Scandinavian Political Studies, Vol. 10 – No. 2, 1987
ISSN 0080–6757

Reputation Pays: Game Theory as a Tool for Analyzing Political Profit from Credibility*

Hannu Salonen, University of Turku
Matti Wiberg, Academy of Finland and University of Turku

The article provides a general account of information problems in politics and games, and an analytical exploration of reputation effects in politics in terms of game theory. It is shown, by three game theoretical examples, that reputation effects may in some circumstances be of essential importance in determining optimal choices.

Reputation plays a central role in many political contexts. Consider for instance the case of spies as information senders, any military or non-military deterrence, bargaining over cabinet coalition alternatives or over wage agreements. Or consider for instance the so-called Confidence Building Measures, which have been on the agenda in international politics at least from the so-called Eden Plan in 1955–1956 and in the follow-up to Eisenhower's 'open skies' proposal in the United Nations. Or consider legislative leadership (Calvert 1986), or OPEC's attempts to maximize oil revenues in the long run (Alt, Calvert, & Humes 1986), or consider any such strategic interaction situation in which there are some differences in the information available to the actors and where beliefs and expectations of the other actors' motivation play some role. Reputation may pay for those actors who take these effects into consideration. Rational political actors may manipulate the strategic interaction situation they face for their own political profit. To maintain reputation and credibility is not costless. It may, however, be quite profitable to use some amount of the resources available in order to maintain an effective record of reputation. Reputational considerations can in some circumstances be major determinants of the choices among alternative decisions. In order to optimize the political profit available, current decisions must optimize the tradeoffs between short-term consequences and the long-run effects on one's reputation. Many political decisions depend on trusting someone whose motives are

* We would like to thank Hannu Nurmi and an anonymous referee of SPS for their comments on a previous version of this paper.

Scandinavian Political Studies, Vol. 10 – No. 2, 1987
ISSN 0080–6757

Reputation Pays: Game Theory as a Tool for Analyzing Political Profit from Credibility*

Hannu Salonen, University of Turku
Matti Wiberg, Academy of Finland and University of Turku

The article provides a general account of information problems in politics and games, and an analytical exploration of reputation effects in politics in terms of game theory. It is shown, by three game theoretical examples, that reputation effects may in some circumstances be of essential importance in determining optimal choices.

Reputation plays a central role in many political contexts. Consider for instance the case of spies as information senders, any military or non-military deterrence, bargaining over cabinet coalition alternatives or over wage agreements. Or consider for instance the so-called Confidence Building Measures, which have been on the agenda in international politics at least from the so-called Eden Plan in 1955–1956 and in the follow-up to Eisenhower's 'open skies' proposal in the United Nations. Or consider legislative leadership (Calvert 1986), or OPEC's attempts to maximize oil revenues in the long run (Alt, Calvert, & Humes 1986), or consider any such strategic interaction situation in which there are some differences in the information available to the actors and where beliefs and expectations of the other actors' motivation play some role. Reputation may pay for those actors who take these effects into consideration. Rational political actors may manipulate the strategic interaction situation they face for their own political profit. To maintain reputation and credibility is not costless. It may, however, be quite profitable to use some amount of the resources available in order to maintain an effective record of reputation. Reputational considerations can in some circumstances be major determinants of the choices among alternative decisions. In order to optimize the political profit available, current decisions must optimize the tradeoffs between short-term consequences and the long-run effects on one's reputation. Many political decisions depend on trusting someone whose motives are

* We would like to thank Hannu Nurmi and an anonymous referee of SPS for their comments on a previous version of this paper.

uncertain. If an agent is uncertain about the goals, motives, and preferences of someone upon whom he must depend, either to provide information or to make decisions, then the extent to which he trusts the other will be based on the partner's earlier actions. Thus there is an incentive for an enemy to behave like a friend in order to increase his future opportunities, and for partnerships to last until someone cashes in (Sobel 1985). Reliability can only be communicated through actions. An agent becomes credible by consistently providing accurate, valuable information or by always acting responsibly.

