Game-Theoretic Interpretations of Commitment

by
Jack Hirshleifer*

Working Paper Number 799
Department of Economics
University of California, Los Angeles
Bunche 2263
Los Angeles, CA 90095-1477
December 20, 2000

*Chapter draft for forthcoming volume edited by Randolph M. Nesse, *Natural Selection and the Capacity for Commitment* (Russell Sage Press).

GAME-THEORETIC INTERPRETATIONS OF COMMITMENT

Abstract

Two types of commitment need to be distinguished. Pre-emptive commitment corresponds to taking the opening move in sequential play. Reactive commitment occurs when the decision-maker who will be acting last pledges to respond, in a specified contingent way, to the opponent's earlier choice. Both types of commitment can be regarded as pre-play moves, as distinguished from the parties' actual execution moves that generate payoffs.

Having the first move is not always desirable. There is a first-move advantage when the parties' interests are relatively harmonious, as in intra-alliance negotiations (epitomized by the payoff environment BATTLE OF THE SEXES), but a second-move advantage when strong opposition of interests exists (as in the constant-sum environment LAND OR SEA).

Reactive commitments take the form of promises, threats, or combinations thereof. Threats and promises can only be effective if credible, though 100% credibility is not essential. If a threat and promise can be combined, their respective credibilities can reinforce one another.

There are rational and non-rational mechanisms of commitment. Rational commitments take the form of unilaterally reducing one's own payoffs from the options not to be chosen. On the non-rationalistic level, a psychic predisposition toward impulsiveness supports making pre-emptive commitments (taking the first move), whereas avoiding the first move can be aided by a predisposition toward dithering and procrastination. For reactive commitments, "hot" anger serves to deter undesired behavior by making the threat of a punishing response credible, and passionate gratitude may serve the same function by guaranteeing a reward for desired behavior. In evolutionary terms, these attitudes might be triggered in contexts in which the indicated behavior would indeed be profitable if rationally thought through.

GAME-THEORETIC INTERPRETATIONS OF COMMITMENT

Jack Hirshleifer

In his contribution to this volume, Thomas C. Schelling disclaims having originated the concept of commitment. Even if that disclaimer is accepted, a strong case remains for recognizing him as inventor at least of the game-theoretic approach to the commitment problem (Schelling 1960). In this paper I follow Schelling's lead, employing the tools of game theory in the hope of providing more precision and rigor to a concept often discussed rather too loosely.

More specifically, I will be addressing the following topics:

(1) How does the concept of commitment relate to the categories of game theory: strategies, payoffs, moves, information sets, and so forth?

(2) Is commitment one single thing, or have two or more concepts been conflated together that game theory can help us distinguish?

(3) In what environments and under what "rules of the game" is commitment possible? Where commitment is possible, when is it profitable as well -- to the committing party, to the targeted party, or possibly to both of them together?

(4) Finally, although this topic has considerable overlap with discussions elsewhere in the volume, I will be going beyond pure theory to say something about the mechanisms whereby a party's commitment can be effectuated and communicated.

The title of this volume refers to the evolution of commitment. It is plausible, though not certain, that when commitment is profitable it will also be favored by selection -- natural or cultural. However, I will not be attempting to model the evolutionary process whereby commitment actions might or might not become part of the behavioral repertory of a population.

I. PAYOFF ENVIRONMENT VERSUS PROTOCOL OF PLAY

The discussion that follows presumes familiarity with basic concepts of game theory such as

payoffs, strategies, moves, and the like. However, I will expand on one point not usually brought out in textbook presentations: the distinction between the payoff environment and the protocol of play (or "rules of the game").[1]

Most readers will be acquainted with payoff matrices like those illustrated in Table 1 and Table 3. The names of several such matrices, among them PRISONERS' DILEMMA, CHICKEN, and BATTLE OF THE SEXES, have in some cases entered the common lingo of academic discourse. Each of these patterns can be regarded as a distinct ecological environment within which social interactions take place. The protocol of play is a somewhat less familiar concept. The idea is that, within any given environment, in order to have a properly structured game the players must be constrained to follow a definite set of procedures – the "rules of the game". The rules might provide, for example, only for a one-time encounter after which play is terminated and the contenders go their separate ways. Or alternatively, perhaps the players will be interacting with one another two or more times in succession. In the latter case the choice options becoming available in later rounds might well influence decisions in round one. However, only single-round interactions will be dealt with in this discussion.

