The theory of games and the formation of defensive alliances

Lieshout, R.H.

The theory of games and the formation of defensive alliances

Robert H. Lieshout*

1. Introduction: alliance formation as a problem of collective action

In this essay I wish to provide a game-*theoretic* (Snidal 1985) analysis of the conditions under which states that face a common enemy will form a defensive alliance that will dissuade the common enemy from attacking them. Put somewhat differently, I use the instruments provided by game theory to trace the conditions under which states will decide to balance, or to appease a common enemy. My analysis has been inspired by a number of books and articles on game theory and collective action that appeared during the eighties, which I consider to provide particularly helpful insights in the substantive problem I wish to deal with in this article, as well as to demonstrate most clearly the relevance of game theory to the study of international relations. (especially Hardin 1982, Wagner 1983, Taylor 1987 and Nicholson 1990 ed.)

In this section I briefly discuss three subjects. In the first place, I introduce the concept of a non-cooperative game. In the second place, I show that the formation of a defensive alliance can be a problem of collective action. Finally, I state my reasons for restricting my analysis to games between two players.

The situation under consideration concerns two states, *i* and *j* that have to decide whether or not to contribute to the provision of a credible defence against a common enemy, where the decision of the one state depends on the decision taken by the other state. That is to say, it concerns a: 'behavior situation in which each player's best choice of action depends on the action he expects the other to take, which he knows depends, in turn, on the other's expectation of his own.' (Schelling 1958, 205) The states are involved in a game of strategy. The game they are involved in is also non-cooperative. This means, that the states can freely discuss the virtues of an agreement to cooperate, that they may even sign a treaty to

this effect, but that they know, that if one of them breaks the agreement, the cost of enforcement is prohibitive to the other. (de Jasay 1989, 26-30)

A credible defence against a common enemy can be regarded as a collective good. That is to say, a good with the following properties: 1) it exhibits some degree of *jointness of supply*, which means that in case the good is provided by one or more states, its consumption by one state does not make impossible its consumption by other states (Blümel et al. 1986, 245); 2) it is to a certain extent *non-excludable*, which implies that most of the time the good is provided by one or more states, it will be too costly to exclude other states from consuming it too. (Olson 1971 ed., 14) Moreover, as far as the consumers of the good are concerned, it exhibits some degree of *non-rivalness*: a certain number of states can consume the good before 'congestion costs' become noticeable. After this limit has been passed the quality of the good progessively diminishes as more and more states consume it. (Head 1962, 202, Taylor 1987, 7) Another important characteristic of the collective good 'a credible defence against a common enemy' is, that most of the time it is a *step good*. This means that usually it cannot be provided unless contributions to the provision of the good exceed some threshold. (Hardin 1982, 55-59)

Russell Hardin has argued that a problem of collective action exists if each of the actors involved orders the possible outcomes of his decision whether or not to contribute to the provision of a collective good, in such a manner that each of them decides not to contribute whatever the decision the other actors come to, although each of them prefers the outcome that each contributes to the outcome that none contributes (the actors find themselves in a Prisoner's Dilemma, that is). (Hardin 1982, 25) Michael Taylor has taken exception to this 'strong definition'. According to him, it does not take account of the fact that the provision of a collective good also can be a problem in situations in which the actors involved know their best choice depends on the choices the others make, and order their preferences over the various possible outcomes in a different manner. For this reason he proposes the following definition: 'a collective action problem exists where rational individual action can lead to a strictly Pareto-inferior outcome, that is, an outcome which is strictly less preferred by every individual than at least one other outcome.' (Taylor 1987, 19) It will be obvious that this definition covers the problem of the formation of a defensive alliance. World history abounds with situations in which states individually decided to appease a common enemy, even when they realized that collectively they would be better off if they decided to balance that common enemy. Two types of games unequivocally depict collective action problems as defined by Taylor: the Prisoner's Dilemma

and the Game of Chicken. Whether the Assurance Game (or Stag Hunt) also represents a collective action problem, is less clear. But, as I argue in section 3.3, one can easily think of circumstances under which the outcome of this kind of game too, would be Pareto-inferior.

A final point that I wish to take up in this introductory section, is my decision to restrict my analysis to games between two players, and not to consider *N*-person games. Many will doubt the relevance of an analysis of this type for the study of collective action problems, which generally would seem to involve more than two players. The following considerations have led me to the conclusion that a game between two players nevertheless does provide an adequate model in the case of defensive alliances.

Michael Nicholson has noted, that as far as the formation of alliances is concerned, we can distinguish three categories of states (I have slightly adapted Nicholson's characterization of the three categories). In the first place there are the *crucial states*. Without the participation of at least one of these states the alliance would not work. These states constitute either alone, or together the *minimal contributing set*, k, of the alliance. Next there are the *significant states*. These are the states that separately, or together would not be able to provide a sufficient amount of the collective good, but in case the collective good is produced by the member(s) of k, their contribution significantly increases the amount (and quality) of the collective good available. The third category consists of the *remainder states*, which may benefit from the alliance, but the participation of which in the alliance hardly increases the quantity (and quality) of the collective good available. (Nicholson 1990 ed., 133) It is my contention, first, that as a matter of fact defensive alliances initially never contain more than two crucial states (either $k = 1$, or $k = 2$) and, second, that a crucial state lets its decision whether or not to contribute to a defensive alliance only be subject to the decision the other crucial state takes (provided there is another one), and not to the decisions the significant states take, not to speak of the decisions the remainder states take. (A possible exception might be the situation in which a crucial state is barely able to provide the collective good, and the significant state only just not; see section 5).

