COOPERATION AND INSTITUTIONAL ARRANGEMENTS

Kenneth A. Shepsle
Washington University, St. Louis
December, 1985

Prepared for the Harvard Conference on International Regimes and Cooperation. Endicott House. Dedham, Massachusetts. February 13-15, 1986.

COOPERATION AND INSTITUTIONAL ARRANGEMENTS

Kenneth A. Shepsle
Washington University, St. Louis
December, 1985

The formal study of cooperation is something of a cottage industry these days in political science, primarily because it is central to so many vital issues. These issues, of course, are not new. Rousseau's famous "stag hunt dilemma," Hume's "worry over commons problems and public goods supply (as in neighbors draining a meadow), and Hobbes's generalized concern over how human societies might avoid the dire consequences of life in the state of nature, all suggest that voluntary cooperation in social settings is a commodity in considerable demand and, presumably, In short supply. Some, like Hobbes, have made this disparity between the demand for and the supply of voluntary cooperation the basis of an elaborate rationale for authoritative coercion In the form of the State.

The Hobbesian solution to the cooperation problem—institutions of coercion—is a severed coercion is often more than is necessary (though, writing in an historical period of anarchy and violence, it may have seemed to Hobbes like the only hope). While I shall not dwell here on some three hundred years of discourse on Hobbes (for which 1 am distinctly ill-equipped in any event), it may be observed that a number of late 18th and 19th century scholars and men of practical affairs, Adam Smith being perhaps the most eminent, argued in a very different manner. For them, institutions of coercion were part of the problem, not part or the solution. Guarding the guardians of coercion is not a trivial matter, since officials in state institutions have interests of their own, interests the pursuit of which do not necessarily further broader social purposes. These "anti-Hobbesians" argued, instead, in behalf of **enforcement institutions** for agreements voluntarily arrived at, not institutions of coercion. Such institutions would encourage and permit voluntary cooperation by reducing the costs of negotiating such agreements, by assuring costless, reliable, disinterested enforcement of such agreements, and by delivering swift and certain punishment in the event of cheating or reneging.

The idea of third-party exogenous enforcement is a powerful one because it permits one to study cooperation processes without dwelling on enforcement aspects (which are assumed to be perfect). The law of contracts and the neoclassical economics of exchange are grand intellectual edifices built on this foundation. By suppressing the details of enforcement—indeed, by assuming away all such problems—the obstacles to voluntary cooperation vanish. Whenever there are prospective gains from cooperation (in economics, in the form of production and exchange), they are always captured. In short, there are no costs to transacting—to negotiating, coming **to** agreement, and enforcing the terms of the agreement.

Put this way, however, it should be apparent that the foundation is a shaky one. In the real world, enforcement institutions are costly to employ, are not entirely reliable, and need not be disinterested. Contemporary students of transactions costs (Coase, Barzel, Cheung, North) are critical of neoclassical formulations precisely because they suppress such considerations. Some details of this critique are articulated in the paper prepared by Douglass North for this conference.

Partially in response to criticisms against theories based either on institutions of coercion or institutions of enforcement, there has been much interest recently in whether such institutions are absolutely essential. Might there not, it is asked, be mechanisms that encourage cooperation with only limited reliance on courts and judges, guns and police? That is, are there circumstances from which cooperation arises endogenously as a by-product of otherwise self-interested behavior? Does cooperation "emerge" or "evolve" (Axelrod, 1981, 1984)? In anarchic settings in which third-party enforcement is infeasible or impractical, for example the criminal underworld (Laver, 1980), the underground economy, or the international political environment (Keohane, 1984), these questions are often regarded as central.

****

In the remainder of this paper I describe four institutional features which may affect the prospects for endogenous cooperation: sequence, interconnection, reputation, and repeat play. I argue, though only in a preliminary way since the hard research has not yet been done, that each of these characteristics may have a salutary effect on the prospects for cooperation in conflict-of-interest settings. Since repeat play has received considerable attention in the game theory literature—and has recently been brought to the attention of political scientists in the elegant work of Axelrod—I treat it first. By turning, then, to reputation, interconnection, and sequence—which I will define more formally below—I want to emphasize that repeat play is but one characteristic of a circumstance that may encourage self-interested agents to cooperate; there may be others. My intuition tells me that repeat play has been asked to carry a lot of water, that the circumstances it describes may not obtain always and everywhere, and that in receiving so much attention it has discouraged consideration of alternatives.

1. Repeat Play

Game theorists have, for some time, been interested in the game (called a supergame) that results when a constituent game is played repeatedly for a known or unknown, finite or nonfinite, number of iterations. Of special interest in the context of cooperation are those games in which the Nash equilibrium or the dominant strategy equilibrium of the constituent game is inefficient. The Prisoners' Dilemma (PD), of course, is the classic example of the latter. Such games, by the very inefficiency of their equilibria (i.e., both of the players are better off with some other outcome), provide prima facie opportunities for gains from cooperation. But the latter require players to deviate from their respective equilibrium strategies. In one-shot plays of the constituent game, incentives to deviate from equilibrium strategies do not exist. In repeat play, on the other hand, such incentives may exist because all the equilibrium strategies of the supergame may not coincide with the equilibrium strategy of the one-shot constituent game.

A variety of equilibria for supergames have been identified which involve some degree of cooperation, i.e. of not repeatedly playing the Nash (dominant) strategy of the constituent game. The conditions under which this may occur depend on several factors—the relative payoffs, the frequency of repetition, the time horizons and discount rates of the players, etc. Formal statements may be found in the work of Axelrod and Calvert (1985), so I won't dwell on them here. The idea can be summed up simply (if imprecisely): A cooperative pattern can be sustained if each player values the discounted flow of future benefits from cooperation more than the one-time windfall derived from reneging on the cooperative arrangement followed by noncooperation thereafter. If time horizons are short, if they are known with certainty, if discounting of the future is heavy, if windfall "profits" are high or cooperative benefits modest, then repeat play may be insufficient to sustain cooperation.