[Table 1 about here]

What will be essential for the analysis here is a different aspect of the protocol of play: the sequence of moves. Do the rules require the players to make their choices simultaneously, or do they take turns? And if they take turns, who moves first? (Note: In game theory, simultaneity refers not to clock or calendar time, but to the state of information. So long as neither side, when making its own move, is aware of the opponent's choice their two actions are regarded as simultaneous.)

---

[1] See Schelling 1960, 46. An exposition appears in my textbook (Hirshleifer and Hirshleifer 1998, 284-291). Although just about every work on game theory covers alternative payoff environments, systematic treatments of alternative protocols of play are relatively rare. The most thorough analysis is in the work of Steven J. Brams (Brams 1994).

II. WHAT IS COMMITMENT?

"/C/ommitment is a device to leave the last clear chance to decide the outcome with the other party" (Schelling 1960, 37). So the concept of commitment turns quite centrally upon the protocol of play, and in particular upon the sequence of moves. However, there are at least two distinct ways in which the "last clear chance" can be left to the opponent. Schelling distinguishes between commitment pure and simple, and commitment that takes the form of a threat:

> The threat differs from the ordinary commitment, however, in that it makes one's course of action conditional on what the other player does. ... The commitment is a means of gaining first move in a game in which first move carries an advantage; the threat is a commitment to a strategy for second move. (Schelling 1960 124)

What Schelling calls "ordinary commitment" corresponds to taking the opening move in sequential play ("seizing the high ground" in military usage). Or, equivalently, forcing the opponent to take the final move. I will call this pre-emptive commitment. To constitute pre-emption, a player's opening action must be irrevocable: as long as he keeps his options open, he is not committed. (In fact, in game-theory terms he has not yet moved at all.) Pre-emptive commitment, taking the first move, is not necessarily profitable. One topic to be taken up here concerns the conditions under which a player might prefer not to move first, but to "have the last word" instead.

Schelling is too limiting, however, in referring to the second type of commitment as "threat". Not only threats but also promises, and perhaps certain other types of conditional engagements as well, fall into a general category I will call reactive commitment. Reactive commitment occurs when the decision-maker with the last move convincingly pledges to respond, in a specified contingent way, to the opponent's earlier choice.

To avoid possible confusion it is helpful to distinguish between what I will call "execution moves" and "pre-play moves". Consistent with Schelling's usage, the terms "first move" and "last

move" here will always refer exclusively to the execution moves -- the choices that actually generate the payoffs. In contrast, a commitment comes even earlier: it is a pre-play move. A pre-emptive commitment is a pre-play move that allows one to take the first execution move. In contrast, a reactive commitment, although also a pre-play move, can be made only by the player who has the last execution move. In either case, by giving up his freedom of choice the committing player leaves the opponent with (in Schelling's words) "the last clear chance to decide the outcome".

In Table 1 consider the matrix associated with the game of CHICKEN.[2] In each cell of the matrix the first number is the payoff to the Row player, the second the payoff to the Column player. (These are to be interpreted as ranked rather than quantitatively scaled payoffs.) Row's highest payoff 4 is in the lower-left cell; Column's 4 is at the upper-right. CHICKEN takes its name from a hypothetical test of nerve. Two teenagers are steering their jalopies toward one another at full speed; the first to swerve aside must suffer the scorn of being the chicken. Maintaining the ornithological metaphor, the alternative strategy options are often termed Dove (the R2 and C2 strategies for Row and Column respectively) and Dove (the R1 and C1 strategies). As the matrix shows, Row as Hawk does best if Column chooses Dove, the achieved payoff-pair being (4,2). But if both sides choose Hawk lots of feathers fly (the jalopies crash). The payoffs are (1,1) – each side suffers its worst possible outcome.

Consider labor-management negotiations. Hawk corresponds to readiness to go on strike or to impose a lockout (as the case may be) if one's demands are not met. Dove represents willingness to come to a reasonable compromise. The side playing Hawk will get the better of an opponent playing Dove; a tough negotiator might win big. But if both sides insist on playing Hawk, the result may be a mutually undesired work stoppage.

---

[2] More precisely, the matrix displays only the payoff environment of CHICKEN. A fully defined game requires specifying not only the payoff environment but also the protocol of play.

