

## CHAPTER 8

---

### Commitments Based on Honor

A constant problem for national leaders is to make their commitments believable. One way is bridgeburning—visibly eliminating the option of backing down (Schelling 1960). In the 1990 Gulf crisis, when George Bush stationed several hundred thousand troops in Saudi Arabia, he made it clearly harder for himself to drop his demands on Saddam Hussein and bring the troops home. Another technique for credibility is to leave the implementation of one's threat to chance or to someone else's control. During the Cold War, U.S. policy was to respond with nuclear weapons to a Soviet nonnuclear attack on Western Europe, but implementing this threat would have risked the destruction of the United States. To make it more credible, German troops were trained in the use of some tactical nuclear weapons. In a crisis the United States could hand over its codes, turning Germany into an instant nuclear power.

These are the dramatic extremes. The usual way that leaders commit themselves is simply by their words. Someone who backs away from a clear statement will lose future credibility and reputation (Pitt-Rivers 1968; James 1986).

*(E9) Honor cultures frequently include the institution of oath taking, which allows protection to be extended to interests beyond those already specified as points of honor.*

This chapter describes a mechanism of verbal commitment based on honor. (Chap. 10 will discuss how the same function can be achieved by manipulating one's own social face.) The first model involves making promises on one's honor and possibly keeping them. Two plausible equilibria arise, reflecting the fact that the rules of promising vary by culture. Honor-based commitments have other functions than threat credibility: making a credible promise allows one to make an alliance, or an agreement that settles a conflict. A second model deals with the kind of commitment that one makes to an ally, which, in

contrast with a promise, imposes an obligation to act only when it is challenged. The model of commitments based on honor is contrasted with recent game models of deterrence and crisis bargaining.

Keeping a commitment is supposed to show that one values honor and to create an expectation that one will keep other commitments in the future. The advantages of commitment making entail a problem. Why should the group draw the conclusion that future commitments will be kept, when any person, honorable or not, would benefit from developing a reputation for honor? This question will be treated here. Other issues are the following: Can a state make a commitment to an entity that is seen as outside the honor group of national states, like the United Nations or the Palestinian people? Which kinds of societies develop honor-based commitments and which do not, and what does this suggest about the kinds of countries that use commitments?

### **The Difference between Commitments, Promises, and Threats**

Schelling (1960) states that a promise is costly when it succeeds and a threat is costly when it fails.<sup>1</sup> The definitions here will be slightly different. My promise means that I will do something you apparently want, and a reason will be that performing the communicative act obliges me to do it. The message incurs the obligation not necessarily because of moral reasons, particularly in the case of a threat. The connection might arise from my sense of honor, the law, people's opinions, or some other element of the context.

**DEFINITION:** *X's commitment* to Y to action A is a communicative act whose meaning is that X will do A and that X has an obligation to do A by virtue of performing the communicative act. *X's promise* to Y is a commitment where X believes that Y wants A done. *X's threat* to Y is a commitment where X believes that Y does not want A done.

A commitment to defend an ally is both a promise and a threat—it is a promise to that ally and a threat to the potential attacker. It involves two receivers in the role of Y: one wants the commitment fulfilled, and the other does not. The action A in a threat is usually conditional, to be undertaken only if some other event happens, typically a deed by Y; promises can be conditional as

---

1. See also Schelling (1989) and the philosophical literature on promising and threatening (e.g., Downie 1985, Seligman 1995.) Klein and O'Flaherty (1993) use game models to analyze the difference between promises and threats.

well. It might be claimed that a promise also has a moral element, involving the promiser's obligation to keep it. This enters the definition as possible grounds for the receiver to believe the promise will be kept. There may be nonmoral grounds for the same inference—the promiser might feel constrained by public opinion or by the law.

### Promising on One's Honor

Figure 16 shows a simple game of promising as a step to the more complicated model of commitment making. If you make a promise, the receiver gives you benefit  $b$ , but keeping your promise costs you  $c$ . Breaking the promise will not take away the benefit but might diminish your perceived honor. Only one player has a move.

