

## CHAPTER 12

---

### Prestige, Normative Regimes, and Moral Authority

Someone possesses *prestige* in a group if the members believe that the person is generally admired in the group. The admiration must be seen as having some grounds, such as the individual's deeds or possessions. A third element is that the members expect the person to gain influence in the group from the admiration. A soldier who dies in battle can be said to have gained honor or glory, but not prestige.

Prestige thus is at the second level, of beliefs about beliefs about the facts. As a matter of objective reality, a person is said to have a good quality; at the first level of beliefs the person has a good reputation—the group thinks that the person has a good quality. Prestige means that everyone thinks that everyone thinks the person has the quality. Perhaps no one admires the person, but if each person thinks that the rest do, that constitutes prestige. As in a trick done with mirrors, someone can gain prestige by convincing everyone that he or she has a good reputation—there is no need to possess the quality in question. This possibility recalls the word's Latin root, *praestare*, meaning to create an illusion, as in "prestidigitation."

Since it can confer rights and benefits in the group, prestige functions like face and honor. Roughly, it is in between them—like honor, it involves an innate quality of the person, but like face, it depends on beliefs about what is in others' minds. A person might possess personal honor while others do not know about that quality, but the idea of having unnoticed prestige or face makes no sense.

Prestige involves what each member believes about others' attitudes toward the individual; face is about what each believes about others' disposition to act toward the individual. Prestige and face also differ according to what produces them and who confers them. For face, respectful treatment comes from a record of the same treatment by the group, but prestige usually starts with the indi-

vidual, with deeds done or objects acquired. For this reason, we speak of prestige being “gained” or “lost.” This is in contrast to face, where the individual is in a reactive position. Face is said to have been “saved” or “lost,” the latter by failing to respond to a threat to it. As well as these differences in the sources of the two concepts, there are differences in their consequences. Someone with high face gets deference in direct on-the-record interactions; but prestige produces perceived admiration by the whole group and, from this, influence of various kinds.

A distinction can be made between prestige and status. Status suggests a hierarchy with clearly defined positions and an institution that set it up. Instead of looking to others’ attitudes as with prestige, with status each party looks to the rules and decisions of the organization. If moral authority derives from normative prestige, the analogue for status is legitimacy. It is the moral right of an actor to be obeyed by virtue of his or her position.

There is no formal structure for the international system, and accordingly *prestige* is usually the better term in that context. “International prestige” is a more common usage than “international status.” Midlarsky (1975) and Luard (1992), however, use “status,” and just why they do so clarifies the difference. Luard was interested in well-defined rules of diplomatic protocol and Midlarsky in the hierarchy of states based on the ranks of diplomatic officials sent to them by other countries. A state’s rank came from a relatively formal system accepted by the whole group, so the choice of these writers is reasonable for their purposes.

This chapter will focus on *normative prestige*, the particular kind that comes from abiding by norms. An important aspect is that good behavior confers influence on related issues.<sup>1</sup> Those who have followed the norms in the past have more influence in pronouncing on what the norms are now in a situation where ambiguity arises. This is called *moral authority* (Lasswell and Kaplan 1950).

### The Relationship of Prestige to Symbolism

Prestige can be gained by doing deeds or possessing things. The latter kind of prestige usually involves concrete, publicly visible tokens that suggest power and ability. A nation holds on to its colonies, acquires a naval fleet, buys modern jets for its national airlines, or hosts an important conference (Sagan 1996,

---

1. Weber (1924), Mills (1951, 239), Parsons (1952, 132), Lasswell and Kaplan (1950), and others have analyzed prestige in similar terms, and Nicolson (1937) discusses it in the international context. McGinn (1972) gives the most systematic theory of it so far in the international context, and Milliken (1996) analyzes its use in justifying U.S. persistence in Vietnam.

1997; Eyre and Suchman 1996). In 1987, when India leased a nuclear-powered submarine from the Soviet Union, the national news agency circulated a photograph of Prime Minister Rajiv Gandhi standing on the conning tower. Although the submarine was Soviet built and stayed under Soviet control, it symbolized India's technological progress.

A so-called prestige symbol usually combines three mechanisms. It can be simply a demonstration that the country is able to do something, in which case it is not a symbol at all in the present usage, but an index in Jervis's sense (1971). A prestige symbol can also rely on the mechanism of message symbolism. The message might be from the grantor of the symbol, not its possessor—a large office is the boss's way of saying that the person has a high rank. It is meant to be on the record, in that the boss is letting everyone know that everyone knows that the person has high rank, and so on. The symbol helps establish the person's face in the hierarchy. This is an example of a typical connection among three central concepts of this work, message symbols, prestige and face.

Some prestige symbols rely on the mechanism of focal symbolism. High technology weapons are salient parts of the typical modern and powerful state. The possessor expects to get its way in future games of power politics, and expects others to recognize this and be more ready to defer.

One way to gain prestige is by winning a symbolic contest. This combines the mechanism of a demonstration with that of a focal symbol. The context must have a structure that makes the symbol analogous to a larger contest for which it is meant to set expectations. It functions like the prenegotiation debates about the shape of the conference table (chap. 4), whose outcome could be a self-fulfilling symbolic indicator of the outcome of the real struggle. One example was the space race. The Soviet Union won the first round by launching unmanned and manned earth satellites. The United States landed on the moon and soon retired. America's motive had been the symbolic competition, a demonstration that it was not slipping into technical inferiority after the surprise of Sputnik. It won a round, but after that, other practical benefits did not justify further effort. The Olympic Games are another costly contest over a prestige symbol. Many countries put their funds into them instead of into athletic programs that would benefit significant numbers of their population. Another example of a symbolic contest involved Taiwan and the People's Republic of China competing for diplomatic recognition from other countries. The next chapter will suggest that the nuclear arms race was also largely symbolic.