There is no denying that the "shadow of the future" is often long, so that repeat play suffices to induce cooperation. Many norms of interaction and cooperation inside families, clubs, firms, and small towns, for example, are sustainable by the prospects of repeated contact. Whether it manifests Itself in an unwillingness of a merchant to cheat a long-standing customer, an agreement by members to observe etiquette in the clubroom ("disagreeing without being disagreeable"), a reciprocal arrangement between spouses on the allocation of family chores, or whatever, cooperation (or the honoring of agreements) Is possible without coercion or third-party enforcement because incentives in the supergame are different from incentives in the one-shot constituent game.

But surely this will not be the case always and everywhere. Do congressmen stop swapping favors when one of them announces, say a year in advance, that he will not seek reelection? The answer is undoubtedly in the negative, but then some kinds of swaps that might once have been feasible are probably no longer so. Does an employer (say, a department chairman) overlook occasional lapses and lowered productivity in an employee (say, the departmental secretary) who is about to retire? Probably, but these so-called "last period problems" (in finitely-repeated games) are often anticipated in advance and, if they get too serious or begin too much in advance of the retirement date, then they come to be regarded as genuine breaches of what might have been a previously established cooperative equilibrium. In short, the conditions necessary for repeat play to induce cooperation (particularly either the nonfiniteness of or the uncertainty surrounding repetition) may not always obtain. Yet, while it may have some effect on the terms of cooperation, It need not mean an end to cooperation altogether. Repeat play, in short, may not be the only basis on which cooperation occurs.

I have tried to make three points here. First, the shadow of the future may be sufficient to induce cooperation in repeat play that would not have been forthcoming in one-shot play. Second, such conditions need not obtain. But, third, there may be other aspects of the circumstances of play that have a bearing on whether cooperation will occur. It is to some of these other aspects that I now turn.

2. Reputation

The idea of reputation has only recently appeared in the literature of game theory. It has been stimulated by a powerful example provided by Selten (1978), known as the "chain store paradox." According to this story, a chain store is contemplating entering 20 markets in each of which there is the possibility that a local competitor might enter. The chain store (C) may behave cooperatively toward the local competitor (L), thereby permitting each to share duopoly profits, or it may behave competitively,' thereby producing a price war. We refer to these strategies as **soft** and **hard,** respectively. L, on the other hand, must commit itself to **enter** or to **stay out,** and must do so before C decides what to do.

The game for the i^{th} market is given in Figure 1, where payoffs for C and L, respectively, are given in parentheses at the end of each branch. Thus, if L enters and C plays soft, they each receive a payoff of 2 (million dollars profit); if, on the other hand, C plays hard after L chooses to enter, the ensuing price war drives their payoffs to zero. If L should choose to stay out, then regardless of what C does, he enjoys a return of 1 on some other investment while C earns monopoly profits of 5. From either the extensive form or the outcome matrix, it may be observed that C has a (-weakly) dominant strategy. He is always at least as well off playing soft as he is playing hard. **And L can anticipate this, even though he** must move

first. Thus, the (subgame perfect) Nash equilibrium to this game is (enter, soft) for L and C, respectively.[1]

An examination of the i^{th} market in isolation, however, is misleading. If C plays each of these games independently, allowing his dominant strategy in each individual market to dictate his behavior, then he will end up with a payoff of 40 (2 from each of 20 markets), while $L_1, L_2, \ldots, L_{20}$ each receive a payoff of 2. But even if C regards his play as interdependent, and even if each of the L_1 anticipates this feature of the situation, it is difficult to arrive at a different conclusion. In the 20th market, after decisions have been reached in all of the others, C and L_{20} are involved in a one-shot play of the game in Figure 1—there is no "future." So, L_{20} can rationally anticipate that C will use his dominant strategy (why would he do anything different?) and the Nash equilibrium (enter, soft) will obtain. But then L_{19} can regard the 19th market as effectively the last (since we already know what will happen in the 20th market) and enter, expecting the Nash equilibrium to prevail.. This thinking applies to each of the local entrants in turn. In short, even though there is an interdependent structure here, the finiteness of the play (only 20 markets) permits the logic of backward induction to unravel the supergame into 20 independent plays of the chain store game. The problem here is similar to that of finite repeat play among the same two players when there is a dominant strategy.

While the logic of backward induction causes the chain store supergame to unravel into 20 independent plays of the constituent game, this

[1]The outcome (stay out, hard) is also a Nash equilibrium, but it is imperfect. For the distinction between these Nash equilibria see Selten (1975).

result runs counter to our (and Selten's) intuition. A similar disparity between logic and intuition holds in other finite repeat-play settings, e.g., the PD supergame, so our discomfiture is not unique to Selten's example. Put more dramatically, is it plausible that a player like C, or one of the players in a PD supergame, would select different strategies depending upon whether the game were played one million times or an infinite (or unknown) number of times? In the former the logic of backward induction (and concomitant unraveling) holds, whereas in the latter it need not. Yet our intuition tells us that there should not be a dramatic discontinuity In behavior In these two setttings.

Returning to the chain store game, suppose C reasoned as follows: Despite my dominant strategy (playing soft), I might be able to discourage (some) subsequent entry if I initially play hard and develop a hard reputation. Suppose, for example, that the first five times an Li enters, I play hard, and suppose further that such behavior on my part discourages L_6 through L_{15} from entering (though L_{16} through L_{20} might enter because, toward the end of the supergame, the logic of backward induction becomes increasingly plausible). My payoff with this supergame strategy would be no worse than 50 (0 for five rounds, 5 for ten rounds, and no *worse* than 0 for the last five rounds). Since this payoff exceeds what I would earn if I played my dominant strategy (40, the result of receiving 2 in each of the twenty rounds), there is clearly some rationale for trying to develop a reputation.