Suppose the Row player can make a pre-emptive commitment, that is, he seizes the chance to make the first execution move. Then he should logically play Hawk (the lower row in the matrix). Column must rationally respond with Dove (the left-hand column of the matrix). The strategy-pair is [R2,C1] with payoffs (4,2). The advantage is to Row -- his pre-emptive commitment has paid off.

What does such a commitment entail, behaviorally speaking? The key point is that a committing player irrevocably surrenders freedom of action. So long as the teenager driver retains his option to swerve or not, he has not yet actually made his move. How can he give up the option to swerve? Schelling has suggested one possibility: demonstratively tossing the steering wheel out of the speeding vehicle! In labor-management negotiations, a union leader might become committed by making himself vulnerable to being voted out of office unless the contract satisfies his publicly proclaimed demands. (I will be discussing these and other possible mechanisms for achieving commitment later on.)

So much for pre-emptive commitment, seizing the first move. Suppose instead that, in pre-play negotiations before any execution move has been made, Column can make a reactive commitment. There are a number of different possible reactive commitments, but let us assume Column convincingly conveys to Row that: "If you choose Hawk I will respond with Hawk; if you choose Dove I will respond with Dove." Now, although Column will be making the last execution move she has given up her freedom of action: it is Row who has the "last clear chance" to determine the outcome. So, when the time comes to take action, Row should rationally choose Dove. After Column carries out her commitment by playing Dove in response, the achieved payoffs will be (3,3). Thus Column's reactive commitment has won her equality of payoff, despite Row's being able to move first in action terms. (Is Column's commitment here a threat or a promise? As will be seen later, it has elements of both.)

III. PRE-EMPTIVE COMMITMENT: WHEN IS IT PROFITABLE?

Pre-emptive commitments -- ways of seizing the first move -- can be profitable for the committing side in one of two ways: (1) They may allow the committing side to gain a relative advantage over the opponent, or else (2) conceivably, both parties may benefit. Which of these outcomes eventuates, if either does, depends also upon the payoff environment.

The payoff matrices of Table 1 are ordered in terms of decreasing opposition of interests. Of the patterns shown, the parties' goals are the most discordant in LAND OR SEA and the most harmonious in the COORDINATION GAME.[3] More specifically:

LAND OR SEA (LOS): This constant-sum pattern represents total disharmony of interests. Each side's gain is the mirror-image of the other side's loss. The title recalls the Paul Revere story. A mismatch of strategies would have been best for the British. If attacking by sea they would prefer the Americans to defend by land; if attacking by land they would like to have the American forces lined up to defend by sea. Paul Revere's task was to ensure this did not happen. Biological predator-prey interactions represent a somewhat similar situation. The predator wants to prowl wherever the prey locates itself, the prey hopes for the reverse.[4]

CHICKEN (CH): This payoff environment has already been described. On the one hand both parties want to avoid the worst-for-both (1,1) payoffs associated with the [R2,C2] strategy-pair. Apart from that each side prefers a particular discoordination outcome: best for Row is the strategy-pair [R2,C1] with payoffs (4,2), but this is next-to-worst for Column. The reverse applies to the strategy-pair [R1,C2] with payoffs (2,4). The Chicken payoff pattern is characteristic of

---

[3] A quantitative scale for measuring divergence of interests is proposed in Axelrod 1970, Chs. 2-3.

[4] Note that I am speaking of constant-sum patterns defined only in terms of ranks. Even when the ranks are constant-sum, the actual quantitative payoffs need not be. Victory or defeat may have differing significance for the two sides. E.g., if the prey escapes, the predator loses a dinner; if the predator wins, the prey loses its life (Dawkins 1982, Ch. 4.)

bargaining in the shadow of potential conflict, as in labor-management negotiations, provided that actual conflict is regarded as the worst possible outcome by each side.[5] The strategy-pair [R1,C1] might seem like a reasonable compromise, but may not be easy to achieve since neither of the contenders wants to be perceived as a soft bargainer.

BATTLE OF THE SEXES (BOS): Here the players' interests are mainly in accord, though not perfectly so. Of the two mutually advantageous patterns, strategy-pair [R1,C2] slightly favors Row (payoffs 3,2) whereas [R2,C1] favors Column (payoffs 2,3). The BOS pattern is typical of choices arrived at by alliances. Members of an alliance share a strong common interest in defeating the common enemy, but each partner might want to jockey for advantage against its associate. In the traditional story, the two allies are husband and wife going out for the evening. Although they have a strong desire for togetherness, one prefers they both attend the opera while the other favors a boxing match. As a military example, after breaking out of the Normandy perimeter in 1944 the British and Americans needed to concentrate logistically upon a single line of attack against Germany. The Americans preferred supporting a drive by General Patton in the south, whereas the British favored supplying Field Marshal Montgomery in the north. (The latter was the action chosen, seemingly the wrong decision as it led to an Allied disaster at Arnhem.)

