

Efficient Regulation¹

Andrei Shleifer

January 2010

Abstract

Regulation of economic activity is ubiquitous around the world, yet standard theories predict it should be rather uncommon. I argue that the ubiquity of regulation is explained not so much by the failure of markets, or by asymmetric information, as by the failure of courts to solve contract and tort disputes cheaply, predictably, and impartially. The approach accounts for the ubiquity of regulation, for its growth over time, as well as for the fact that contracts themselves are heavily regulated. It also makes predictions, both across activities and across jurisdictions, for the efficiency of regulation and litigation as strategies of enforcing efficient conduct.

¹ Department of Economics, Harvard University and NBER. I am grateful to Nicola Gennaioli, Oliver Hart, Joshua Schwartzstein, and Holger Spamann for helpful comments.

Ubiquitous Regulation

The American and European societies are much richer today than they were 100 years ago, yet they are also vastly more regulated. Today, we work in jobs extensively regulated by the government, from hiring procedures, to working hours and conditions, to rules for joining unions, to dismissal practices. We live in houses and apartment buildings whose construction – from zoning, to use of materials, to fire codes – is heavily regulated. We eat food grown with approved fertilizers and hormones, processed in regulated factories, and sold in licensed outlets with mandatory labels and warnings. Our cars, buses, and airplanes are made, sold, driven, and maintained under heavy government regulation. Our children attend schools that teach material authorized by the state, visit doctors following regulated procedures, and play on play-grounds certified based on government-mandated safety standards.

Government regulation is extensive in all rich and middle income countries. It transcends not only levels of economic development, but also cultures, legal traditions, levels of democratization, and all other factors economists use to explain differences among countries. There is surely a lot of variation across countries, but it pales by comparison with the raw fact of ubiquity. Why is there so much government regulation?

To a student of traditional Pigouvian (1938) welfare economics, such extensive government regulation makes perfect sense. Markets fail, a Pigouvian would say, because of externalities, asymmetric information, and lack of competition, and governments need to regulate them to counter these failures. Regulation is ubiquitous because market failures are.

This view, however, has lost much ground over the last half century, under relentless intellectual pressure from the law and economics tradition originating with Coase (1960). This tradition holds that competition is merciless in driving firms toward efficiency, that markets

exhibit tremendous ingenuity in dealing with potential failures, that contracts enforced by courts get around most externalities, and that even when for some reason contracts do not take care of all harmful conduct, tort law addresses most of the rest. The space left for efficient regulation is then very limited. From the efficiency perspective, the ubiquity of regulation is puzzling.

In fact, it is even more puzzling than the Coasian logic would suggest. In Coase's view, contracts are a substitute for regulation. If potential externalities can be contracted around, no regulation is necessary. Yet, contrary to this prediction, we see extensive government regulation of contracts themselves. Employment terms are delineated in contracts, yet these contracts are heavily regulated by the government. Purchases of various goods, from homes, to appliances, to stocks, are governed by detailed contracts, yet these contracts too are restricted by government mandates. The regulation of contracts goes much beyond mandatory disclosure, which suggests that asymmetric information is not at the heart of the problem. The fact that contracting itself is so heavily regulated severely undermines both the Pigouvian and the Coasian theories of regulation. The Pigouvian theory is undermined because market failures or information asymmetries do not seem to be necessary for regulation, yet those are seen by the theory as the prerequisites for government intervention. The Coasian position is undermined because free contracts are expected to remedy market failures and eliminate the need for regulation, yet regulation often intervenes in and restricts contracts themselves, including contracts with no third party effects. The puzzle of ubiquitous regulation remains.

These considerations have led many economists to accept the position that regulation is driven not by efficiency but by politics. Under the most prominent version of this theory, proposed by Stigler (1971), industries or other interest groups organize and capture the regulators to raise prices, restrict entry, or otherwise benefit the incumbents. Alternatively, regulation is

just a popular response to an economic crisis, introduced under public pressure whenever market outcomes are seen as undesirable, regardless of whether there are more efficient solutions (Hart 2009). Yet the political theories are not entirely persuasive, as they fail to come to grips with the fairly obvious facts which opened this paper, namely that regulation is ubiquitous in the richest, most democratic countries, with most benign governments, and seems to support the highest quality of life. Extensive regulation seems to be embraced in nearly all corners of these societies, which seems inconsistent with the view that regulation is inefficient.

In this paper, I revisit the case for efficient regulation. My basic point is simple. The case against regulation relies on well-functioning courts. Courts are needed both to enforce contracts and to provide remedy for torts, and hence are central to the basic private mechanisms for curing market failures. In so far as courts resolve disputes cheaply, predictably, and impartially, the efficiency case for regulation is difficult to make in most areas. Efficient regulation would be an exception, not the rule. But when litigation is expensive, unpredictable, or biased, the efficiency case for regulation opens up. Contracts accomplish less when their interpretation is unpredictable and their enforcement is expensive. Liability rules would not cure market failures if compensation of the victims is vulnerable to the vagaries of courts. In short, the case for efficient regulation rests on the failures of courts.