2. The Prisoner's Dilemma

2.1. *The international system* – The international system under consideration consists of several states, which all try to retain their independence in an environment that has something of an anarchical society (Bull 1977).[1]

There are three classes of states in the system. First there are two superpowers, *g* and *h*, which are the crucial states as far as the international system is concerned. Then there are two great powers, *i* and *j*, which are significant states with regard to the international system. If *i* and *j* would somehow manage to pool their resources, they would become a crucial state too. Finally, there are various remainder states. Assume moreover, that:

1) g and *h* are capable of deterring each other, as well as capable of defending their own territory, if the other would try to invade it[2];
2) *i* and *j* are capable of defending themselves against one of the superpowers, provided they coordinate their defence efforts, but they are not able to deter a superpower. Alone, neither *i* nor *j* is capable of defending its territory against one of the superpowers;
3) the remainder states are neither capable of defending themselves against the superpowers or the great powers, nor are they capable of deterring these powers, so that:
$cap_g = cap_h > cap_i = cap_j > cap_{r,1...n}$;
4) the geographical distance between *g* on the one hand, and *i* and *j* on the other, is greater than the geographical distance between *h* on the one hand, and *i* and *j* on the other. Consequently, *g* will be *i* and *j*'s natural ally against *h*. (Walt 1985) Also *i* and *j* are neighbours, so that they find themselves in that relationship of: 'mutual emulation and hatred which generally prevails among nations in close neighbourhood.' (Hume 1985 ed., 404) Accordingly:
$\{D_{ig}, D_{jg}\} > \{D_{ih}, D_{jh}\} > \{D_{ij}\}$;
5) g will assist *i* and *j* in their defence against a possible invasion by *h*, and this assistance will be significant, but will not use its deterrent capability against *h*, in case the latter invades *i* or *j*, or both countries. (compare Ordeshook 1986, 138-141) If *i* or *j* alone has to fight *h*, *g* will come to the aid of the sole resister, but *g*'s aid will be too little and too late, and will not save the sole resister from defeat. So that *i* and *j* constitute the minimal contributing set as far as their own defence against *h* is concerned.

2.2. *The Prisoner's Dilemma as a one-shot game* – Consider the following situation. Both *i* and *j* have been presented with an ultimatum by *h*, in which the latter demands that *i* and *j* in the future will coordinate their decisions with respect to their foreign policy and trade with *h*. If *i* and *j* do not accept this demand within thirty days, *h* will declare war on them, and will fight until they have accepted its demand. In the course of the subsequent secret negotiations *i* and *j* separately (on the request of *h*) conduct

with *h*, it becomes clear that in case either *i* or *j* is prepared to sign a treaty of friendship with *h*, the latter will reward this sole appeaser by allowing it to take part in the campaign against the sole resister, and after the latter's defeat to annex a part of its territory, while *h* will annex the rest of that country. If both *i* and *j* accept *h*'s demands however, appeasement is far less attractive, as then *h* will only be satisfied when its demands are accepted in full.

I assume (as in all one-shot games discussed in this essay) that *i* and *j* can only choose between two strategies: either to cooperate with one another, C, in order to balance *h*, or to defect, D, with the object of appeasing *h*.[3] I further assume that *i* and *j* will have to choose their strategy C or D before knowing the choice the other has made. A final assumption is that the structure of the game -available strategies, outcomes, and the utilities *i* and *j* attach to these outcomes- is *common knowledge*.

The most attractive outcome to *i* as well as *j* would be if it were the sole appeaser. Failing that, the next best outcome would be for them to cooperate in a defensive alliance against *h*, which outcome, although costly, they would prefer to the outcome that both of them would appease *h*, as this would involve some loss of independence. Both *i* and *j* would prefer this latter outcome however, to the one in which it would be the only state to resist *h*'s demands. In view of the utilities *i* and *j* attach to the four possible outcomes, they find themselves in a Prisoner's Dilemma.[4] This game can be represented in *normal* form as follows:

		j	
		D	D
i	C	b_i, b_j	d_i, a_j
	D	a_i, d_j	c_i, c_j

Diagram 1: Prisoner's Dilemma

Where *a*>b>c>d, and the first entry in each cell represents *i*'s payoff, and the second *j*'s payoff.[5] Under the circumstances both *i* and *j*'s choice of strategy will be D, and the outcome of the game will be (c_i,c_j) although (c_i,c_j) is Pareto-inferior to (b_i,b_j). Both players will nevertheless choose D, because in choosing D they do: 'no worse and, in general better, regardless of the strategy chosen by the other player,' and accordingly D will *dominate* C. (Rapoport and Guyer 1966, 14) Moreover, the outcome (c_i,c_j) is an *equilibrium outcome*. Which means that neither player can profit from unilaterally shifting to the other strategy.

2.3. *The Prisoner's Dilemma Supergame* – When I now turn to a discussion of the chances of cooperation in an iterated Prisoner's Dilemma, the first thing to notice is, that in the case where the players know how many times the game will be played, D will be their choice of strategy in each constituent game. The logic behind this is, that the players know that in the last constituent game D will be their dominant strategy, which makes the penultimate game 'in strategic reality' the last, so also in this game D will be their dominant strategy, and so on, backwards to the first constituent game.[6] (Luce and Raiffa 1957, 98-99)

Consider again the situation of i and j, as set out in the previous section, but assume this time that both states realize that their mutual defection will not be the end of the story. They both know that after a certain period of time h will make new demands, and after they have complied with these demands, there will be other demands, and so on. For the same reason, being the sole appeaser seems far less attractive to them, although in each round of new demands this will remain the most attractive outcome to both i and j individually (as a matter of fact their preference order over the four possible outcomes in each round of play will remain the same as in the one-shot Prisoner's Dilemma). At the same time the future is less important to them than the present, not only because they discount future payoffs, but also because they are unsure about the future. Perhaps h's demands should be interpreted as a sign of weakness, as a sign that h will not be able to maintain its position as a superpower much longer. Why not acquiesce to them now, if it will not be long before they can be ignored with impunity? So the future casts shadows in two opposite directions. Cooperation does look more attractive to i and j when they take the future into account, but so does defection.

In this situation a state will choose C, if the discounted value of the string of payoffs of mutual cooperation exceeds the value of a single defection while the other state cooperates. Where, following convention, I assume that future payoffs are discounted exponentially by a discount factor w, and that w is a number between zero and one. That is to say, both states will choose C if for both of them the condition holds that:

$$b + b(w + w^2 + w^3 + \ldots + w^n) > a$$

and if so (b_i,b_j) will be the equilibrium outcome.

My representation of the options available to i and j is extreme (but not altogether unrealistic in some cases, I think) in the sense that being the sole resister spells complete disaster for the state that finds itself in that position. It loses its independence, and it will not be able to play again. Suppose therefore, that being the sole resister is still the worst outcome to both states, but that this outcome is only slightly more worse than mutual

appeasement. The sole resister will have to agree to additional demands by h, and to cede a piece of territory to the sole appeaser, but it will retain its position as a great power to a large degree, and in the future may even regain it in full.