There is, however, a flaw in C's reasoning. The game, as defined, is one of perfect information. Thus, each of the players knows the payoff matrix and the sequence of play with certainty. So long as there is no uncertainty, there is no reason for the past history of play to affect the beliefs of any of the Li5. Each knows that C has a dominant strategy and, whatever C might think he is doing in trying to create the impression that he is hard, from this point on that threat is not credible. That is, for any i (= l,...,20) the backward induction argument holds for the supergame that commences with the i^{th} market and proceeds to the end. This provides us with an important general conclusion: reputational effects can play no role in a game of **perfect information.**

Before proceeding with some more constructive observations about reputational effects, let me summarize the argument to this point. In supergames of infinite or unknown length, there are conditions in which a player's optimal supergame strategy differs from his optimal strategy in the one-shot constituent game. This may entail "cooperation" in a repeat-play setting that might not otherwise be anticipated in a static analysis. If, however, the repeat play is finite and known, then this argument no longer is appropriate. Nevertheless, intuition and observation tell us that, for a sufficiently large number, of repetitions, a finite supergame should smoothly approximate its nonfinite counterpart. In order to square intuition with logic, a role for the development of reputations might be imagined. A hard reputation for a chain store, or a tit-for-tat reputation in a PD supergame, would yield higher payoffs than playing the dominant strategy of the constitutent game at each stage. But this argument, while superficially plausible, is itself flawed. The culprit is perfect information.

Returning to the chain store game, suppose the *Li5* are not perfectly informed. Specifically, suppose each believes at the outset that there is some positive (though possibly very small) probability, δ, that C has the payoffs given in Figure 2. These payoffs differ from those in Figure 1 only in the southwest cell, thereby rendering **hard** a dominant strategy. Thus, each L_i believes there is a probability δ that **hard** is a dominant strategy for C and a complementary probability $1-\delta$ that **soft** is his dominant strategy. Consider L_1's calculations. If he chooses **so**, then he receives a payoff of 1 no matter which type C is. If, however, he chooses **e**, then, assuming C chooses his dominant strategy whichever type he is, L_1 receives an expected payoff of $2(1-\delta) + 0(\delta) = 2(1-\delta)$. His payoff from **so** exceeds his expected payoff from **e** if and only if $1 > 2(1-\delta)$, or $\delta > 1/2$. Thus, if δ is small enough, the uncertainty will not be sufficient to deter L_1 from entering.

C suffers no imperfections of information. He knows that he is playing the game with payoff matrix given in Figure 1. He also is assumed to know δ. If $\delta > 1/2$, then it is clear that L_1 will not enter, so it does not matter what C does (along the equilibrium path, C will never have to choose between **soft** and **hard** since the L_i will not enter). But what if $\delta < 1/2$ and L_1 therefore enters? If C plays **s**, then all of the remaining L_is will have their information imperfection completely resolved--they will know for certain that C's payoffs are governed by the matrix in Figure 1, since no Figure 2-type would ever play **s**. Thus, playing **s** at any point in the sequence completely resolves any uncertainty for the L_is who have not yet moved. At the first play of the sequence, then, if C plays **s** when L_1 enters, then C assures himself of a total payoff of 40, since all subsequent plays of the game will produce entry by the remaining L_is.

It is this prospect that may induce C to consider imitating a Figure 2-type player (even though he knows his type is given by Figure 1). If we assume that the L_is update their beliefs on the basis of previous experience and observation in accord with Bayes Rule, then (as already seen) δ goes to zero whenever s is observed, but δ grows whenever **h** is observed.[2] Thus, in a repeat-play game with imperfect information, imitation can affect beliefs. And this is what is meant by building a reputation. As Wilson (1983, pp. 2, 11) puts it,

> ...one's reputation is a state variable affecting future opportunities; moreover the evolution of this state variable depends on the history of one's actions. Hence, current decisions must optimize the tradeoffs between short-term consequences and the longer-run effects on one's reputation....The key ingredient is that a player can adopt actions that sustain the probability assessments made by other participants that yield favorable long-term consequences. Whenever it is feasible to imitate the behavior one would adopt if one's private information [i.e., one's "type"] were different than it is, and this would affect others' actions favorably, there is a potential for reputational effects.

In the chain store game, the repetition over many markets, when conjoined with uncertain beliefs of the L_is as to C's type, encourages strategic posturing by C as he builds a reputation by imitating a Figure 2-type. Thus, in this particular example, while there is a demonstrable reputation effect, it has the consequence of **diminishing** cooperative behavior by the chain store. In another well worked out example involving the PD game, where the uncertainty of one of the players involves not knowing for sure whether his

[2]According to Bayes Rule, the value of δ is changed from its prior value on the basis of observation. That is, a posterior δ is conditioned on what C did at his previous move. My colleague, Randall Calvert, notes that there is some ambiguity here. L_i need not take what C actually did in the previous move at face value since he (L_i) knows that C might be an impostor. Thus, the manner in which updating of probabilities occurs is unclear. While this will not affect updating if an s is observed, it probably depresses increases in δ when an **h** is observed.

opponent is a "normal" player with a dominant strategy or is "addicted" to a tit-for-tat strategy, reputation has a salutary effect on the evolution of cooperation, **even though the supergame has a known, finite** number **of stages.** This result is demonstrated by Kreps, Milgrom, Roberts and Wilson (1982).