COORDINATION GAME (CG): Here there is perfect correspondence of interests: the players' payoff ranks are equal in each cell. Both parties do best if they can coordinate on the strategies [R1,C1] with payoffs (3,3), next best coordinating upon [R2,C2] with payoffs (2,2), but worst of all (1,1) if their strategies fail to match. Examples of CG situations are choice of a common language, of systems of measures and weights, of rules of the road. In some cases it makes no

---

[5] If a contender prefers "death with honor" to "surrender" the mutually destructive [R2,C2] strategy-pair is not the worst possible outcome for him or her. Should this be true for both players, the payoff environment is converted from CHICKEN to PRISONERS' DILEMMA.

difference which particular pattern of coordination or convention is chosen, provided that one of them is. Everyone driving on the right may be just as good as everyone driving on the left. But in the environment postulated by the CG game, by assumption one pattern of coordination is better for both sides. E.g., the agreed convention of changing over to Daylight Saving time in the summer season appears to be more efficient than remaining on standard time.

We now ask, in what circumstances can pre-emptive commitment be profitable, to one side or conceivably to both? For each of the four payoff environments we have been considering, Table 2 indicates: (1) the degree of opposition of interests between the two players, (2) the strategy-pair and numerical payoffs arrived at in rational play (assuming it is the Row player who commits to moving first), and (3) where the advantage lies.

[Table 2 about here]

In game-theoretic language, each of the outcomes shown is the Subgame-perfect Equilibrium (SGPE) for the corresponding game. This equilibrium concept is based on the hypothesis that the first-mover will be making his rational (payoff-maximizing) choice, on the assumption that the second-mover will be doing the same when it comes to her turn.

Setting aside the COORDINATION GAME, where there is no conflict of interest at all, the table suggests an inverse association between opposition of interests and first-mover advantage. Or putting it the other way, it pays to pre-emptively commit to the first move when the conflict of interest is only slight but not when it is strong. Intuitively, when the opposition of interests is strong, a second-mover doing best for herself will also be automatically harming her opponent. Whatever the first-mover's strategy choice might be, he remains vulnerable to such a countermove by the opponent. So in LAND OR SEA, the environment with the most extreme opposition of interests, it is vital to have not the first but the last move. In the Paul Revere story, for the British to proclaim a pre-emptive commitment to attacking by sea (or by land) would have been fatal for their chances of victory, and

similarly for the Americans to commit to defending by land or defending by sea.

In contrast, if the parties' interests are highly correlated, the second-mover cannot punish her opponent without injuring herself to some degree as well. Thus in BATTLE OF THE SEXES, where the there is only slight opposition of interests, an ally having the first move can profitably commit to his more advantageous option -- because, rather than letting the mutual enemy win, his confederate does better accommodating to his choice as best she can.

Without getting into the mathematical and logical considerations making for strong or mild opposition of interests, notice that in the relatively "friendly" BATTLE OF THE SEXES environment the parties share an interest in avoiding either of the two discoordination outcomes in which both of them incur the worst possible payoffs (1,1). CHICKEN is less "friendly" an environment, but nevertheless the parties still share one strong interest: avoiding the Hawk-Hawk strategy-pair with its (1,1) payoffs. So again there is a first-mover advantage. But in LAND OR SEA, there is no strategy-pair they <u>both</u> want to avoid: any combination of strategies that is bad for the one side must be good for the other. Hence the advantage shifts to the second-mover.

