OXFORD SOCIO-LEGAL STUDIES

RESPONSIVE REGULATION

Transcending the Deregulation Debate

IAN AYRES
JOHN BRAITHWAITE

2

The Benign Big Gun

In this chapter, we show a convergence between a rational choice analysis of what works in securing compliance with regulatory laws and some sociological analyses that reject the assumptions of the rational choice model. The convergence is about the efficacy of tit-for-tat (TFT) enforcement—regulation that is contingently provokable and forgiving. Building on this convergence, we argue that regulatory agencies are often best able to secure compliance when they are benign big guns. That is, regulators will be more able to speak softly when they carry big sticks (and crucially, a hierarchy of lesser sanctions).[1] Paradoxically, the bigger and the more various are the sticks, the greater the success regulators will achieve by speaking softly.

We argue for a minimal sufficiency principle in the deployment of the big and smaller sticks: the more sanctions can be kept in the background, the more regulation can be transacted through moral suasion, the more effective regulation will be. The key conclusions are derived from qualifying the assumptions that the regulatory behavior of firms is rational and unitary. Among the alternative claims advanced are the following:

1. To understand regulation, we need to aggregate firms into industry associations and disaggregate firms into corporate subunits, subunits into individual corporate actors, and individuals into multiple selves. Regulatory agencies advance their objectives in games at each of these levels of aggregation by moves in games at other levels of aggregation.

2. Some corporate actors will only comply with the law if it is economically rational for them to do so; most corporate actors will comply with the law most of the time simply because it is the law; all corporate actors are bundles of contradictory commitments to values about economic rationality, law abidingness, and business responsibility. Business executives have profit-maximizing selves and law-abiding selves, at different moments, in different contexts, the different selves prevail.

3. A strategy based totally on persuasion and self-regulation will be exploited when actors are motivated by economic rationality.

4. A strategy based mostly on punishment will undermine the good will of actors when they are motivated by a sense of responsibility.

5. Punishment is expensive; persuasion is cheap. A strategy based mostly on punishment wastes resources on litigation that would be better spent on monitor-

ing and persuasion. (A highly punitive mining inspectorate will spend more time in court than in mines).

6. A strategy based mostly on punishment fosters an organized business subculture of resistance to regulation wherein methods of legal resistance and counterattack are incorporated into industry socialization (Bardach and Kagan, 1982). Punitive enforcement engenders a game of regulatory cat-and-mouse whereby firms defy the spirit of the law by exploiting loopholes, and the state writes more and more specific rules to cover the loopholes.

This chapter is divided into two parts. In the first part we explore the multiple motivations and objectives of the regulated. We argue that firms and the decision makers within them are actuated by different motives. Our larger argument is that sound public policy must and can speak to the diverse motivations of the regulated public. We suggest that TFT regulatory strategies may work well not only in constraining noncompliance of purely economic actors, but also in engendering the inculcation of trust and civic virtue.

In the second part, we build on these insights to argue that pyramids of increasingly stringent enforcement measures are needed to respond to the diverse objectives of the regulated firms. An enforcement pyramid subjects regulated firms to escalating forms of regulatory intervention if they continually refuse to respond to regulatory demands.

Mixed Motives and Tit-for-Tat

Deterrence Versus Compliance Models of Regulation

The first step on the road to our conclusions is to understand that there is a long history of barren disputation within regulatory agencies, and more recently among scholars of regulation, between those who think that corporations will comply with the law only when confronted with tough sanctions and those who believe that gentle persuasion works in securing business compliance with the law. Reiss (1980) has dubbed these two competing models of regulation "deterrence" versus "compliance." Justice Holmes long ago adopted the deterrence model in arguing that the law needed to be tailored with a mind not toward "good men" (who would look to law as a guide for proper action), but with a mind toward "bad men" (who would try to evade the legal strictures of society).[2] Today most, although by no means all, regulators are in the compliance camp (see the studies cited by Hawkins, 1984: 3; Grabosky and Braithwaite, 1986), whereas most regulation scholars are in the deterrence camp.

Many of the academics, be they conservative economists or radical criminologists, interpret this state of affairs as evidence of how captured the regulators are in their analysis, in contrast to the independence and worldly wise cynicism of their own position. The regulators, on the other hand, are often inclined to construe the disagreement as evidence of how out of touch academics are with how to achieve results in the real world.

Happily, this era of crude polarization of the regulatory enforcement debate between staunch advocates of deterrence and defenders of the compliance model is beginning to pass. Increasingly within both scholarly and regulatory communities there is a feeling that the regulatory agencies that do best at achieving their goals are those that strike some sort of sophisticated balance between the two models. The crucial question has become: When to punish; when to persuade?

The Game Theorist's Answer

The game theory literature stood ready with some answers to this question. It was John Scholz (1984a, b) who made the great contribution of translating this work into the regulatory arena. In the next chapter a more detailed treatment of this contribution lays the foundations for the game theoretic analysis of capture and tripartism therein. For the moment, we just overview its conclusions.

Scholz models regulation as a prisoner's dilemma game wherein the motivation of the firm is to minimize regulatory costs and the motivation of the regulator is to maximize compliance outcomes. He shows that a TFT enforcement strategy will most likely establish mutually beneficial cooperation, under assumptions he believes will be met in many regulatory contexts. TFT means that the regulator refrains from a deterrent response as long as the firm is cooperating; but when the firm yields to the temptation to exploit the cooperative posture of the regulator and cheats on compliance, then the regulator shifts from a cooperative to a deterrent response. Confronted with the matrix of payoffs typical in the enforcement dilemma, the optimal strategy is for both the firm and the regulator to cooperate until the other defects from cooperation. Then the rational player should retaliate (the state to deterrence regulation; the firm to a law evasion strategy). If and only if the retaliation secures a return to cooperation by the other player, then the retaliator should be forgiving, restoring the benefits of mutual cooperation in place of the lower payoffs of mutual defection. Drawing on the work of Axelrod (1984), Scholz contends that in the prisoner's dilemma game TFT has been pitted against other strategies to demonstrate mathematically, experimentally, and through the use of computer-simulation tournaments that TFT will often maximize the payoffs of players.

> As a "nice" strategy (one that does not use deterrence until after the firm defects), TFT gains the full advantage of mutual cooperation with all firms pursuing nice strategies. As a vengeful strategy which retaliates immediately, it gets stuck with the sucker payoff only once against firms that evade in every round. Yet as a forgiving strategy it responds almost immediately if a previous evader begins to comply, thereby restoring the benefit of mutual cooperation rather than the lower payoffs of mutual defection. Furthermore, the simplicity of TFT makes it easily recognized by an opponent. (Scholz, 1984a: 192)

Alternative Motivational Accounts

The TFT policy prescription is for the regulator to try cooperation first. This conclusion is not grounded in any assumption that business people are cooperative in nature; rather, the payoffs in the regulation game make cooperation rational until the

other player defects from cooperation. The motivational account of the firm is of a unitary actor concerned only with maximizing profit.

Braithwaite's empirical work on corporate offending has led him to posit some alternative motivational accounts. Throughout this chapter, data from this work will be used to illustrate these alternative accounts. The fieldwork on the impact of adverse publicity on corporations is summarized in Fisse and Braithwaite (1983), on pharmaceutical companies in Braithwaite (1984), on coal mining companies in Braithwaite (1985), on Australian regulatory agencies in Grabosky and Braithwaite (1986), and a very preliminary report on the work on nursing home regulation is provided in Braithwaite et al. (1990). The data consists of interviews with executives, observation of corporate decision making, and observation of regulatory encounters. We argue that a strong case for TFT can be made from the alternative motivational accounts revealed from these studies. The first stage of our argument is, therefore, that TFT is an unusually robust policy idea because radically divergent accounts of regulation converge on the efficacy of TFT. First, however, we must go to the conclusions from Braithwaite's empirical work.

The first strand of this work is the research with Brent Fisse in *The Impact of Publicity on Corporate Offenders*. Interviews with executives of large corporations that had been through adverse publicity crises concerning allegations of corporate wrongdoing showed that both individual executives and the corporation collectively generally valued a good reputation for its own sake. There was some concern that adverse corporate publicity might do serious damage to profits, but neither this subjective concern nor the objective fact of economic damage to the corporation from adverse publicity was widespread. Nevertheless, the informants cared deeply about the adverse publicity; they viewed both their personal reputation in the community and their corporate reputation as priceless assets.

The implications of this first empirical questioning of the pure economic rationality model of regulatory encounters do not seem too profound. If individual and corporate actors are deterred not only by economic losses but also by reputational losses, then consideration can be given to adverse publicity sanctions for regulatory offenders. These are precisely the policy solutions considered by Fisse and Braithwaite (1983). TFT can operate perfectly well with adverse publicity supplying the punishment payoff.

Other data suggest, however, that we need to delve deeper into the limitations of a rational calculus (see Yeager [1990] on deontological reasoning in business organizations). Corporate actors are not just value maximizers—of profits or of reputation. They are also often concerned to do what is right, to be faithful to their identity as a law abiding citizen, and to sustain a self-concept of social responsibility. During Braithwaite's fieldwork, business informants repeatedly argued that the common characterization of them as motivated only by money was a simplistic stereotype. Conceding that their primary motivations were economic, they claimed that they and their colleagues took seriously business responsibility, ethics, and obligations to abide by the law and to be responsive to nonshareholding stakeholders in the corporation. This was true even of respondents who admitted widespread law breaking in their company or their industry.

Some economists will read all such claims as humbug, and there are many

instances where Braithwaite would not want to disagree with such a reading of his fieldwork notes. The problem is, however, that to discount every incident of nonutilitarian reasoning as a delusion or a smokescreen to conceal some deeper rational pursuit of interests is to be tautological about the proposition that human beings always pursue their economic interests or their reputational interests, or whatever conception of interest is advanced as self-evident. It is one thing to analyze these fieldwork notes and observe that there is other evidence that this actor is not sincere in a belief about doing what is right whatever the cost. It is another to insist that business executives are motivated only by money when they say otherwise, and when there is evidence of economically irrational compliance with the law. Such insistence will build a science on foundations immune from empirical refutation. We should not scoff at a top pharmaceutical executive who says that concern for improving human health motivates her and her staff to maintain high standards more than the fear of regulatory sanctions, where this is a company that can be observed to maintain fairly high standards in a Third World country that effectively has no pharmaceuticals regulation.

Nor should we scoff at the Director of Nursing at a large nursing home who said to one of Braithwaite's co-fieldworkers: "You become family to them. You are their mother, their sister." The reason the fieldworkers took this account seriously is that they talked to residents who agreed with the assessment; they observed the Director of Nursing interacting lovingly with residents; and inspectors reported that they too had always observed this.