- STAGE 1: The player knows his sense of honor  $h$ .
- STAGE 2: The player makes or does not make the promise, the former producing a utility increment  $b$ .
- STAGE 3: A player who has made a promise chooses whether to keep it or not; keeping it produces a utility decrement of  $c$ . The audience revises its estimate of the player's honor, according to whether the promise was made and kept, made and broken, or not made.

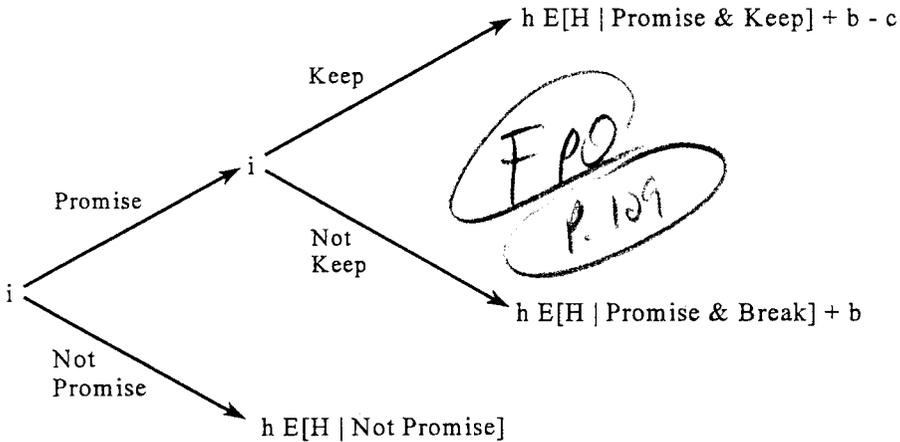


Fig. 16. Individual  $i$  making a promise based on honor and keeping it or not

8.4

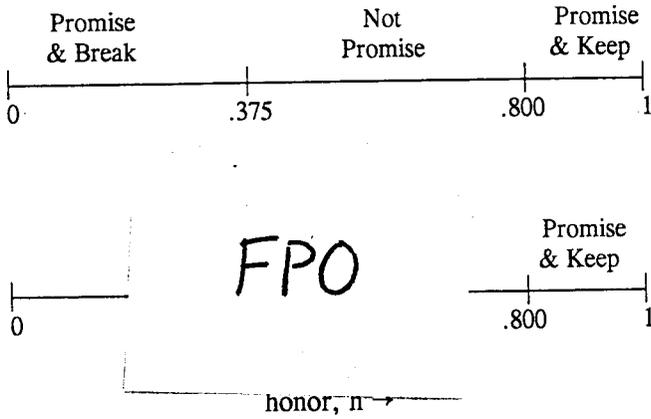


Fig. 17. Two equilibria in a game of promising

To derive some results, it is assumed that  $h$  is selected from a uniform distribution on  $[0,1]$ , that if the audience sees an action inconsistent with the expected equilibrium, it assigns zero honor (this is usually not essential), and that the utility parameters are  $b = .3$  and  $c = .4$ . The game has several equilibria, but some can be skipped because they render promises unbelievable. At one, for instance, promises are always broken but a player still gets benefit  $b$  for making one. Requiring some credibility for a promise leaves two equilibria (fig. 17), derived in appendix C. In one equilibrium, individuals of low honor ( $0 \leq h < .375$ ) make it and break it; those of middle honor ( $.375 \leq h < .800$ ) do not make it; and those of high honor ( $.800 \leq h \leq 1$ ) make it and keep it. In the other, everyone makes the promise; those of low honor ( $h < .800$ ) break it, and those of high honor ( $h > .800$ ) keep it.<sup>2</sup>

The existence of two equilibria means that the institution of promising is consistent with different social customs. Promises are more credible in the first society, where people sometimes refuse to make them: 35 percent are kept there, compared to 20 percent in the second society. Why do some people refuse to promise? We are familiar with positions based on moral principles, like refusing to take a loyalty oath, but the motive here is different. The player is simply

2. The two thresholds are  $b/2c$  and  $2c$  in the first case and  $2c$  in the second. As before, equilibria that differ only in behavior at the endpoints are not distinguished.

weighing the benefit of making the promise against the reputational costs of breaking it and deciding that overall it pays not to promise.