In what follows, I show that this approach explains the ubiquity of regulation, but also its growth over the last century. The approach also helps shed light on the patterns of regulation and litigation across activities, as well as across jurisdictions. I am not suggesting that regulation is universally desirable; regulators often suffer from far deeper problems than courts. The point is that there are tradeoffs between the two. Indeed, if the approach is correct, it suggests that the growth of regulation reflects an efficient institutional adaptation to a more complex world.

Perfect Courts

To fix ideas and to illustrate the arguments, consider the example of workplace safety regulation, an important area of government intervention in markets. Workplace safety is especially informative because the traditional objection to the Coase Theorem, namely that contracting is impractical because many parties are involved (as with pollution), does not apply. Nor is it plausible that asymmetric information between firms and their workers, who are specialists and interact over time, limits contracts (or has third party effects). Indeed, the puzzle of ubiquitous regulation is most dramatic in areas, such as workplace safety, where there are no obvious limitations on or externalities from contracts and tort law is well developed.

The explosion of workplace safety regulation is indeed puzzling from the Coasian perspective. To begin, market forces should work in this area, even with spot labor markets and without complex contracts. Because wages adjust in risky occupations, employers have an economic incentive to control accident risks so as to reduce the wage premium they have to pay. Firms would also want to establish reputations as safe employers to attract better workers, and to pay them less. Competition for labor provides strong incentives to take care of safety.

Extensive contracting opportunities are available as well. Employees, through collective bargaining agreements or even individual employment contracts, can require firms to take safety precautions. Firms can likewise require certain levels of care from their employees by specifying that they follow safety procedures. Private insurance is available to both workers and firms to insure the damages to health and property resulting from accidents. Insurance companies can then demand, as part of the insurance contract, that firms and workers take specific precautions. With knowledgeable firms, knowledgeable workers, and knowledgeable insurance companies, one might think correct incentives could be worked out. Moreover, the

parties interact over time, and are able to learn where the risks are, mitigate them, and adjust their contracts accordingly. It seems compelling, in this context, that private solutions provide parties with correct incentives to take efficient precautions.

Should one of the parties fail to follow the terms of the contract, the other can go to court. Indeed, it can do so even before an accident occurs if contractual terms regarding precautions are violated. After an accident, likewise, the victim can demand in a lawsuit a contractually specified compensation. Courts can then enforce the contracts, by requiring the insurance company or the firm to pay, or alternatively by finding that the worker had not taken contractually agreed upon precautions. No government authority beyond courts is needed.

If the necessary contracts are too elaborate to negotiate up front, insurance companies and industry associations can produce recommendations for safety standards, and contracts can incorporate those. An individual firm, a union, or even a worker bears few incremental costs of figuring out what is appropriate by opting into industry standards. Standardization also reduces the costs of compliance by creating standard safety equipment, standard safety procedures, etc.

Finally, even if contracts do not cover some eventualities, tort law deals with accidents not covered by contracts. Courts develop precedents and guidelines for addressing questions of liability and damages, and can also rely on industry standards for reaching conclusions. As the law develops over time, these precedents and other rules developed by courts cover more and more situations, leaving ever smaller uncertainties. Indeed, as courts complete the law, there will be no need for actual litigation as parties will know what to expect and settle before trial.

With so many protective mechanisms for both workers and firms available through markets and courts, and so many incentives for efficient precautions provided by these mechanisms, why would anyone need regulation?

Once the puzzle is framed in this way, it becomes clear where to look for the answer. Start with the forces of competition on the spot markets, without contracts or insurance. It is probably true that, in the world of well-healed and well-established firms, with access to capital markets and expectation of long run survival, the savings from taking efficient precautions outweigh the immediate costs. But many firms operate in a very different world, in which capital is scarce, downward pressure on prices is relentless, and incentives to cut costs today are strong. In such a competitive world, the firm may face huge pressure to undersupply precautions relative to the efficient level and to accept incremental accident risks. Should an accident happen, the firm might be able to fight its liability in court (see below), settle for a small sum with a desperate victim, or go bankrupt. To hold the firm accountable for causing accidents, there need to be effective courts. The incentive to undersupply precautions is even greater when competitors undersupply them and so face lower short run costs, perhaps because they come from different countries. Competition without courts and contracts does not do much for safety.

If competition does not lead firms to take efficient precautions, it must be contracts, including insurance contracts, as well as tort rules, that do the job. But those fundamentally rely on courts. Suppose for concreteness that the firm and its employees have agreed on a contract that delineates the precautions that need to be taken, and suppose further that the firm has taken out an insurance policy compensating workers who are hurt. After that, an accident happens. Neither the insurance company nor the firm wants to pay the victim, so the victim has to sue for damages. Most accidents occur because of some combination of bad luck and lack of precautions on the parts of both the employer and employee (or, to make it more complex, an employee other than the one who got hurt). Each litigant blames the other, often sincerely. And even if the “true” facts of the case are clear to an omniscient observer, and even if the litigants

know what happened, they each have a story for why it is not their fault, but the other party's. The insurance company likewise has a story for why the particular accident is not covered, or if covered not to the full extent of the damages. A court, or some substitute such as an arbitration board, has to ascertain the facts and interpret the contract or apply the law. The question then becomes how cheaply, predictably, and impartially the court can do so.