There is another path this debate can take. Smith says one reason he pays the taxes he owes the government is that he believes it is right to do so. The reader accepts that Smith is sincere in this belief. Furthermore Smith asserts that it is economically irrational for him not to seize a variety of opportunities to cheat that he chooses not to exploit. For example, Smith, an Australian citizen, says he travels to the United States every year, with the American Bar Foundation meeting his expenses. If he were to present his airline ticket and other receipts to the Australian Taxation Office as tax deductible research expenses, he could collect payment twice with little risk of detection. Yes, the reader accepts that Smith is both sincere in his commitment to being an honest taxpayer and is also irrational not to seize certain opportunities to be dishonest that are known to him.[3]

The cynic says, however, if we go back far enough into Smith's life, we will understand the psychic rewards he has been given for honesty. We might be able to see the praise that his mother heaped on him when he returned the money he had taken from her purse, and a learning theory analysis of such incidents would enable us to understand why Smith will continue to derive satisfaction from being honest for the rest of his life. Ultimately, the cynic alleges, Smith only chooses honesty because of his self-interested approach to reaping those psychic rewards on his mother's knee?[4] Smith concedes the cynic may be right in this diagnosis. Even though he may be right such speculative reduction to ultimate interest pursuit makes for an unilluminating social science. If we expand backward with self-interest as an explanation until it absorbs everything, including even altruism, then it signifies nothing—it lacks explanatory specificity or power.

What matters in understanding what Smith does today is the motivational

structure that is in place in Smith today. If Smith's motivation is substantially about rejecting the appeal of self-interested reasoning in favor of deontological reasoning,[5] seeking to change his behavior by strengthening the appeal to this self-interest will not work. It might even insult him, provoking him to dig in deeper with his principles. This empirical fact about the here and now is quite unaffected by the other fact—that ultimately it was self-interested motivations in Smith's past that explain his contemporary spurning of economic self-interest. To do good social science, we are best to be informed by empirical evidence about what motivates actors at a particular moment in history so that we can understand the choices those actors make at that moment. If Smith is more strongly motivated by honesty than by money in a particular context, then in that context appeals to honesty are more likely to move him than opportunities for more money, or indeed appeals to what a self-interested little brat he was when on his mother's knee.

So we think it important to know empirically that in the context of regulatory choices, business executives, with varying degrees of apparent sincerity of commitment to action, explain their motivation in the discourse of social responsibility. The variants of that elevation of social responsibility above economic interest are many. It can be the responsibility to obey the law whatever the cost, the responsibility to scientific integrity for pharmaceutical industry scientists, the responsibility to patients for nurses working for a nursing home corporation, the responsibility to professional ethics for company lawyers, or the responsibility a coal mining executive says he has to never put profit ahead of the lives of his workers. There is something about the context of social encounters concerning choices to comply with regulatory laws that makes social responsibility discourse of one sort or another rather prominent, at least in the diverse regulatory arenas that Braithwaite has studied.

Quite often in this research, the rhetoric about putting social responsibility ahead of profits is not matched by responsible action. But, also quite often, the nursing home manager, for example, will be observed to do what she sees as responsible even when she knows that it is costly and when the legal risks from not doing it are perceived as zero, and in fact *are* near zero. Go out with nursing home inspectors in a jurisdiction that never prosecutes, never takes legal action for noncompliance with a standard, and you may be surprised at how frequently profit-making organizations agree to do costly things to comply with the law. When you ask them why, they say: "because it is the law" or "because I agree with the lady from the Health Authority when she says that this is in the best interests of the residents."

So Braithwaite concluded in *To Punish or Persuade* that you could not develop a sound regulatory enforcement policy unless you understood the fact that sometimes business actors were powerfully motivated by making money and sometimes they were powerfully motivated by a sense of social responsibility. He, therefore, rejected a regulatory strategy based totally on persuasion and a strategy based totally on punishment. He concluded that business actors exploit a strategy of persuasion and self-regulation when they are motivated by economic rationality. But a strategy based mostly on punishment will undermine the good will of actors when they are motivated by a sense of responsibility. This will be true of any version of responsibility that is construed by actors as a more noble calling than making money. When actors see

themselves as pursuing a higher calling, to treat them as driven by what they see as baser motivation insults them, demotivates them:

> ADMINISTRATOR: We don't do anything right. The industry is like a bad girl. . . . Why can't you put down something we do right. . . . People don't go into this industry unless they have some compassion. . . . Nurses are not machines. . . . They get burnt out by inspectors criticizing them. . . . All these psychologists with books on praise—from Dr. Spock up—praise children and they do the right thing. But no one thinks to apply that to us.
>
> * * *
>
> INSPECTOR TO ADMINISTRATOR: Your staff were real nice to work with.
>
> ADMINISTRATOR: So were you. Not like Public Aid [another regulatory agency]. They're so demoralizing.
>
> BRAITHWAITE: What's the difference?
>
> ADMINISTRATOR: Public Aid nurses always assume the worst. They treat you as if you are doing something wrong. . . .

A crucial danger of a punitive posture that projects negative expectations of the regulated actor is that it inhibits self-regulation. This is not something peculiar to business regulatory encounters. Lansky (1984) makes the same point about the dangers of treating violence in patients as an eruption that must be held down by regulation of movement, physical or chemical restraint. A model of "holding-down" both inhibits dialogue about the interpersonal vulnerabilities that lead to the violence and justifies "a type of overregulation that humiliates the patient and complicates the return of self-regulation" (Lansky, 1984: 23). When punishment rather than dialogue is in the foreground of regulatory encounters, it is basic to human psychology that people will find this humiliating, will resent and resist in ways that include abandoning self-regulation. The point is not new; it is made in a passage from Confucius that every educated Chinese used to know by heart: "If people be led by laws and uniformity is sought to be given them by punishments, they will try to avoid punishments but have no sense of shame" (Tay, 1990: 160–61).

Individual rebellion against being stigmatized as controllable only by punishment is aggregated within business communities into collective forms of resistance. Bardach and Kagan (1982) identify one of the problems of a mostly punitive policy is that it fosters an organized business subculture of resistance to regulation—a subculture that facilitates the sharing of knowledge about methods of legal resistance and counterattack. When regulators wade in with a punitive model of human beings as essentially bad, they dissipate the will of well-intentioned actors to comply when they treat them as if they are ill intentioned. The problem with the persuasion model, however, based as it is on a typification of people as basically good—reasonable, of good faith, motivated to abide by the law—is that it fails to recognize that there are some who are not good, and who will take advantage of being presumed to be so.

To reject punitive regulation is naive; to be totally committed to it is to lead a charge of the Light Brigade. The trick of successful regulation is to establish a synergy between punishment and persuasion. Strategic punishment underwrites regulatory persuasion as something that ought to be attended to. Persuasion legiti-

mates punishment as reasonable, fair, and even something that might elicit remorse or repentance.

We reject Holmes' notion that the law should solely adopt a "bad man" perspective. Going in with punishment as a strategy of first choice is counterproductive in a number of ways. First, punishment is expensive; persuasion is cheap. Therefore, if persuasion is tried first and works, more resources are left to expand regulatory coverage. In contrast, a mining inspectorate with a first preference for punitive enforcement will spend more time in court than in mines. Second, punitive enforcement engenders a game of regulatory cat-and-mouse whereby firms defy the spirit of the law by exploiting loopholes, and the state writes more and more specific rules to cover the loopholes. The result can be: (1) rule making by accretion that gives no coherence to the rules as a package, and (2) a barren legalism concentrating on specific, simple, visible violations to the neglect of underlying systemic problems (Bardach and Kagan, 1982; Braithwaite, 1985). Third, heavy reliance must be placed on persuasion rather than on punishment in industries where technological and environmental realities change so quickly that the regulations that give detailed content to the law cannot keep up to date.

Given these problems of punitive enforcement, and given that large numbers of corporate actors in many contexts do fit the responsible citizen model, *To Punish or Persuade* argued that persuasion is preferable to punishment as the strategy of first choice. To adopt punishment as a strategy of first choice is unaffordable, unworkable, and counterproductive in undermining the good will of those with a commitment to compliance. However, when firms which are not responsible corporate citizens exploit the privilege of persuasion, the regulator should switch to a tough punitive response.

By a very different route from the economic rationality calculus in the work of Scholz and Axelrod, *To Punish or Persuade* came to essentially the same conclusion—that TFT was the best strategy. Regulator defection to a punitive strategy with recalcitrant companies, or companies that defied the spirit of the law by exploiting loopholes, would underwrite the authority of the regulator through victories in court. It would also support the sense of fairness of responsible companies who eschewed the temptations of regulatory cat and mouse. Preserving that perception of fairness is important to nurturing voluntary compliance. Chester Bowles (1971) concluded from his experience with the U.S. Office of Price Administration during World War II that 20 percent of all firms would comply unconditionally with any rule, 5 percent would attempt to evade it, and the remaining 75 percent are also likely to comply, but only if the punitive threat to the dishonest 5 percent is credible (Bardach and Kagan, 1982: 65–66; see also Levi, 1987).

TFT is the best strategy for Scholz because, in maximizing the difference between the punishment payoff and the cooperation payoff, it makes cooperation the most economically rational response. TFT is the best strategy in *To Punish or Persuade* because it holds out the best hope of nurturing the noneconomic motivations of firms to be responsible and law abiding. Paradoxically, diametrically opposed motivational accounts of business can converge on the same enforcement prescription.

TFT resolves the contradictions of punishment versus persuasion outlined earlier. By cooperating with firms until they cheat, regulators avert the counterproductivity of

undermining the good faith of socially responsible actors. By getting tough with cheaters, actors are made to suffer when they are motivated by money alone; they are given reason to favor their socially responsible, law-abiding selves over their venal selves. In short, they are given reason to reform, more so because when they do reform they find the regulator forgiving. When they put reforms in place, they find that the forgiving regulator treats them as if their socially responsible self was always their "real" self. For Scholz, forgiveness for firms planning to cooperate in the future is part of maximizing the difference between the cooperation and punishment payoffs. In *To Punish or Persuade,* forgiveness is advocated more for its importance in building a commitment to comply in future.

By nurturing expectations of responsibility and cooperation, the regulator can coax and caress fidelity to the spirit of the law even in contexts where the law is riddled with gaps or loopholes. In this way TFT resolves the loophole-opening contradiction of punitive regulation.

In all of these ways analyses of what makes compliance rational and what builds business cultures of social responsibility can converge on the conclusion that compliance is optimized by regulation that is contingently ferocious and forgiving.

The Lexical Ordering of Money and Responsibility

Braithwaite and his colleagues came to realize during their fieldwork on nursing home regulation that another distinction apparent in the modes of economic reasoning about regulatory compliance is recognized in philosophical discourse. This is the idea of two principles being lexically ordered. A lexical order is one "which requires us to satisfy the first principle in the ordering before we can move on to the second" (Rawls, 1973: 43). Some of our subjects were remarkably explicit in their attachment to the idea of lexical ordering:

> The advice I give to the Directors of Nursing in my nursing homes is to list their priorities under the following headings: 'Must do. Should do. Could do. Won't do.' Then they should start at the top of the list and work down until the money runs out.

In the nursing home fieldwork, the goals of making money and caring for residents have been found to be differently lexically ordered between different actors. Before moving on to these cases, however, we first consider actors more singlemindedly motivated by either profits or what they define as caring for the patients.

There are actors, individual and corporate, whose behavior can reasonably be modeled as a rational pursuit of profit without any concern for resident care except insofar as improving resident care will contribute to profit. In contrast, there are many individual actors in the nursing home industry whose motivation can reasonably be modeled as the pursuit of what is best for the care of the residents without any concern for maximizing profits: economic considerations are viewed as obstacles that less caring others put in the way of those concerned to do best by the residents.