### **Committing Oneself to Defend an Interest**

Commitments have been prominent in U.S. foreign policy (Jentleson 1987), but game models have tended to pass them over.<sup>3</sup> The model deals with using commitments to deter injuries to one's interests. Defender (D) holds a prize desired by a potential Attacker (A) and tries to deter A by announcing an honor-based commitment to defend the prize. The game tree is shown in figure 18.

STAGE 1: D knows his honor  $h_D$ , and A knows his value  $v_A$  for the prize; each is uncertain about the other's value. D is commonly known to value the prize at  $b$ ; A is commonly known to be unconcerned with honor.

STAGE 2: D makes a commitment to defend the prize or does not make one.

STAGE 3: If D commits, A can attack or not; if A does not attack, then D keeps the prize.

STAGE 4: If A attacks, both lose cost  $c$ , and D chooses between defending or not; if D defends, both suffer another decrement  $c$  and get the prize with equal probability.

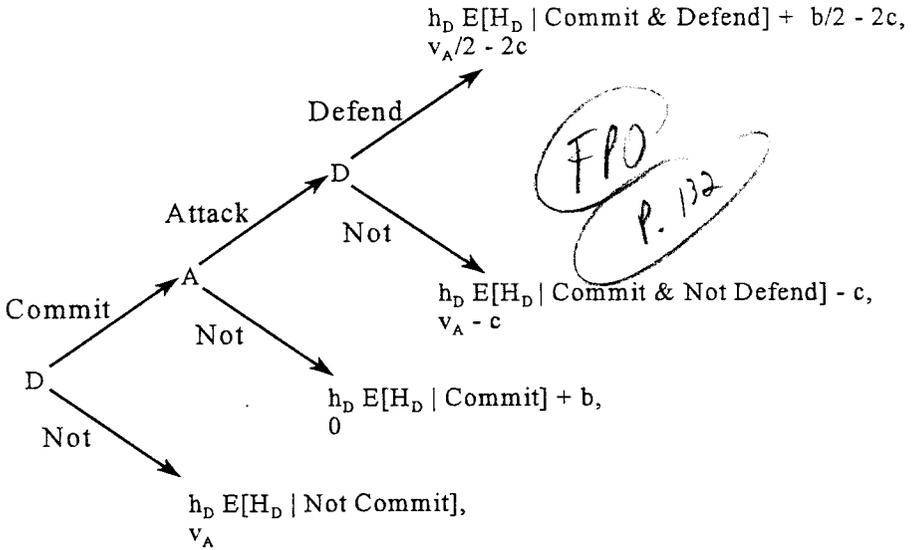
STAGE 5: The audience reevaluates A's honor in the light of A's observed behavior: committing or not and defending or not.

The assumption that player D is uncertain about A's prize value when it decides whether to make a commitment, seems reasonable in that commitments can be made long before the crisis, perhaps before the committer knows who the adversary will be.

For some parameter values, the equilibria in the game lead one to always making a commitment no matter what one's honor or perhaps to never making one. The values here are chosen to yield a more interesting equilibrium. The Defender's honor is uniformly distributed on  $(0, 1)$ , the Attacker's prize is uniformly distributed on  $(0, 2)$ , and  $b$  and  $c$  are set at  $.4$ . One equilibrium, derived in appendix C, has Defenders of low honor ( $0 \leq h_D < .078$ ) not committing,

---

3. Some international relations authors use the distinction between *situational* and *nonsituational* commitments (Weinstein 1969). The former arise because of preexisting objective facts—a country feels compelled to defend its source of oil. For the latter, which are the kind of interest here, the state chooses to make a commitment and binds itself by its declaration.