Finally, in the COORDINATION GAME there is no opposition of interests at all. Here allowing <u>either</u> party to have the first move leads both sides to the Pareto-efficient payoff-pair (3,3). So if the baseline situation for purposes of comparison were the simultaneous-play protocol, either side should be happy to convert play of the game to a sequential protocol, no matter which side is empowered to make the pre-emptive commitment (move first).[6]

The upshot is that ability to move first tends to be of <u>differential</u> advantage in interactions with slight (BOS) or moderate (CH) opposition of interests between the players, and it is in those

---

[6] Under the standard Nash equilibrium concept for the simultaneous-move protocol, there are three possibilities: convergence upon the Pareto-superior strategy-pair [R1,C1], upon the inferior coordination pattern [R2,C2], or adoption by both sides of mixed strategies (which imply some likelihood of total coordination failure). Only the first of these three possibilities is as efficient as the outcome under sequential play.

environments that we would expect the ability to pre-empt to be cultivated (or to emerge by evolutionary selection). Ability to move first is also of advantage when there is no conflict of interest at all, although the advantage redounds to the benefit of both players. With very strong conflict of interests, as in the constant-sum (LOS) game, we would expect to see contention for the position of second-mover.[7]

IV. THREATS AND PROMISES

We now turn to the second commitment concept. Reactive commitment typically takes the form of a threat or a promise (or combination thereof). The protocol of play goes as follows: first comes the commitment itself (a "pre-play move"), then the first execution move by the target player, and finally the reactive execution move by the committing player.

To illustrate the nature of threats and promises, consider the PRISONERS' DILEMMA (PD) and CHICKEN (CH) payoff matrices in Table 3. In each case the first row and column represent the "more cooperative" action, which it will be convenient to symbolize here as c, and the second row and column the "less cooperative" action denoted as d. As before I will be using the Subgame-Perfect Equilibrium concept (SGPE): first-mover will be making a rational (own-payoff-maximizing) choice, in anticipation that second-mover will be doing the same in her turn. But under the reactive-commitment protocol it is the second-mover (the Column player) who makes the pre-play threat or promise -- the first actual execution move remaining with the Row player.

[Table 3 about here]

Promise:

In the PRISONERS' DILEMMA context, absent any commitment possibility for Column, if

---

[7] Baik and Shogren (1992) analyze a somewhat different aspect of the possible contention between decision-makers as to which will be the first-mover.

Row moves first the SGPE would be the [d,d] strategy-pair at lower right with payoffs (2,2).[8] To escape this mutually unsatisfactory outcome, the committing Column player might announce a strategy in which her play is contingent upon Row's opening move, as follows:

(a) If Row cooperates (chooses c), my reply will be c -- payoffs (3,3)

(b) If Row defects (chooses d), my reply will be d -- payoffs (2,2).

Row, if he regards Column's conditional commitment as credible, will then rationally prefer option (a). Part (a) of Column's strategy is a promise. It is a commitment by Column to choose her more cooperative action c, leading to the payoff-pair (3,3), even though action d leading to payoffs (1,4) would be more advantageous for her in terms of immediate profitability. But Part (b) of her announced strategy is not a threat, properly speaking – because responding to d with d is just what Column would be doing in any case, even if there were no question of commitment.

Threat:

Threat commitments stand out more clearly in the CHICKEN environment. Given the CH matrix of Table 3, in the absence of any reactive commitment on the part of Column the SGPE would be the [d,c] strategy-pair at lower left, with payoffs (4,2). Row, having the first action move, comes out ahead. But Column can reverse the advantage if she has the power to commit to a threat. She does so by announcing the following strategy:

(a) If Row chooses c, my reply will be c -- payoffs (2,4)

(b) If Row chooses d, my reply will be d -- payoffs (1,1).

Assuming credibility, Row of course prefers option (a). Part (b) is Column's threat: her commitment to take action d leading to payoffs (1,1) even though, in terms of immediate profitability, action c leading to payoffs (4,2) would be more advantageous for her. But Part (a) here is not a

---

[8] Actually, in PD the non-cooperative [d,d] strategy-pair is the equilibrium outcome for the sequential protocol regardless of who moves first, and for the simultaneous-play protocol as well.

promise, properly speaking – since responding to c with d in the CHICKEN environment is just what Column would be doing in any case, even if there were no question of commitment.

Note that if a promise succeeds in eliciting the desired behavior then -- absent deception -- it requires costly performance by the committing party. But a threat, if it works, does not have to be carried out. (Thus, Row, if he accedes to it, would not be able to determine whether in fact the threat was real and would actually have been fulfilled.)

Threat-and-Promise:

Finally, a reactive commitment might involve both a threat and a promise (stick and carrot). Still in the CHICKEN environment, the committing Column player announces the strategy:

(a) If Row chooses c, my reply will be c -- payoffs (3,3)

(b) If Row chooses d, my reply will be d -- payoffs (1,1).