No examples have been discovered of organizations that could be modeled as seeking to maximize the care of residents without any concern for profits. However, one case was encountered of an unusually profitable American nursing home corporation where several senior executives espoused the philosophy that the way for

this corporation to make profits was for its middle managers and staff not to be concerned about money. Attracting wealthy private-pay residents (as opposed to Medicaid–Medicare residents) was seen as necessary for high profits. The best way to attract them was to engender an atmosphere within the nursing home that whatever was needed for the resident to be comfortable and happy would be done, regardless of cost. One head office executive made the analogy to a 5-star hotel, in which "when you want something, you're likely to get it." So there was a conscious effort to tell staff and middle managers that it was not their job to worry about money; it was their job to provide the best possible care for those residents. The senior managers believed that if they succeeded in engendering this attitude, "the profits automatically follow." As soon as you have staff who are calculating over the need to cut corners on care to save cost, the ideology of excellence, of doing whatever it takes to provide the best quality care possible is threatened. In the view of these senior executives, profit was the goal, but paradoxically the best way to achieve that goal was to have staff who were constrained not to care about it (see generally, Pettit and Brennan, 1986; Pettit, 1988).

A more common type of nursing home management is intensely concerned with profits but sets itself a minimum standard of basic care that must be met for each resident before the pursuit of maximum profits is countenanced. This minimum standard of care on the one hand, and profit on the other, are lexically ordered. Different nursing homes of this type set very different levels of minimum standards of acceptable care that they are first constrained to meet. Some nursing homes in Australia and the United States will not tolerate physically restraining any of their residents. For them, it is unacceptable to cut costs by applying a restraint to a resident. This can be quite a sacrifice because a good way of reducing the hours of staffing needed to run a nursing home smoothly can be to tie up all the residents who might fall, annoy other residents, tear off their dressings, or cause disruption by tugging at furniture or curtains. At the other extreme, U.S. Health Care Financing Administration survey data for 1986 showed that 77 percent of nursing home residents in Nevada were physically restrained.

Other nursing homes that have minimum standards of care and profits in this lexical order set the minima very low: "We provide basic medical care. We give them enough food. We shovel it in one end and keep their bums clean when it comes out the other end." Such nursing homes can be horrible places, lacking joy, activity, love, and any spark of what it is to be human. But impeccable basic care may be provided and healthy profits made. And depressing as these nursing homes are, low as the minimum care standards are, these standards are minimum requirements on the responsible conduct of the business that are taken seriously. Braithwaite encountered two cases of depressing homes of this sort that had residents who could no longer pay their fees. Their savings were gone and, although destitute, they had no entitlement to nursing home benefits. For example, one was a recent Vietnamese boat person who was not entitled to Australian retirement benefits. In both cases, the proprietors of these nursing homes of otherwise quite low standards continued to care for these residents, at a substantial loss. The minimum standard of responsible business conduct for those proprietors did not sink to putting "residents of mine in need of care out on the street."

Instead of nursing home management feeling constrained to meet a minimum

standard of care and then maximizing profits, other nursing homes seek to maximize the quality of care up to the point where they run up against a financial constraint. Many church and charitable nursing homes are like this. They see their goal as providing the best possible care for their residents, constrained only by the need to stay in business. The financial constraint can be to break even, or to make a loss no greater than they can manage to make up through the annual fete, to make a "steady profit," or to meet loan repayments.

What strategy can the regulator implement to deal with actors some of whom are: (1) exclusively motivated by money, (2) exclusively motivated by caring goals, (3) virtually exclusively oriented to caring because they think this is the best way to make money, (4) lexically ordered—minimum care constraint/maximum money, and (5) lexically ordered minimum money constraint/maximum care? We argue that the best strategy to deal with such motivational diversity is TFT.

We have already argued that if an actor is motivated by social responsibility goals—in this context, resident care goals—then persuasion rather than punishment is the best strategy to further cultivate that motivation. This will be true irrespective of whether the caring motivation is itself motivated by profit seeking, nursing professionalism, love, religion, or what the carer learnt on his mother's knee. In the case where the commitment to caring is up to only a depressingly low minimum, the regulator will still do best to build on that minimum, to seek to define the requirements of the law as part of that minimum level of care that the actor feels responsible to meet. Only when it fails in getting the acceptance of such a definition should it shift to the tough enforcement response.

In the obverse case where the organization pursues maximum quality of care until a financial constraint is struck, in most cases the regulator will be rewarded by a presumption that persuasion will be positively responded to. However, such organizations will at times confront financial crises that constrain their maximizing of care below a level acceptable to the regulator. Then, even with this type of caring organization, the regulator must defect to punishment to defend the integrity of the standards in the law.

What then of the uncaring actor concerned only to maximize profits? The reasons for playing the regulatory game TFT even here are supplied by the economic rationalist analysis in the work of Axelrod and Scholz. By sustaining the rewards that the compliant profit-maximizer gets from repeated cooperation, TFT can be the best strategy here too.

A final type of actor we must consider is neither totally committed to money or responsibility goals, nor committed to any lexical ordering of the two. Rather, these actors have a trade-off function for choosing between being responsible and making money; when the money involved passes a certain threshhold, responsibility is forgotten. Up to a certain point on the trade-off function, these actors behave the same as one totally motivated by responsibility. Beyond that point, they behave the same as actors totally motivated by money. Since we have shown that TFT is the best strategy for actors totally motivated by social responsibility and for actors totally motivated by money, it follows that TFT is the best strategy for actors who trade off the two concerns.

It may be that there are two types of cases that challenge this general conclusion that TFT is the best way to go. The first is a type of case that Braithwaite and his

colleagues have not encountered in their nursing home fieldwork—an organization that is concerned neither about being socially responsible in caring for residents nor about behaving in an economically rational way in its dealings with the regulatory agency. Let us call this the pathological irrational organization. Although Braithwaite did not encounter them, some of his regulatory informants claimed that they had—these were nursing homes that the regulators had been forced to close down. Their proprietors resented the government telling them what to do to care for their residents, so much so that they allowed themselves to be put out of business rather than toe the line. It may be that such irrationally unresponsive organizations that are still in business are so rare, and so difficult to identify in advance of the making of regulatory requests, that TFT will not necessarily cause substantial regulatory failure. The costs of falsely identifying organizations as pathological may be so great that the "suck it and see" quality of TFT may still be the best option.

But TFT will certainly cause regulatory failure here if its sanctions are limited to purely deterrent sanctions (e.g., fines). To deal with irrational actors, incapacitive sanctions are needed (e.g., license or charter revocation; removal of nursing home residents). The law must have sanctions designed to cope with irrational actors as well as rational actors, because where irrational actors exist they are likely to be the loose cannons on the deck that can do the greatest damage.[6] Just as it is a minor change to TFT based on economic deterrence to accommodate reputational deterrence, so it is a minor change to provide for sanctions that supply incapacitation rather than deterrence.

The second case where trying persuasion first will fail is with the determinedly profit-maximizing actor in a regulatory context where, unlike the nursing home case, the regulator and the firm are in a single-round regulatory encounter, the one-off customs offender in a foreign country. Handler (1986: 4) may be right that "continuity of relation is probably the norm in the modern state." Nevertheless, there are certainly many areas where regulatory encounters are not continuous, but one-off (for example, the regulation of many types of fraud). Persuasion will clearly fail here because the economically rational firm that does not feel constrained to be law abiding will cheat every time. Because persuasion will fail here, it does not follow that automatic recourse to punishment will succeed. The fiscal costs of such a policy are so high that the state can only afford it with the most serious types of offending; for less serious offenses, the state may still do better by seeking voluntary remediation from those actors who do have a commitment to being law abiding.

Braithwaite's nursing home data are adequate to show that the disparate orderings of concerns about money and social responsibility discussed in this section do exist. However, the data do not estimate their respective frequency or even which is the most common. In the next section we see why any such estimation might in any case be misguided positivism. For the purposes of the present analysis, moreover, the relative frequency of different orderings of profit and people concerns does not matter, as our hypothesis is that TFT is the wisest strategy, whatever the ordering.

Disorder in the Multiple Self

In the previous section, we considered different forms of lexical ordering of concern about money and responsibility that are manifest in different regulatory encounters.

What is also manifest, however, is a great deal of disorder. The corporation that the regulator feels should be dealt with as an unscrupulous profit maker this month will be dealt with as a socially responsible corporate citizen the next month. Part of the explanation for this lies in the fact that the corporation is not a monolith: "When it is 'Fred' you have to deal with, steel yourself for the worst. But when 'Harry' is on duty on the other side of the desk, you can work things out without any need to act tough."

However, more than this, the disorder we observe in regulatory typifications even of individual business actors arises because most business actors are bundles of contradictory commitments to values of economic rationality, law abidingness, and business responsibility. Business executives have profit-maximizing selves and law-abiding selves; at different moments, in different contexts, the different selves prevail.

> INSPECTOR 1: She (the administrator) cried at the exit conference when we had some adverse findings. She's crying saying: 'I give good care to my patients. I care about them.'
> INSPECTOR 2: And she does. She does care.
> INSPECTOR 3: [unconvinced] In her own way.
> INSPECTOR 4: [aside to Braithwaite]: She was in Dachau in the war and some of the surveyors call her nursing home Dachau.

Later Braithwaite spent 3 days observing this administrator interacting with another nursing home inspection team and talking with her staff and residents. He saw her caring self, her dictatorial self, and her economically rational self. She was a woman motivated both by her deep concern for her residents and by power. She enjoyed having control over an institution and its people and she bridled at any questioning of her commands. Although the inspectors sought to effect change by appealing to her caring self, they were constantly on guard against an explosive encounter with her dictatorial self.

Situations arise when the money-making self and the responsible self are forced to stare each other in the face. The following incident illustrates two devices used to dissolve such a showdown: (1) laughter—the two selves discover humor in their contradictory predicament, and (2) timing—the selves take turns in prevailing. The incident occurred during an in-house Quality Assurance Committee meeting where the nursing home administrator was alluding to the fact that Medicaid reimbursement is graduated according to the care needs of residents:

> ADMINISTRATOR: If we get 'Mr. Jones' independent on this program, we lose money. They punish us for being successful. So don't be successful in a hurry [general laughter around the table].

Regulatory actors also have multiple selves: they can be nice guys or tough guys, self-interested or public-spirited, professional or unprofessional, diligent or lazy, intelligent or confused. The stuff of regulatory disasters is where the tough, unprofessional, confused self of the regulator encounters the irresponsible profit-driven self of the business executive. Such disasters come to everyone's attention—the head of the regulatory agency cops flak from them, the industry association publicizes them as evidence of the unreasonableness of regulatory demands, the public interest group publicizes them as evidence of industry rapaciousness and

irresponsibility, and scholars write books about them. Compound this with the fact that many regulators are not well qualified to do their jobs—they are not the cream of the labor market—and there are good reasons to be cynical that good can come of regulatory encounters.