The focus of the rest of this chapter is on a certain kind of prestige—normative prestige. It is gained from doing appropriate deeds, and very often does not involve symbolism.

### **Normative Prestige and Moral Authority**

Normative prestige yields moral authority. When a country speaks out on a normative issue, its voice has greater legitimacy if it possesses prestige in that context. Costa Rica had a stable democratic system and gained a role in the negotiations for peace in Nicaragua. It acquired moral authority, the self-fulfilling expectation in the group that the group would follow the possessor's normative suggestions. Conversely, when a country commits a normative lapse, it loses moral authority on related issues.<sup>2</sup> In December 1993, Hungary changed its election laws to a system that made it harder for ethnic Germans living there to elect representatives to the National Assembly. According to the BBC, the group Alliance of Germans in Hungary charged that this cast a shadow on Hungary's minority policy and diminished its international prestige. The Hungarian leadership would have no moral grounds, the Alliance said, to raise its voice for the rights of Hungarians beyond its borders.

Normative prestige is illustrated by parallel controversies in Canadian defense policy, 25 years apart. In the early 1960s, Canada was debating whether to acquire nuclear warheads for its fighter-bombers and antiaircraft missiles (Cox 1985). The United States would supply the warheads, and, by U.S. law, would keep ownership and veto power over their use. Proponents argued that unless Canada accepted a nuclear role, it would lose its sway in NATO. Opponents held that acquiring nuclear weapons would weaken Canada's voice against proliferation and the arms race. Both sides were citing moral authority, but they focused on different reference groups, one on the military milieu of NATO, where acquiring the weapons was normative, the other on diplomatic circles striving for arms reductions, where it was not. Lester Pearson's government accepted the weapons, but soon afterward Canada began dropping its nuclear involvement piece by piece, and after a decade, the antinuclear position prevailed. Sagan (1996) describes a similar Ukrainian dilemma about the strategic nuclear weapons left on its territory after the end of the Soviet Union.

In 1985, a dilemma arose for Canada between normative prestige and focal symbolism. A government committee advocated the purchase of 10 nuclear-powered attack submarines. Opposition voices criticized the proposal for undermining Canada's stand against proliferation, and submarine advocates stressed the need to patrol the Arctic. Canada had 20 diesel submarines, but the newer ones could stay submerged long enough to operate under the ice. Even

---

2. As in message and focal symbols, contrast with the past increases the symbol's power. Haglund (1989) suggests the existence of an "Elmer Gantry phenomenon" around nuclear proliferation that the holier you have been against it, the more harm you do if you start proliferating.

so, the concept of patrolling was a symbolic one, since the submarines' chance of detecting an encroaching submarine was nil. The rationale was that Canadian claims in Arctic territory disputes with the United States would be stronger if Canada had a military presence there (Desjardins and Rauf 1988; Haglund 1989). The normative prestige argument against the submarines competed with an argument based on symbolic dominion. In the end, the plan was rejected as too expensive.

### Social Norms and Normative Regimes

In these examples, moral authority came from "good behavior," and the latter concept requires an analysis of social norms and normative regimes. The first thesis concerns a prototypical feature of social norms: a social norm is part of a normative regime. This is a system of norms connected to one another in relationships of mutual support.<sup>3</sup> Compliance with the given norm is bolstered by other norms in the regime, which put individuals under obligation to grant rewards to a norm follower and to punish a norm violator. Others have a positive duty to reward or punish, and sometimes an action that is forbidden in regular circumstances becomes required after a violation. Treating someone inequitably is wrong, unless the person violated a norm, and then he or she deserves it. The 1970s and 1980s saw the growth of an international norm against racism, and states like Britain, that did not act against South Africa, were themselves criticized (Klotz 1995). This support structure is a normative regime, and being part of one is a core feature of a norm.

Apologizing and forgiving are part of a normative regime. When you forgive someone, you are making a promise, and if you do not keep it, it is the forgiven person who has a grievance. Whether a social system of honor constitutes a regime is controversial. Honor is supported by further norms—the members of society are supposed to ostracize a dishonorable person, and if they do not, they too will be punished. Their punishment is not necessarily a loss of honor, but it will take some aversive form, like criticism. Whether honor systems really qualify depends on whether one sees the attitude behind them as a moral concept or as one closer to the kind of self-esteem one has from being athletic or speaking several languages. Gerrard (1994) argues that the attitude supporting honor is close to an aesthetic one.

---

3. Krasner (1983) defines international regimes in general as "sets of implicit or explicit principles, norms, rules and decision making procedures around which actors' expectations converge in a given area of international relations." Kydd and Snidal (1993) give a summary of game-theory approaches.

The norms invoked when a norm is kept or violated can be called its *supporting norms*.<sup>4</sup> This leads to a puzzle: What keeps the supporting norms in place? If there is a hierarchy of supporting norms, then, there is either an infinite regress or a highest norm with nothing to back it up. In the analysis here, norms do not go off to infinity but rather form loops of support. Diagrams will show how this works.