Once again Row prefers option (a). Here Part (b) is the same threat as before. It is a commitment on Column's part to take action d leading to payoffs (1,1) even though, in terms of immediate profitability, action c leading to payoffs (4,2) would be more advantageous for her. But Part (a) now is indeed a promise – Column does not respond to c with d leading to payoffs (2,4), which would maximize her immediate profit, but instead commits to respond to c with c leading to payoffs (3,3). Here, if the commitment was truthful and it works, only the promise but not the threat needs to be carried out.

One might wonder why Column would ever commit to a threat-and-promise, since it does no more for her than a simple promise and does less for her than a simple threat. As a possible explanation, consider the problem of credibility.

Credibility

To be effective, a commitment must not only be made but conveyed: the targeted player must

find the threat or promise to be credible (Klein and O'Flaherty 1993). But 100% credibility may not be necessary.

Returning to the PRISONERS' DILEMMA matrix of Table 3 and the associated promise commitment, recall that the purpose of Column's pre-play move was to induce Row to shift his first action from the less cooperative d strategy to the more cooperative c strategy. Let us suppose now that Row does not find the promise fully credible, and indeed only assigns the likelihood k_P to its being faithfully executed (on a scale from 0 to 1). How high does the partial credibility k_P have to be for the promise to be effective?

We can readily calculate: If Row accedes to the inducement (and chooses the upper row c of the matrix), his expected payoff will be the average of his reward over the two possible outcomes:

$$V_R(\underline{c}) = 3k_P + 1(1 - k_P)$$

If he does not respond to the promise and continues to choose the lower row d of the matrix, his payoff is simply:

$$V_R(\underline{d}) = 2$$

In the latter case Column will simply pursue her own advantage on the final move, so the parties end up at the lower-right cell of the PRISONERS' DILEMMA matrix with payoffs (2,2).

Simple calculation shows that $V_R(\underline{c}) > V_R(\underline{d})$ if and only if $k_P > ½$. Thus, given the illustrative payoff numbers of the PRISONERS' DILEMMA payoff environment in Table 3, a credibility of only 50% is required for the promise to be effective.

Dealing in exactly the same way with the CHICKEN environment, for the specific numbers employed in Table 3 Column's pre-play threat will be effective -- that is, Row will rationally accede to it rather than defy it -- if and only if its level of credibility is $k_T > 2/3$.

That brings us finally to the rationale for a reactive commitment ever taking the form of a threat-and-promise. In general, these two elements of the conditional commitment might have distinct

degrees of credibility k_P and k_T. If Row chooses the c strategy (accedes to the promise and threat), Column will carry out her promise and his payoff will be:

$$V_R(\underline{c}) = 3k_P + 2(1 - k_P)$$

If he chooses not to accede, Column will carry out her threat so that the expected payoff from his d strategy will be:

$$V_R(\underline{d}) = k_T + 4(1 - k_T)$$

The first expression allows for the partial credibility of Column's promise; the second allows for the partial credibility of Column's threat. It follows that Row will rationally choose his c strategy if and only if:

$$k_T + k_P/3 > 2/3$$

Recall that, to be effective, the threat alone required credibility $k_T>2/3$. So the credibility of an associated promise may to some extent make up for a deficiency in credibility of a threat. The upshot is that a threat-and-promise may be effective in circumstances where a threat alone would not be.

V. MECHANISMS OF COMMITMENT

What is it that makes commitments feasible and (wholly or partially) credible?

The essence of commitment is foregoing certain options, closing off a portion of one's own choice set. Just how can that be done? In dealing with these issues it will be useful to distinguish -- although the terms are somewhat unsatisfactory -- between rationalistic and non-rationalistic considerations. Under the former heading, patterns of payoffs and/or protocols of play might exist (or might perhaps be created by an appropriate pre-play move) in which rational players would be enabled to make a commitment. Under the latter heading, it may be that certain non-rational psychic patterns, notably the "emotions", are capable of being exploited for this purpose. That is, to guarantee that one or more otherwise available actions would not in fact be chosen.