Elsewhere Kreps and Wilson (1982) and Milgrom and Roberts (1982) make the even subtler point that player uncertainty about the payoffs of their fellows is not even necessary. "[I]t does nearly as well if there is no uncertainty about player payoffs, but there is uncertainty about whether this is so." (Kreps and Wilson, 1982, 266). That is, reputational effects may emerge, either enhancing or diminishing the prospects for cooperation, if player payoffs (or what I have earlier called player types) are not common **knowledge.** Common knowledge is assumed in most game settings, by which is meant that every player knows the payoffs of every other player, every player knows that every player know this, every player knows that every player knows that every player knows this...ad infinitum. Common knowledge may be violated (as In the modified chain store game above) If one or more players are not certain about the payoffs of their fellow players. But it may also be violated if every player is certain about *every* other player's payoffs, but is **not** certain that *every* other player Is certain about this.

In sum, I have attempted in this section to characterize a reputational effect and to suggest some circumstances in which it might emerge. A reputation is a belief by a player about his opponent. That belief (the opponent's reputation) is affected by the priors a player brings to the situation, by his opponent's subsequent behavior, and by the manner in which the player updates his beliefs on the basis of his opponent's subsequent play. In finite-stage supergames in which one of the players has a dominant strategy, I have pointed out that perfect information unlinks, and makes independent, the otherwise interdependent stages of the supergame through backward induction. A (sometimes modest) amount of information imperfection, or lack of common knowledge, gives reputations a role to play, and plausibly accounts "...for strong intertemporal linkages along a sequence of otherwise independent situations ." (Wilson, 1983, 1). Reputation-building may (but need not) affect cooperative behavior in settings where repeat play alone would fail to provide such incentives.

3. Sequence

In each of the constituent games comprising the chain store supergame (see Figure l), a specific sequence of moves was given. The local competitor must first commit to enter or stay out and, if the former, then the chain store must determine whether to play cooperatively or aggressively. One could, however, associate a different sequence (that is, a different extensive form) with the same outcome matrix. In this alternative, it is the chain store that moves first, revealing whether it will meet entry cooperatively or aggressively. For example, the chain store might build a store big enough to handle the larger quantities required in more competitive situations, indicating its willingness to play aggressively **(hard).** The payoffs are the same as those given in the outcome matrix of Figure 1, but the extensive form is now given in Figure 3. C still has a dominant strategy (soft), and the outcome, **(soft, enter),** is still a Nash equilibrium. But it is an

Imperfect equilibrium (Selten, 1975) in the following sense: There is another Nash equilibrium, **(hard, stay out),** which is preferred by C and can be effected by C's first move. That Is, C can force L to choose **stay out** by his initial choice of **hard.**

In the original version of the chain store game, the entrant moved first and, only subsequently could the chain store move, punishing or cooperating. And, as the previous section argued, only if there were some uncertainty about the chain store's 'type' in the minds of prospective entrants could C attempt to affect their behavior by building a reputation for aggressiveness. In this new version, however, C may affect the behavior of any prospective entrant by reversing the sequence of moves. In the case of Figure 3, the paradox disappears, even if their is no uncertainty about C's type, because he can credibly act as though he were a Figure 2-type by choosing **hard** at his move.

The point of this apparently superficial change In the chain store game Is that sequence matters. Any given outcome matrix can 'support' different extensive forms of a game, and it is the latter that affects the equilibrium. To make this point clear in the context of cooperation, let me present one additional example, given In Figure 4. In this example there are two players, A and B, each of whom have two strategies, Up and Down. Their payoffs are given in the outcome matrix. The first extensive form in the figure has A moving first. If he chooses U, then he knows that B will also choose U (since B prefers the outcome (l,0) to (l,-l)). If, on the other hand, A chooses D, then he knows that B will choose D as well (since B prefers (-1,2) to (2,1)). In effect, then, A may choose between the outcomes (1,0) and (-1,2), since he can forecast B's reaction. Obviously, he will choose U and the outcome is (1,0).

It turns out that this is a Pareto-inferior outcome. Each player is made better off by the outcome (2,1), the result of the strategy pair (D,U). But, as noted, if A chooses D, then B will not choose U. Since we are assuming a noncooperative game, i.e., no exogenous enforcement, this means there is no way for A and B to "cooperate." No promise by B that he will play U if A plays D is credible; he has every incentive to renege on his promise. In the absence of exogenous enforcement institutions (or repeat play, or reputation), this "society" is unable to extract all of the advantage from the situation.

This society, however, has another degree of freedom. The outcome matrix is akin to a production possibility array. Society's technology can produce any of four outcomes depending upon four different combinations of inputs. The first extensive form, with A moving first, yields a Pareto-inferior outcome. But, there is another extensive form consistent with these production possibilities with B moving first. B, like A in the previous paragraph, can anticipate his opponent's reactions to his strategy choices. If B chooses U, then A will choose D, yielding the outcome (2,1). If, on the other hand, B chooses D, then A will choose U, yielding the outcome (1,-1). Since B prefers the former to the latter, the Pareto-optimal outcome will be realized. That is, cooperation in a noncooperative setting may be realized (at least in this particular example) by appropriate sequencing.

I do not know how general this prospect is or, put differently, the conditions under which Pareto-improvements may be achieved through alternative sequencing. To the extent, however, that societies may choose their Institutions, and given that the sequence of moves is seen as an alterable property of the game defined by an institutional arrangement, then It follows that we have In sequence another handle on cooperation. If A and B are not dealt a particular sequence ex ante that they cannot rearrange, then we would expect them to opt for the latter sequence in order to capture the gains from cooperation.