### Normative Regimes for a Repeated Prisoner's Dilemma

The concept of a normative regime will be clarified by comparing it with an equilibrium of a repeated game. The game is a repeated Prisoner's Dilemma, and the equilibria discussed here are Always Defect, the Grim Strategy equilibrium, and Fight-and-Forget.<sup>5</sup> One criterion for a normative regime lies outside game theory and involves feelings of moral approval and disapproval. However other conditions can be stated within the theory and show some of the equilibria to be poor candidates for normative regimes. Always Defect cannot be one, the Grim Strategy is a middling candidate, and Fight-and-Forget is a good one.

#### A Repeated Prisoner's Dilemma

Suppose the Prisoner's Dilemma game of matrix 1 is played repeatedly, at times  $t = 1, 2, 3, \dots$  to infinity. Each player chooses to Cooperate (C) or Defect (D).

Matrix 1 is the *stage game*, and the whole sequence is the *supergame*. The

---

4. The concept has appeared under different names. Crawford and Ostrom (1995) refer to these as "monitoring" or "sanctioning" norms, on the grounds that people have a duty to watch and punish each other, and Axelrod (1986) terms them *metanorms*.

5. The well-known strategy of Tit-for-Tat (start cooperatively, then duplicate the other's previous move) is not one considered here. Tit-for-Tat generates an equilibrium that is typically not subgame perfect. That is to say, it prescribes non-utility-maximizing behavior in some hypothetical situations. (It still qualifies as an equilibrium since the situations that would yield this behavior never arise if the equilibrium strategies are played, so the player's payoff is not decreased.) If a player chose D at some point, the two would go back and forth: CD, DC, CD, DC, ... forever. Unless the game has payoffs fixed to give precise equalities at the choice points, either the Row chooser should not cooperate to produce CD, since Row knows that Column will defect, or, if Row is choosing C in hopes of prompting a return to the cooperative path, then Row should not choose D on the next move. One way or the other, Row is not playing optimally. It might be objected that if the players had stuck to Tit-for-Tat as they were supposed to, they would not reach this CD-DC cycle. However, each player's reason for avoiding D is based on the expectation of what would happen after choosing it, and Tit-for-Tat suggests that they expect this illogical play.

		C	D
		10, 10	3, 12
C	C	10, 10	3, 12
	D	12, 3	7, 7

**Matrix 1. The stage game, a Prisoner's Dilemma.**

rules are the standard ones: the players are told each other's move immediately after each stage game and before they choose their next moves. Future payoffs are discounted by a factor  $\delta$  between 0 and 1, meaning that 10 units received one period from now are valued equal to  $10\delta$  units received now, 10 units two periods from now are valued  $10\delta^2$ , and so on.<sup>6</sup> Each player's goal is to maximize his or her total discounted payoff. Thus, if the players' strategies resulted in cooperation at every stage, each would receive a stream of payoffs 10, 10, 10, … and would assess this at present utility  $10 + 10\delta + 10\delta^2 + \dots = 10/(1 - \delta)$ . For  $\delta = .9$ , a player would value this repeated game at 100. The whole game is thus defined by the stage game, the goal of playing, and the information known at each point.

### The Always-Defect Equilibrium

This equilibrium has each player use D on every play, no matter what has happened so far. Both get a stream of payoffs 7, 7, 7, … . This pair of strategies is an equilibrium for any discount factor, since no player could do better by switching unilaterally to the other move. The equilibrium is subgame perfect.

### The Grim Equilibrium

The Grim Strategy says to Cooperate if no one has chosen Defect so far. If there has been a D chosen, then use D forever. For matrix 1 both players using the Grim Strategy constitutes a subgame perfect equilibrium on the condition that  $\delta \geq .4$  and produces an indefinite string of CC pairs, with the payoff stream of 10, 10, 10, … . This equilibrium leads to mutual cooperation at as low values of  $\delta$  as any other equilibrium. Players have the greatest incentive to choose cooperation, since doing otherwise causes the worst consequence.

---

6. The discount factor can also be viewed as the probability that the game will continue after each play, as opposed to simply ending.

### Fight-and-Forget

The final equilibrium discussed can be termed *Fight-and-Forget*. It states:

*At the beginning of the game or if both Cooperated last time, Cooperate;  
if exactly one player Defected last time, Defect;  
if both Defected last time, Cooperate.*

In other words, the players Cooperate if they played identically last time, either with CC or DD. If they played differently, they get even, then get back to cooperating. A fight would last only a single round, but the negative prospect of this keeps them cooperating each time, as long as the discount factor is high enough. In the previous game, Fight-and-Forget is a subgame perfect equilibrium for  $\delta \geq \frac{1}{3}$ . The strategy has its advantages and disadvantages. For some Prisoner's Dilemma games it does not give an equilibrium for any discount factor. However, it is an equilibrium in some other types of games, including repeated Chicken. One of its merits is that it responds to a criticism of the Grim Strategy that if a player made a chance error in the Grim Strategy the two players would be stuck indefinitely in defections. With Fight-and-Forget, players who slip from mutual cooperation would return to it in two moves.<sup>7</sup>

The next section will amplify the concept of a normative regime, then these equilibria will be discussed for their properties as normative regimes.