In this situation i and j can choose between more than two strategies, in fact the number of strategies available to them is infinitely large. Following Taylor, I assume however, that i and j will only choose between four strategies: 1) all-D, defect in each constituent game whatever the other state does; 2) all-C, cooperate in each constituent game regardless what the other state does; 3) 'Tit-for-Tat', choose C in the first constituent game, and in successive games choose C if and only if the other player chose C in the preceding game, otherwise choose D; and 4) 'Tat-for-Tit', choose D in the first constituent game, and in successive games choose C if and only if the other player chose C in the preceding game, otherwise choose D.[7] (Taylor 1987, 65-67)

Both players in this situation will choose a strategy of conditional cooperation, in this case Tit-for-Tat, if for both of them the condition holds, that:

$$w \geq \frac{a-b}{a-c} \quad \text{and:} \quad w \geq \frac{a-b}{b-d}$$

The latter condition reflects the fact that both players should prefer cooperating in each constituent game to taking turns exploiting each other in the constituent games (which condition was definitely satisfied in my first representation of a Prisoner's Dilemma supergame!). If these conditions are satisfied, the outcome wil be (b_i,b_j), and it will be an equilibrium. (Taylor 1987, 65-69)

A few words to conclude this section on Robert Axelrod's operationalization of w (not only in his famous computer tournament, but also in his article with William D. Hamilton). This is such that w no longer stands for the present discounted value of future payoffs, but for 'the probability the players will meet again in the future.' (Axelrod 1984, 42, 65 & 93) Although it is far from me to deny that both concepts have a lot in common, they are nevertheless not identical (if only for the 'technical' reason that p can take on the value of zero and one, whereas w cannot). And it may even be true, as Michael Nicholson maintains, that: 'the interpretation of the model in terms of probabilities is really closer to the point than a model interpreted narrowly in terms of discount parameters.' (Nicholson 1990 ed., 40) But why then not cast the analysis in terms of probabilities right away? However that may be, an analysis of a Prisoner's Dilemma supergame in which it is assumed that the players not only

discount future payoffs, but also make probability estimates of these payoffs, seems to me both to be more realistic, and to shorten the shadow of the future considerably (see also section 3.3).

2.4. *Repeated choice in a single Prisoner's Dilemma* – In order to substantiate his claim, that: 'not even the Prisoner's Dilemma supergame is an accurate model of many of the situations to which the Prisoner's Dilemma has been applied,' R. Harrison Wagner draws attention to:

> three other conditions that must be met for a situation to be represented accurately by the Prisoner's Dilemma game: 1) There must be only two actors; 2) each must have one and only one opportunity to choose between the alternatives C and D, before payoffs are received; and 3) each must choose in ignorance of the choice made by the other. (Wagner 1983, 331)

Since I have already dealt with the two-actors assumption in my introductory remarks, I shall not comment on it here. As far as the third condition is concerned, I agree with Wagner that very often this condition will not be very realistic. But, as Wagner himself observes, replacing this assumption with the assumption that the players choose sequentially, does not get them out of a Prisoner's Dilemma. The introduction of this assumption, and the corresponding representation of the game in *extensive* form, does make clear however, that the second assumption *is* genuinely crucial to the existence of the dilemma. The dilemma exists because the Prisoner's Dilemma is a game of, what I should like to call, the 'Red Queen type'. A game in which the players already know -what Alice only finds out when she tries to correct herself after she has said that thunder is the cause of lightning- that they will not have a second chance. 'It's too late to correct it,' said the Red Queen: 'when you've once said a thing, that fixes it, and you must take the consequences.' (Carroll 1954 ed., 224)

In the situation I considered in section 2.2, Wagner's critique could be taken care of by assuming that *i* and *j* know that *h* will not hold on to the thirty days time limit in its ultimatum, provided that by that time serious negotiations with respect to *h*'s demands have begun, and that these negotiations may even drag on for months, but that in case these negotiations do not lead to the desired result, *h* eventually will start hostilities without further notice. An additional assumption must be, that the secret negotiations *i* and *j* conduct separately with *h*, are not so secret that both *i* and *j* would not find out if the other tried to appease *h*. Assume moreover that *i* will have to choose its policy first (because the mobilization of its armed forces takes more time), but that *j* cannot postpone its choice of policy much longer, and that after *j*'s choice *i* gets an opportunity to reverse its

policy, should this be necessary in the light of *j*'s choice. The payoffs of *i* and *j* in this game can be represented in extensive form in the following manner:

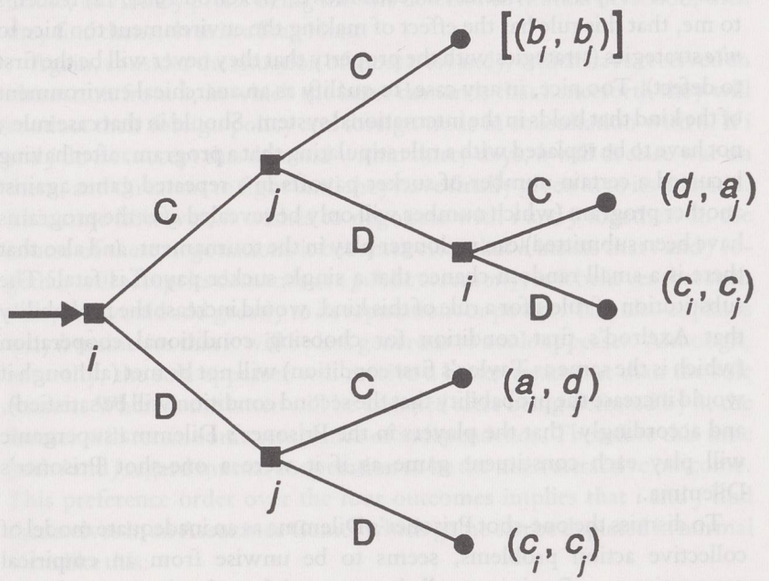

Figure 1: Repeated choice in a single Prisoner's Dilemma

In this situation there are two equilibrium outcomes – (b_i,b_j) and (c_i,c_j) – and *i* can safely try to reach the Pareto superior outcome. It can take the risk of choosing C the first time, because it can change to D if it turns out that *j* has chosen D, and so *i* chooses the conditional strategy 'cooperate, then defect if the other defects.' Now *j* effectively faces the choice between (b_i,b_j) and (c_i,c_j), and as a result, chooses C. This result can be generalized: 'in any situation in which players choose one after the other, with full knowledge of each other's choices, conditional cooperation will be the optimal strategy as long as no one can count on having the last choice.' (Wagner 1983, 333)