Consider this question in steps. Begin with courts as assumed in law and economics. In those courts, 1) verification is relatively straightforward and inexpensive, 2) judges are motivated to exert effort to enforce contracts and laws, 3) judges are knowledgeable enough to verify the facts, and 4) judges are impartial. I argue later that all four of these assumptions are dubious descriptions of reality, and that the failure of each gives rise to a distinct argument for regulation. But for now consider this extreme.

One might think that, under these assumptions, courts could easily verify which contractual terms apply. But even here several issues hamper adjudication and make it uncertain. First, judges do not witness the accident and need to figure out what happened. They can only do so imperfectly. The litigants have different perceptions of what had happened, even if they are honest, and the judge needs to piece the story together. Second, the contract may not cover the exact facts of the dispute: in an accident, both litigants are often at fault. Moreover, language is often unclear, and vulnerable to alternative interpretations. What are best efforts, for example? The litigants then disagree on how the contract allocates the costs of an accident. The judge has to decide what the contract means and how tort law applies.

This leads to a third set of issues, namely that both contractual interpretation and tort liability are governed by multiple conflicting principles, and judges need to pick which ones to apply. Judges reason by analogy to precedents, and a case is often similar to multiple precedents

with conflicting results. Lawyers argue that the precedent favoring their clients is the closest one. Judges then decide. It might be difficult to tell in advance which of the potentially governing precedents the judge will pick, especially when the facts are close to the line.

With factual, contractual, and legal uncertainty, the judge must exercise at least some discretion in resolving a dispute. Aspects of such discretion have been called fact discretion, referring to the judge's flexibility in interpreting facts, and legal discretion, referring to room to maneuver in applying the law to the facts. Pistor and Xu (2005) aptly call this "incomplete law." Posner (2008) refers to this as "open area" uncertainty. Posner recognizes the existence of such uncertainty, but seems to believe that this open area is usually small. I return to this issue below.

Even assuming that judges are unbiased, knowledgeable, and properly motivated, judicial discretion imposes risk on the litigants. The litigants can settle and avoid the risk, but the prospect of such settlement distorts incentives and contracts (Gennaioli 2009). Judicial discretion, which follows from legal, contractual, and factual uncertainty, is an essential feature of litigation, and one from which many consequences follow.

Recent research has begun to uncover systematic evidence of judicial discretion. A large empirical literature discussed by Posner (2008) documents the effect of the judges' political party affiliations on their decisions. Chang and Schoar (2007) use a sample of 5,000 Chapter 11 filings by private companies in the US, and find that in their motion granting practices, some bankruptcy judges are systematically more pro-creditor than others. Niblett (2009) examines interpretation of very standard arbitration clauses in contracts by California appellate courts. He finds that judges make arbitrary distinctions in their contract interpretation even in the simplest of cases, for instance focusing on the size of the print or the location of the arbitration clause in the contract. This evidence is noteworthy because the cases Niblett selects are so similar.

Beyond judicial decision making, there is also the problem of enforcement. A firm, especially a small firm, might not have the money to pay to compensate the employee, might not have bought insurance, and might even go bankrupt. In fact, such a firm might ex ante choose to skimp on precautions and go bankrupt after an accident occurs. This problem, identified by Summers (1983) and Shavell (1984), plagues contract and tort law enforcement. Even without bankruptcy, damage payments for negligence might be high, especially when they are jacked up to compensate for imperfect detection (Becker 1968). Although such penalties may provide strong incentives for firms to take precautions, they may also deter socially useful activity when courts make unavoidable mistakes in assigning liability. Firms reluctant to bear such risks may exit, or not enter in the first place. This aspect of imperfect enforcement leads to inefficiency because it reduces desirable business activity (Schwartzstein and Shleifer 2008).

These aspects of the legal process interact prominently with the effectiveness of competition. A firm facing significant price competition seeks to reduce costs. It knows that, should an accident occur, the trial may take some time and it may wiggle out of paying under either contract or tort. It can also settle with the victim of an accident, who may need money and be less patient. If all goes badly in court, it can go bankrupt and still avoid paying. Facing competitive pressure today, such a firm might take fewer precautions or buy less insurance. When justice is not certain, competition leads firms to economize on worker safety measures.

All these problems arise in even the simplest of circumstances. Their effect is to make contract enforcement expensive and unpredictable, leading workers and firms to bear unnecessary risks. This of course is just the beginning of the story: we need to return to our four assumptions about judges.

What Do Judges Do?