### Normative Regimes as Game Equilibria

A normative regime and an equilibrium of a repeated game are similar in certain ways. Both tell people what to do in social situations, and both prescribe behavior that accords with incentives in some sense. A game equilibrium cannot prescribe suboptimal behavior for a player, and somewhat similarly, a culture whose norms urge people to take a certain course and that then punishes them for doing so would be under pressure to change. Also, game equilibria and

---

7. Fight-and-Forget seems to have been published first by Rapoport and Chammah (1965), who called it "Simpleton," on the grounds that getting one of the two better payoffs leads a player to repeat the strategy and getting one of the worse two causes a switch. A stochastic version goes by the name "Pavlov" for the same reason: that it tends to repeat a successful move (although Pavlov's theory of learning actually involved association rather than reinforcement, so a better name would be "Skinner"). Some biologists have found that it does well in computer simulations (Kraines and Kraines 1987; Nowak and Sigmund 1993). Binmore and Samuelson (1993) investigated its evolutionary development under the name Tat-for-Tit, in a version where the initial move was D to discourage the spread of a simpler mutant species that cooperates every time. Fudenberg and Maskin (1990) called it "Win-Stay, Lose-Switch."

normative regimes rules for the current situation are backed up by expectations of hypothetical situations: in a game, players stay on the equilibrium path because of the prospective consequences of going off it, and in a normative regime, people stay on the straight and narrow for fear of being punished. As with the common knowledge assumptions that are often made in solving games, group members recognize that others know the norms.

These are the similarities, but game equilibria and normative regimes are different in several ways.

*Norms are linked to feelings of moral approval and disapproval; equilibria do not necessarily involve such feelings.* Past definitions of a norm have included four elements, which have been emphasized to different degrees: behavior, beliefs, intentions, and moral attitudes.<sup>8</sup> Axelrod's definition (1986, 1097) stressed behavior, beliefs, and intentions: "a norm exists in a given social setting to the extent that individuals act in a certain way and are often punished when seen not to be acting in this way." Elster's definition involved belief and moral attitude (1989, 105): "A norm . . . is the propensity to feel shame and to anticipate sanctions by others at the thought of behaving in a certain way." The analysis here focuses on beliefs, behavior, and intentions, without negating the importance of attitudes like approval, disapproval, pride, guilt, or shame.

*Equilibria apply to actions chosen by people in situations; norms apply to types of actions chosen by types of people in types of situations.* In the repeated Prisoner's Dilemma, a norm would not just prescribe cooperation at a certain point in a particular game but rather cooperation in similar points in games of that type. Strictly speaking, norms apply to "game forms" that group situations, people, and moves. They are parallel to scientific laws in this regard. Moral obligation in a normative system corresponds to nomological necessity in an empirical one. According to Hempel (1965), a scientific law cannot refer to a specific place or time. To say that a certain force grows stronger near the North Pole, or as the year 2000 comes, may be a true statement, but it does not have the form of a possible law. Similarly, a norm cannot refer to specific situations or people. A social group might hold different norms for men and women, or for kings and commoners, but the norms cannot refer to a specific king or commoner. Norms must be personally neutral.

*Equilibrium behavior is never violated within the model, but normative behavior sometimes is.* Within a game model, choosing a nonequilibrium strategy is a contradiction, given the appropriate assumptions on the players' knowledge

---

8. Many definitions have been proposed in political science (Crawford and Ostrom 1995), and others are available elsewhere in the social sciences, in particular sociology.

and utility functions. Situations are grouped for norms, and sometimes two situations that end up together are strategically different. The norm will call for a certain type of action in that type of situation, but players may have different utilities for the outcomes and make a different choice. They would be violating the norm.

*A normative regime is socially beneficial; an equilibrium is not necessarily so.* In general, a social norm is supposed to benefit the group. In a repeated Prisoner's Dilemma game, the equilibrium of Always Defect cannot be termed a norm since it leads to inefficient payoffs.

*A norm is supported by the prospect of intentional reward or punishment by others; players stay with an equilibrium for its costs and benefits, but these do not necessarily arise through others' intentions to reward or punish.* Punishment means more than the violator losing benefit. Others must impose the loss deliberately because of their moral disapproval of the violation. Norms require a causal sequence:

- i complies or violates →
- others observe i's action and approve or disapprove →
- others reward or punish i.

This marks the difference between norms and the equilibria generated by many conventions, in Lewis's sense of the word (1969). Suppose Jim and Bob go to an amusement park, arranging that if they are separated they will go to the Ferris wheel. They do get separated, but for some reason Jim goes to the entrance booth, and they have a bad time for the rest of the fair. Jim's violating the convention did not induce Bob to go to the Ferris wheel to punish Jim. Bob had no intention to punish. Their agreement was a convention, not a norm. Simple conventions are *self-enforcing*, which is one reason that we do not hear of "regimes of conventions."<sup>9</sup>

### Diagramming a Normative Regime

The Fight-and-Forget equilibrium will be the first example. It can be shown to satisfy all the conditions for a normative regime that can be formulated within game theory. (If it were also true in a certain context that people acted with an

---

9. Behavioral definitions of norms, like Axelrod's and the present one, have others doing the punishing and rewarding. In Elster's definition and similar ones, the norm follower acts from conscience. The approaches are different, but their relation is clear. In some norms, socialization has internalized the punisher.

moral attitude, the situation would be a full normative regime.) First, one must define the action types and the situation types. A game equilibrium analysis would distinguish the moves available at the first stage of play from those at the second and third, but this normative analysis groups moves across stages. The natural choice for the two action types is {Cooperate, Defect}. As to situation types, a reasonable grouping has three: {Start of the game or two C's last time, one C and one D last time, two D's last time}. Fight-and-Forget can then be stated by giving the rule for each situation type:

- Start, or two C's last time: *choose C*;
- One C and one D last time: *choose D*;
- Two D's last time: *choose C*.