2.5. *The relevance of the one-shot Prisoner's Dilemma* – In the foregoing I have shown myself to be critical of the results of Axelrod's computer Prisoner's Dilemma tournament. Explicitly, when I questioned his operationalization of the discount factor *w*, as well as implicitly, when I discussed the possibility of cooperation in a Prisoner's Dilemma supergame in which a

player can be eliminated, as that possibility is ruled out by rule 3 of the tournament. This rule states that: 'there is no way to eliminate the other player or run away from the interaction. Therefore each player retains the ability to cooperate or defect on each move.' (Axelrod 1984, 12) It seems to me, that this rule has the effect of making the environment too nice to *nice* strategies (strategies with the property that they never will be the first to defect). Too nice, in any case, to qualify as an anarchical environment of the kind that holds in the international system. Should in that case rule 3 not have to be replaced with a rule stipulating that a program, after having incurred a certain number of sucker payoffs in a repeated game against another program (which number will only be revealed *after* the programs have been submitted) can no longer play in the tournament, and also that there is a small random chance that a single sucker payoff is fatal? The substitution of rule 3 for a rule of this kind, would increase the probability that Axelrod's first condition for choosing conditional cooperation (which is the same as Taylor's first condition) will not be met (although it would increase the probability that the second condition will be satisfied), and accordingly, that the players in the Prisoner's Dilemma supergame will play each constituent game as if it were a one-shot Prisoner's Dilemma.

To dismiss the one-shot Prisoner's Dilemma as an inadequate model of collective action problems, seems to be unwise from an empirical perspective too. For it is a well-documented fact, that in times of acute crisis the future closes in on the participants. (see, for instance, Janis 1982) The implication of this finding is, that under these circumstances the value of w apparently decreases sharply, and correspondingly, that for a certain class of decisions, not uncommon in international relations, the one-shot Prisoner's Dilemma can be an adequate model after all.[8]

3. Games with a coordination equilibrium

3.1. *The Assurance Game* – One of the conditions implicit in the Prisoner's Dilemma is, that both players are unable to provide any amount of the collective good by themselves. This condition also holds for the class of games that I shall discuss in this section. The most important difference between the Prisoner's Dilemma and these so-called coordination games is, that in the latter the actors prefer the outcome that they both cooperate to any other outcome. The most important similarity is, that also in these games each of them considers the outcome that it alone contributes to the provision of the collective good, to be the worst outcome.

Their preference order over the other two outcomes determines the kind of coordination game the actors are playing. If they prefer mutual defection to being the sole defector, then they are playing an Assurance Game, but if they prefer being the sole defector to mutual defection, then they find themselves in a Stag Hunt.

Again, consider the situation of the states i and j, which have received an ultimatum from h, in which the latter demands that henceforth they will conduct their foreign policy and foreign trade in consultation with h. If i and j do not accept this demand within thirty days, h will declare war on them, and threatens to fight until they have finally given in to its demand. This time however, h wishes to negotiate with i and j together. In the course of these negotiations it becomes clear that h insists that i and j together will accept its demands at a public ceremony, as it believes that such a display would add greatly to its reputation for power. This also explains why h maintains that it will be less generous to a sole appeaser. Although, naturally, the sole appeaser will receive a better treatment than the sole resister. Whereas the latter will be annexed after being defeated by h, the former will retain some semblance of independence. Therefore this time both i and j regard mutual cooperation to be the most attractive outcome. This preference order over the four outcomes implies that i and j find themselves in an Assurance Game, which game can be depicted in normal form like this:

		j C	j D
i	C	a_i, a_j	d_i, c_j
	D	c_i, d_j	b_i, b_j

Diagram 2: Assurance Game.

Both i and j will choose C, and the outcome will be (a_i,a_j). This outcome is an equilibrium, and also a coordination equilibrium. The strategy vector (a_i,a_j) is: 'such that either player is made no better off, not only if he himself unilaterally changes strategy, but also if the *other* player changes strategy.' (Taylor 1987, 71)

3.2. *Stag Hunt* – Due to the work of Kenneth Waltz and Robert Jervis, Rousseau's Stag Hunt is to students of international relations almost as familiar as the Prisoner's Dilemma. Take Jervis's illustration of the effects of anarchy on the possibility of cooperation between states (in this case

with respect to mutual disarmament) with the help of the Stag Hunt:

> Anarchy encourages behavior that leaves all concerned worse off than they could be ... This is true of the men in Rousseau's "Stag Hunt". If they cooperate to trap the stag, they will all eat well. But if one person defects to chase a rabbit -which he likes less than stag- none of the others will get anything. Thus, all actors have the same preference order, and there is a solution that gives each his first choice: (1) cooperate and trap the stag (the international analogue being cooperation and disarmament); (2) chase a rabbit while others remain at their posts (maintain a high level of arms while others are disarmed); (3) all chase rabbits (arms competition and high risks of war); and (4) stay at the original position while another chases a rabbit (being disarmed while others are armed).[9] (Jervis 1978, 167)

Which preference order can be realized in the situation I set out in the previous section, by stipulating that in the course of the negotiations with *h* it appears that the latter state again will be more generous in case there is a sole appeaser, than if both *i* and *j* decide to appease *h*, but that the conditions *h* will impose on the sole appeaser are still such, that to *i* and *j* mutual cooperation remains their most favoured outcome. The preference order of *i* and *j* in this situation can be represented in normal form as in diagram 3.

		j C	*j* D
i	C	a_i, a_j	d_i, b_j
	D	b_i, d_j	c_i, c_j

Diagram 3: Stag Hunt

Also in this situation *i* and *j* will choose C, and the outcome of the game will be (a_i,a_j), which is again a coordination equilibrium.