But this pessimistic analysis misses one important observation that is inescapable if you spend time actually watching regulatory interactions. It is that mediocre people of middling morality and intelligence tend to put their best foot forward when they enter a regulatory encounter. Rather, they have multiple selves, referred to above, and they put their best self forward. The tense, inherently conflictual nature of a regulatory transaction means that the agent of the firm is likely to confront a regulatory agent who has psyched himself to be as professional, intelligent, compassionate, and empathic as he is capable of being. Have a drink with him afterward and he may slip into the unprofessional name-calling mode that is his more relaxed self. But when it matters, when the deal is done, he will amaze at how he can put his best self forward. Obversely, the agent of the firm is likely to put aside his economic self. He will not say "Greed is good," as in Hollywood's portrayal of "Wall Street." He does say:

> NURSING HOME PROPRIETOR TO INSPECTION TEAM: We're pleased to have you here. We've been doing this for such a long time. There are probably some ways that, because we have been at it for so long, that we can't see the wood for the trees. If you find things we are doing wrong, we'll play the game.

The emergence of the business executive's socially responsible self just when it is most needed—at the point where the regulatory deals are done—is not just a matter of the executive choosing alone to put his best self forward. Regulatory agents actively work at encouraging business agents to put their best self forward.

> INSPECTOR TO NURSING HOME ADMINISTRATOR: What you want and what we want more than anything else is to improve the quality of care your residents are getting . . .
> (later) INSPECTOR TO BRAITHWAITE: When you say to them that we all agree that the care of the resident is what we are all concerned about, you know that's not true, that they're concerned about making money. But what are they going to say? They can't turn around and say, 'Hell no, I don't care about the residents; all I care about is profits so I'm not going to do it.'

Sophisticated regulators are practitioners of achieving their goals by manipulating vocabularies of motive (Sarat and Felstiner, 1988). C. Wright Mills has explained how a key research challenge is the understanding of how power can be grounded in the sophisticated deployment of vocabularies of motive:

> Men discern situations with particular vocabularies, and it is in terms of some delimited vocabulary that they anticipate consequences of conduct . . . In a societal situation, implicit in the names for consequences is the social dimension of motives. . . . We influence a man by naming his acts or imputing motives to them or to "him." . . . The research task is the locating of particular types of action within typal frames of normative actions and socially situated clusters of motive. (Mills, 1940: 908,913)

Even without such manipulation, our empirical claim is that business actors are likely to put forward a self that they, the regulator and the researcher observing them, are all likely to view as their socially responsible self. There is an obvious explanation for this. It is good for your survival in the regulatory game, both emotionally and economically (firms often fire employees who upset working relationships with regulatory agencies), to learn how to put your best self forward and how to encourage the other players to put their best self forward. At the end of a day's regulatory negotiation, inspectors sometimes say: "I'm so sick of being nice to people." Even though keeping your guard up by being nice, being empathic, is exhausting, it is less so than the consequences of letting that guard down and unleashing an emotional blood-letting.

Perhaps when the tension abates, when the regulator goes home, the executive's profit-maximizing self will be reasserted to frustrate the achievement of the regulatory goals. But perhaps that profit-maximizing self will think it best to stick with the deal once it is done because this will create a better negotiating position next time. And sometimes, actually quite often, executives will be sincerely persuaded that the inspector has asked them to do something that they should do. At the very least, the tendency to put the best self forward should temper the standard pessimism about the possibility of good coming of regulatory encounters.

What is the implication of all of this for the question of TFT? It is again that TFT is the right way to go. The lesson from the experience of regulators who are sophisticated manipulators of vocabularies of motive is that lurking even within business executives who most of us would typify as ruthless, there is a glimmer of a socially responsible self that can be drawn to the fore in a regulatory encounter.[7] Therefore, we should not succumb too readily to an analysis that says that we can identify certain actors as bound to be recalcitrant, and for whom it would be a waste of time to use persuasion as a strategy of first choice. Furthermore, we should avoid a premature assumption that certain offenders are incorrigible, an assumption that a strategy that is forgiving will only be exploited. Forgiveness, giving the wrongdoer a second chance, can bring out the best in the worst of us—an idea prominent in Christian teaching and in the common sense of those schoolteachers who are competent disciplinarians. So we should be cooperative at first to give others a chance to put their cooperative self forward; we should be tough with cheaters to give them reason to favor their cooperative selves; and we should extend forgiveness to those who show sign of abandoning cheating in favor of cooperation.

Many Players, Many Selves, Many Games

We have argued that firms are not monolithic. Not all of the relevant actors have the same interest in profit maximization as those at the top may have. Regulators do not have their greatest effects by directly sanctioning law breakers; more important is the indirect effect of lending authoritative support to law-abiding constituencies within the organization.

> ADMINISTRATOR: Regulation keeps the proprietor on his toes. They know the inspectors are going to come. It helps us. Without them we'd have no power with the proprietor.

Moreover, it has become clear that not only do we need to disaggregate the organization, but also to disaggregate individuals within it who have multiple selves. Furthermore, it should be clear that we need to think about the regulatory game as something played at even higher levels of aggregation than the firm (with the industry association) and at intermediate levels of aggregation between the individual and the firm (the subunit, the subsidiary, the Australian environmental protection division). Although it may be sensible to model the pharmaceutical company as motivated by profit, it may be better to model its research division as motivated by scientific glory, and the industry association lobbyists who work for it as motivated by a desire to sustain a reputation among the political elite as formidable but principled lobbyists.

Cooperative, tough, and forgiving regulatory routines might, therefore, be played *simultaneously* with different audiences in mind as much as *sequentially* as the TFT account implies.[8] Really, however, all of the foregoing recommends that the regulator is best to play TFT in a number of simultaneous games. Thus, the regulatory agency may be in confrontation mode with an industry association that is urging its members to resist a new regulation. At the same time, it is in cooperative mode with one of the member firms of that association that believes the regulation is right. This is a firm that already has corporate policies in place to require compliance. Still at the same time, the agency may be confronting the manager of a particular plant belonging to that firm who has been a recalcitrant offender against the new regulation. The agency may then seek to conspire with the cooperative corporation to sacrifice that plant manager on the altar of an individual criminal prosecution (in which the corporation is not charged). Or the regulator may hint at the desirability of the cooperative firm dismissing the uncooperative manager.

It is in fact very common to observe cooperative games being played at one level to advance imposition of a punishment payoff in a game being played at another level. Braithwaite has observed many encounters where coal mine safety directors or directors of nursing take the inspector aside to say: "I'm concerned about the lack of safety involved in [these company practices]. I've raised it with head office. They say it will cost too much. But if you look into it and insist that they fix it, they'll spend the money." The manager will then direct the inspector to the site of the problem and to the files where failure to address it is documented. With nursing home inspections in the United States and Australia, it is rare for an inspection to take place without one staff member or another giving the inspection team a tip-off of some value.

The irony here is that playing of the game in a cooperative way can generate the information to play the game in a punitive way at another level. Where inspectors walk into a workplace with the demeanor of a tough law enforcer, they get little information. Where they walk in with the demeanor of a friendly persuader, they get the information that can empower them as tough enforcers. TFT is, therefore, better than consistent punitiveness. But it is also better than consistent persuasion at getting information. One reason that actors feed information to regulators about the crimes of other actors is that they fear punishment themselves; they correctly calculate that if they enter into a cooperative game with the regulator, they will be the players least likely to be punished. They would have no reason to make this calculation, however, if they were dealing with a regulatory agency that never punishes. They have maximum reason to do so with an organization that plays TFT with informers. In the

whistle-blowing game, as in the other games, players often have mixed motives: they can supply information partly because they think this is right, partly because they fear punishment.

Pyramid Strategies of Responsive Regulation

The Enforcement Pyramid

The more we consider the motivational complexity in regulatory encounters, the more we disaggregate and aggregate to unpack the complexity of agency and action, the more attractive TFT looks. Still we have qualified TFT as it appears in classic economic rationalist formulations. In this classical form, sanctions are conceived only as deterrents, usually monetary deterrents. We have stressed the importance of reputational deterrents; in the next chapter we argue further that the importance of adverse publicity directed at wrongdoing is not so much its deterrent effect as its effect in constituting consciences, in moral education. We have also contended that the law must be designed to incapacitate irrational actors as well as to deter rational ones; hence, other types of sanctions, designed with incapacitation rather than deterrence in mind, are needed.

In this section we take this suggestion of the need for a range of sanctions a radical step further. This step was taken in Braithwaite's *To Punish or Persuade* where the argument was first made that compliance is most likely when an agency displays an explicit enforcement pyramid. An example of an enforcement pyramid appears in Figure 2.1. Most regulatory action occurs at the base of the pyramid where attempts are initially made to coax compliance by persuasion. The next phase of enforcement

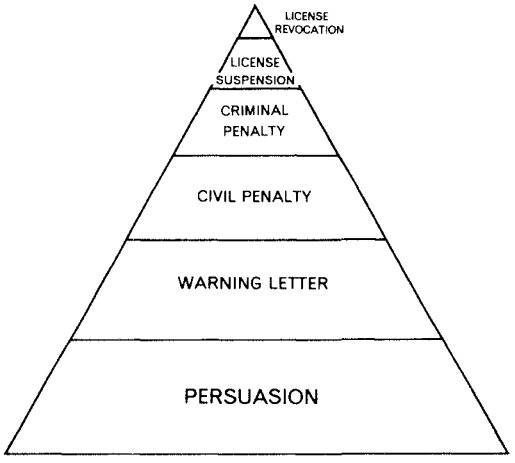

Figure 2.1. Example of an enforcement pyramid. The proportion of space at each layer represents the proportion of enforcement activity at that level.

escalation is a warning letter; if this fails to secure compliance, imposition of civil monetary penalties; if this fails, criminal prosecution; if this fails, plant shutdown or temporary suspension of a license to operate; if this fails, permanent revocation of license. This particular enforcement pyramid might be applicable to occupational health and safety, environment or nursing home regulation, but inapplicable to banking or affirmative action regulation. It is not the content of the enforcement pyramid on which we wish to focus during this discussion, but its form. Different kinds of sanctioning are appropriate to different regulatory arenas.

Defection from cooperation is likely to be a less attractive proposition for business when it faces a regulator with an enforcement pyramid than when confronted with a regulator having only one deterrence option. This is true even where the deterrence option available to the regulator is maximally potent. Actually, it is especially true where the single deterrence option is cataclysmic. It is not uncommon for regulatory agencies to have the power to withdraw or suspend licenses as the only effective power at their disposal. The problem is that the sanction is such a drastic one (e.g., putting a television station off the air) that it is politically impossible and morally unacceptable to use it with any but the most extraordinary offenses. Hence, such agencies often find themselves in the situation where their implied plea to "cooperate or else" has little credibility. This is one case of how we can get the paradox of extremely stringent regulatory laws causing underregulation (Mendeloff, 1979; Sunstein, 1990: 91–92). Regulatory agencies have maximum capacity to lever cooperation when they can escalate deterrence in a way that is responsive to the degree of uncooperativeness of the firm, and to the moral and political acceptability of the response. It is the same point as in strategic deterrence in international affairs; a country with a nuclear deterrent but no conventional forces may be more vulnerable than one that can bargain with a limited range of conventional escalations.[9] And it is the same point that has been demonstrated empirically in the domain of criminal justice: if death is the sentence for rape, juries that think this excessive will not convict rapists; if mandatory imprisonment is provided for drunk drivers, many police officers will decline to arrest them (see generally Feeley, 1983: 126–138).