For a game, these are rules that define an equilibrium; for a moral situation, they are norms constituting a normative regime.

One can now draw figure 26, the *consequence graph*. Its nodes (circles) are norms. The diagram takes the viewpoint of one player and shows the results of that player's actions. By making a move, the player keeps the situation type as it was or shifts to a new one. The possible shifts are shown by arrows, of which there are two kinds: a solid one showing what happens if the player keeps the norm and a dotted arrow showing the result of violating it. Violating the norm "If two C's last time or the start of the game, *choose C*," for example, would move play to the situation "One D last time." The arrows are drawn under the assumption that the other player keeps the norm.<sup>10</sup>

Next is constructed figure 27, the *normative regime graph*. The nodes are the same, but the arrows have a new meaning—instead of consequences, they show incentives. If the graph represents a subgame perfect equilibrium, then keeping the norm at each situation is as good as or better than violating it. The incentive to keep or violate the norm is the immediate payoff from doing so, combined with the prospects of the next situation resulting from what one does (possibly the same one as before). A situation resulting from keeping the norm provides *inducement* to keep it, and violating the norm should move the individual to a situation that is less desirable and that, therefore, provides a *deter-*

---

10. Note that the graph is not a depiction of an equilibrium in that it makes the assumption that the other is following a certain strategy. It does not show one necessary element in an equilibrium, what would happen if both players violated the strategy. It resembles the diagrams of Moore machines of automata theory, used by Osborne and Rubinstein (1994) and others to portray evolutionary game strategies, but it is richer than these in that it includes what a player would do after that very player violated the recommended strategy. Moore diagrams do not show this, since automata are not supposed to make mistakes or think hypothetically.

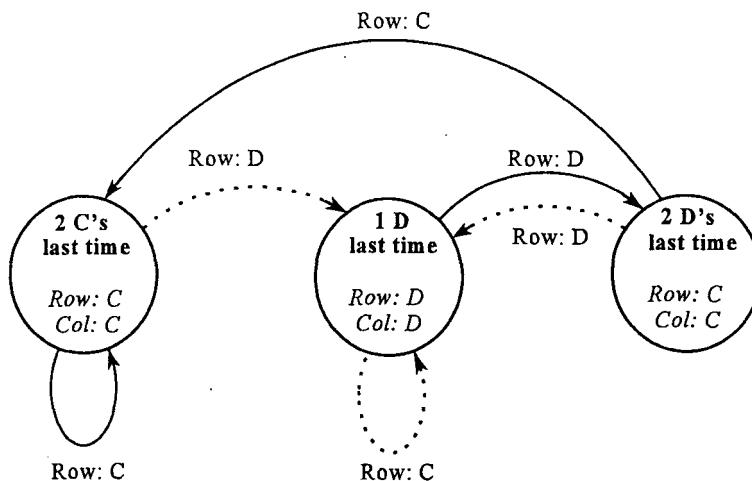


Fig. 26. Consequence graph for the Fight-and-Forget equilibrium. Norms are in circles, including the situations and their rules. The graph is from Row's viewpoint and assumes that Column acts normatively. Solid lines show the consequences of normative actions by the Row player, and dotted lines show those of counternormative actions.

rent that motivates the player to keep the current norm. Deterrent and inducement support are the two kinds of normative support, and they are shown by labeled arrows.

For the case of Fight-and-Forget, the normative regime graph, figure 27, involves taking the consequence graph, figure 26, and reversing all its arrows. Those arrows that represented normative action in a situation now provide normative support, and the ones that are counternormative move the situation to an undesirable one, and so provide deterrent support. At the middle situation, one player or the other has just chosen D. Row is now supposed to choose D, and expects Column to do the same. (As before the graph is from Row's viewpoint.) This will move Row to the third situation, where Row expects mutual cooperation. If Row does not act normatively in the middle situation, Row will stay there. Thus the third situation provides inducement support for Row to keep the norm in the middle, and the middle situation provides inducement support for itself.

Figure 27 shows how Fight-and-Forget satisfies the criterion listed for a normative regime in that each norm gets support from others. Norms are supported directly by a certain set of norms and indirectly by other norms further removed. A norm can be part of its own supporting set, either directly or indi-