I have argued that judicial work is quite complex, and judicial outcomes uncertain, even in relatively simple circumstances, and with well intentioned, hard-working, and unbiased judges. Reality, of course, is less idyllic. Several of the assumptions I made must be revisited.

The first assumption -- that verification is straightforward and inexpensive -- is typically false. To protect the system from manipulation, legal procedure is itself heavily regulated, burdensome and expensive. Discovery is extensive, invasive, and expensive, including both the collection of records and the examination of witnesses. Judges consider multiple cases at once, so cases drag on for years, consuming resources and postponing compensation of the victim. Djankov, La Porta, Lopez de Silanes, and Shleifer (2003) examine the regulation of legal procedure and efficiency of courts in 109 countries by focusing on the simplest cases: the eviction of a non-paying tenants and the collection of a bounced check. Based on surveys of legal experts in these countries, they find that the judicial procedures governing such litigation are extremely cumbersome and time consuming, yielding highly uncertain payoffs to plaintiffs.

Beyond the slowness and expense of court operations, the facts are often complex. Witnesses lie or shade the truth. Courts must rely on representations by attorneys, who are paid for advocacy just short of deception. When issues are complex, courts rely on experts to interpret contracts and testify as to appropriate precautions, remedies, and damages. These experts as well lie or shade the truth when they are hired by the litigants. In some areas, even with expert advice, it might take a judge an enormous effort to understand liability and damages.

Part of the reason is that disputes are highly idiosyncratic, and consequently so is litigation. There are two sides to most arguments. Legal scholars tend to think that most cases are routine, and the law can be easily applied to established facts, but litigants obviously do not

think so. The court needs to familiarize itself with the details of each case, and assess whether particular conduct crosses the threshold of liability, be it negligence, gross negligence, recklessness, or some other standard. This threshold typically depends on many factual circumstances, including those that are extremely difficult, perhaps impossible, to verify, such as intent and knowledge of the defendant. To sort out these issues, the court often cannot rely on documents, but must instead interview witnesses and decide whom to believe when the evidence is conflicting. The point, again, is that litigation is both expensive and uncertain, which reduces the effectiveness of contract and tort law in providing socially correct incentives.

The second assumption is that judges are motivated to understand the issues of the case. In reality, judges face weak incentives (see Posner 2008). Judges cannot be fired. They do not receive promotions with sufficient likelihood to elicit effort, and promotion need not depend on diligence. Judges are not paid for performance. Some judges are elected, but it may be not diligence but humoring the community that improves their election chances. The weak incentives of judges to work hard are particularly important when the cost of verification is high, as when the facts are complex. And when judges do not bother to verify, litigants bear the risk of judicial error. Firms might fail to take precautions, for example, hoping to confuse the judge.

The third assumption about judges is that they are knowledgeable enough, at least with the assistance of court experts, to get to the bottom of the relevant issues. This is a tall order, especially in the modern world. Judges are trained as lawyers, not safety experts. Their work is fundamentally general: they consider large numbers of cases in multiple areas of law. In a complex case, judges must rely on lawyers and experts. The goal of lawyers and experts, however, is to seek judicial favor not enlightenment. It would often take a rather brilliant judge to get to the bottom of the issue when persuasion takes this form.

The fourth assumption is the most interesting, namely that judges are impartial. Judicial partiality may derive from many sources, including political biases, intrinsic preferences over litigants, incentives such as those coming from re-election, or vulnerability to persuasion by the litigants, appropriate or not. Partiality would not be problematic if judicial discretion were minimal. But when discretion in finding fact, interpreting contract, or applying legal rules is substantial, it can massively amplify the effects of partiality.

Start with judicial preferences over litigants, or over their lawyers. Legal realists such as Frank (1930) thought that those are crucial in shaping the outcomes of trials. These preferences might be over individuals, but also over issues, perhaps because of the political preferences of judges. Some judges sympathize with workers injured in accidents and believe that, absent overwhelming evidence of worker malfeasance, companies should pay. Other judges feel that workers employed in dangerous occupations accept the risks and are put on notice to be extra-careful, so absent overwhelming evidence of company malfeasance, they should not collect. When the facts of the case are uncertain, and judges exercise discretion over which testimony or expert analysis to accept, these biases, even if relatively minor and even unconscious, can translate into substantially biased decisions.

Recent research has begun to unravel, at least theoretically, the crucial interaction between judicial preferences and fact discretion. Gennaioli and Shleifer (2008) argue that the selection of “relevant” facts is the crucial mechanism by which judges satisfy their biases, especially because, as we discuss below, the finding of fact is rarely vulnerable to appeal. Once the facts are found, the application of the law to the facts is typically uncontroversial. Likewise, when contractual terms are uncertain, judges can interpret contracts to favor the party, or the issue, to which they are sympathetic, perhaps by choosing one of the several conflicting

principles of contractual interpretation (Gennaioli 2009). In either case, adjudication is biased. And when the contracting parties do not know the judge's preferences up front, they bear substantial risks which they can mitigate by taking inefficient actions and signing inefficient contracts that protect them from judicial discretion in the first place.