A regulatory agency with only a sanction that cannot politically or legally be used in a particular situation is unable to deliver a punishment payoff. When a regulatory agency has a number of weapons in its armory, for any particular offense the rational offending firm will calculate that the regulatory agency will in practical terms be unable to use some of the weapons theoretically at its disposal. For those sanctions that can practically be used, it will calculate that the regulator can choose sanctions ranging in severity from s_1 to s_n with probabilities that these sanctions can actually be delivered ranging from p_1 to p_n. But the information costs of calculating these probabilities will be high even for a large company with the best legal advice. These information costs imply that the regulator with an enforcement pyramid may have superior resources with which it can bargain and bluff, a subject we return to later.[10]

Figure 2.2 lays out the dilemma more explicitly. It depicts a regulatory arena where five offenses of increasing seriousness are possible—A,B,C,D, and E—and where the law provides for only two possible punishments for these offenses—X andY. X and Y are punishments that community sentiment judges to be suitable to offenses C and D. However, they are unacceptably severe sanctions for the less

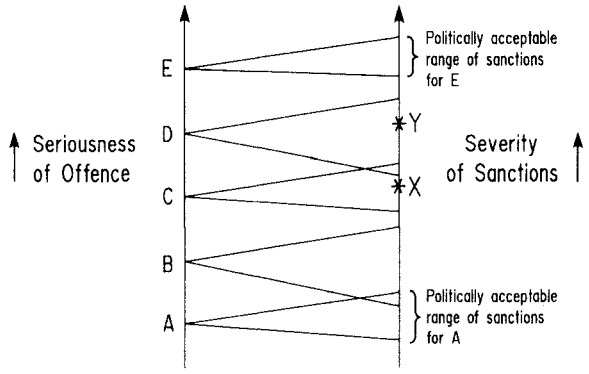

Figure 2.2. A regulatory arena with five offenses of seriousness (A–E), but sanctions with only two severities (X and Y).

serious offenses A and B. There is, therefore, no politically acceptable way of punishing these offenses. There is another type of problem with E. E can be punished (by X or Y) but only at a level that the community judges to be unacceptably lenient. A further implication of this problem may be that if the firm is bold and irresponsible enough to push noncompliance beyond certain limits, it will find a zone where the benefits of noncompliance will exceed the costs of all possible sanctions.

The solution to the dilemma depicted in Figure 2.2 is an appropriately designed enforcement pyramid that makes available to the regulatory agency at least three extra sanctions that fall within the bounds of political acceptability for offense types A, B, and E. Then every escalation of noncompliance by the firm can be matched with a corresponding escalation in punitiveness by the state. Again the strategic analogy is appropriate. If there is any level of escalation in either conventional or nuclear mobilization that is available to the Soviet Union but unavailable to the United States, then the United States is at a significant tactical disadvantage (Schelling, 1966). The disadvantage arises from the fact that the United States must either overreact, with the risk or escalating the conflict beyond bounds where it can be contained, or underreact, risking defeat.

> The idea of "graduated deterrence" and much of the argument for a conventional warfare capability in Europe are based on the notion that if passive deterrence initially fails, the more active kind may yet work . . . if the aggressive move takes time, if the adversary did not believe he would meet resistance or did not appreciate how costly it would be, one can still hope to demonstrate that the threat is in force, after he begins. If he expected no opposition, encountering some may cause him to change his mind. (Schelling, 1966:78)

The enforcement pyramid solves a serious problem with the negotiation of win-win solutions through simple TFT, a problem demonstrated by Langbein and Kerwin (1985). Langbein and Kerwin show that, contrary to the conclusions from static economic modeling of environmental and safety regulation, it is often irrational for

firms to comply with laws when the costs of compliance are less than the benefits. If enforcement is the outcome of a process of negotiation, as it generally is, rational firms will avoid immediate compliance when it will be cheaper to negotiate a compliance deal with the agency. At the least, negotiation delays the firm's compliance costs; at the best, the firm extracts concessions that reduce compliance costs. This is only true, however, if holding back on compliance does not cause an escalation of penalties. The appeal of the enforcement pyramid is that it solves the problem with simple TFT bargaining raised by Langbein and Kerwin's model. Firms that resist initial compliance will be pushed up the enforcement pyramid. Not only escalating penalties, but also escalating frequency of inspection and tripartite monitoring by trade unions (see Chapter 3) can then negate the returns to delayed compliance.

The Pyramid of Regulatory Strategies

The pyramid of sanctions in Figure 2.1 is pitched at the target of the single regulated firm. But there is a more fundamental enforcement pyramid pitched at the entire industry. This is a pyramid of regulatory strategies.

To Punish or Persuade argued that governments are most likely to achieve their goals by communicating to industry that in any regulatory arena the preferred strategy is industry self-regulation. When self-regulation works well, it is the least burdensome approach from the point of view of both taxpayers and the regulated industry. When the state negotiates the substantive regulatory goal with industry, leaving the industry discretion and responsibility of how to achieve this goal, then there is the best chance of an optimal strategy that trades off maximum goal attainment at least cost to productive efficiency. But given that an industry will be tempted to exploit the privilege of self-regulation by socially suboptimal compliance with regulatory goals, the state must also communicate its willingness to escalate its regulatory strategy up another pyramid of interventionism. The pyramid suggested was from self-regulation to enforced self-regulation (see Chapter 4) to command regulation with discretionary punishment to command regulation with nondiscretionary punishment (Fig. 2.3). Command regulation with nondiscretionary punishment has its military analogue in the burning of bridges. If the bridges that are an army's only route of retreat are burned, the enemy knows that it must fight a bloody battle if it advances beyond a certain point. Burning bridges and enacting a policy of nondiscretionary punishment both have the effect of demonstrating commitment—of communicating to an adversary an intention never to give in.[11]

Again, this is just one example of the particular strategies that might be installed at different layers of the strategy pyramid. One could conceive of another pyramid that might escalate from self-regulation to negative licensing (see Grabosky and Braithwaite, 1986: 188), to positive licensing, to taxes on harm (Anderson et al., 1977). Another kind of pyramid from a totally free market to various forms of partial-industry regulation to industry-wide regulation will be considered in Chapter 5.

Escalation up this pyramid gives the state greater capacity to enforce compliance but at the cost of increasingly inflexible and adversarial regulation. Clear communication in advance of willingness by the state to escalate up the pyramid gives incentives

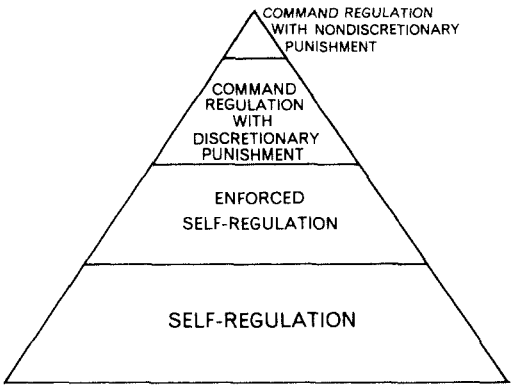

Figure 2.3. Example of a pyramid of enforcement strategies.

to both the industry and regulatory agents to make regulation work at lower levels of interventionism. The key contention of this regulatory theory is that the existence of the gradients and peaks of the two enforcement pyramids channels most of the regulatory action to the base of the pyramid—in the realms of persuasion and self-regulation. The irony proposed was that the existence and signaling of the capacity to get as tough as needed can usher in a regulatory climate that is more voluntaristic and nonlitigious than is possible when the state rules out adversariness and punitiveness as an option. Lop the tops off the enforcement pyramids and there is less prospect of self-regulation, less prospect of persuasion as an alternative to punishment.

Modeling of regulatory deterrence tends to fall into the trap of considering only the option of passive deterrence—deterrent credibility shaped by the potency of the sanctions waiting to be used. The modeling of deterrence in warfare, however, has long involved recognition of the importance of active escalation—a more dynamic modeling of deterrence as an unfolding process. The finesse with which escalation is executed can be as crucial to the efficacy of deterrence as the potency of passive sanctions.

The idea of the pyramid of regulatory strategies underlines the importance of transcending models of regulation as games played with single firms (e.g., Scholz, 1984a, b). The importance of business subcultures of resistance to regulation means that we must understand the significance of industry-wide forces beyond the agency of the single firm. In some respects industry associations can be more important regulatory players than single firms. For example, individual firms will often follow the advice of the industry association to cooperate on a particular regulatory requirement because if the industry does not make this requirement work, it will confront a political backlash that may lead to a more interventionist regulatory regime. Hence, the importance of the pyramid of regulatory strategies (Fig. 2.3) as well as the pyramid of sanctions (Fig. 2.1). Regulatory cultures can be transformed by clever signaling by regulatory agencies, public interest groups, and political leaders that an escalation of the interventionism of regulatory strategy may be in the

offing. As even bigger costs and more unfathomable probabilities are involved in such threats, the potential for bluff is even greater. So much so that industry associations can often be coopted into disciplining and bluffing individual firms that free ride on the regulatory future of the industry.

The Benign Big Gun

The possibility that the range and the nature of the sanctions and strategies at the disposal of regulators may matter is suggested from the application of a variety of multivariate techniques to taxonomize ninety-six Australian regulatory agencies according to patterns of enforcement behavior (Grabosky and Braithwaite, 1986; Braithwaite et al., 1987). A "benign big gun" cluster of agencies emerged from this research. The benign big guns were agencies that spoke softly while carrying very big sticks. The agencies in the benign big gun cluster were distinguished by having enormous powers: the power of the Reserve Bank to take over banks, seize gold, increase reserve deposit ratios; the power of the Australian Broadcasting Tribunal or the Life Insurance Commissioner to shut down business completely by revoking licenses; the power of oil and gas regulators to stop production on rigs at extraordinary cost; the power of drug and motor vehicle safety regulators to refuse to allow a product on the market that has cost a fortune in development. The core members of this cluster of agencies had such enormous powers but never, or hardly ever, used them. They also never or hardly ever used the lesser power of criminal prosecution. Commentators in the past have described the Australian Broadcasting Tribunal's strategy as "regulation by raised eyebrows," and the Reserve Bank strategy as "regulation by vice-regal suasion."

The data from this study are not adequate to measure the relative effectiveness of these ninety-six agencies in achieving their regulatory goals. Nevertheless, the very empirical association of speaking softly and carrying big sticks is an interesting basis for theoretical speculation. The pyramid of enforcement idea suggests that the greater the heights of punitiveness to which an agency can escalate, the greater its capacity to push regulation down to the cooperative base of the pyramid. Graduated response up to draconian final solutions can make passive deterrence formidable (even if the final solution has never been used, as in nuclear deterrence) and can give active (escalated) deterrence room to maneuver. Thus, the theory would be that the successful pursuit of cooperative regulation is predicted by:

1. Use of a tit-for-tat strategy;
2. Access to a hierarchical range of sanctions and a hierarchy of interventionism in regulatory style (the enforcement pyramids); and
3. Height of the pyramid (the punitiveness of the most severe sanction).