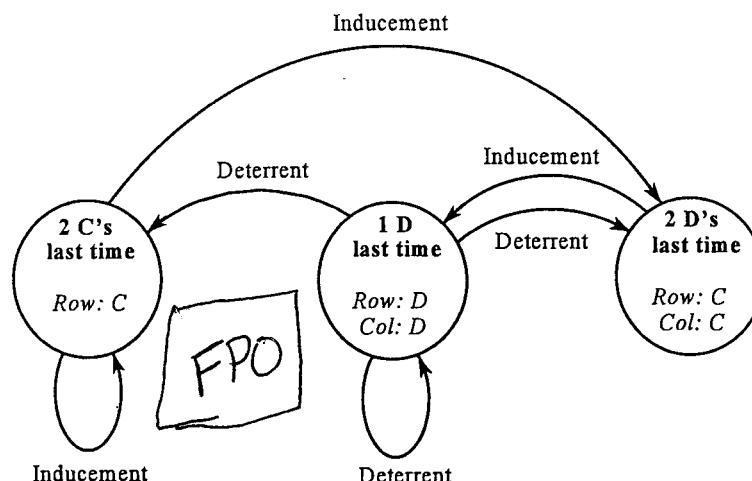


Fig. 27. Normative regime graph for the Fight-and-Forget equilibrium. Norms are in circles, with their situations and rules. Deterrent and inducement support relationships among the norms are shown for the Row player.

rectly, and this answers the earlier puzzle about how norms can support each other without an endless hierarchy. If the structure of normative support in a game is arranged properly to give correct incentives, then at equilibrium the norms can maintain each other in loops.

### Why the Always Defect and Grim Equilibria Do Not Yield Normative Regimes

The Always Defect equilibrium is an equilibrium for the repeated Prisoner's Dilemma, but it is not the basis of a normative regime. Going through the mechanics of constructing its normative regime graph shows why this is so. The graph is figure 28 (*bottom right*) and one finds rules that are self-supporting, not based on the prospect of the rewards and punishments of others. Both players are to play D, but it is not the prospect of future punishment that induces them to do it. As with Jim and Bob at the fairgrounds, it is the current payoff that makes them stay with D. Norms require others to punish a violator in response, and that is not happening here. No matter what a player does in the situation “At least one D,” the result is the same situation, so others are not differentially rewarding or punishing based on the player’s behavior. If Row were to choose C instead of D, Row would be sorry, but not because Column was punishing Row in reaction to the deviation. Column would have defected anyway.

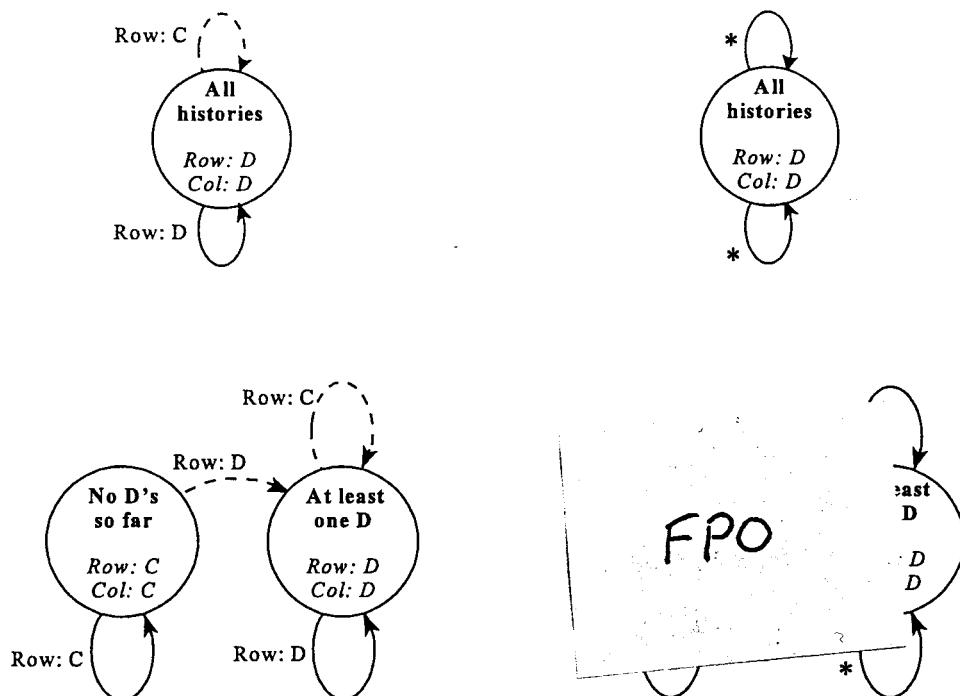


Fig. 28. Consequence graphs and resulting normative regime graphs for the Always Defect (top) and the Grim Strategy equilibria (bottom). In a normative regime, arrows marked (\*) would involve deterrent or inducement support, but here the situation always leads to itself and so cannot provide differential reward or punishment.

Regarding the qualifications of the Grim Strategy as a normative regime, first note that it has two situation types. One, “At least one D” is not supported by a norm—it is a subgame that is identical to the Always Defect equilibrium, and the preceding arguments hold for it. One cannot make the argument for the other situation, “No D’s so far,” so the Grim Strategy could be called a partial normative regime. Players continue to cooperate for fear of switching to the second kind of situation, that is, for fear of a regime breakdown.

#### A Normative Regime Using Optimal Penal Codes

Another example satisfies the conditions for a full normative regime and shows some more subtle ways that a normative regime can maintain its loops of sup-

		Low	Med	High	
		Low	10, 10	3, 15	0, 7
		Med	15, 3	7, 7	-4, 5
		High	7, 0	5, -4	-15, -15

**Matrix 2.** The stage game for Abreu's game, a Prisoner's Dilemma supplemented with a dominated row and column

port. It is an especially effective regime, since for the particular game, no equilibrium maintains cooperative behavior at a lower discount factor. Matrix 2, from Abreu (1988), is played repeatedly at  $t = 1, 2, 3, \dots$  with a discount factor  $\delta$ . Each time, the players choose among three levels of competition, Low, Medium, or High. The matrix is a  $2 \times 2$  Prisoner's Dilemma game with an added third move. Choosing Low corresponds to cooperation in a normal Prisoner's Dilemma, Medium is defection, and the new High level is an especially uncooperative move. Competing at the High level is so intense that it is dominated by the other two moves. On that account, this move would appear to do no good, but it will turn out to be important as a part of a punishment threat.