Many legal scholars do not like this kind of argument. Some, like Posner (2008) see judicial biases as relatively minor (except on the highly political Supreme Court) because judges are selected from a relatively uniform population: "The pool from which our judges are chosen is not homogeneous, though neither is it fully representative; it is limited as a practical matter to upper-echelon lawyers, almost all of whom are well-socialized, well behaved, conventionally-minded members of the upper middle class (p. 155)". Posner is surely correct, but it is far from clear how much such selection limits the variety of views. Politicians who appoint judges wish to be confident that judges agree with them on particular issues. Such selection may well bias away from moderation. For example, a politician moderately concerned with worker safety might choose not a judge who is centrist on that issue but rather one who is far left of center, just to be sure. Similarly, lawyers who accept judgeships often give up considerable income as private attorneys. In part, they do so to win the respect of their peers, but in part because, like academics, they have strong beliefs about influencing the world. As with academics, such beliefs are not always conducive to moderation.

Another argument is that review by appellate courts constrains trial judges. If judges were automata applying unambiguous law to unambiguous facts, this argument would compel. But, as I already indicated, much of the time judges are interpreting incomplete contracts in light of uncertain facts, or else applying uncertain law. In such circumstances, the role for appeal is more limited, especially since appellate courts do not review the facts except for "egregious

error.” Fact discretion gives trial judges enormous flexibility. Indeed, appellate review may cause a trial judge to further distort his rendition of facts, so as to render the application of the law to those facts uncontroversial and thus invulnerable to appeal (Gennaioli and Shleifer 2008).

In many jurisdictions, judges are elected, which raises the question of whether this mechanism bolsters impartiality. The electoral process for judges, like that for other local officials, selects individuals whose views are representative of the communities they serve. Berdejo and Yuchtman (2009) examine judicial elections in the State of Washington, and find that judges increase sentences prior to elections, in line with voter preferences for harsher criminal sanctions. More generally, if the community is large and diverse, the median voter is likely to be fairly centrist. On the other hand, when a community is neither large nor diverse, the views of its median voter might be quite biased relative to a broader group. The United States Congress is full of representatives diligently articulating the parochial views of their constituents.

The final source of judicial bias is historically the most important one, and that is judicial vulnerability to persuasion or subversion. Such vulnerability follows from litigants having different access to resources. Defendants in workplace accident cases have access to substantial financial resources, including better lawyers, while the victims might be poor, in part because they are injured. Some adaptations, such as contingency fees for attorneys, ameliorate this problem, but probably only in selected cases. If judges do not correct for the inequality of weapons, litigants with more resources have a substantial advantage in court.

Better resources may take the form of better lawyers and court tactics, of delay when plaintiffs cannot wait and settle for less, but they may also take the form of bribes. Corruption may be only modestly relevant for the US courts today (how one thinks of this depends in part on the distinction between bribes and campaign contributions), but bribing judges was evidently

common in 19th century United States, and is pervasive in large parts of the world today. Corruption helps the richer litigants, and would lead to fewer precautions and excessive injuries.

Judicial bias interacts with the evolution of law over time, and the consequent predictability of the legal rules themselves. It is one of the core beliefs of law and economics that common law is fairly complete and provides unique legal answers to most patterns of case facts. A stronger version of this thesis holds that common law converges to efficient legal rules, i.e, these answers encourage efficient behavior. Yet recent scholarship begins to question the belief in convergence to efficiency, both theoretically and empirically. Plausible models do not suggest that sequential decision making by appellate courts with preferences over the shape of the law brings the law to efficient rules (Gennaioli and Shleifer 2007a,b). From time to time, judges overrule existing precedents because their preferences are different from those of their predecessors. Such overruling undermines convergence. More frequently, judges do not overrule the existing precedents, but rather distinguish cases from precedents, based on possibly material, and sometimes immaterial, facts, with the result that they reach different conclusions, which of course reflect their own preferences, based on these new facts. Such distinguishing can refine and complete the law over time, but need not do so. Even if there is improvement on average, the law need not converge or become predictable (Gennaioli and Shleifer 2007a).

Recent empirical evidence casts doubt on the proposition that common law converges to efficient rules over time, even in relatively simple situations. Niblett, Posner and Shleifer (2009) look at the evolution of the Economic Loss Rule, a well-known common law doctrine limiting tort claims when plaintiffs only suffer financial losses, using the universe of state appellate court decisions in a very homogeneous group of construction disputes. The authors find that, over the last 30 years, different US states treated the Economic Loss Rule in construction disputes very

differently, and have achieved no agreement on the scope of its applicability and exceptions. Nor is there evidence of convergence over time in the acceptance of this rule or of its exceptions. Legal certainty looks like a myth even in standard situations after decades of legal evolution.