Figure 2.4 represents the predicted change in the shape of the enforcement pyramid as the most powerful sanction in the regulatory agency's armory increases in potency. As the capacity for regulatory escalation increases to greater heights as we move from left to right in Figure 2.4, we move to agencies that speak ever more softly (pushing more of the enforcement activity down to the cooperative base of the pyramid) while carrying ever bigger sticks. The bigger the sticks, the less they use

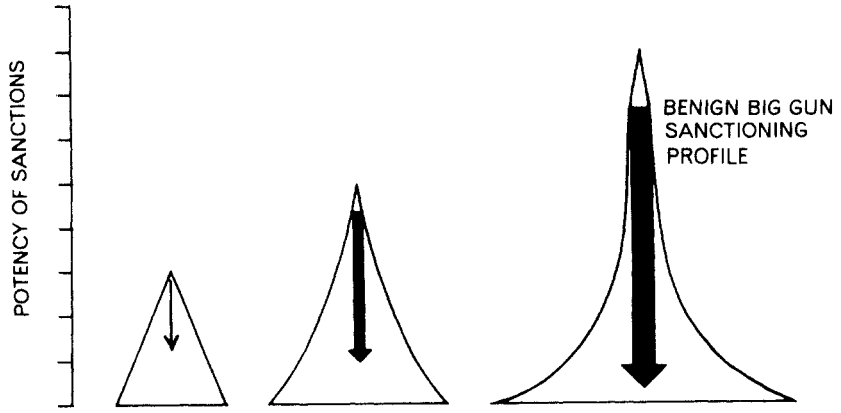

Figure 2.4. Model of the effect of increasing potency of maximum sanctions on the tendency of an enforcement pyramid to push enforcement down to the cooperative base of the pyramid.

them. Indeed, we could conceive of the pyramids in Figure 2.4 as regulatory "density functions" depleting the proportion of regulated firms governed by a particular form of regulation (with the area of the entire pyramid summing to 1 or 100 percent). Our argument is then that higher peaked pyramids will be skinnier at the top (a lower percentage of severe intervention) and fat at the bottom (a larger percentage of gentle intervention).

The empirical claim is not an implausible one. Again it is the same as that commonly made in international strategic debate. It is often said that the period between World War II and World War III is proving to be much longer than that between World Wars I and II because the superpowers have such big sticks.[12] Since the Cuban missile crisis, the superpowers have behaved like benign big guns toward each other. We might take comfort from this accomplishment were it not for the fact that the sticks are so big that one swipe could end us all.[13]

The Theory of Super-Punishments

Goodin (1984) has suggested that social life is not plausibly modeled either as a single-round prisoner's dilemma game where the players interact and then walk away for good, or by a repeatedly iterated game. Social life is, instead, a series of iterated but self-contained games. So conceived, how is it best to play the games? Goodin suggests that the best way is to follow two well-known political precepts, "Don't get mad, get even" and "Forgive but never forget." These precepts need not be contradictory when political life is viewed as a series of iterated but self-contained episodes.

> Politicians should indeed try to 'get even' within a single episode, 'forgive' each other between episodes, but 'never forget' the useful information they acquired in previous episodes about the others' styles of play. Consider the case of one

> perennially large subset of American politicians, Democrats with presidential ambitions. Candidates would be mad not to try to get even with their opponents during the course of a primary election campaign or nominating convention. They would be equally mad to continue grinding that axe once their party's nomination is settled. Yet they would also be mad to forget all they had learned in the course of that campaign about who was trustworthy and who was not. (Goodin, 1984: 130)

An interesting implication Goodin draws is that vindictive tit-for-tat (VTFT), a strategy that proves ineffective in an endlessly iterated game, might be effective under the episodic model. VTFT is an unforgiving strategy that matches every one of the other player's defections from cooperation with multiple punitive defections of its own.[14] VTFT has been found not to be a successful strategy in iterated games: when VTFT plays TFT, a downward spiral occurs into more and more punitive defections (Taylor, 1976; Axelrod, 1984). This need not happen, however, when there is a "partial restart" after each discrete episode of iterations. Certainly the first episode should fall into the mutually destructive cycle of defections. But with the restart, grudges are canceled.

> [In the new episdoe] the TFT player will have marked his opponent as a 'don't tread on me' player, inclined toward vindictiveness and overreaction. He knows that unless he treats his VTFT opponent right he will fly off the handle. And knowing that, he *does* treat him right, even if—indeed, precisely because—they got into a fight last time. The consequence is that TFT will *never* thereafter defect when playing against VTFT [Goodin goes on the qualify this claim]. There should, therefore, be an endless run of mutually profitable co-operation throughout all subsequent episodes. Compare this with the ordinary outcome of a series of games between TFT and TFT players. They regularly revert to defection every thirty rounds or so, just to make sure their opponent is still playing TFT. Their opponent inevitably retaliates with costly defections, and both take some moderately heavy losses while the co-operative equilibrium re-establishes itself. (Goodin, 1984: 131)

Much could be said about the conditions under which this analysis is right or wrong. We will not delve into this except to say that the empirical evidence of business subcultures of resistance to regulation (Bardach and Kagan, 1982) suggests that in some contexts business executives who are dealt with vindictively may be more reluctant to set aside their grudges against government regulators than are Democratic presidential candidates.

Nevertheless, it does seem useful to model much regulation as a series of episodes. For example, the Australian nursing home inspection team experiences a sequence of regulatory encounters as it seeks to extract information and cooperation with changes needed to comply with nursing home standards. The episode of iterated opportunities to cooperate or defect continues as a draft report on the compliance rating of the home is negotiated and then as "Agreed Action Plans" are negotiated. It may continue over the next month or two through follow-up visits to check that these plans are being implemented. The episode then ends; the inspection team will not return for a year or two, when the next episode begins.

Similarly with dealings at the level of the industry association. This year the consumer protection agency deals with the bankers' association over consumer protection concerning aspects of Electronic Funds Transfer Systems. Will the

industry control the abuses in this area through self-regulation or will escalation of regulatory intervention occur in the course of this extended episode? A year or two later a fresh start will be made between these two players in a new episode concerning misleading advertising of interest rates by banks. And so on.

If VTFT runs too great a risk of engendering a business subculture of resistance, other strategies for demonstrating that the regulator is a "don't tread on me" player may not. Making an example of a cheat in a single episode by iterated escalation of punishment up to the point of firing the big gun may be the better way to get the "don't tread on me" message out than VTFT. Strategic publicizing of the escalation to super-punishment against one player in one episode will get this message out to all players for all episodes. The industry grapevine readily assures this even in a loosely knit industry. Moreover, if the regulatory agency is patient and fair in its escalation, giving plenty of warning of the inevitablility of escalation and sufficient time to comply before moving up a rung in the enforcement pyramid, then it will enhance its reputation for justice as well as its reputation for ruthless retaliation against recalcitrance.

VTFT is not a desirable strategy where a government regulator is concerned because vindictive repeated punishment of citizens who are now complying with the law will undermine the legitimacy of the state and thereby threaten voluntary compliance (cf. Levi, 1987). However, we can accept Goodin's episodic analysis as applicable to much regulation and accept the advantages that Goodin demonstrates as flowing from a "don't tread on me" demonstration effect. Given the need for the state to be just, the better way for it to achieve this effect is by patient but inexorable escalation up an enforcement pyramid that ultimately crushes firms that persist in flouting the law.

The plausibility of VTFT regulatory strategies is broadly consonant with the recent literature on super-punishments. Dilip Abreu (1986), for example, has shown that oligopolists may be able to sustain greater compliance with cartel regimes by instituting threats of "extremal equilibria" or super-punishments. Such super-punishments are often characterized by:

1. The punishment consisting of a short "stick" period of discomfort followed by a longer "carrot" period of reintegration; and
2. The punished party itself being induced to cooperate with its punishers during the "stick" period.

The "stick and carrot" nature of super-punishment encourages cooperation of the punished firm and even self-punishment because by cooperating, the punished firm can more quickly move from the more painful "stick" to the less painful "carrot."

Super-punishments may be of use to agencies seeking regulatory compliance. By engendering a firm's cooperation in its own punishment, agencies can radically reduce the costs of punishing. This increases the fiscal feasibility of costly super-punishments in the most extreme cases. This is illustrated by the use of plea bargaining. By cooperating with punishment on one charge, the defendant may get the carrot of immunity from further prosecution on other, more serious, charges. If defendants "take their medicine," they can more quickly move to the carrot period of reintegration. If they do not, instead of stick–carrot, they get stick and more stick—an

escalation up the pyramid. A functional equivalent in the business regulation domain has been the U.S. Securities and Exchange Commission stewardship of its scarce punitive resources by persuading firms such as E. F. Hutton (Orland and Tyler, 1987: 887–907) and Gulf Oil (McCloy, 1976) to pay for independent counsel to produce public reports on corporate malfeasance and on the need to dismiss or demote responsible senior managers.

Super-punishments thus broaden the notion of a responsive regulatory pyramid. First, super-punishments increase the height of the regulatory pyramid. Furthermore, the threat of the super "stick and stick" punishment preserves scarce regulatory resources by channeling violators to "stick and carrot" punishments (where violators cooperate in implementing self-sanctions). Also, by increasing the credibility of regulatory responsiveness it more effectively channels industry behavior to more cooperative paths to regulatory compliance. With effective super-punishments, agencies can more credibly deter noncompliance because they can more convincingly say "if you violate it is going to be cheap for us to hurt you (because you are going to help us hurt you)." The notion of escalating super-punishments even further broadens the notion that pyramids can engender cooperation—because super-punishment theory shows that, even within the most punitive portions of the enforcement pyramid, eliciting firm cooperation can enhance the channeling effects of responsive regulation.

The Image of Invincibility

There surely are good reasons to be pessimistic about the capacity of regulatory agencies to stand up to powerful industries. Yet the pathologically pessimistic might consider the regulatory power of a creature more socially accomplished than ourselves—the dog. How is it that a single Australian sheep-dog or cattle-dog can exercise unchallenged command over a large flock of sheep or herd of cattle every member of which is bigger than herself?[15] How is it that a dog can force the retreat of a man, even a man with a knife—when the man is bigger, more intelligent, and more lethally armed?

The first point to make about the regulatory accomplishments of the dog is that dogs are delightfully friendly to other creatures who cooperate with them. Second, dogs are convincing at escalating deterrent threats while rarely allowing themselves to play their last card. They bark so convincingly that a bite seems more inevitable and more terrifying than it is. And they know how to escalate interactively—in a way that is strategically responsive to the advance or the retreat of the intruder. Friendliness can turn to a warning bark, then a more menacing growl, posture and raising of fur transforms her—she is bigger and seems ready to pounce, teeth are bared, slightly at first, the dog advances slowly but with a deliberateness that engenders irrational fears that a sudden rush will occur at any moment. The dog's remarkable regulatory accomplishments are based on a TFT strategy (the intruder will be extended friendliness when reintroduced as a friend; the sheep will be protected, led to food and drink when they cooperate). The success is also based on finesse at dynamic interactive deterrent escalation, and at projecting an image of invincibility.

In this section we must now begin to ponder the possibilities for regulators to

project an image of invincibility to industries that may be more powerful than themselves. Before erecting any more than our three propositions around the TFT, enforcement pyramid and benign big gun ideas, we require more empirical work to better ground theory construction. The starting place is to ponder Hawkins' (1984) seminal study of British water pollution inspectors. Hawkins found that much of the activity of the inspectors was interpreted by them as maneuvering to preserve their authority and in particular to sustain the myth that compliance with their requirements was inevitable. Now these water boards were anything but benign big guns. Their field officers played a game of regulatory bluff because the fines that actually flowed from the supposedly awesome prospect of prosecution were derisory. Puny penalties were dealt with by a degree of misrepresentation of the terrible consequences of prosecution and by inspectors alluding to the humiliation of a court appearance and adverse publicity rather than emphasizing the fine. All in all, "Negotiating tactics are organized to display the enforcement process as inexorable, as an unremitting progress, in the absence of compliance, towards an unpleasant end" (Hawkins, 1984: 153).

