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 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.

To a student of traditional Pigouvian (1938) welfare economics, this explosion of government regulation makes perfect sense. Markets fail, a Pigouvian would say, because of externalities and lack of competition, and governments need to regulate them to counter these failures. 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, even before one gets to all the failures of regulation in the real world (Stigler 1971). From the efficiency perspective, the ubiquity of regulation is puzzling.
In this paper, I revisit the case for efficient regulation, but from the law and economics rather than the Pigouvian perspective. My basic point is simple. The law and economics argument against regulation relies heavily on well-functioning courts. 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. And if we think that the ubiquity of regulation, including in areas with extensive contracting and well-developed tort law, reflects efficient institutional adaptation, my argument indirectly points to extreme inefficiency of litigation. Regulation is ubiquitous because courts fail.

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. Indeed, the puzzle about the ubiquity of regulation is most dramatic in areas, such as workplace safety, where there are no obvious limits on contracts and tort law is well developed.

The explosion of workplace safety regulation is indeed puzzling from the law and economics perspective. To begin, market forces should work in this area, even without complex contracting. Wages can and do adjust in risky occupations, and employers have an economic incentive to control accident risks so as to reduce the wage premium. Firms would also want to establish reputations as safe employers to attract better workers. Private insurance is available to both workers and firms to insure the damage from accidents to health and property.
These market pressures apply in the environment of extensive contracting opportunities. 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. Insurance companies can demand that firms and workers take particular precautions as part of the insurance contract. Moreover, the parties interact over time, and are able to learn where the risks are and mitigate them. It seems compelling, in this context, that private solutions provide parties with correct incentives for taking efficient precautions.

Should one of the parties fail to follow the agreement put down in the contract, the aggrieved party can go to court. Indeed, they can do so even before an accident occurs. After an accident, likewise, the victim can insist in a lawsuit on contractually specified compensation. Courts can then enforce the contracts, by forcing the insurance company or the firm to pay, or alternatively by finding that the worker did not take contractually agreed upon precautions. No government authority beyond courts is needed.

Even if necessary contracts are too elaborate to negotiate up front, industry associations can come up with recommendations for safety standards, and contracts can simply rely on those. An individual firm or worker bears few incremental costs of figuring out what is appropriate by opting into industry standards. Such standardization also reduces the costs of compliance, since there will be standard safety equipment, standard safety procedures, etc.

Finally, even if contracts do not cover some eventualities, tort law deals with accidents and questions of liability. Courts develop precedents and guidelines for addressing questions of liability, and can also rely on industry standards for reaching conclusions. Courts also develop rules, such as strict liability in some circumstances, to reduce the costs of adjudication.
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? In fact, I believe the above account fairly summarizes the standard law and economics logic for why hardly anyone would. One is then naturally led to the conclusion that the only reasons we have so much regulation are political.

Once the puzzle is framed in this way, it becomes clear where one looks for the answer. Start with the forces of competition. 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 today relative to the efficient level and to take 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. The incentive to undersupply precautions is all the greater when other firms in the industry undersupply them and so face lower short run costs, perhaps because they come from different countries.

If competition does not lead firms to take efficient precautions, they take excessive risks, and accidents happen. The next stop is the court. 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 quite sincerely. And even if the “true” facts of the case would be clear to an omniscient observer, and even if the litigants knew 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. A court, or some substitute such as an
arbitration board, has to ascertain the facts and apply the contract or the law. The question then
is how cheaply, predictably, and impartially the court can do so.

We consider this question in steps. Begin with courts as assumed in law and economics.
In those courts, 1) verification is straightforward and inexpensive, 2) judges are motivated, 3)
judges are knowledgeable enough to verify the facts, and 4) judges are impartial. Such courts
could in principle easily verify the contractual terms and figure out which ones apply. But even
here several issues hamper the efficiency of adjudication. First, judges do not witness the
accident and need to figure out what happened. The litigants shade the truth, so the judge needs
to decide whom to trust. Second, the contract does 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. The litigants then disagree on how the contract allocates the costs of
an accident. The judge has to decide what the contract means, and if the contract does not cover
what happened, 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
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 pick. Their choices are not predetermined by the facts of the case.

With factual, contractual, and legal uncertainty, the judge must exercise at least some,
and possibly a lot of discretion is resolving a dispute. Such discretion has been described as fact
discretion and legal discretion. Even assuming that judges are unbiased, knowledgeable, and
properly motivated, judicial discretion imposes risk on the litigants. The litigants can settle and avoid this risk, but the prospect of such settlement distorts incentives to take precautions. Judicial discretion, which follows from legal, contractual, and factual uncertainty, is an essential feature of courts, and one from which many consequences follow.

Legal uncertainty deserves a special mention. It is one of the fundamental beliefs of law and economics that common law is fairly complete and provides unique answers to most factual questions. A stronger version of this thesis holds that common law converges to efficient legal rules. Yet recent scholarship begins to question this belief, both theoretically and empirically. At the theoretical level, plausible models do not suggest that sequential decision making by appellate courts brings the law to efficient rules (Gennaioli and Shleifer 2007, 2008). There are plausible reasons the law may improve over time as judges refine it, but no compelling case for convergence. Empirically, even over relatively long periods of time some common law legal rules fail to converge, and substantial dispersion in judicial interpretation of basic rules survives (Niblett, Posner, and Shleifer 2009). As Niblett (2009) shows, judges use discretion to make arbitrary distinctions in contractual interpretation even in the simplest of circumstances. Legal certainty looks like a myth even with settled facts.