This suggests that, in substantive applications of game theory, we take care not to focus exclusively on the outcome matrix of that game. Students occasionally confuse the outcome matrix with the normal form of a game. In fact, this is not correct. U and D are behavioral alternatives for the players, not strategies. They happen, in this case, to be strategies for the first mover. The second mover's strategies are intentions contingent on the first mover's actual move:

I.	**U** If **U**; **U** if **D**	**(ALWAYS U)**
II.	**U if U**; **D** if **D**	**(IMITATION)**
III.	**D** If **U**; **U if D**	**(OPPOSITE)**
Iv.	**D if U**; **D if D**	**(ALWAYS D).**

In the example, If A Is the first mover, then B's optimal **strategy** Is li; he should imitate A's choice. If, however, B is the first mover, then A's optimal **strategy** Is III; he should do the opposite of B.

Let me reiterate the point of this discussion. As a student of Institutions, I am. content to define an institution as a game in extensive form. With this definition I seek to account for regularities by examining the equilibria of the game that characterizes this institutional setting. I have elsewhere referred to these as institutional equilibria or **structure-induced equilibria** (Shepsle, 1979, 1985). A limitation of game theory, and consequently of the analysis of structure-induced equilibria, is that the game may not be given ex ante. Certain constraints may exist (behaviors, production possibilities) and may be fixed for the duration of a prospective game. But the form that the game takes may be malleable, as in the alternative sequencing possibilities of Figure 4. For students of cooperation this prospect should not be overlooked. It seems to me that sequencing (or, more generally, the structuring of a game from extant conditions and constraints) stands alongside repeat play, reputation, exogenous enforcement, and exogenous coercion as another possibility for enhancing or diminishing the prospects of cooperation. The same is true of interconnection, the last topic to be taken up.

4. Interconnection[5]

Some years ago a handful of economists became interested in legislative roll-call behavior. In papers (for example, Kalt and Zupan, 1984) and books (for example, Kau and Rubin, 1982), they puzzled over what was for them an anomaly. Try as they might to provide an empirical case for "economic" voting by legislators, i.e., that legislators voted in accord with some measure

[5] This section is based on some lengthy conversations with William Riker during the 1983-84 academic year when he was a visiting professor at Washington University. The examples reported here are drawn from these discussions, and the interpretation 1 give were inspired by them. I do not, however, want to implicate Professor Riker for any errors contained here.

of economic interest in their constituencies, large residuals remained in their econometric analyses. This is not to say there were not strong economic effects; there were. But, unable to make the residuals go away, they decided to adopt them as their own and give them a name. Thus, they suggested that, in addition to economic interests of constituents, legislators entertained personal ideological preferences. Because their own behavior was only very imperfectly monitored by the folks back home, legislators could indulge these ideological tastes in their voting behavior in the legislature. Thus, "ideology" (formerly a puzzling residual in an econometric model) was given the status of an explanatory variable.

I have never been a fan of this form of explanation, despite the fact that some of the work in which it figures prominently is very carefully crafted. I was stimulated to think of alternative ways to reason about this anomaly. In most any econometric study, it is pretty easy to criticize specifications and to express dissatisfaction with the way particular variables are measured; the studies cited above are no exception. But my interest is less in these criticisms than in the general conceptualization of legislators as naive translators of constituent economic concerns. The anomaly derives, in my opinion, from this mechanical formulation.

As Denzau, Riker and I (1985) have argued, the legislator is simultaneously embedded in two games—one in the legislative arena and the other in the electoral arena. In the former he is engaged in institutional career advancement; in the latter he is animated by the Fiorina-Mayhew "electoral connection." For some (but certainly not all) legislators the two games are indistinguishable; and, for nearly all legislators, the two games bear some relationship to one another (hence, the econometric support for a strong relationship between constituency economic interests and legislative' voting). There is, however, one interdependency which I want to take up here in a rather extreme form.

Let me abstract from the above substantive context and consider an example involving three players—A, B, and C. A is simultaneously engaged in separate games with each of the other two players. Let me assume for now that neither B nor C is aware that A is playing against anyone else. Each player is interested only in maximizing his own return from the game. But, for A this means that he is maximizing the **sum** of the payoffs from the two games. Finally, and this is the kicker, suppose that A is restricted in his strategy choice. Whatever strategy he chooses, he must play the same one in **both** games. Thus, to return to the legislative example, a legislator cannot vote one way in the legislature and then turn around and suggest to his constituents that he actually did something different. Rather, a strategic consistency is imposed which requires that he "play the same strategy" in both the institutional arena and the electoral arena. The games are said to **be interconnected** because of this strategic restriction.

An example, given in Figure 5, will give this concept some concreteness. **The** first two matrices give payoffs for the game between A and B and between A and C, respectively. It happens that both B and C have dominant strategies (b_2 and C_1, respectively). The Nash Equilibrium for each of the games is indicated by a *. Although neither B nor C realize it, A is playing the "sum game," given by the third outcome matrix in Figure 5. For each configuration of strategy choices, the payoffs for B and C are the same as In their component games with A. But the payoff for A is the sum of his payoffs in the two component games. Since B and C have dominant strategies, the outcome that results from those being played, and A picking the strategy that maximizes the sum of payoffs to him contingent on those dominant strategies being played, is also a Nash equilibrium. Thus, the ex ante outcome, (a_1, b_2, c_1), is stable, though B will undoubtedly be (unpleasantly) surprised by A's choice. Nevertheless, because of the dominance in this example, each player's strategy choice is a best response to what the others have done.

This example is revealing because, despite the imperfect information (for B and C) and the interconnected structure, there is a well-defined equilibrium. Admittedly, B will be surprised and pu22led by the outcome. But, because of the dominant strategy for B, this has no consequences in the one-shot game. Interestingly, an observer of the game between A and B, who is also unaware of the interconnected structure, would likewise be surprised and puzzled. Indeed, one might imagine the observer attempting to rationalize the outcome by postulating some unobservable factor, like A's "ideology," to account for the disparity between outcome and expectation. The point, of course, is that the interconnected structure produces a **different** equilibrium than might otherwise be anticipated.