We can read Hawkins as instructing us that even though British water boards are not benign big guns, in a sense they struggle to give the appearance of being just so. There are costs of managing such an appearance, in backsliding and cross-negotiation to extricate the agency from the risk of an appeal or an unsuccessful prosecution. In a more litigious business regulatory culture such as in the United States, one wonders whether a regulatory agency could sustain for long such a fragile image of invincibility.

Notwithstanding this, the Hawkins study does sensitize us to the possibility that there may be a weak relationship between the reality and the image of the enforcement powers of regulatory agencies. What may predict cooperative, compliant outcomes is whether regulatory agencies are perceived by industry as benign big guns. That is, if the agency is perceived as cooperative even though it may be quite adversarial; if it is perceived as willing and able to repay uncooperativeness with awesome sanctioning even though it is not, then firms will cooperate and comply.

A number of possibilities follow. A regulatory agency that is legally able but politically unwilling to fire its big guns might get enormous mileage in management of the appearance of invincibility from a single famous case where it had political support to fire its big guns and where it brought a powerful corporation to its knees. Conversely, an agency that skates on thin ice in the management of these appearances (such as the British water boards) is vulnerable to a litigious firm determined to shatter its myth of invincibility. Equally, a benign big gun with awesome powers it is rarely willing to use is vulnerable to a legal defeat or to being forced into political capitulation on the first occasion it seeks to fire its big gun. In international relations we know that the image of superpowers as invincible is vulnerable to Vietnams and Afghanistans when political and moral exigencies mean that the level of escalation required for victory is impossible or imprudent.

An understanding of political power requires a grounding of how actors come to be granted the credibility of being benign big guns. Firing big guns causes a lot of fallout in resentment among those who are bullied by brute force. It can be a delegitimating business for the state, especially when the firing mechanism on the gun

is untested or rusty. The superpower does better to coax and caress compliance from weaker nations whose very insignificance adds to the damage they will do if they successfully challenge the superpower's image of invincibility. When great powers do bully smaller countries, they must be sure to win. And so with political leaders. They must seek never to lose at the hands of a less significant adversary within the party. While being routinely benign, they must choose their moment to show that they hold the big guns by ruthlessly crushing an opponent. It must be believed that they are both good to those who are loyal, but invincible when they decide to bully the disloyal.

So there might be a more general benign big gun theory of power. If there is, it must stem from an understanding of how actors build appearances of invincibility by displaying their firepower in strategic contests. Certainly Hawkins' work suggests that this might be so in the arena of regulation.

To develop a benign big gun theory of regulation, we must build upon Hawkins' work on the day-to-day interaction of field officers with firms. We must move beyond this to study how agencies handle the crucial tests of their strength that occur at watersheds in their history—how they handle their Cuban missile crisis, their Vietnam. In addition we must study how regulatory agencies go about the business of preemptive efforts to construct an image of invincibility. This might include the study of how agencies pick the "right case" to show that an amendment to an act has "teeth." It might include research on how industry reacts to publicity campaigns calculated to create the impression that agencies with relatively weak powers are benign big guns.

Finally, we need empirical work on what happens when agencies that do have big guns say and do things that threaten their appearance of regulatory clout. For example, in 1987, the Chairman of the Australian Broadcasting Tribunal declared that certain legislative deficiencies made the agency "a toothless tiger." While the agency continued to have very big legal teeth with respect to other matters, one wonders whether across the board the agency came to be perceived as one that could only gum the industry as a result of headlines like "ABT is Toothless Tiger—O'Connor [the Chairman]" (*Australian Financial Review,* 9 November, 1987) and editorials like "Farce at the Broadcasting Tribunal" (*Australian Financial Review,* 10 November, 1987). Do senior executives of regulatory agencies in some ways foster a demeanor of confidence among their own staff, keeping doubts about the fragility of their powers to themselves and nurturing a culture of invincibility within the organization? Or are they frank with their staff on these matters, training them in how to bluff while skating on thin legal ice, how to bargain by making the law appear to cast a bigger shadow than it does?

We also need more empirical work on how actors in business socially construct what it is they fear from regulators (see, for example, the start in Fisse and Braithwaite, 1983). Then we can begin to ground a theory of compliance in the ways of thinking about compliance, the various forms of bounded rationality about regulatory threats and legal obligations, that regulated actors actually manifest. Here, we must explore how compliance constituencies *within* business exploit an exaggerated spectre of an all-powerful regulator to increase the authority or resources of their own safety, auditing, environmental or legal department (for evidence of this

exaggeration of the regulator's power by corporate safety departments, see Rees, 1988: 56–59). This is the kind of program of research needed to ground a theory that might take up the ideas of TFT, the enforcement pyramid and the benign big gun approach to secure cooperative compliance.

Keeping Punishment in the Background

A case will be put in the next chapter for republican institutions that are oriented to persuading actors to deliberate in a socially responsible way (see also Braithwaite and Pettit, 1990). Under such institutions, the public interest is fostered as a result of such public-regarding modes of deliberation. These republican institutions are contrasted with market institutions that seek to foster the public interest by aggregating the endeavors of actors who all pursue their private interests.

We have illustrated in this chapter some of the variety in the modalities of what the republican considers socially responsible deliberation—directors of nursing who deliberate in terms of the well-being of patients instead of self-interest, mine managers pondering how to secure maximum safety for their employees, factory managers weighing how to minimize the pollution caused by their operation.

An interesting way of reframing our hypothesis is that without the spectre of sanctions in the background, such social responsibility concerns would not occupy the foreground of our deliberation. In contrast, where punishment is thrust into the foreground, it is difficult to also sustain public-regarding modes of deliberation in the foreground. Hence the attraction of TFT as a strategy that keeps punishment in the background until there is no choice but to move it to the foreground.[16]

To put some flesh on these bones, let us return to Smith filling out his tax return. As he sits surrounded by receipts, a method of cheating to reduce his tax enters his head. Almost simultaneously, the idea of getting into trouble with the tax authorities crosses his mind. This is enough to stop the cheating idea in its tracks. He does not go on to calculate rationally the chances of the spectre of "trouble with the Tax Office" eventuating or what the actual penalties might be. Rather, within an instant of this spectre crossing his mind, another thought has occupied the foreground of his deliberation. This is that it is morally wrong to violate the tax laws; he remembers his self-concept as an honest taxpayer; turning his mind to careful calculation of the risks and benefits of cheating would be a distasteful mode of deliberation for him. This rapid sequence of ideas in the foreground of Smith's consciousness—cheating, trouble and honesty—is perhaps more likely with a law where the enforcement agency enjoys both an image of invincibility and an image of fairness (Levi, 1987; Stalans et al., 1990; Tyler, 1990). Remember that Smith in this thought sequence never gets to the point of rational calculation. It is just the spectre of awesome and authoritative power that matters, not any realistic assessment of the possibility that such powers would actually be used and with what effect. This is what we mean by keeping awesome powers in the background to motivate socially responsible deliberation in the foreground.

A month later, Smith is audited. He had misunderstood the requirements of the tax law in one respect and claimed a larger deduction than was his entitlement. Smith is taken aback when the government auditor treats him as a cheat; he threatens Smith

with punishment. This causes him to dig in his heels and fight. His motivation is now to beat the tax man, to retaliate against the person who threatens him. The socially responsible motivation to understand the law and comply with it (normally in the foreground of Smith's deliberation) has been driven off his deliberative agenda by accusation, innuendo of dishonest intent and threat of punishment. His preoccupation is simply to fight and to win.

Imagine, in contrast, that the auditor had attributed a different vocabulary of motive to Smith: "Here is a responsible citizen who has just misunderstood the law; I do not need to threaten punishment." A more likely response from Smith then is apology, remorse, offer of immediate repayment. By keeping punishment in the background instead of the foreground of the encounter, the auditor is more likely to keep Smith's law-abiding self to the fore. And given the economic and moral costs of consistently enforcing punitive law in such cases, sustaining a willingness of taxpayers to settle voluntarily is critical.

Is responsibility in the shadow of the axe really responsibility? Interestingly, we believe that it is usually so construed by human actors. When my boss trusts me with $50 to buy some flowers for a sick colleague, it may be trust against a background of a variety of forms of retribution should my boss detect its breach. Yet I may not view the trust any differently than I would if the $50 were entrusted to me by a person who had no power over me. If my boss reminds me of power ("If you don't spend every penny of that $50 on flowers, you're fired"), then I will not feel trusted. So whether actors feel trusted is a matter of subtle social construction about whether the principal is invoking threat.

It may be that trustworthiness is best secured when the obligations of trust are owed to principals with great power over the agent, where that power is threatening but never threatened. Because retaliation is not threatened, trustworthy agents can be motivated by the positive regard of being trusted. Because the power of the principal is nevertheless threatening, untrustworthy agents may calculate that it is best not to breach trust. The trick of domination is the presentation of a social reality wherein threat will be disguised for those who wish to be trusted, and threat becomes apparent to those who deliberate untrustworthiness. When domination works, compliance is seen as natural and right rather than compelled. My boss does not need to threaten retaliation if the obligations of the trust she confers are violated; her power to so retaliate is clear from her authority. In contrast, the stranger in the street may have to threaten to "come after me" if I do not spend his $50 on flowers. But as soon as he does that, he shatters any illusion that I am trusted.

We should not underestimate the capacity of human actors (like Smith as he completes his tax return) to opt for a prosocial interpretation of their own action—trustworthiness—in preference to interpreting their action as submission to compulsion, even in the face of evidence of compulsion. So there is nothing baffling about the capacity of trust in the shadow of the axe to persuade actors that they are trusted, while also persuading them at other moments that they had better watch out if they are untrustworthy.

Considered together, therefore, these accounts suggest that regulators should not do without an image of invincibility in the background, and should be reluctant to

push punishment to the foreground of day-to-day regulatory encounters. They do best when they are benign big guns.

Poorly conceived regulatory strategies, like that of the U.S. Occupational Health and Safety Administration (OSHA), do just the opposite of this. They constantly nip at firms with flea-bite fines. In most encounters with OSHA inspectors, petty punitiveness is in the foreground and no big guns are in the background (see Bardach and Kagan, 1982; Rees, 1988). The result of routine flea-biting is that cooperation is destroyed without any of the benefits that can flow from tough enforcement being secured. When scholars point to an agency like OSHA to conclude that punishment and persuasion are incompatible, they have not understood the foregrounding of cooperation and backgrounding of punishment that benign big guns can accomplish.

The Minimal Sufficiency Principle

Unfortunately, when people think about social control, they tend not to grasp the importance of subtlety and salience in how we underwrite persuasion with punishment. Instead, in most respects, lay psychology involves a crudely economistic model of social control. A series of experiments by Boggiano et al. (1987) showed that parents and college students adopt a "maximal-operant" theory of how to motivate students to work at academic tasks. The subjects believed that long-term commitment to academic tasks varies positively with the size of short-term extrinsic rewards. This belief is wrong.