3.3. *Coordination games as a problem of collective action* – In the Assurance Game as well as in the Stag Hunt there are two equilibria: (a_i,a_j) and (b_i,b_j) in the Assurance Game, and (a_i,a_j) and (c_i,c_j) in the Stag Hunt. But since *i* and *j* in both cases prefer mutual cooperation to mutual defection, and, consequently, neither of them expects mutual defection to be the outcome, the unique Pareto-optimal outcome (a_i,a_j) will be the result. (Taylor 1987, 39) The same result obtains in case of an iterated Assurance Game or Stag Hunt. So neither of these games, whether in the one-shot

version, or in the iterated version, does imply a collective action problem.

In view of this conclusion it seems rather puzzling that the Stag Hunt has aroused so much interest among students of international relations. How can it be reconciled with Kenneth Waltz's view that 'the simple story' of the Stag Hunt has 'tremendous implications'?; (Waltz 1959 ed., 168) or, for that matter, with Robert Jervis's decision of all games to use the Stag Hunt to illustrate the possibility, that: 'although actors may know that they seek a common goal, they may not able to reach it'? (Jervis 1978, 168) Are they simply mistaken, or may there be something in the story that is left out in the usual representation of the Stag Hunt? As a matter of fact there is. Again it concerns the condition that payoffs are received immediately after the players have chosen their strategy.

Consider first the situation that it will take the protagonists quite some time before the issue between them could be settled by military means. So *h* has put a time limit of two years on its ultimatum, and *i* and *j* now face the decision whether or not to take the necessary measures to prepare for a contingency that only will materialize in two years time. Mutual cooperation is still their most preferred outcome, but it is also the case that: 'la prévoyance n'étoit rien pour eux, et loin de s'occuper d'un avenir éloigné, ils ne songeoient pas même au lendemain.' (Rousseau 1964, III, 166) Which implies that they discount a future common defence so heavily that it appears to them to be smaller than (not preferred to) instantaneous appeasement. (de Jasay 1989, 51) If it is further assumed, as in a Prisoner's Dilemma with repeated choice, that the state that finds itself to be the sole resister, will have the opportunity to switch to appeasement too, then there is not even a coordination problem. But if so, the implications of *this* Stag Hunt cannot be called tremendous either. Also the outcome of mutual defection can hardly be regarded to be Pareto-inferior, and accordingly this interpretation does not imply a collective action problem. In this interpretation Rousseau's parable turns out to be no more than a comment on the frailty of human nature. All of us should prefer the future stag, but things being as they are, we are all quite happy with our present hares. Notice, by the way, that this interpretation implies that such a thing as an iterated Stag Hunt simply cannot exist. (Hardin 1982, 189)

Consider now the situation in which a credible defence against *h* is again discounted by *i* and *j* (but not so heavily as in the foregoing interpretation) but where each state also takes into account that as the length of the time period between *t*o and *t*n increases, also the chances increase that the other, being short-sighted, will not make the necessary preparations for a successful common defence, and, as a consequence, that from a certain point in time the other state will have no other feasible option than to

appease *h*. Each state also realizes, that it will not be able to detect the true state of affairs with respect to the other's preparations, so that it must reckon with the possibility that only at *tn* it will find out that the other has decided to appease *h* (the properties of the game are no longer common knowledge). (compare Nicholson 1990 ed., 98-102)

This Stag Hunt does present a collective action problem, and this must be the situation Waltz has in mind, considering his characterization of the Stag Hunt: 'To allow in my calculation for the irrational acts of others can lead to no determinate solutions, but to attempt to act on a rational calculation without making such an allowance may lead to my undoing.'[10] (Waltz 1959 ed., 169) Under these circumstances both *i* and *j* might simply decide to choose the strategy that maximizes their minimum payoff, in this case D, but they could also let their decision depend on the subjectively estimated probability that the other will in fact cooperate, the cost of contributing to the collective good, and the (discounted) benefit of mutual cooperation. In this case each state faces the set of contingencies as shown in table 1 from *i*'s perspective (which table I have adapted from Taylor 1987, 46).

Table 1: Outcomes and *i*'s payoffs

i		$P_{i,<s-1}$	$P_{i,s-1}$
	C	no provision *-c*	provision *wb-c*
	D	no provision 0	no provision 0

Where $P_{i<s-1}$ stands for *i*'s subjectively estimated probability that *j* will defect, and $P_{i,s-1}$ for the estimated probability that *j* will cooperate, *c* for the costs of C, and *wb* for the discounted value of the provision of the collective good at t_n (I assume that the cost of cooperation is incurred right away). In these circumstances state *i* will choose strategy C as long as its subjectively estimated probability of it being pivotal exceeds the cost/discounted benefit ratio, otherwise it chooses its *maximin* strategy D (and the same goes for *j*): $P_{i,s-1} > c/wb$.

The subjectively estimated probability of being pivotal, clearly is a function of the length of the time period between t_o and t_n, and the degree of predictability in the international system of which *i* and *j* are part (in *N*-person games $P_{i,s-i}$ would also be a function of the *N*-1 number of players involved and the number of possible *s*-1 coalitions that can be formed,

Taylor 1987, 47). The longer the period of time between t_o and t_n, and the higher the level of uncertainty in the international system, the smaller a state's subjectively estimated probability that its contribution will be pivotal, and the more likely that it will choose to defect. (compare Nicholson 1990 ed., 40)

Now consider a third interpretation of the Stag Hunt. In this interpretation both *i* and *j* have a second chance. They are able to detect whether the other state defects or not, and they have the opportunity to change their strategy if this becomes necessary. Only this time there is the problem that *h* has put a premium on being the first appeaser (although to *i* and *j* this outcome is still not as attractive as mutual cooperation). (Wagner 1983, 339-340) This type of Stag Hunt too, poses a genuine collective action problem. The set of contingencies facing each state now looks as follows (again from the point of view of *i*):

Table 2: Outcomes and *i*'s payoffs with a premium on first defection

i		$P_{i,<s-1}$	$P_{i,s-1}$
	C	no provision *-c*	provision *wb-c*
	D	no provision 0	no provision *b'*

Where $wb\text{-}c > b' > 0$. In this situation *i* will cooperate as long as $P_{i,s-1} > c/wb\text{-}b'$, otherwise it will defect, and the same goes for *j*.