Rational Commitments

The supporting mechanisms are much the same for both the pre-emptive and the reactive forms of commitment. To begin with, consider first the CHICKEN game (in which, as we have seen, there is a first-mover advantage). A famous technique for achieving pre-emptive commitment, for gaining the first-mover advantage, is burning your bridges behind you. Suppose the options are attack (the "less cooperative" d move) versus retreat (the "more cooperative" c move). Burning your bridges behind you amounts to telling the opponent: "Since you now know I cannot retreat, you can be sure of having a costly fight on your hands unless you give way." In game-theory terms, you have drastically reduced your own payoff from choosing the c or retreat strategy -- or even made it absolutely impossible. This having been visibly achieved, the opponent faces a fait accompli: pre-emption is effective. In contrast, there is no motivation for pre-emption in PRISONERS' DILEMMA, where there is neither a first-move nor a last-move advantage. And certainly no motive to do so in LAND OR SEA where the advantage lies with the second-mover. Similarly, effectuating reactive commitment requires penalizing oneself for failure to carry out the associated threat or promise.

Various techniques that might guarantee performance of a commitment have been suggested by Schelling [1960] and others. I can only briefly describe a few of them here.

As already mentioned, in labor-management negotiations (the CHICKEN payoff environment) a union leader might pre-emptively commit to a tough strategy by promising his constituency, in a pre-play move, that he will resign his position if he does not get his way. Knowing that the opposing negotiator is so constrained, management may have to accept a bad bargain as an alternative to an even worse work stoppage. As a related possibility, a negotiator might make a side-bet with some third party that he will not give way – thus reducing his own payoffs in the event of his failing to stand fast.

Also of interest are the reverse techniques, pre-play actions that commit the player not to make the first execution move. The U.S. Constitution provides that a treaty signed by the President becomes

effective only if ratified by the Senate. At least when facing an opponent not so constrained, and provided that a second-move advantage exists, an ratification requirement might strengthen the President's hand in treaty negotiations. As previously indicated, a second-move advantage is associated with payoff environments characterized by strong conflict of interests, e.g, the LAND OR SEA payoff environment. So we would expect the Senate to be insistent upon retaining this power over negotiations with hostile nations -- and the President ought not object. On the other hand, where the conflict of interest is not nearly so strong, as in international trade negotiations, the Senate is more likely to grant the President freedom to conclude Executive Agreements without requiring ratification.

In a pre-play contest for sequential advantage, differing time-urgencies may play a role. Supposing there is a second-move advantage, a player may so arrange matters that he is in no great hurry to terminate the interaction by an execution move. So an opponent under time-pressure may be forced to take the first move. In battle encounters where the advantage lies with the defense, a better-supplied general may be able to stand and wait until the opponent risks going on the attack.

When it comes to reactive commitments, pledging one's fortune or reputation as surety for your promise or threat is very much like making a side-bet that you will not default. Giving hostages or placing monetary collateral at risk also serve the same function.

Beyond Rationality

Stepping outside the bounds of rationality, there are possible psychic predispositions that may promote the ability to pre-emptively commit (take the first move) or the reverse. Some people are by nature impulsive, rushing in where angels fear to tread. But then there are the ditherers and procrastinators. ("/T/he native hue of resolution is sicklied o'er with the pale cast of thought" – Hamlet.) Is it possible that, without rationally thinking it through, people have evolved an ability to turn on the former psychic mode when there is a first-move advantage, and the latter when the balance

of benefit tilts the other way? I will have to leave that question open here.

Turning now to reactive commitments, it has been proposed that what are called emotions -- or more properly passions, such as anger and gratitude -- may have evolved to guarantee the execution of threats and promises (Hirshleifer 1987, 1993, Frank 1988). Under the influence of passion, an individual may no longer calculate profitability; he or she "loses control".

The passions are intrinsically reactive in nature; they are triggered by the other party's prior move. Outrageous behavior, even on the part of your beloved child, may precipitate a furious response. On the other hand a surprising act of chivalry, even on the part of an enemy, may trigger an unwarrantedly magnanimous rejoinder. Economic experiments have shown that subjects are indeed typically willing to respond favorably to generous actions and to punish ungenerous ones, even when not "rationally" warranted in terms of benefits and costs to themselves (Fehr and Gachter 2000).

VI. SUMMARY

Pre-emptive commitment corresponds to taking the opening move in sequential play, or equivalently forcing the opponent to take the final move. To constitute pre-emption, the opening action must be irrevocable; the opponent must face a fait accompli. Reactive commitment occurs when the decision-maker who will be acting last pledges to respond, in a specified contingent way, to the opponent's earlier choice. Both types of commitment can be regarded as pre-play moves, as distinguished from the parties' actual execution moves that generate payoffs.