What the maximal-operant principle forgets is that what may be best for short-term compliance might also be counterproductive for long-term internalization of a desire to comply. And this long-term internalization is the more important matter in almost any domain of social control because it is usually impossible for society to organize its resources so that rewards and punishments await every act of compliance or noncompliance.

As opposed to the maximal-operant principle, a great deal of empirical evidence supports a minimal-sufficiency principle: the less salient and powerful the control technique used to secure compliance, the more likely that internalization will result (Lepper, 1973, 1981, 1983; Lepper and Greene, 1978). Experimental research on children and college students demonstrates the counterproductive effect salient rewards and punishments can have: long-term internalization of values like altruism and resistance to temptation is inhibited when they view their action as caused by a reward or punishment (Lepper, 1973, 1983; Lepper and Greene, 1978; Dix and Grusec, 1983; Hoffman, 1983).

> Over 50 studies examining the effect of extrinsic incentives on later intrinsic motivation indicate that inducements that are often perceived as controlling (e.g. tangible rewards, surveillance, deadlines), depending on the manner in which they are administered, reduce feelings of self-determination and undermine subsequent motivation in a wide variety of achievement-related activities after the reward is removed. (Boggiano et al., 1987: 867)

The minimal-sufficiency principle seems to be of fairly general import, being supported in a variety of domains including moral behavior, altruism, personal

interaction, aggressive behaviors, and resistance to temptation (Lepper, 1973; Dienstbier et al., 1975; Dix and Grusec, 1983; Boggiano et al., 1986). Just as strong external incentives retard internalization, using reasoning in preference to power-assertion tends to promote it (Cheyne and Walters, 1969; Parke, 1969; Hoffman, 1970; Baumrind, 1973; Zahn-Waxler et al., 1979).

There is one important sense, however, in which Boggiano et al. (1987) found that citizens were not narrowly economistic in their views on social control. They did understand the positive effects of attributing positive intrinsic motivations to encourage desired behavior. Therefore, adults rated the strategy of telling children that they (children) really seemed to "get a lot of pleasure" out of being helpful, or that they (children) were "the kind of person" who enjoys helping others, as a better strategy than reward, punishment or noninterference for increasing helping behavior. Recall that Braithwaite's nursing home inspectors also frequently use positive attribution as a compliance strategy. And the experimental evidence is encouraging that positive attribution works (Pittman et al., 1977; Grusec et al., 1978; Grusec and Redler, 1980; Rushton, 1980).

Together, the minimal-sufficiency principle and the positive attribution principle ground in tested psychological theory why it is best to have a TFT strategy wherein: (1) escalating to punishment is a last resort, and (2) the resort will be only to a point up the enforcement pyramid that is minimally sufficient to secure compliance. At the same time it can be argued that the big gun idea shakes these theoretical foundations. However, in the previous section we attempted to solve this problem by showing how the benignness of the big gun can mean that the gun has low salience for actors who are intrinsically motivated to comply (being kept in the background of deliberation). At the same time the big gun can be made salient for actors who have no intrinsic motivation to be resonsible that is worth nurturing, and who can only be moved by bringing forward the extrinsic incentive.

Effective regulation is about finesse in manipulating the salience of sanctions and the attribution of responsibility so that regulatory goals are maximally internalized, and so that deterrence and incapacitation works when internalization fails.

Rational choice theorists like to urge the design of institutions that economize on virtue (Brennan and Buchanan, 1985). They believe that institutions are less likely to fail if they minimally depend on citizens being virtuous. The benign big gun institution economizes on motivation, not just virtuous motivation. It does not depend on citizens being virtuous. If they are not virtuous, guns are ready to be fired. But because guns are not fired at the virtuous, and because the threat of the gun is kept so far in the background that people are not forced to think just in terms of it, virtue is saved from being undermined; indeed virtue is nurtured. The benign big gun institution does not depend on citizens being economically rational. If the threat of the gun—say a monetary penalty—does not motivate them, a variety of forms of persuasion are available—appeals to social responsibility or law abidingness, positive attribution, praise, holding up other organizations as models of how to solve a problem successfully.

If both money and a sense of responsibility fail to motivate, the benign big gun institution can mobilize an incapacitative sanction such as temporary license suspension or permanent corporate capital punishment. If noncompliance is a result of

neither greed nor irresponsibility, but of incompetence, the benign big gun institution can seek to persuade management to accept consultancy advice. As an institution economizing on motivation, it has a solution to the problem of noncompliance regardless of what motivates the noncompliance. And it avoids the profound danger of institutions that economize only on virtue—the danger that they will not only fail to nurture virtue but will actively crush it.

Conclusion

We have shown that analyses of what makes compliance rational and what builds business cultures of social responsibility can converge on the conclusion that compliance is optimized by regulation that is contingently cooperative, tough and forgiving. In Scholz's (1984a, b) work, forgiveness for firms planning to cooperate in the future is part of maximizing the difference between the cooperation and confrontation payoffs. In Braithwaite's (1985) work, forgiveness is advocated more for its importance in building commitment to comply in the future. In Scholz's formulation, punishment is all about deterrence. Braithwaite places greater importance on the moral educative effects of punishment (Braithwaite, 1989a; see also the next chapter), and on the role of punishment in constituting an image of invincibility within a regulatory culture. When analyses grounded in very different accounts of human motivation can converge on the virtue of the same policy idea, then there is some hope that this may be a robust idea.

If there is a theme beyond transcending the debate between regulation and deregulation that permeates this book, it is the pursuit of such convergence. Much of contemporary social science is a stalemate between theories assuming economic rationality on the part of actors and theories counterposing action as variously motivated by a desire to comply with norms, to maintain a sense of identity, to do good or simply to act out a habituated behavioral sequence. We think robust policy ideas are most likely to be discovered when we pursue areas of convergence between analyses based on *Homo economicus* and those based on *Homo sociologicus*. Unfortunately, however, our disciplinary predilections are to seek battlegrounds for new clashes between them. Perhaps such convergence will be rare, but where we do find it we might be in the domain of policies that have intended effects instead of the more common unintended effects. The pursuit of such convergence is taken up in a different way in the next chapter.

In this chapter we have seen how an economic and a social defense of TFT move away from the notion of an optimum static strategy, and instead move toward the idea of an optimum way to play a dynamic enforcement game. This too contributes to robustness because the enforcement game is of course dynamic rather than static.

For Scholz, it is the fallback deterrence payoff that determines the level at which firms will voluntarily comply. We have made a number of points about this fallback deterrence payoff here. First, the setting of the deterrent payoff is also a dynamic rather than a static matter. What deterrent can be imposed will depend on interactions between the regulatory agency and other actors in its environment—prosecutors, courts, political leaders—a fact that sophisticated firms understand. When regulatory

agencies have maximum deterrents that are beyond their political capacity to deliver, firms are not likely to take note of the maximum in calculating a deterrence payoff. What they may take greater note of is whether the agency has at its disposal such a range of sanctions (including sanctions that can be imposed without going to court) that it is always likely to be able to respond to evasion with a punitive reply. Furthermore, if firms note that the agency has in place a hierarchy of enforcement response, they may calculate that even though they can tolerate the politically feasible deterrent response from a given level of evasion, escalation to the next round of the enforcement pyramid in a subsequent evasive encounter may be less bearable.

So we might think of the literature as winding toward a more sophisticated account of the possibilities of regulatory enforcement being used to steer corporate behavior away from socially harmful forms of conduct. First, there was the assumption that the more stringent the enforcement of tougher laws to protect say, the environment, the better the environment would be protected. Then there was the caveat that tougher laws would reduce environmental degradation up to the point where the costs of compliance exceeded expected punishment costs. But once an optimum level of stringency is passed, as the law becomes even tougher, compliance falls (see Viscusi and Zeckhauser, 1979). When the compliance standard is set so high as to threaten the survival of firms, many will decide that it is better to be hung for a sheep than for a lamb. Ultimately, the environmental gain from higher standards is outweighed by the environmental loss from lower compliance with those standards. So the economist's advice to the environmental activist is to do best by the environment by lobbying for optimum stringency rather than maximum stringency.

But then Scholz (1984a, b) takes the economists to task for thinking statically in terms of an optimum compliance level in the standard. Rather, it is the fallback deterrence payoff put in the context of a dynamic enforcement strategy that predicts compliance. Now we are taking Scholz to task for adopting a static view of that fallback payoff. Different fallback payoffs will have differential credibility depending on the social context in which an attempt is made to apply them. There is no simple linear relationship between the potency of "a" fallback deterrence payoff and compliance under a TFT strategy. Rather compliance is most likely when regulators (1) have access to an armory of deterrent and incapacitative weapons, and (2) when they avoid both the mistake of selecting a sledge hammer to swat a fly and selecting a flyswatter to stop a charging bull. Compliance is predicted by both the existence of an awesome armory and by the avoidance of clumsy deployment of it.

In summary, the possibility has been advanced that compliance is responsive to the existence of TFT strategy, the existence of an enforcement pyramid appropriate to the particular regulatory domain, and the potency of the upper limits of sanctioning within that pyramid. Empirical work is needed to build an understanding of how regulatory agencies succeed or fail in maintaining the appearance that noncompliance is a slippery slope that will eventually lead to these ultimate sanctions unless the firm reforms.

A worry about a policy that delivers to regulatory agencies discretion to shift the degree of regulatory intervention or nonintervention up and down two pyramids, such as shown in Figures 2.1 and 2.2, is that there will be abuse of discretion—perhaps

capture by the malevolent or victimization of the innocent. The next chapter explores republican tripartism as a strategy for addressing this worry.

In emphasizing this participatory solution for assuring the accountability of discretion, we do not mean to pretend that it is enough. In any shift from autonomous to responsive law, as Nonet and Selznick (1978) point out, there is a risk of jettisoning the major advances that autonomous law involved as a movement away from coercive law (where the law is no more than an extension of the administrative power of the state). Therefore, it is essential to preserve rights of appeal against regulatory decisions to courts that enjoy substantial autonomy from the state. Although we have seen the virtues of giving regulatory agencies big guns, it is crucial that the state set limits on the maximum sanctions that can be imposed and on the offenses for which they can be applied. Obviously, the rule of law is needed as protection against the excesses that we have seen from regulators with the backing of the ruling party in countries such as China, excesses that have included execution and arrest without trial. The rule of law is not only essential to a republican regulatory order, it is definitional of it (Braithwaite and Pettit, 1990: 54–136). And in a republican state, however participatory and open is the context in which regulatory discretion is exercised, the process must be monitored by representatives elected by the entire citizenry who can legislate to forbid or constrain a particular kind of discretion.

Should our regulatory institutions, therefore, be designed for knaves or should they be designed to foster civic virtue? Our answer has been that they should be designed to protect us against knaves while leaving space for the nurturing of civic virtue. This requires the design of dynamic rather than static institutions. First, we should seek to solve our regulatory problems by appeals to the social responsibility of the firm. If that fails, the dynamism of a responsive strategy enables us to shift our motivational appeal to rational self-interest by escalating through deterrent threats. If that fails, if we confront a firm that irrationally resists deterrent threats or that commits an offence so lucrative as to be beyond deterrence, the control strategy shifts from deterrence to incapacitation. Ultimately, this means corporate capital punishment when we withdraw the licence or charter of the company. This is how we can design institutions that avoid the mistake of assuming that actors will be rational or that they will be virtuous.