4. The Game of Chicken

4.1. *One-shot Game of Chicken* – Consider again the situation of the states *i* and *j*, as presented in section 2.2, but assume this time that with *g*'s assistance *i* or *j* alone is able to put up a credible defence against *h*, and that both states consider the cost of mutual appeasement to exceed the cost of providing the collective good by itself (with *g*'s assistance, of course). The most attractive outcome to *i* and *j* is to adopt a policy of strict neutrality, while the other state incurs the cost of building up a credible defence against *h*. But each state prefers the outcome that it has to carry the burden of a credible defence alone, to the outcome that neither of them tries to balance *h*. Therefore, the next best outcome to them is, that they both decide to balance *h*, and share the burden of a credible defence against their common

enemy. Their preference order over the four outcomes can be represented in normal form in the following way:

		j	
		C	D
i	C	b_i, b_i	c_i, a_j
	D	a_i, c_j	d_i, d_j

Diagram 4: Game of Chicken

The reason why the Game of Chicken can lead to the Pareto-inferior outcome of mutual defection, is that with both *i* and *j* able to put up a credible defence against *h* (again, thanks to the help of *g*) they constitute a group that with respect to the provision of this collective good is 'doubly privileged'. (Taylor 1987, 36) Both *i* and *j* are tempted to commit themselves one way or another to a policy of non-cooperation, and to take a free ride on the efforts of the other state. But in the conventional representation of the game this temptation is more than matched by the fear of the consequences if they both decided to free-ride. Accordingly, both *i* and *j* choose C, and (b_i,b_j) is the outcome, although not an equilibrium outcome.

The danger that both states may fail to contribute to the provision of the collective good, is anything but removed by assuming that they have a second chance, as in the Prisoner's Dilemma with repeated choice. In particular since the attractions of pre-commitment are such, that no player wishes to be thought of as having a second chance. The only situation in which repeated choice definitely implies the elimination of the danger of mutual defection, is if it is assumed that the state that moves first, say *i*, always has the last choice before payments are received, and that both states are aware of each other's choices. In this case *j* can safely choose D, as that leaves *i* with the choice between the worst and the next to worst outcome, so that *i* chooses C and (c_i,a_j) will be the outcome. This outcome *is* an equilibrium outcome. However that may be, it will be clear that Wagner's conclusion, that whenever actors choose one after the other, with full knowledge of each other's choices, and no actor can count on having the last choice, conditional cooperation will be their optimal strategy, does not hold in situations in which more than one actor is able to provide some amount of the collective good by itself.

4.2. *Iterated Game of Chicken* – The first thing to notice when analysing the possibility of cooperation in a Chicken supergame, is, as Hugh Ward has pointed out, that most of the time Games of Chicken are games of ruin. Games with the property that after a certain number of rounds of play in which both players defect, they reach 'the brink'. If not at least one player decides to cooperate in the next constituent game, then the game will end, as the players are no longer able to form a minimal contributing set with respect to the provision of the collective good in question, which good might even have been lost forever.[11] (Ward 1987, 28)

Consider again the situation of *i* and *j*, which states fully realize that giving in to *h*'s demands will lead to future rounds of demands, and that these, if the demands are accepted by both in each round, will eventually lead to the point that *i* and *j*, either alone or together, will no longer be able to put up a credible defence against *h*. But with each new round of demands there is the temptation for *i* and *j* not to cooperate with the other state in resisting *h*'s demands. Especially because non-cooperation in this round increases the probability that the other state may think that cooperation also will not be forthcoming in the next round, and accordingly increases the probability that the other state feels that it has no other option than to provide by itself some amount of the collective good in the rounds of the game that yet have to follow. So again the future casts its shadow in two opposite directions. Not only in the direction of cooperation in order to prevent the collective good from getting irretrievably lost, but also in the direction of non-cooperation with a view to acquire a reputation for toughness.

Now assume that both *i* and *j* know only by approximation when they will be at the brink. Accordingly, they allow for a certain margin of safety, which for the moment I assume to be the same for both states, and they will choose C if this margin is reached. Moreover, to both *i* and *j* the attraction of choosing D increases with the estimated number of constituent games left in the game before their margin of safety is reached. As a result, both *i* and *j* choose D in each constituent game until their margin of safety is reached, provided *h* demands only small concessions of them at a time. (Ward 1987) After that, they choose C, but they revert to D if they think the margin of safety is re-established, and so on. A quite different picture emerges, if it is assumed that one state, say *i*, allows for a larger margin of safety than the other state. In this situation, both players choose D, until *i*'s margin of safety is reached, after which it chooses C, while *j* continues to choose D. After that, the Chicken supergame stays at *i*'s margin of safety, and *j* enjoys an undisturbed free ride on *i*'s efforts. (This may not be true if *i* reasons that *j* must realize that this outcome is 'unfair', see the next section.)

5. Conclusion: the problem of the free rider

In this section I wish to draw attention to two general points that have emerged from my fact-deficient exercise, but that I have ignored thus far. The first may be nothing more than a curiosity, and accordingly I shall only note it in passing before I turn to discuss the second point, which appears to me to be well worth closer inspection. My first general point then, concerns the fact that the assumptions that can brighten the prospects of mutual cooperation between actors that have a preference order similar to the players in the single-shot Prisoner's Dilemma – for instance, that the actors find themselves in a supergame, or in a single game with repeated choice – can actually worsen the prospects of mutual cooperation if introduced in the analysis of the Game of Chicken and the Stag Hunt. Games that is, that in their standard representation have mutual cooperation as their outcome.

My second general point is that the common proposition, that: 'the core of the problem of collective goods, is the classic problem of the free rider,' as Nicholson has phrased it, is generally speaking untenable. (Nicholson 1990 ed., 131) It does not apply to the actors that find themselves in a Prisoner's Dilemma, or in a Coordination Game. These actors know that each of them is unable to provide any amount of the collective good by itself. Consequently, their main worry is not that the other may take a free ride on their efforts, but that the consequences of an unsuccessful contribution to the provision of the collective good may be their undoing. Also this proposition does not apply to the actors involved in a one-shot Chicken Game. These are definitely bothered by the fact that the other may be able to free-ride, but such thoughts do not prevent them from contributing to the provision of the collective good, since not contributing could mean an even worse outcome.