In this article, we present some game-theoretical results which throw light on the effect of reputation in political contexts. The article begins with a general account of information problems in politics in general and in games in particular. These introductory remarks are in order to motivate the game theoretic study of reputation effects. We then select a few important cases, discuss the nature and structure of the possible strategic interaction situations in these cases, and show how some recent results in game theory nicely make rather complex things more illuminative. The article should be read as an attempt to model reputation and credibility considerations in an exact way in politically relevant strategic interaction contexts by synthesizing a number of unnecessarily separated perspectives.

What then is *game theory*? It is a mathematical theory to explicate optimal choices in interdependent decision situations, wherein the outcome depends on the choices of two or more actors. Game theory concentrates in any interaction situation on (1) the set of actors, called players; (2) a set of strategies available to each player; (3) a set of outcomes, each of which is a result of particular choices of strategies made by the players on a given play of the game; and (4) a set of payoffs accorded to each player in each of the possible outcomes. Hence, *bi- and multilateral decision making* is the essence of game theory. The actors in game theory are thought to be *rational* in the sense that they try to maximize net benefits, and choose courses of action based on their own preferences and expectations of how others will behave. Political actors are generally concerned with the implications of current actions for the future. If we want to make political science into a science proper, we should switch from ad hoc expectation-generation to the theoretically more attractive approaches.

Several arguments in favor of the use of game theory have been presented (Schelenker & Bonoma 1978). Firstly, game theory can serve as a skeletal analogy of many social situations and contexts. Games can be seen as structural analogies of real-world interactions. In constructing a game analogy, an attempt is made to dissect from the complexities of real social interactions some fundamental structural aspects that can be employed to facilitate our understanding of the actual situations. The game is, of course, distinguished from the complexity and completeness of the reality; yet it

lays claim to capturing the broad *structural* outlines of interaction, though not the *process* by which people make choices.

Once constructed, the analogy can be manipulated by the user and applied to many aspects of the world that were not in its original domain; this *heuristic* function is a second major advantage of games. Games serve the heuristic function of helping the user construe phenomena from new and unique perspectives. Games permit perspective shifts.

Thirdly, games can also be used to study the degree to which actors depart from normative criteria of rationality. Essentially, the game is used as a structure within which 'rational' or 'irrational' behavior can take place, as determined by one or another mathematical model of decision behavior. Hence, games facilitate an analysis of rational – irrational action. They reveal how rationality is disturbed by social factors.

Fourthly, games facilitate testing theoretical propositions. A game can be used to provide a researcher with a high degree of control over the situation confronted by subjects. When a theory is addressed to situations that have the properties of a game, the game can be employed to test the specific predictions made by the theory. When games are used in this role, in one sense the utility of the findings is not affected by the degree to which the game situation is isomorphic to the actual situation being modelled – rather, the crucial question is how well the game setting allows the investigator to test the hypotheses of interest. In another sense, though, the problem of articulation with the real world situations does not go away, it just gets properly put back one step to ask if the theoretical principles under consideration, however supported, are applicable to a wide range of situations. The use of games allows theoretically-derived hypotheses to be tested.

Game theory has been criticized on several grounds. Of all the criticisms of games, the most prevalent is that games and real strategic interactions do not correspond isomorphically. But this is surely something a proponent of game theory is willing to admit. By definition, an analogy is different from the real thing – otherwise it would not be an analogy but would be the thing itself (Schlenker & Bonoma 1978, 21).

Another point of criticism centers around the demand for information in game theory. Game theory often seems to demand more information that can feasibly be obtained. Ironically, it cannot always adequately incorporate other important available information – including relevant historical details about the context of interaction, insights into the personalities and the behavior of decision makers, and understanding of the nature of the real world processes under consideration. These shortcomings of game theoretic analysis have led some analysts to conclude that its usefulness as a theoretical guide to the empirical study of politics is seriously impaired. This conclusion shows a misunderstanding of the power of game

theory by treating it as a descriptive and not as an analytical tool, as Snidal (1985, 26) points out. It is not useful nor is it even fair to evaluate the importance of game theory by the worst applications of it. There surely are various examples of misguided applications of game theory, but the potential usefulness of game theory as an analytical tool should not be measured according to these. Game theory as any other theory should be evaluated by its most useful and fruitful or otherwise best uses.

Much of earlier game theory seemed (fairly or unfairly) to be irrelevant to the scientific study of politics, because so much of it concentrated on static considerations only. It is, of course, quite obvious that the most interesting strategic interaction situations in politics are dynamic. The most significant relationships between political actors can be adequately studied only in a dynamic frame. In real world politics, there is always a tomorrow to be taken into consideration.