Posner (2008), recognizes all these concerns with the exercise of judicial discretion, but in the end appeals to judicial professionalism. “To regard oneself and be regarded by others, especially one’s peers, as a good judge requires conformity to the accepted norms of judging. One cannot be regarded as a good judge if one takes bribes, decides cases by flipping a coin, falls asleep in the courtroom, ignores legal doctrine, cannot make up one’s mind, bases decisions on the personal attractiveness or unattractiveness of litigants or their lawyers, or decides cases on the basis of “politics” (depending on how that slippery word is defined).” (p. 61.) I think that Posner is exactly right. So long as a judge does not take bribes, acts deliberately, refers to precedents and legal principles, renders decisions with only modest delays, does not flirt with lawyers, and is not overtly and exorbitantly political, he will be regarded as professional and remain unchecked. The open area uncertainty referred to by Posner is vast.

The Efficiency Case for Regulation

The implication of this analysis is that the case for efficient regulation is essentially the case for the failure of courts. When courts are expensive, unpredictable, and biased, the public will seek alternatives to dispute resolution in courts. The form this alternative has taken throughout the world is regulation. Indeed, each of the problems with litigation discussed in the previous section gives rise to a separate argument for efficient regulation.

Begin with the point that litigation is idiosyncratic, so the facts relevant to the establishment of liability are costly to verify. Regulation tends to homogenize the requirements

for appropriate conduct by both employees and firms. Such homogenization is often excessively rigid, but it reduces enforcement costs because the items that need to be verified are standardized. Does the factory floor have the required number of fire exits? Is there proper spacing between machines? Are the workers wearing helmets? Even if the determination of violation and of damages is left to courts, regulation may reduce the costs of litigation because it effectively provides a judge with a checklist of items that need to be verified, as opposed to leaving open the scope of issues to be debated to the litigants. By narrowing the range of issues to be debated, regulation may render outcomes, and hence behavior, more predictable.

What about incentives? In contrast to judges, the incentives of regulators can be manipulated by their superiors, or even by legislation. Regulators can be forced to specify precautions, to verify whether they are taken, and to investigate in detail after an accident occurs. Unlike judges, regulators can be asked to go through checklists of items to be verified, and incentivized to follow these rules. This possibility of compelling the regulators to investigate and check, perhaps by rewarding them for finding violations, was one of the crucial New Deal arguments for regulation (see Landis 1938, and Glaeser, Johnson, and Shleifer 2001).

Unlike the generalist judges, regulators also tend to be specialized, and are expected to understand more. Courts, of course, can also be specialized (Posner 2008), but perhaps not to the same extent as regulators. In principle, such specialization lowers the costs of understanding the facts in a given situation, as well as of applying the rules to the facts. Specialization of the regulators is the central efficiency argument in their favor, particularly in the areas, such as finance and the environment, where the issues are enormously complex (Landis 1938).

Finally, one can make a case for regulation as a mechanism for reducing the vulnerability of law enforcers to subversion. Unlike judges, regulators are experts, and hence might be less

vulnerable to persuasion by the skilled but disingenuous litigants. In some situations, because they have limited job security, they may also be less susceptible to corruption than the judges (I am skeptical that this argument is general). Historically, inequality of weapons has been the crucial factor behind the rise of the regulatory state in the United States. The mechanism was democratic politics at the end of the 19th and the beginning of the 20th century. As industrialization changed the economic landscape, the country saw a sharp rise of industrial and railroad injuries. Evidently, workers could not find adequate compensation for these injuries in courts, because companies exercised what many saw as undue influence on judges. As muckraking journalists exposed the problem, it became a political issue in several presidential campaigns, including those of Theodore Roosevelt and Woodrow Wilson. Regulation became a central feature of Wilson's *New Freedom* program. Glaeser and Shleifer (2003) summarize these historical developments, and argue that the rise of regulation was indeed a political – and efficient – response to the failure of courts to adjust to economic changes in the country.

Regulation can take a variety of forms (at the extreme, the government can take ownership of firms if it believes nothing short of complete control can get around the consequence of market or contractual failure – think about the ownership of Air Force One). In some instances, the government can lay down the rules for required precautions, conduct inspections, and impose penalties both up front for failure to comply before an accident occurs, and after an accident happens if the failure to comply is recognized only then. In other instances, the government can lay down the regulations, such as disclosure rules and procedures for dealing with conflicts of interest, but then leave the enforcement to private action in court, as in the case of many financial regulations. The purpose of such regulations, very much in the spirit of the present argument, is to reduce the costs of litigation, so that both courts and litigants know more

precisely what constitutes liability. A particular version of this approach is the regulation of contracts, which makes perfect sense as a strategy of facilitating enforcement by courts when judges exercise discretion, but not if all contracts are interpreted equally predictably.

This, in sum, is the argument. I should stress that the analysis is not in any way intended as an endorsement of all regulation and of its expansion. At the level of implementation, all the complaints leveled at judges apply to regulators as well. Enforcement effort, expertise, and absence of bias in the public sector can all be fairly questioned. Regulators are public sector employees, and as such often lack incentives for hard work, know less than they ought to, exhibit policy preferences inconsistent with efficiency, and are vulnerable to subversion by those they regulate. Academic studies and news stories are replete with accounts of regulatory failures.