Moreover, the proposition does not hold in case one actor is able to provide a certain amount of the collective good by itself, and orders it preferences in the same manner as a player in a Game of Chicken, and the other actor is not able to provide a certain amount of the collective good alone, and has the same preference order as a player in a Prisoner's Dilemma. Clearly there exists no collective action problem under these circumstances. The Chicken-player will provide the collective good knowing full well that the PD-player will take a free ride on its efforts. Although one can conceive of a set of circumstances under which the 'exploitation of the great by the small' (Olson 1971 ed., 29) in this Chicken/PD-game may indeed lead to a Pareto-inferior outcome. This may happen if the power differential between the Chicken-player and the

PD-player has been established only recently, and is still such, that the former is barely able to provide a credible defence by itself, and the latter only just not. In that case the Chicken-player's 'coefficient of sensitivity to gaps in payoffs' may still be so large as to induce it to choose D instead of C after all. (Grieco 1988, 500–501)

The other situation I can think of where the problem of the free rider can prevent the provision of a collective good, is in a Chicken supergame in which one of the players has successfully established a reputation for toughness (as state *j* succeeded in doing in the situation I set out at the end of section 4.2). This equilibrium can be destroyed in case the sucker player reasons that, as more and more rounds of the game are being played, the tough player must begin to realize that the sucker player may reach the point that it decides that any outcome is better than the certainty of a string of sucker-payoffs of indefinite length. That is to say, that the tough player assumes, not unnaturally under the circumstances, and perhaps because it is familiar with the history of Japan's decision to declare war on the United States in December 1941, (Sagan 1989) that the value of the sucker player's discount parameter is inversely related to the number of constituent games played, and eventually will reach a value such, that to the latter the discounted value of an indefinite string of sucker-payoffs will be less than the value of mutual defection. And that in the light of this reasoning, the sucker player thinks that with each constituent game the probability increases that the tough player will choose C in the next constituent game in order to prevent the supergame from going over the brink. The sucker player will then choose to defect in the next constituent game if the expected value of a single defection (the value of outcome *a* multiplied by the subjectively estimated probability that the tough player will cooperate) exceeds the discounted value of the indefinite string of sucker-payoffs, and, in case it turns out that it is mistaken about the value of *p*, the problem of the free rider will have prevented the collective good from being provided.

Notes

* I am grateful for comments by Kees Aarts, Michael Nicholson, Huib Pellikaan and Dick Ruiter on earlier drafts of this article.

1. Together with a fourth assumption, to the effect that the environment is continuously undergoing changes that have negative or positive repercussions on the capabilities of the states in the system, these assumptions constitute the hard core of the realist tradition in the study of international relations.

2. Most of the time in the literature a distinction is made between dissuasion by

deterrence on the one hand, and dissuasion by defence on the other. Deterrence implies that a state is able to frighten an adversary out of attacking by threatening unacceptable punishment. Defence means that a state is able to make an invasion too costly for the invader. Deterrence is a more pure collective good than defence, as it is less divisible, it is more difficult to exclude another state from consuming it, and its degree of non-rivalness is higher. (Waltz 1981, 4-7, Thies 1987, 300-307 & 326)

3. A strategy is a complete plan for playing a game that takes account of all possible contingencies, including the choices of the other relevant actors. (see Ordeshook 1986, 108) In this article I assume that the players only use pure strategies.

4. Huib Pellikaan has drawn my attention to the fact that the derivation of the utilities *i* and *j* attach to the four outcomes, is anything but straightforward. So that, strictly speaking, it must be an assumption that *i* and *j* attach these utilities to the four outcomes.

5. I use this notation, instead of the more common one of $4>3>2>1$, because the introduction of numbers such as these appears to be: 'a source of trouble, for one must develop an almost inhuman self-control not to read into these numbers those properties which numbers usually enjoy.' Luce and Raiffa 1957, 16.

6. However, Kreps et al. have demonstrated that: 'a "small amount" of the "right kind" of incomplete information' (in this case, that at least one of the players is unsure about the other player's payoffs, so that the properties of the game are no longer common knowledge): '*must* yield a significant measure of cooperation in equilibrium.'(italics in original) In this case a so-called *sequential* equilibrium. (Kreps et al. 1982, 246)

7. Other possible strategies are Ordeshook's *a+*: choose C for the first trial; choose C if the other person chooses C in the previous trial. Once the other person chooses D, choose D thereafter. (Ordeshook 1986, 445) And Shubik's *Ak*:

C is chosen in the first game, and it is chosen in each subsequent game as long as the other player chooses *C* in the previous game; if the other Defects in any game, *D* is chosen for the next *k* games; *C* is then chosen no matter what the other player's last choice is; it continues to be chosen as long as the other player choose *C* in the preceding game; when the other player next Defects, *D* is chosen for $k+1$ games; and so on; the number of games in which the other player is 'punished' for a Defection increases by one each time; and each time there is a return to *C*. (Taylor 1987, 73)

As *a+* fails the test of forgiveness (Axelrod 1984), and strategy *Ak* seems to me to be a bit too intricate for ordinary states with short-term memories, I shall not take these strategies into consideration.

8. This also means that the conventional assumption in the Prisoner's Dilemma supergame, that *w* is constant, is inappropriate if we want to model as part of a constituent game in such a supergame a state's decision whether or not to contribute to a defensive alliance in times of acute crisis. Note moreover, that this assumption effectively turns the Prisoner's Dilemma supergame into a static game.

9. It seems to me that the way in which Jervis orders the preferences of the individual states in an anarchical environment, is open to doubt. Would each state not prefer the outcome that it would have a (clear) superiority in arms, to the outcome that all states find themselves in a state of equal disarmament? Even if the former

outcome would be more costly than the latter? If one accepts this line of argument, then the preference order of the players would be identical with the one of the players in the Prisoner's Dilemma. George Quester comes to the same conclusion: 'The first choice of any nation may well be to sneak into a position of superiority even if deceit is required; compared with this, substantial disarmament may be only the second preference, ahead of a mutually continued arms race.' (Quester 1977, 102)

10. Rousseau appears to think – and Waltz seems to agree with him – that a single defector can destroy whatever cooperation there is in a society: 'all the citizens without exception would have to be [equally] good Christians; if by ill hap there should be a single self-seeker or hypocrite ... he would certainly get the better of his pious compatriots.' (Waltz 1959 ed., 169) If this would be true, than a strategy of conditional cooperation could not be, to use Axelrod's expression, 'collectively stable'. But things are not as hopeless as that. Axelrod has shown that, given certain conditions, Tit-for-Tat cannot be invaded by all-D, and accordingly can be collectively stable. (Axelrod 1984, 207-215) Taylor moreover, has demonstrated that under certain conditions a mixture of conditional and unconditional strategies (including all-D) can be an equilibrium. (Taylor 1987, 88-96)

11. This is certainly a property of the original version of the Game of Chicken, where already in the first round mutual defection implies mutilation or even death.