**Fig. 18. A model of deterrence through commitments based on honor. Uncertainty involves the Defender's honor and the Attacker's value for the prize.**

those of middle honor ( $.078 \leq h_D < .434$ ) committing but not defending if attacked, and those in the high range ( $.434 \leq h_D < 1$ ) making a commitment and defending if attacked. The first group lets the prize go from worry about an attack and disinterest in their reputation; the middle group makes a commitment for the sake of reputation and deterrence but backs down when attacked; the high group defends the prize for its value and for their honor. Attackers with prize value less than .932 do not attack in the face of a commitment, and those with value between .932 and 2 attack.

### Models of Commitment versus Crisis Signaling Models

The commitment-on-honor approach can be contrasted with crisis signaling models, sometimes called deterrence or crisis bargaining models. They usually involve the following elements (O'Neill 1993): (1) The players are an Attacker A looking for a gain and a Defender D preferring the status quo. (2) A can initiate the crisis, and D then resists or not. (3) A is unsure how motivated D is to

fight, but D's initial response resolves some of this uncertainty. (4) This evidence influences A's further decisions of whether to press the attack.

The crisis signaling model begins after a crisis is under way, started by A's initial probe. Each player starts with uncertainty about the other's value for the prize, and this becomes less and less as the player observes the other's moves. In the language of the literature, D's resistance shows that D is "tough" or "hard" or "resolved." However, each side's continuing carries a cost each time. An equilibrium tells the player when to stop depending on his value for the prize.

This kind of generic crisis signaling model can be compared with the honor-based commitment approach. The model focuses on setting up deterrence, not responding to a deterrence failure. In the latter, the commitment decision is made before D knows that there will be a crisis; it is made to prevent one. Both approaches depend on incomplete information about the other side's values but the locus of the uncertainty about the Defender is different. In the crisis signaling model, it is the value for the prize, or sometimes the cost of a conflict. In the commitment model, it is D's honor. Crisis signaling models rely on the argument for costly signaling that willingness to sacrifice cost shows one's motivation. The honor model makes D's initial move the announcement of a commitment, which in itself is cost free unless it is challenged. It reveals D's incentive to defend the interest, but, different from the standard models, it also increases that incentive. This is bridge burning more than costly signaling.

A question for the crisis signaling approach is the following: Why should A take D's earlier resistance as evidence of D's resolve on the main issue? Why does it not reveal D's unexpectedly high utility for winning the earlier conflict, or low cost for engaging in it, or perhaps its optimism about winning on it? The answer is simply that the model is set up that way—A is assumed to be uncertain about only one thing, D's value for the prize, so D's resistance can convey information on only that issue. The conclusion is logical within the model, but in reality there would be many sources of uncertainty, and historical analyses of international signals have found them to be repeatedly misinterpreted (e.g., Thies 1980; Mercer 1996). It would be bad methodology to include all of the uncertainties—the game would be so complicated that one could learn nothing from it—but it is worth considering alternatives like the honor-based commitment model.

### **Real versus Virtual Honor**

If honor pays, then anyone would want an honorable reputation. That puts any claim to honor in doubt, since when a state sacrifices by keeping a commitment,

onlookers may suspect that the real motive was the practical benefits of appearing honorable. This section distinguishes two ways in which the quest for honor becomes impure and discusses how one can tell them from the real thing.

To represent an honorable reputation that carries practical benefits, one can start with the basic game of honor. The individual had two choices:

incurring the cost, with utility  $h E[H | S] - c$ ,

not incurring it, with utility  $h E[H | N]$ .

Here  $E[H | \text{—}]$  is the audience's estimate of  $H$  given what they observe, and  $S$  and  $N$  stand for making some Sacrifice, like keeping a commitment, or Not Sacrificing, respectively. Finally,  $c$  is the cost of the sacrifice.

