whether by market exchange if transaction costs are low, or by legal exchange if they are high.

Having described the general structure of the above framework, I now offer several specific examples of how it applies to various situations from the law of torts, contracts, and property as a way of illustrating its ability to unify various areas within a single framework.33 As noted, cell II corresponds to the Pigovian approach to externalities, and is best exemplified by the well-known case of *Boomer v. Atlantic Cement Co*.34 That case involved a group of landowners who sought injunctive relief against a large cement plant because of the noise and smoke that it produced. The court, however, denied the injunction and instead awarded damages—in other words, the court awarded the plaintiffs the right to be free from pollution but protected it with a liability rule rather than a property rule. The court’s reasoning was that issuing an injunction (choosing cell I) would have forced the plant to close because high transaction costs would have likely prevented the defendant from successfully entering into negotiations with all of the plaintiffs in order to keep the plant open. (It is the same problem the railroad would have faced if it had to negotiate with multiple farmers.) Based on this logic, the court properly placed the case in cell II.

Breach of contract cases in which the plaintiff (promisee) is awarded money damages also fit into cell II if we recognize that the failure of the breaching party (the promisor) to carry

---

33 This discussion draws on Miceli (2009: 164-165).
out his or her contractual promise imposes costs on the recipient of the promise. The payment of damages thus compensates the promisee for the resulting harm. As argued above, this remedy is justified as a way to induce efficient breaches by the promisor, defined to be breaches for which the cost of performance exceeds the gains. Some writers, however, have criticized this argument on the grounds that post-contract negotiation costs between promisors and promisees should be low because, after all, they have already demonstrated their ability to bargain. Thus, the conditions for property rule protection of the promisee’s right to performance would seem to be present. In this view, the remedy of specific performance, essentially an injunction ordering performance, is preferred to money damages because it would allow the promisee to bargain with the promisor over the terms for non-performance (Ulen, 1984; Friedmann, 1989). In other words, breach of contract cases would be better placed in cell I than cell II.

Cells III and IV involve cases where the rancher prevails. Although the rancher is the “cause” of the harm in a literal sense, remember Coase’s insight that both the rancher and farmer must be present for harm to occur—straying cattle would cause no damage in the absence of crops. Further, suppose that the rancher was there first. If the farmer then shows up, plants crops, and complains about damages, we might have little sympathy for him. And in fact, the common law has a doctrine called “coming to the nuisance” that protects pre-existing land uses against just such nuisance claims (Wittman, 1980). The well-known case of Spur Industries v. Del E. Webb Development Co. provides an example of this doctrine. The case involved a developer who encroached upon a pre-existing cattle feed lot and then sued to have the lot shut down as a nuisance. The court granted the developer’s request, but, invoking the coming-to-the-

---

35 On optimal remedies for breach focusing on damage measures, see Shavell (1980, 1984).
36 Note that this is an example of normative analysis. The actual use of specific performance is usually limited to contracts for unique goods (like land) for which monetary compensation would not be a good substitute for performance.
nuisance doctrine, ordered the developer to pay the feed lot’s re-location costs. The court thus assigned the right in question to the feed lot but protected it with a liability rule (cell IV). The developer therefore had the option to pay the feed lot to move, or to allow it to continue in its current location, depending on which involved a lower cost.

The *Spur* case provides a counterpoint to *Boomer* in the sense that the two courts chose opposite assignments of the right in question, although in both cases it protected the assignment with a liability rule. In *Boomer*, the defendant had to pay damages to the plaintiff to remain in operation (the Pigovian solution), while in *Spur* the plaintiff had to pay damages to the defendant to force the latter to shut down. One could quibble with the *Spur* decision by noting that, in contrast to *Boomer*, there were only two parties involved (the developer and the feed lot owner), so transaction costs should have been low enough for the court to use a property rule.\footnote{For example, Coleman (1988: 31, note 5) makes such an argument. It is worth noting, however, that small numbers do not ensure low transaction costs. Sometimes, bargaining costs between two parties can be quite high if they have no alternative but to bargain with each other. Such a situation is referred to as a bilateral monopoly. Posner (2003: 71), for example, argues that the high transaction costs in the *Boomer* case were due to bilateral monopoly.} In either case, however, the efficient outcome would have been achieved provided that the court set the damages correctly.

A final case illustrates the potential inefficiency of using a liability rule when the court cannot accurately measure damages. The case of *Peevyhouse v. Garland Coal & Mining Co.* involved a breach of contract between a strip mining company and the owners of a large tract of undeveloped land.\footnote{382 P.2d 109, *cert. denied*, 375 U.S. 906 (Okla. 1961).} The contract called for the mining company to restore the land to its original state after completing the mining operation, but the company breached this clause when it learned that the cost of restoration would be $29,000 compared to the mere $300 market value of the restored land. The court allowed the breach and ordered the mining company to pay the
Peevyhouses $300 in damages. The court thus awarded the Peevyhouses the right to performance and protected it with a liability rule (cell II).

Although this appears to be a clear-cut case of efficient breach, given the vastly higher cost of performance compared to the value of performance, the problem is that the market value of the restored land may not have been an accurate measure of the true value of performance to the Peevyhouses. The reason is that market value represents what the highest bidder is willing to pay for a piece of property (as measured by recent sales of comparable properties), not what the owner is willing to accept. The difference between the owner’s true value (the amount he or she would accept in a consensual sale) and market value is sometimes called subjective value, and it is a legitimate economic value. If it weren’t, then an owner’s right to refuse to sell for any reason would be an inefficient impediment to market exchange, thus nullifying the primary economic benefit of property rules.