Literature

Axelrod, R., *The evolution of cooperation* (New York: Basic Books, 1984);

Blümel, W., R. Pethig and O. von dem Hagen, "The theory of public goods: a survey of recent issues", *Journal of Institutional and Theoretical Economics* 142, 1986, p. 242-309;

Bull, H., *The anarchical society. A study of order in world politics* (London and Basingstoke: Macmillan, 1977);

Carroll, L., *Through the looking glass* (1871, London: J.M. Dent & Sons, 1954 ed.);

Grieco, J.M., "Anarchy and the limits of cooperation: a realist critique of the newest liberal institutionalism", *International Organization* 42, 1988, p. 485-507;

Hardin, R., *Collective action* (Baltimore: The Johns Hopkins University Press, 1982);

Head, J.G., "Public goods and public policy", *Public Finance* XVII, 1962, p. 197-219;

Hume, D., *Essays, moral, political and literary* (Indianapolis: Liberty Press, 1985 ed.);

Janis, I. L., *Groupthink* (Boston: Houghton Mifflin Company, 1982 ed.);

Jasay, A. de, *Social contract, free ride. A study of the public goods problem* (Oxford: Clarendon Press, 1989);

Jervis, R., "Cooperation under the security dilemma", *World Politics* XXXI, 1978, p. 167-214;

Kreps, D.M., P. Milgrom, J. Roberts and R. Wilson, "Rational cooperation in the finitely repeated Prisoners' Dilemma", *Journal of Economic Theory* 27, 1982, p. 245-252;

Luce, R.D. and H. Raiffa, *Games and decisions. Introduction and critical survey* (New York, etc.: John Wiley & Sons, 1957);
Nicholson, M., *Formal theories in international relations* (1989, Cambridge etc.: Cambridge University Press, 1990 ed.);
Olson, M., *The logic of collective action. Public goods and the theory of groups* (1965, Cambridge, Mass.: Harvard University Press, 1971 ed.);
Ordeshook, P., *Game theory and political theory: an introduction* (Cambridge, etc.: Cambridge University Press, 1986);
Quester, G.H., *Offense and defense in the international system* (New York, etc.: John Wiley & Sons, 1977);
Rapoport, A. and M. Guyer, "A taxanomy of 2x2 games", *Peace Research Society (International) Papers* VI, 1966, p. 11-26;
Rousseau, J.-J., *Oeuvres complètes*, III (Paris: Gallimard, 1964);
Sagan, S. D., "The origins of the Pacific war", in: Rotberg, R.I. and T.K. Rabb (eds.), *The origin and prevention of major wars* (Cambridge, etc.: Cambridge University Press, 1989), p. 323-352;
Schelling, T., "The strategy of conflict. Prospectus for a reorientation of game theory", *Journal of Conflict Resolution* II, 1958, p. 203-264;
Snidal, D., "The game theory of international politics", *World Politics* XXXVIII, 1985, p. 25-57;
Thies, W.J., "Alliances and collective goods", *Journal of Conflict Resolution* 31, 1987, p. 298-332;
Taylor, M., *The possibility of cooperation* (Cambridge, etc.: Cambridge University Press, 1987);
Wagner, R.H., "The theory of games and the problem of international cooperation", *American Political Science Review* 77, 1983, p. 330-346;
Walt, S.M., "Alliance formation and the balance of world power", *International Security* 9, 1985, p. 3-41;
Waltz, K.N., *Man, the state, and war. A theoretical analysis* (1954, New York, etc.: Columbia University Press, 1959 ed.);
Waltz, K.N., *The spread of nuclear weapons: more may be better* (London: International Institute for Strategic Studies, 1981);
Ward, H., "The risks of a reputation for toughness: strategy in public goods provision problems modelled by chicken supergames", *British Journal for Political Science* 17, 1987, p. 23-52.

Toekenning van de Daniël Heinsius-scriptieprijs 1991

De kernredactie van *Acta Politica* heeft eind 1990 besloten een prijs in te stellen voor politicologische scripties, omdat, naar de mening van de redactie, veel kwalitatief goed werk dat door studenten in het kader van hun afstuderen wordt verricht, nauwelijks buiten de kleine kring van beoordelaars bekend wordt. Juist om meer bekendheid te kunnen geven aan kwalitatief hoogwaardig werk van studenten heeft de redactie gemeend, dat bekroning van de beste scriptie het meest geëigend is door publikatie van een op de scriptie gebaseerd artikel in *Acta Politica*, het wetenschappelijk tijdschrift van de Nederlandse Kring voor Wetenschap der Politiek.

Het is de bedoeling van de redactie de Daniël Heinsius-scriptieprijs jaarlijks uit te reiken aan een student die, naar haar oordeel, in zijn of haar scriptie in het voorgaande academisch jaar op de meest aansprekende wijze met behulp van politicologische theorieën en/of methoden een antwoord heeft gegeven op een politicologisch relevante probleemstelling.

De prijs is vernoemd naar Daniël Heinsius, die kan worden beschouwd als de eerste hoogleraar in Europa met een leeropdracht op het terrein van de politieke wetenschap. In 1612 werd hij aan de Leidse universiteit benoemd tot hoogleraar *Politices*. Hij was zozeer geliefd om zijn uitnemende colleges dat hij bekend stond als het *lumen academiae*. Door middel van de instelling van de Daniël Heinsius-scriptieprijs wil de redactie nieuwe generaties politicologen aanmoedigen hun licht niet onder de korenmaat te zetten.

Ter mededinging naar de Daniël Heinsius-scriptieprijs 1991, voor scripties geschreven in het academisch jaar 1990-1991, heeft de redactie vier scripties ontvangen. Het kleine aantal wijt zij vooralsnog aan de relatieve onbekendheid van de prijs. De redactie was zeer gelukkig over het kwalitatieve niveau van de ingezonden scripties. Na rijp beraad heeft de redactie besloten de prijs toe te kennen aan mevrouw drs. Karin Wittebrood voor haar scriptie *Kennis van Zaken: Een onderzoek naar de Politieke Kennis van de*