A somewhat more involved example is given in Figure 6. Neither of the component games has a dominant strategy for either player; but each has a pure strategy equilibrium point (indicated by a *). If A anticipates that B and C will play their component-game equilibrium strategies (b_1 and c_2, respectively), then his best response (in column two of the sum game) is a_2, which is labeled the ex ante outcome. In this case B will, like in the preceding example, be (unpleasantly) surprised. But, because his is not a **dominant** strategy, b_1 is no longer a best response. He will switch to b_2, which is the Nash equilibrium of the sum game. Should the process go all the way to equilibrium, then an external observer, again unaware of the Interconnected structure, would be doubly surprised (both by A's and B's strategy choices).

These two examples do not exhaust all the possibilities. Indeed, they have been selected because each of the component games, as well as the sum game, has a pure strategy equilibrium point. This feature, of course, need not hold. Similarly, the information condition—in which B and C are ignorant of the interlocking structure—is a special circumstance. As knowledge becomes more common, the strategic possibilities become richer. Nevertheless, two general results can be reported for this setting.

First, If the same strategy is an equilibrium for A in both component games, then It is also an equilibrium in the sum game *(a fortiori,* if it is a dominant strategy). In these settings, there will be no "surprises" deriving **from** the interconnected structure. This will be true even if B and C should learn about it. Such structures may be said to be **separable.** They may be analyzed separately with no loss of information, and they may be played separately with no loss of advantage.

Second, if $(\alpha_i, \beta_j, \chi_k)$ is an equilibrium point of the sum game, then either (α_i, β_j) or (α_i, χ_k) is an equilibrium point of one of the component

games. Thus, at least half the time, there will be no surprises to observers and players of the component games, despite the interlocking structure.

At this early stage it is a bit difficult to assess the consequences of interconnectedness for cooperative play. It is difficult to generalize on so subtle a matter on the basis of a few examples. But it must surely be the case that interconnection affects the manner in which such instruments as repeat play, sequencing, and reputation may be exploited for cooperative purposes.

5. Concluding Remarks

The purpose of these notes has been to suggest some tools or building blocks with which to construct institutional regimes. Repeat play has received a disproportionate amount of attention recently. 1 should like to encourage the analysis of some additional building blocks—sequencing, reputation, and interconnection. There are undoubtedly others.

One thing, it seems to me, is clear. The game-theoretic approach Is an extremely valuable one for students of Institutions. It provides a basis for anticipating how rational agents will adapt to the environment in which they find themselves. It therefore allows us to examine how the institutional setting, through its effects on both incentives and Information, affects behavior and outcomes.

As noted earlier, however, institutional arrangements are not foreordained. They are created by men; they may be altered by men. Game theory is a useful tool for examining equilbrium outcomes and behavior, once the game structure Is provided. What I am suggesting in these notes is that we seek to endogenize the choice of the game, itself. Not all margins are available, of course—preferences are often unchangeable in the short-term; so, too, are production possibilities. But, given certain fixed features of a situation, It Is entirely possible to design and construct instruments or institutions to accomplish particular purposes. To do so requires that we understand the building materials we have available to us. My purpose has been to suggest what some of those building materials might be.

6. References

Axelrod, Robert. 1981. "The Emergence of Cooperation Among Egoists." **American Political Science Review** 75: 306-318.

. 1984. **The Evolution' of Cooperation.** New York: Basic.

Calvert, Randall. 1985. "Reciprocity Among Self-interested Actors." Mimeo. Washington University, St. Louis.

Den2au, Arthur, William Riker, and Kenneth Shepsle. 1985. "Farquharson and Fenno: Sophisticated Voting and Home Style." **American Political Science Review** 79: in press.

Kalt, Joseph and Mark Zupan. 1984. "Capture and Ideology in the Economic Theory of Politics." **American Economic Review** 74: 279-301.

Kau, James and Paul Rubin. 1982. **Congressmen, Constituents, and Contributors.** Boston: Martinus Nijhoff.

Keohane, Robert. 1984. **After Hegemony.** Princeton: Princeton University Press.

Kreps, David and Robert Wilson. 1982. "Reputation and Imperfect Information." **Journal of Economic Theory** 27: 253-279.

Kreps, David, Paul Milgrom, John Robert, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." **Journal of Economic Theory** 27: 245-252.

Laver, Michael. 1980. "The Politics of Crime and Vice Versa." Delivered at the Public Choice Workshop, Washington University, St. Louis.

Milgrom, Paul and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." **Journal of Economic Theory** 27: 280-312.

Selten, Reinhard. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." **International Journal of Game Theory** 4: 25-55.

______s. 1978. "The Chain Store Paradox." **Theory and Decision** 9: 127-159.

Shepsle, Kenneth. 1979. "Institutional Arrangements and Equilibrium in Multidimensional Voting Models." **American Journal of Political Science** 23: 27-59.

______, 1985. "The Positive Theory of Legislative Institutions." Delivered at Carnegie Conference on Political Economy, Pittsburgh.

Wilson, Robert. 1983. "Reputations in Games and Markets." Technical Report No. 434. Institute for Mathematical Studies in the Social Sciences. Stanford University.