The Peevyhouse case represents the cost of using liability rules to protect entitlements. Although performance may not have been efficient in this case, use of a property rule (specific performance) would at least have allowed the Peevyhouses to negotiate a buy-out on more favorable terms. And since there were small numbers involved, the transaction costs should not have been prohibitively high. Thus, a reasonable interpretation of this case suggests that it might have been better assigned to cell I than cell II; that is, as argued above, the court should have ordered specific performance rather than money damages as the remedy.

3.4. Criminal law again

I conclude the discussion of the General Transaction Structure by suggesting how it could be extended to incorporate criminal law. Although the economic analysis of crime has long been a prominent sub-field of law and economics, it has been something of a nagging question to
explain why a separate category of “crimes” is needed at all.\textsuperscript{40} Both criminal and tort law, after all, are concerned with the same problem—namely, internalizing unwanted harms—so one might ask why tort law alone is not sufficient. It is not an adequate answer to simply say that crimes are intentional and torts are accidental because the economic model of tort law as outlined above can easily accommodate intent (see, for example, Landes and Posner, 1987: Chapter 6). Economists thus seek a deeper reason.

One interesting perspective suggested by Calabresi and Melamed (1972) is that criminal law is needed to enforce the General Transaction Structure.\textsuperscript{41} As has been argued, that structure is designed to enforce a particular arrangement of legal rights and to facilitate the efficient exchange of those rights, but what if some individuals seek to violate that structure? Consider, for example, a thief who steals someone else’s property. When the thief is caught, why not simply make him pay the victim the value of the stolen property? If the thief values the property more than the owner, this “transaction” would produce an “efficient theft” which seems, at first glance, no different than a court’s imposition of liability on polluters, or requiring the government to pay just compensation for takings. In other words, once we accept the validity of coercive exchanges as efficient in those contexts, must we not be prepared to accept them in all such cases of forced exchange? What makes the act that we call “theft” different from the other involuntary transfers of entitlements that are allowed within the General Transaction Structure? The question turns out to be a difficult one that is sidestepped by the BPS model of crime, which simply takes the existence of the criminal category as given and proceeds to derive the optimal enforcement policy.

\textsuperscript{40} See, for example, Friedman (2000: Chapter 18) and Posner (1983).
\textsuperscript{41} Also see Klevorick (1985), and Coleman (1988: Chapter 6).
One response to this dilemma is to argue that thieves are not caught with certainty, and so damages will undercompensate victims. This problem, however, is easily handled by simply inflating the amount of the penalty by the inverse of the probability of capture. (The economic theory of punitive damages is based on precisely this logic (Cooter, 1982).)

A better answer is to recall the reason why liability rules are not universally allowed as the basis for exchange. Liability rules require courts to estimate the value of the entitlement to the owner after the fact, which means that there is no guarantee that the resulting exchange is efficient. In contrast, property rules ensure that only efficient exchanges occur because the owner has the right to refuse the transaction. Thus, when transaction costs are low, property rules are preferred. But what if a thief violates an owner’s property rule by stealing the protected item? If the legal remedy is simply to require him to pay compensation, then he has succeeded in transforming the property rule into a liability rule. To prevent this, some further sanction (a “kicker”) is needed to discourage the thief from violating the transaction structure. This enhanced sanction represents the criminal penalty.

If the logic of this argument makes sense in the context of property transfers, it becomes even more compelling when applied to bodily harm or other violent acts, where court efforts to estimate the dollar cost to victims are bound to be inadequate (or impossible). In these cases, criminal sanctions are meant to discourage exchanges that should never happen at all—those that are protected by inalienability rules. The specific form of the criminal sanction, whether a fine, imprisonment, or some combination, is not important. What is important is that that the sanction in question is intended as a punishment for an illegitimate transaction, as opposed to compensation for a legitimate transaction (Cooter, 1984). It is this sense of criminal law that represents the common understanding of most people.
The preceding argument is attractive to those who seek a unified economic approach to law because it plausibly explains how criminal laws, viewed as punishments, fit within the same framework used to explain private laws, viewed as a prices. At its base is a coherent view of the role of the law in facilitating efficient economic exchange by channeling transactions into the market when it is expected to function well, and through the legal system when markets fail. At the same time, some will find this view of law as impoverished, or at least incomplete, because it ignores moral dimensions (Coleman, 1988: 165). Most economists would probably agree.

4. Concluding Remarks

The application of economic theory to the law has proven, over the past half century, to be a remarkably successful research program. The chief reason for this success is that economics and law are both concerned with the same essential problem—namely, establishing incentives for guiding human behavior in socially desirable directions. For this reason, economics provides a coherent theory of law based on the application of economic tools and methods founded on the axiom of rational behavior.

Despite this natural affinity between the two disciplines, economists who study law do not necessarily claim that economics offers a complete understanding of the law. What they do believe is that economics can explain broad areas of the law based on the pursuit of efficiency, and that it can suggest how other goals (like justice) can be pursued in more economical ways. At the same time, economists can learn a lot about how the economy functions by thinking more carefully about the role of law in facilitating economic activity. Highlighting this interrelationship represents the greatest success of the economic analysis of law.
References


Figure 1. The General Transaction Structure.

<table>
<thead>
<tr>
<th>Farmer</th>
<th>Property Rule</th>
<th>Enforcement Rule</th>
<th>Liability Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farmer can enjoin ranching</td>
<td>I.</td>
<td>Farmer can seek compensation for lost crops</td>
<td>II.</td>
</tr>
<tr>
<td>Rancher is free to impose harm</td>
<td>III.</td>
<td>Rancher can seek compensation for abating harm</td>
<td>IV.</td>
</tr>
</tbody>
</table>