Beyond judicial decision making, there is also the problem of enforcement. A firm, particularly 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 may 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 mistakes in assigning liability. Firms reluctant to bear such risks may exit, or not enter in the first place. This aspect of imperfect enforcement also leads to inefficiency (Schwartzstein and Shleifer 2008).

These aspects of the legal process interact prominently with the effectiveness of competition. A firm operating in a competitive environment faces significant pressure on prices and seeks to reduce costs. This firm knows that, should an accident occur, the trial may take some time and it may wiggle out of paying because it has a colorable defense. It can also settle with the victim of an accident, who may need money and be less patient. In the worst case scenario, if everything goes badly, it can go bankrupt, and still avoid paying. Facing competitive pressure today, such a firm might decide to take fewer precautions, to buy less insurance, and to otherwise save money. 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. In principle, regulation might improve matters by delineating precautions that must be taken ex ante by both workers and firms, including buying insurance, by verifying with public resources whether these precautions had been taken, by imposing modest penalties if they had not before an accident occurs, by insisting that the errors be corrected, and by giving guidance to judges ex post should an accident occur as to the necessary ingredients of liability. This of course is just the beginning: we need to return to our four assumptions about judges.
What Do Judges Do?

The first assumption -- that verification is straightforward and inexpensive – is often 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. 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 as 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 litigation is highly idiosyncratic. The court needs to familiarize itself with the details of each case, and assess whether particular conduct crosses the threshold of liability. 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 interview witnesses and decide whom to believe when the evidence is conflicting. These factual questions are often addressed by juries, which add further uncertainty to the outcomes. The point, again, is that litigation is both expensive and uncertain, which reduces the effectiveness of competition and torts in providing incentives.

The second assumption is that judges are motivated to get to the bottom of things. In reality, independent judges face very low incentives (e.g., Posner 2008). Judges cannot be fired.
They do not receive promotions with enough 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 chances. The weak incentives of judges to work hard are particularly important when the cost of verification is high, e.g., when the facts are complex. And when judges do not bother to verify, litigants bear risk. 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. This is one of the most important New Deal arguments for regulation (see Landis 1938, and Glaeser, Johnson, and Shleifer 2001).

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. With a few exceptions, such as bankruptcy, there is little judicial specialization. In a complex case, it would be a miracle if a judge knew the pertinent issues ex ante. The goal of a litigant 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. Regulators, in contrast, are specialized, and hence do not face this problem to nearly the same extent. Specialization of the regulators is the central efficiency argument in their favor (Landis 1938).

The fourth assumption is the most interesting, namely that judges are impartial. Judicial partiality may derive from many causes, including 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. 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 be paid. With uncertainty over facts, judges can exercise their discretion to accept the facts as described by the party they favor, thereby reaching biased decisions (Gennaioli and Shleifer 2008). Likewise, when contractual terms are uncertain, judges can interpret contracts to favor the party, or the issue, they are sympathetic to (Gennaioli 2006). In either case, adjudication is biased. And when the contracting parties do not know the judge’s type ex ante, 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 minor (except on the highly political Supreme Court) because judges are pooled from a homogeneous population of middle of the road identically trained conventional thinkers. This may not be so. 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 not choose 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 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 uncertain contracts in light of uncertain facts, or else applying uncertain law. In such circumstances, the role for appeal is much more limited, especially since appeal does 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 (Gennaioli and Shleifer 2008).

In many jurisdictions, judges are elected, which raises the question of whether this mechanism bolsters impartiality. In all likelihood, the electoral process for judges, like that for other local officials, selects individuals whose views are representative of the communities they serve. If the community is large and diverse, the median voter is likely to be fairly centrist. On the other hand, many communities are neither large nor diverse, and the views of their median voters might be quite eccentric. The United States Congress is full of representatives diligently reflecting the rather bizarre 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. The consequence is insufficient precautions taken by firms, exposing workers to unnecessary risks.

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.

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.

The implication of all this analysis is that the case for efficient regulation is fundamentally 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.
This, in sum, is the argument. I should stress that it is not in any way intended as an endorsement of all regulation and of its expansion. Many of the complaints leveled at judges apply to regulators are well. Regulators are public sector employees, and as such often lack incentives for hard work. The best select themselves into the private sector, so the level of expertise of the remaining regulators might fall dramatically behind that of private market participants. The trial-and-error path of common law, while imperfect, might lead to the more efficient legal rules than the rash mandates of the legislators (Ponzetto and Fernandez 2008).

Perhaps most importantly, as Stigler argued, regulators might be captured by the industry to a much greater extent than judges, who do not have long term relationships with firms. Their behavior then 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 evaluated. But this in no way detracts from my basic point: the case for efficient regulation is one against efficient courts.

Institutional Choices

The preceding discussion does not describe the conditions under which each particular form of social control is appropriate. When would a society prefer markets, private orderings, courts, regulators, or for that matter state ownership? Research on these issues is in its infancy. Yet some initial findings are appearing, both in the narrow context of controlling a particular activity in a country, and in the broader context of institutional choices across countries. Such choices are governed not just by efficiency, but also by politics, history, and culture.

My analysis may shed some light on the choice between courts and regulators for a given activity in a country. When problems recur often enough and are standard enough that repeated
utilization of courts is too expensive, regulation might be a socially cheaper alternative. Mulligan and Shleifer (2005) find in a cross-section of both US states and countries that higher populations are associated with more extensive regulation. This might also explain why we see pervasive regulation in areas such as workplace safety, even though contracts and torts are readily available. Standardized regulation might be cheaper 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 may 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. This, too, might account for the ubiquity of regulation 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). 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. 2009), 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 et al. (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. (2009) 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 never be ignored in considering which institutions survive. In the rich countries in particular, the case for efficiency of courts as opposed to regulators is far from universal.
References