An equilibrium is defined by specifying each player's strategies, and here this is done indirectly, by defining *paths of play*. In Abreu's definition, a path of play is a sequence stating the outcome at each stage. Here three paths are necessary to give the equilibrium: the equilibrium path, which the players are supposed to start on, and two punishment paths, one for each player, which they move to following a player's unilateral deviation from the equilibrium path. A player's punishment path is also started if the player deviates unilaterally from a punishment path, either as the punisher or the violator. A punishment path requires certain moves from both the punisher and the violator—the latter must take a role in his or her own punishment. The motivation to cooperate in one's own punishment arises because the path is especially aversive at the beginning, so the violator would rather stay on it than restart it. The simplicity of Abreu's approach is that it does not tailor the punishment to the offense but gives one punishment that serves for all deviations. It is so effective because the punishment is the most severe one that is still credible.<sup>11</sup>

---

11. Fight-and-Forget has this property to some degree, since if either player balks at the mutual Defect outcome, the game lingers in an unresolved state until a mutual defection.

The equilibrium is then as follows.

Players start on the *equilibrium path*, which is the series of outcomes when both play Low forever, giving the stream of payoff pairs: (10,10), (10,10), (10,10), ...

If the Row player deviates unilaterally from the current path, players implement *Row's punishment path*, (-4,5), (3,15), (3,15), (3,15) ...

If the Column player deviates unilaterally from the current path, they implement *Column's punishment path*, (5,-4), (15,3), (15,3), (15,3) ...

Joint deviations are ignored—the current path is continued.

The two-phase character of the punishment paths, with an initial severe punishment and a second part that continues indefinitely, is somewhat like Prisoner's Dilemma's Fight-and-Forget equilibrium, except that there is no forgetting. To make the punishment more powerful, a deviation has permanent consequences. This equilibrium holds for discount factor  $\delta \geq \frac{4}{7}$ , and, Abreu shows, no other equilibrium does better. Using the new row and column makes the difference. If one were to ignore them and treat the game as a  $2 \times 2$  Prisoner's Dilemma, cooperation could be maintained only down to  $\delta = \frac{5}{8}$  by use of the Grim Strategy. The particular punishment paths are *optimal penal codes*, in Abreu's vocabulary. The paths are defined so that a player has an incentive to stay on the current path at any point. As figure 29 shows, there are five norms and each one draws support from one or more of the others, so the equilibrium is a full normative regime. In this example, a normative regime is used to produce maximum cooperation.

## Moral Authority

An important question for peace and justice involves normative change. Norms against slavery, racism, biological weapons, and other evils were not around from the start but developed. Other norms have disappeared, like those against submarine attacks on merchant ships or the bombing of civilian targets (Legro 1997). Some mechanisms of normative change are precedent (Kier and Mercer 1996), argumentation (Crawford 1993), changing influences originating in culture (Legro 1997), and the self-serving definition of the norms by the hegemonic state. The preceding model suggests another mechanism: an actor gains normative prestige by good behavior and can then specify further norms.

An example of a normative regime in the context of nuclear proliferation shows how this might work. The presentation will be informal, without setting

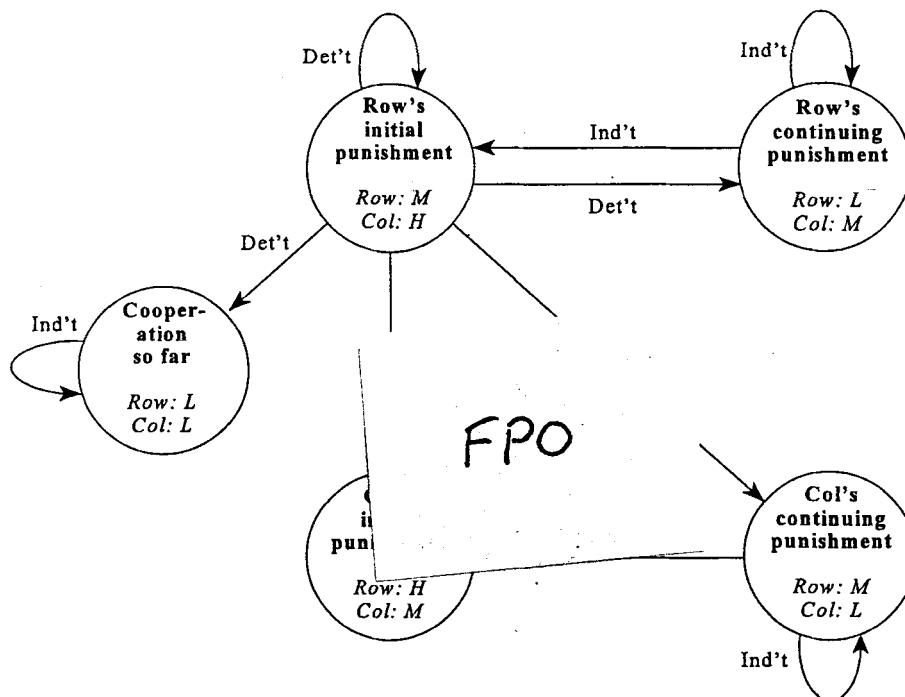


Fig. 29. Normative regime graph for optimal penal codes in matrix 2. The graph is from Row's viewpoint and assumes Column behaves normatively.