With respect to the creation of rules, there are even deeper concerns about regulators than about judges. After all, judges are supposed to be relatively impartial, and legal rules evolve slowly over time, which makes them less vulnerable to improper political influence than regulations (Ponzetto and Fernandez 2008). Regulators, in contrast, might pursue highly political agenda, and create regulations that further the incumbent government's goals, or create opportunities for bribe taking by officials. Djankov, La Porta, Lopez de Silanes, and Shleier (2002) examine the rules for entry regulation by new firms in 85 countries. They find little evidence that these rules further efficiency, but more evidence that they are correlated with poor government performance and corruption. Perhaps more importantly, as Stigler (1971) argued, regulators might be captured by the industry to a much greater extent than judges possibly can, since judges do not have long term relationships with firms. The regulators' behavior may end up considerably more biased against consumers than that of the judges.

The choice between regulators and courts, then, is one between imperfect alternatives, in which the virtues and failings of each must be compared. But this in no way detracts from my basic point: the case for efficient regulation rests on that against efficient courts. And historical trends in the best-governed countries suggest that this efficiency case often wins the day.

Institutional Choices

The comparative perspective on regulation and litigation yields a range of empirical predictions. Some of these turn on the comparative efficiency of the two approaches to enforcing efficient conducts. Other predictions focus on institutional choices shaped by considerations other than efficiency, such as politics, history, and culture.

To begin, the analysis may shed some light on the choice between courts and regulators for a given activity in a country. Courts appear to be particularly appropriate in relatively non-technical yet idiosyncratic situations, such as the interpretation of individual contracts (even if some aspects of these contracts are restricted by regulation) or the determination of liability in torts or fault in crimes. In these situations, flexibility is of great value *ex ante*, and application of reasonably broad standards is of value *ex post*. Such situations are difficult to homogenize through regulation, and indeed there are typically addressed by courts (see Posner 2009).

On the other hand, when similar problems recur often enough that repeated utilization of courts is too expensive or unpredictable, regulation might be a socially cheaper alternative. This would be so if the regulator is the ultimate decision maker, but even if in the end the judge must decide, regulation can delineate the issues that must be addressed. It might be more efficient for the legislature to specify the rules than for courts to sort out the threshold of liability in distinct situations. This argument makes the strong prediction that regulation should be more efficient in

the more common situations. Mulligan and Shleifer (2005) test this prediction. They find, in cross-sections of both US states and countries, that higher populations are associated with more extensive regulation. They argue that the regulation of a particular area requires a fixed setup cost, which can be amortized over a higher number of disputes that comes with more people. With small populations, litigation, while idiosyncratic, is rare enough that fixed costs are not worth paying. This approach might also explain why we see regulation in areas such as workplace safety, where contracts and torts are readily available: disputes occur often enough that standardized regulation is cheaper, and more predictable, than idiosyncratic litigation.

Regulation would also be more common in situations where facts are complex, and fact finding requires expertise and incentives. As the society develops, this criterion might apply to a growing range of activities. This observation might explain the basic fact of growing regulation over time. It might also explain why we see regulation in financial markets or in complex industrial activities. Indeed, as I mentioned earlier, expertise and motivation of the regulators were the crucial arguments for the expansion of regulation in the US (Landis 1938).

Finally, regulation might be particularly relevant in situations of inequality between the injured plaintiffs and the injurer. The rise of regulation might be intimately tied to specialization and the rise of large corporations as organizational forms. Thus, while courts or similar methods of dispute resolution might work when disputants have comparable resources, they fail when inequality of weapons becomes overwhelming. This, too, might account for the ubiquity of regulation, including the regulation of contracts between parties with different resources, in the modern world. In fact, if we go back to the introductory paragraph, this might be the reason for regulation of so many basic aspects of consumption and employment.

When it comes to a comparison of patterns of social control across countries, many additional considerations come into play (see Djankov et al. 2003). Different societies might have different levels of expertise, and hence comparative advantage, at regulation, or litigation, or perhaps other forms of social control. For example, as Glaeser and Shleifer (2003) have argued, poor countries might experience severe failures of all public administration, including both regulation and litigation. In these countries, free markets might be the best approach, even when market failure is pervasive. In more developed countries, in which the capacity to administer laws and regulations is higher, stronger government intervention, whether through courts or regulators, becomes more attractive.

One crucial determinant of the actual choices is specialization. In a series of papers written with Simeon Djankov, Florencio Lopez de Silanes, and Rafael La Porta (e.g., La Porta et al. 2008), I have argued that countries from common and civil law legal traditions exhibit different regulatory styles. Relatively speaking, common law countries tend to rely on private orderings and courts, while civil law countries, particularly French civil law ones, rely more heavily on regulation. We see these differences empirically across a broad range of activities, from the regulation of product and labor market, to the regulation of legal procedure, to military draft. Such specialization in the forms of social control might be efficient, as each legal tradition perfects its approach, or it might be just a consequence of hysteresis. Whatever the ultimate cause, we see substantial variation the reliance on regulation and litigation across legal traditions.