Having the first move is not always desirable. There is a first-move advantage when the parties' interests are relatively harmonious, as in intra-alliance negotiations (epitomized by the payoff environment BATTLE OF THE SEXES), but a second-move advantage when strong opposition of interests exists (as in the constant-sum environment LAND OR SEA).

Reactive commitments take the form of promises, threats, or combinations thereof. The

possibility and potential efficacy of reactive commitments also depend upon the nature of the payoff environment. In the PRISONERS' DILEMMA environment a credible promise can generate a gain for the committing player (and the target player also benefits). In Chicken a promise would be ineffective, but a credible threat would gain an advantage for the committing player. (Here, however, the target player loses thereby.)

Threats and promises can only be effective if credible. In general, something short of 100% credibility may suffice. If a threat and promise can be combined, their respective credibilities tend to reinforce one another in influencing the targeted player.

Mechanisms of commitment can be divided between the rationalistic and non-rationalistic. A player might rationally pre-empt by foreclosing all but his desired choice option ("burning one's bridges behind you"). This corresponds to unilaterally reducing one's own payoffs from the options not to be chosen. To do the same for reactive commitment, the payoffs would have to be selectively reduced; that is, the committing player would have to make it unprofitable for herself not to carry out the promise or threat previously made.

On the non-rationalistic level, taking the first move might be supported by a psychic predispositions toward impulsiveness, and avoiding the first move by a predisposition toward dithering and procrastination. It remains to be seen whether these attitudes are triggered in contexts in which the indicated behavior would indeed be profitable if rationally thought through. When it comes to reactive commitment the picture is somewhat clearer. "Hot" anger serves to deter undesired behavior by making the threat of a punishing response credible, even if executing that punishment is not actually materially profitable. On the positive side, passionate gratitude may serve the same function by guaranteeing a reward for desired behavior.

References

Axelrod, Robert. 1970. *Conflict of Interest*. Chicago: Markham Publishing Co.

Baik, Kyung H. and Jason F. Shogren. 1992. "Strategic Behavior in Contests: Comment." *American Economic Review* 82(1): 359-362.

Brams, Steven J. 1994. *Theory of Moves*. Cambridge: Cambridge University Press.

Dawkins, Richard. 1982. *The Extended Phenotype*. Oxford: Oxford University Press.

Fehr, Ernst and Simon Gachter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90(4): 980-994.

Frank, Robert H. 1988. *Passions within Reason*. New York: Norton & Co.

Hirshleifer, Jack. 1987. "On the Emotions as Guarantors of Threats and Promises." In *The Latest on the Best: Essays in Evolution and Optimality*, edited by John Dupré. Cambridge MA: MIT Press.

Hirshleifer, Jack. 1993. "The Affections and the Passions: Their Economic Logic." *Rationality and Society* 5(2): 185-202.

Hirshleifer, Jack and David Hirshleifer. 1998. *Price Theory and Applications*, 6th edition. Upper Saddle River NJ: Prentice Hall.

Klein, Daniel B. and Brendan O'Flaherty. 1993. "A game-theoretic rendering of promises and threats." *Journal of Economic Behavior and Organization* 21: 295-314.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. London: Oxford University Press.

Table 1
2x2 Matrices, by Decreasing Opposition of Interests

Matrix 1: LAND OR SEA

	C1	C2
R1	2,1	1,2
R2	1,2	2,1

Matrix 2: CHICKEN

	C1	C2
R1	3, 3	2, 4
R2	4, 2	1,1

Matrix 3: BATTLE OF THE SEXES

	C1	C2
R1	1,1	3,2
R2	2,3	1,1

Matrix 4: COORDINATION GAME

	C1	C2
R1	3, 3	1, 1
R2	1, 1	2, 2

Table 2
Equilibria for Matrices of Table 1

	LAND OR SEA	CHICKEN	BATTLE OF THE SEXES	COORD-INATION GAME
Opposition of Interests:	Complete	Substantial	Slight	None
Equilibrium strategy-pair:	[R1,C2]	[R2,C1]	[R1,C2]	[R1,C1]
Equilibrium payoffs:	1,2	4,2	3,2	3,3
Advantage to player moving:	Second	Second	First	Neither

Table 3
Threats and Promises in Two Payoff Environments

PRISONERS' DILEMMA

	c	d
c	3,3	1,4
d	4,1	2,2

CHICKEN

	c	d
c	3, 3	2, 4
d	4, 2	1,1