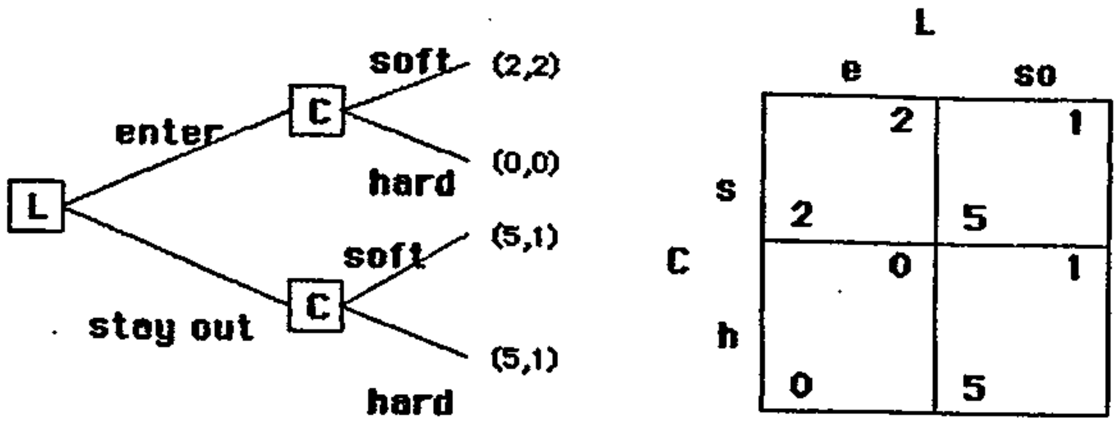

FIGURE 1. CHAIN STORE PARADOX

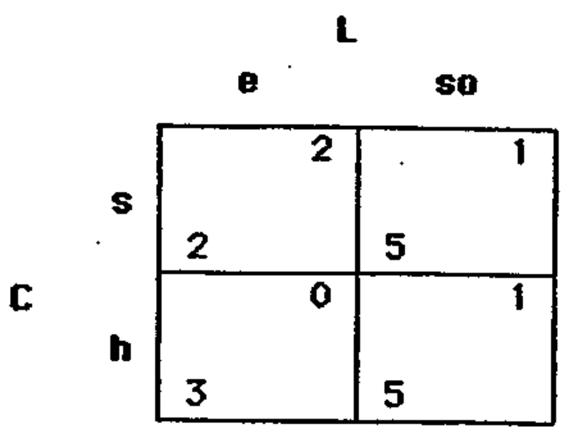

FIGURE 2. REVISED CHAIN STORE PARADOX

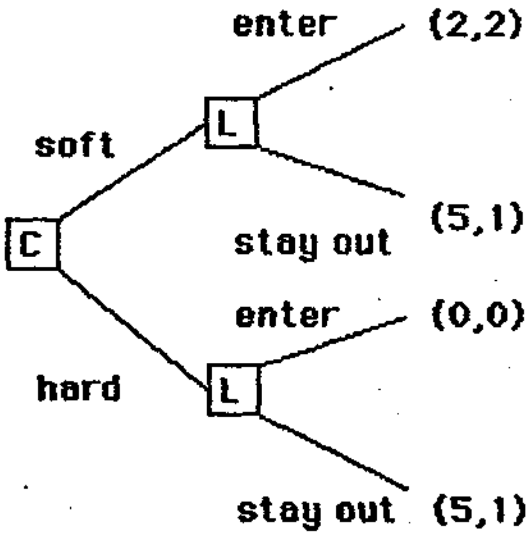

FIGURE 3. CHAIN STORE GAME:
ALTERNATIVE SEQUENCE

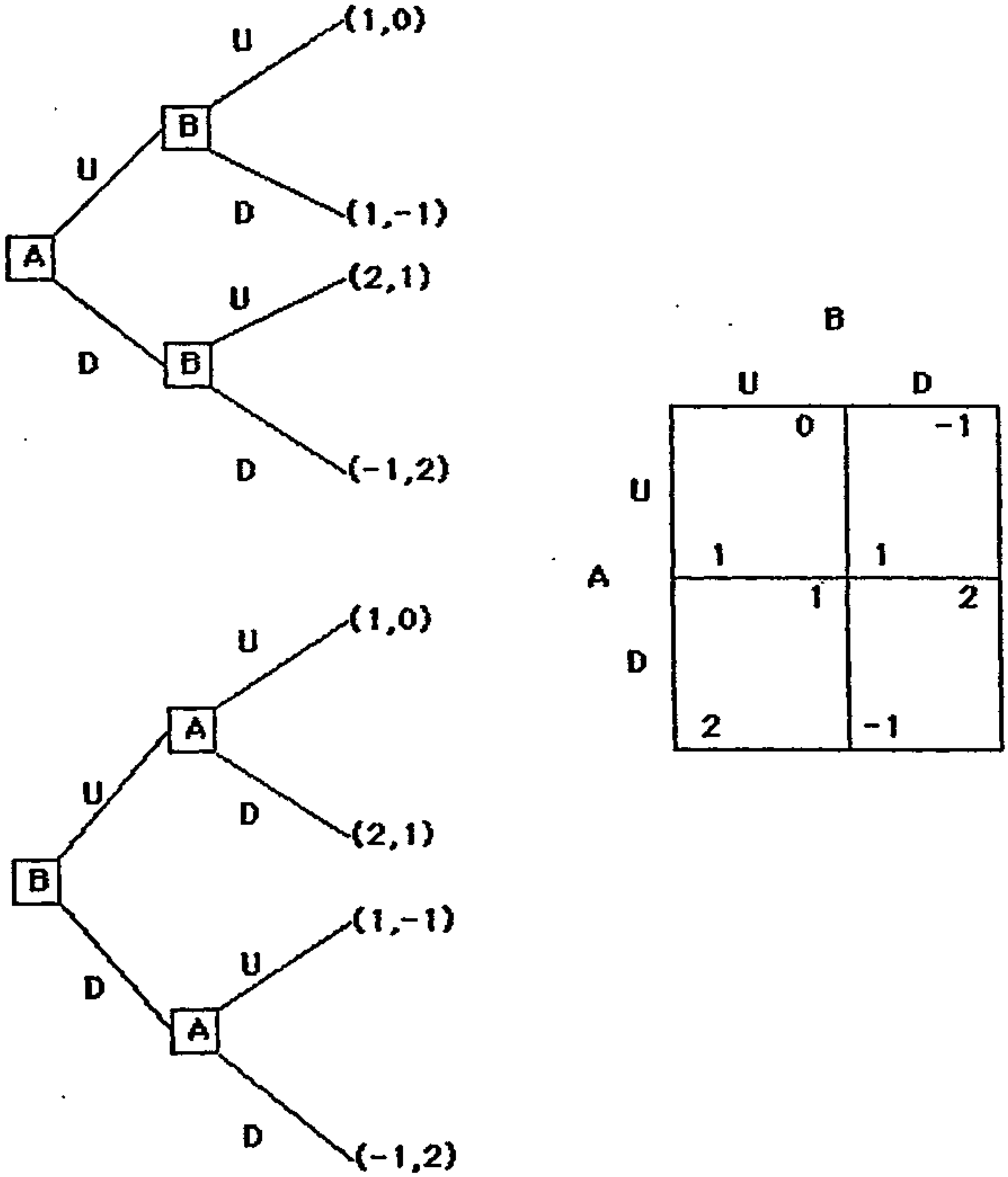

FIGURE 4. ALTERNATIVE SEQUENCES

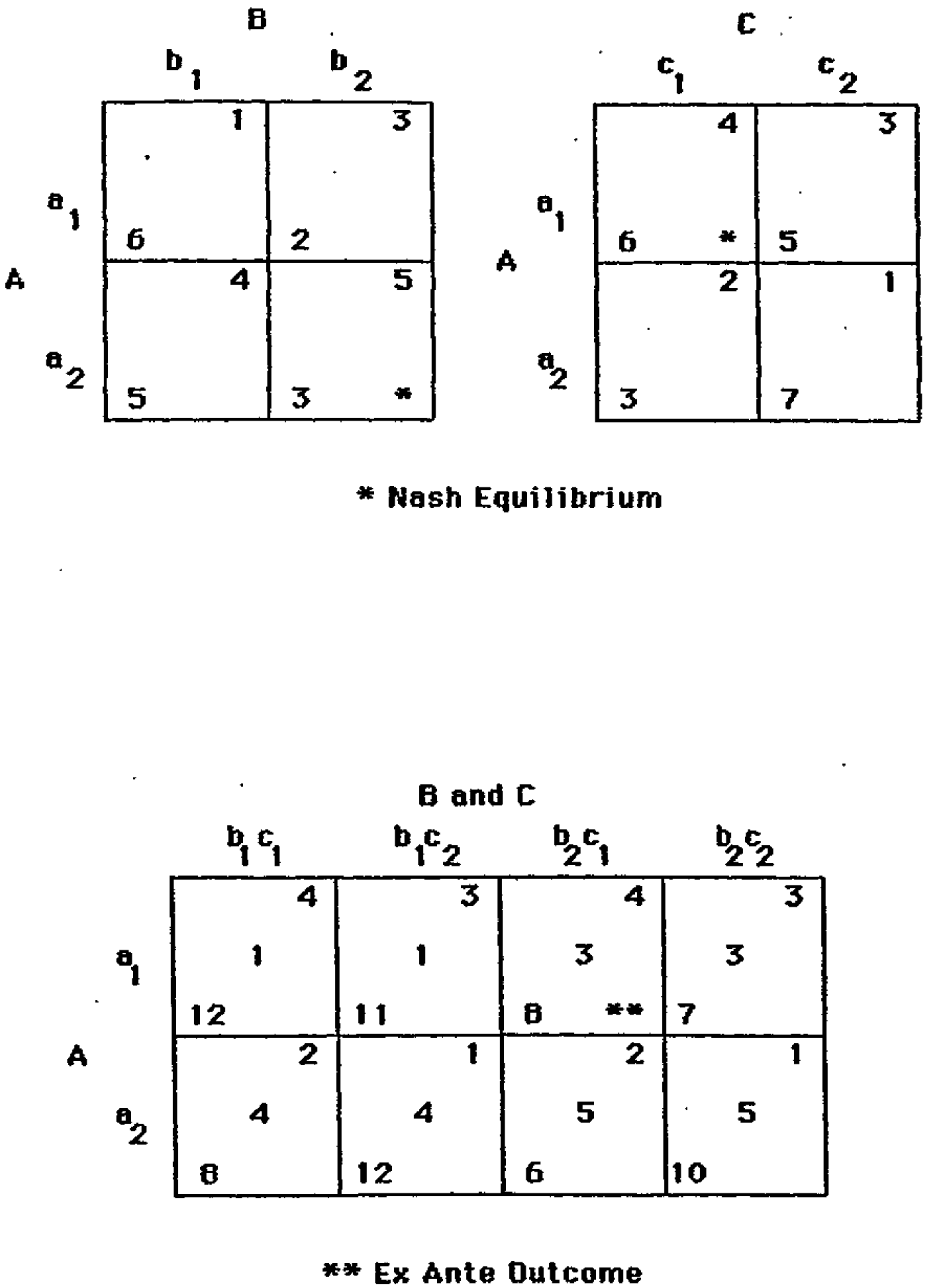

FIGURE 5. AN INTERCONNECTED GAME

B

A	b_1	b_2
a_1	4 6 *	3 2
a_2	0 5	2 3

C

A	c_1	c_2
a_1	0 6	1 5
a_2	3 3	4 7 *

* Nash Equilibrium

B and C

A	b_1c_1	b_1c_2	b_2c_1	b_2c_2
a_1	0 4 12	1 4 11	0 3 8	1 3 7
a_2	3 0 8	4 0 12 **	3 2 6	4 2 10 *

* Nash Equilibrium **Ex Ante Outcome

FIGURE 6. AN INTERCONNECTED GAME WITHOUT DOMINANT STRATEGIES