If an honorable reputation also has a self-interest benefit, these payoffs can be modified to

$$(h_i + k) E[H_i | S] - c$$

and

$$(h_i + k) E[H_i | N].$$

(Here the subscript is added to show that  $h_i$  is specific to the individual  $i$  and  $k$  is not.) Honorable people value external reputation the same as others, it is assumed, so the multiplier is a constant  $k$ , the same for all individuals and commonly known, the importance of the self-interest benefit of being seen as honorable. Assuming a uniform distribution on  $h_i$  on  $(0,1)$ , there is a threshold equilibrium: Sacrifice if  $h_i > 2c - k$ . This holds as long as that quantity is non-negative; otherwise one finds only a pooling equilibrium. Compared to the proportion  $1 - 2c$  in the original game, the proportion of people who sacrifice now is  $1 - 2c + k$ , which is higher than before. The revised estimate of the honor of a sacrificer is  $\frac{1}{2} + c - k/2$ , which is lower. Introducing a practical benefit of being seen as honorable means that more people will sacrifice, and that sacrificing will do their reputations less good, since their act is attributed partly to expediency.

Some individuals would not have sacrificed except for a nonzero  $k$ , and they get a higher reputation than they deserve. The difference in reputation could be termed *virtual* honor. It feeds off true honor in the sense that if the audience were sure that the society were honorless, that  $h_i = 0$  for all  $i$ , sacrificing would

make no difference and no one would pursue virtual honor.<sup>4</sup> Society must believe that true honor is not dead for virtual honor to play a role. In international relations the distinction between true and virtual honor is revealed when there is no tomorrow, as in Herman Kahn's scenario of suicidal nuclear retaliation (chap. 6). No doubt much honor behavior involves caring for the practical benefits of a reputation, but elements of others' behavior and our own introspection suggest that the true version is present. For simplicity, true honor alone is included in the models, with the suggestion that the general conclusions would hold true if the other type were added.

### **Commitments to Those outside the Honor Group**

Gerrard (1994, 6) relates a story from Harriet Jacobs, writing about her life as a young slave in the American South. Her grandmother had loaned money to the mistress.

[She] had laid up three hundred dollars, which her mistress one day begged as a loan, promising to pay her back. The reader knows that no promise or writing given to a slave is legally binding; for, according to Southern laws, a slave, being property, can hold no property. When my grandmother lent her hard earnings to her mistress, she trusted solely to her honor. The honor of a slave holder to a slave!

The debt was not repaid. Element E2 indicates that honor is like group membership and E8 that challenges from outside can be ignored. A further element can be added.

*(E10) Commitments to those outside the group are not binding.*

Some states have been disinclined to keep commitments to parties that were not also states in the system. Backing out can be a delicate matter, since one has to explain why the commitment was made in the first place. One way is to put the refusal to keep the commitment in obscure language. An example, related by Robert Keohane, involved an 1814 controversy between the United States and Britain over the terms of the Treaty of Ghent. Hoping to limit U.S. access to Canada, the British diplomats proposed a condition that the boundaries that the United States had accorded to native peoples in a 1795 treaty

---

4. A different model of how "honor pays" posits that people want to be seen as believing that an honorable reputation has practical benefits (as opposed to the model of virtual honor where they really do believe that). The results are much the same as in the original model.

would be made permanent. John Quincy Adams, Henry Clay, and other U.S. cosigners refused, on the grounds that the United States was determined to buy the tribal lands. The explanation they sent to Britain seems self-contradictory, asserting native rights from the earlier treaty but suggesting that the members of the tribes were American subjects, so they would have to give up their lands: “the United States, while intending never to acquire lands from the Indians other than peaceably, and with their free consent, are fully determined, in that manner, to bring into cultivation every portion of the territory contained within their acknowledged boundaries. . . . If this be a spirit of aggrandizement, the undersigned are prepared to admit, in that sense, its existence; but they must deny that it affords the slightest proof of an intention not to respect the boundaries between them and European nations.” The United States would not consent to “arresting their [the United States’] natural growth within their own territories, for the sake of preserving a perpetual desert for savages (*American State Papers* 1814, vol. 3, 1719).” Not to say that we do not keep our word, but there are treaties with European equals and treaties with “savages.”