up the full model. States randomly confront situations, and for each situation, the prevailing norm telling the state what to do can be Fuzzy or Clear. States are divided into three groups: those of High, Middle, and Low normative prestige on proliferation. States can move up or down in prestige. They move downward in two ways: for those who are High or Middle, committing a Clear violation puts them into the Low group, and for those in the High group, committing a Fuzzy violation puts them in Middle. States can move upward in prestige in two ways: those in the Low group rise to the Middle by some random process, as long as their recent record has been good on Clear norms. A Middle state moves to High in the same way, as long as its recent record has been good on Clear and Fuzzy norms. The final step is to state the incentives for being in each group. The advantage of being High is that the Highs determine appropriate behavior when any state is confronted with a Fuzzy norm. They get to say what is a violation, and they benefit from having this moral authority. Those in the Low group are pariahs. Some of the situations that states meet are opportunities for

pairwise interactions, for example, the sharing of technical information. It is a Clear violation to deal with someone in the Low group.

The key to the analysis of a system like this is the normative regime graph, since it shows an equilibrium's properties as a regime if it is one and shows what is missing if it is not. The examples before this one were unusually tight—in the Prisoner's Dilemma games and in Abreu's game, norms called for a unique action at each situation, and there were no chance events. This example includes these other elements. The normative regime graph must be defined in a different manner, since doing it the previous way would produce something very complicated. First, situations are grouped by ignoring what kind of nation the norm is being applied to. A norm may have different consequences depending on one's prestige group, but the different possibilities are collected in one node. The arrows too are generalized. They now mean that there is some circumstance in which the one norm supports the other (as opposed to the situation in figs. 28 and 29, where it means support at every situation), and they also do not distinguish deterrent from inducement support. The result can be called a *generalized normative regime graph*.

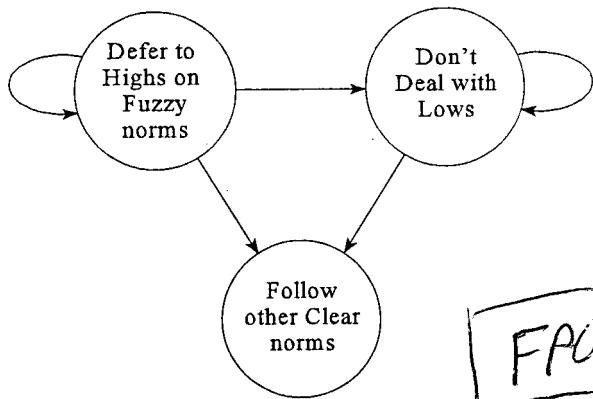
The graph for the proliferation regime is shown in figure 30. It contains five arrows. "Defer to Highs" supports itself because following it sometimes lets a state stay in the High group and keep its moral authority. "Don't Deal with Lows" supports itself since violating it puts a state in the Low group or keeps it there. "Don't Deal with Lows" is supported by "Defer to Highs" since keeping the former norm may allow a state to move into the High group. One cannot say the reverse—that arrow is absent—since not deferring to the authority of the Highs on a Fuzzy norm does not move a state into Low. A state gets motivation to "Follow other Clear norms" (those other than avoiding Lows and deferring to Highs on Fuzzy cases) from the other two norms, and this fact accounts for the final two arrows.<sup>12</sup>

The generalized graph gives a more practical way of looking at support relations and checking for the properties of a normative regime. The activity of constructing it helps one consider the full set of relationships and possibilities.

The benefit of having a High group with this power is that they can sometimes lead the whole group to a better equilibrium. The institution of moral authority can solve the problem of coordination in a group by selecting a leader. At the 1995 conference to renew the Non-Proliferation Treaty, the South African and Canadian representatives became the active leaders (Bunn, personal com-

---

12. One element of moral authority is missing from the model—it does not state that a decision of the Highs on a Fuzzy norm establishes a norm thereafter—but one would expect this to be true by virtue of the precedent set on the occasion.



**Fig. 30.** Support relationships among norms for a regime based on moral authority and exclusion. The support can be deterrent, inducement, or both, depending on whether the state they apply to is High, Medium, or Low.

munication 1995; Dunn 1995). South Africa had been the first nation ever to give up nuclear weapons and had further prestige for its rejection of apartheid and the biracial character of its delegation. Canada had rejected nuclear weapons in spite of having had the technology to make them. The two countries jointly took on roles of proposing initiatives and lobbying the other delegates. Canada collected signatures on a list of countries willing to extend the treaty, hoping to generate a bandwagon effect, and South Africa led the non-aligned nations to a compromise with the nuclear nations. Their diplomatic skills were important, but their prestige on the relevant issues created an expectation that their leadership would be followed.

Honor, Symbols, and War  
Barry O'Neill  
<http://www.press.umich.edu/titleDetailDesc.do?id=14453>  
The University of Michigan Press