More recently, Aghion, Algan, Cahuc, and Shleifer (2009) found that another factor shaping a nation's reliance on regulation is trust. High trust appears to be a substitute for regulation. In high trust societies, individuals do not expect to be mistreated by other individuals or firms, and hence support a lower level of restrictions on others in the form of regulation. In

low trust societies, in contrast, individuals do expect to be mistreated by others, and hence support greater restraint of business activity through regulation. Aghion et al. argue further that these approaches to regulation are self-fulfilling: when levels of regulation are low, people choose to act civically because civic behavior opens up more attractive entrepreneurial opportunities, which would otherwise have been limited by regulation.

These aspects of institutional choice, like Stigler's emphasis on politics, are part of a broader picture of institutional evolution. Yet one point remains central in conclusion: efficiency should not be ignored in considering which institutions survive. In the rich countries in particular, the case for efficiency of courts as opposed to regulators is often tenuous.

References

- Aghion, Philippe, Yann Algan, Pierre Cahuc, and Andrei Shleifer. 2009. "Regulation and Distrust." *Quarterly Journal of Economics*, forthcoming.
- Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76(2): 169-217.
- Berdejo, Carlos, and Noam M. Yuchtman. 2009. "Crime, Punishment and Politics: An Analysis of Political Cycles in Criminal Sentencing." Unpublished manuscript, Harvard University.
- Chang, Tom, and Antoinette Schoar. 2007. "Judge Specific Differences in Chapter 11 and Firm Outcomes." Mimeo, MIT.
- Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1-44
- Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1-37.
- _____. 2003. "Courts." *Quarterly Journal of Economics* 118(2): 453-517.
- Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "The New Comparative Economics." *Journal of Comparative Economics* 31: 595-619.
- Frank, Jerome. 1930. *Law and the Modern Mind*. New York, N.Y.: Brentano's.
- Gennaioli, Nicola. 2009. "Contracting in the Shadow of the Law." Unpublished manuscript. Harvard University.
- Gennaioli, Nicola, and Andrei Shleifer. 2007a. "The Evolution of Common Law." *Journal of Political Economy* 115(1):43-68.

- _____. 2007b. "Overruling and the Instability of Law." *Journal of Comparative Economics* 35(2): 309-328.
- _____. 2008. "Judicial Fact Discretion." *Journal of Legal Studies* 37(1): 1-36.
- Glaeser, Edward L., Simon Johnson, and Andrei Shleifer. 2001. "Coase versus the Coasians." *Quarterly Journal of Economics* 116(3): 853-899.
- Glaeser, Edward L., and Andrei Shleifer. 2003. "The Rise of the Regulatory State." *Journal of Economic Literature* 41(2): 401-425.
- Hart, Oliver. 2009. "Regulation and Sarbanes-Oxley." *Journal of Accounting Research* 47(2): 437-445.
- La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46(2):285-332.
- Landis, James M. 1938. *The Administrative Process*. Westport, Conn.: Greenwood Press.
- Mulligan, Casey, and Andrei Shleifer. 2005. "The Extent of the Market and the Supply of Regulation." *Quarterly Journal of Economics* 120(4): 1445-1473.
- Niblett, Anthony. 2009. "Inconsistent Contract Enforcement." Unpublished manuscript. Harvard University.
- Niblett, Anthony, Richard A. Posner, and Andrei Shleifer. 2009. "The Evolution of a Legal Rule." NBER Working Paper No. 13856.
- Pigou, Arthur C. 1938. *The Economics of Welfare*. 4th ed. London: Macmillan and Co.
- Pistor, Katharina, and Chenggang Xu. 2003. "Incomplete Law – A Conceptual and Analytical Framework and its Application to the Evolution of Financial Market Regulation." *Journal of International Law and Politics* 35(4): 931-1013.
- Ponzetto, Giacomo A. M., and Patricio A. Fernandez. 2008. "Case Law versus Statute Law: An Evolutionary Comparison." *Journal of Legal Studies* 37(2): 379-430.

Posner, Richard A. 2008. *How Judges Think*. Cambridge, MA: Harvard University Press.

_____. 2009. "Regulation (Agencies) versus Litigation (Courts): An Analytical Framework."

Unpublished manuscript.

Schwartzstein, Joshua, and Andrei Shleifer. 2008. "Litigation and Regulation." NBER Working

Paper No. 14752.

Shavell, Steven. 1984. "Liability for Harm versus Regulation of Safety." *Journal of Legal*

Studies 13(2): 357-374.

Stigler, George J. 1971. "The Theory of Economic Regulation." *Bell Journal of Economics*

and Management Science 2(1): 3-21.

Summers, John. 1983. "The Case of the Disappearing Defendant: An Economic Analysis."

University of Pennsylvania Law Review 132 (1): 145-185.