Recent events suggest that violating a treaty is easier if the treaty partner does not have membership status in the group of states. One can compare worries about treaty compliance with the Soviet Union and with the United Nations. In the mid-1980s the U.S. administration was eager to conduct research and development activities that would have violated the Anti-Ballistic Missile Treaty with the Soviet Union. The administration pointed to a Soviet treaty violation, the construction of a large radar site at Krasnoyarsk, but was not willing to abrogate the treaty on that account. In the public debate and in Congress, much emphasis was placed on keeping one’s treaty obligations, but it was proposed to reinterpret the treaty’s language with a unilateral declaration that it actually permitted what had generally been seen as forbidden.

This contrasts with U.S. dealings with the United Nations. In 1994, Congress declared that the United States would pay no more than 25 percent of the cost of UN peacekeeping. This contravened the dues schedule of 31 percent, which followed a formula that the General Assembly had set pursuant to the UN Charter, a U.S. treaty obligation. Concern about keeping this treaty was not reflected in the congressional debate and no excusing rationale was developed. The United Nations, not itself a nation, did not generate a sufficient duty to keep commitments.

### **Honor-Based Commitments and Offensive Advantage**

The discussion so far has suggested that honor serves the purposes of assurance and deterrence. However, it may not be the best way to do that, since it leads to con-

tinual disputes. Is it a self-perpetuating trap? Do single members feel forced to follow it, as Franklin implied, even though all would benefit by simultaneously abandoning it?

One reason to suspect that there are better alternatives than honor is that often one finds two cultures side by side, with comparable economies and environments, one using honor and the other not. The only explanation would seem to be their histories (Nisbett and Cohen 1996). In the United States, Yankee practicality met the Southern culture of honor. Greenberg (1990, 1996) tells the story of the "Feejee Mermaid," P. T. Barnum's invention of the 1840s, the torso of a monkey sewn onto the body of a fish, which came out looking like a tiny grotesque centaur. When Barnum exhibited the mermaid in Northern states, the public debated whether it was a hoax and took the possibility that Barnum was trying to fool them with humor more than resentment. When he brought his marvel to the South, the controversy took a dangerous turn. In one newspaper debate over its authenticity, each party took offense not at the other's opinion but at the lack of appropriate respect in his expressing it. The controversy became self-referential and almost led to a duel.

What is it about a culture, then, that induces a system of honor? Nisbett and Cohen, extending an idea of Campbell's (1964), suggest that honor develops in societies where the formal institutions of law and punishment are weak and where a person's livelihood can be stolen suddenly and irreversibly. These conditions often hold in societies based on animal herding rather than farming. The theft of livestock can mean sudden and thorough ruin, so one has to deter a transgression rather than try to fix it afterward. A private system of deterrence substitutes for law and requires drawing a wide boundary around one's rights and cultivating an expectation of vengeance for its own sake. The same conditions, especially the absence of law enforcement, hold in U.S. inner cities, they argue.

*(E11) Honor systems often arise in societies without an effective law enforcement system, where members face harm to their interests that is sudden, disastrous, and hard to rectify.*

An analogy can be made to the international system. There is no supernational body to enforce justice. When the military environment gives an advantage to taking the offensive, nations look to deterrence rather than defense. An emphasis on national honor may then be a response to an offensive advantage, a variable that some international relations theorists have seen as crucial in a nation's security policy (Questaer 1977; Jervis 1978; Snyder 1984; van Evera 1984). During the Cold War, for example, the United States' worry with credi-

bility was not that the enemy would attack the U.S. homeland—that seemed unlikely—but that U.S. distant interests in Europe, Asia and the Middle East, might be “gobbled up” overnight. The United States took up commitments and other honor-related behavior because it was in the position of a herdsman instead of a farmer.