We must warn against a possible misunderstanding here. It makes perfectly good sense to use a static theory to express dynamic phenomena, also. This can be done by using *strategies*: we reformulate a sequence of decisions, even a whole policy of action, as a single, contingency-laden decision or 'gameplan'. The multimove game can then be reduced to a game in which each player makes just one move – his choice of strategy. Another way to have a static model express a process over time is to reinterpret the static solutions as steadystate solutions of a corresponding model that is in continuous (or periodic) operation, so that quantities of goods, for example, become rates of flow (Shubik 1983, 8–9).

Information Constraints in Politics and Games

Decison making in real world politics is typically decision making under some sort of uncertainty, e.g. the decision makers do not typically know each other's preferences, nor do they typically know even all the possible action alternatives open to themselves or to other political actors. In game theory there is a distinction between on the one hand *complete* and *incomplete* information, and between *perfect* and *imperfect* information on the other. The first distinction refers to the amount of information the players have about the rules of the game. The second distinction refers to the amount of information they have about the other players' and their own previous moves (and about previous chance moves). In a game of perfect information, each player is fully informed about all prior choices when it is his turn to move. Players with complete information know both the rules of the game and the preferences of the other players over the set of outcomes. The rules of the game refer to information about the set of possible outcomes and of the choices available to each player at each move

plus information indicating each player's ability to determine his situation at each move. Thus, complete information obtains when each actor knows (a) who the actors are, (b) all actions available to all actors, (c) all potential outcomes, and (d) that everyone in the strategic interaction situation knows all of these things, (e) how the outcomes are related to choices, i.e. the outcome function, (f) the preferences available. From this we see that there are various degrees of incomplete information ranging from total ignorance to complete information.

Although the assumption of complete information is analytically convenient, it precludes consideration of important features of real world politics. It also precludes the possibility of studying those formal and informal procedural details that political actors use either to estimate how others will act or to effect how others will act. With incomplete information, beliefs about preference can change as the interaction proceeds. If I observe you acting one way in some early stage of our interaction, I may infer one thing about your preferences, whereas if I observe a different choice, I may believe something else. And if I believe something different, then I may act differently in subsequent interactions. Thus actions serve a dual purpose, to affect outcomes and to affect beliefs, and thus future actions. Hence, actions affect outcomes also indirectly by beliefs. In this world of incomplete information, the usual definitions of equilibrium must be augmented with conditions on the stability and consistency of beliefs. Since both strategies and beliefs can vary, an equilibrium in a world of incomplete information consists of a set of beliefs and strategies such that no one has any incentive to change their strategies given their beliefs at any stage of the interaction and all beliefs are consistent with the strategies of other actors and prior assessments about their preferences (Ordeshook & Palfrey 1986, 3–5).

Another important feature in real world politically relevant interactions is almost always that information is unevenly spread among the relevant actors: information is typically asymmetric in political contexts. Hence, in respect to information, we most typically face case 4 in political contexts:

INFORMATION:	Complete	Incomplete
Symmetric	1	2
Asymmetric	3	4

Most of the classical game theory focused on case 1. Quite recently many game theorists have focused on the most realistic situation (case 4), and they have presented some quite surprising results (for a good summary, see Wilson 1984). In these situations the actors may have some 'private information', e.g. information that only one actor knows (usually about

himself), such as how much the actor in question values some potential arrangements. (Note that case 3 is, by definition, empty.)

It is important to make a distinction between the quality and quantity of signals available. Consider the following simple crosstabulation:

SIGNALS:

		Quality	
		+	−
Quantity	+	1	2
	−	3	4

In some situations we have many and good signals (1). In some others we have only a few and of bad quality (4). It is, of course, a different decision situation when we have many signals of bad quality (2) and when we have only a few, but of good quality (3). Hence we have four radically different situations, which all have different consequences for decision making.

Information may, of course, also be used strategically (see Crawford & Sobel 1982). When and how to reveal and when and how to conceal, when to believe revealed information and when not, are questions which should be taken into consideration because their answers may be used as a means to improve one's outcomes.

As we have seen, lack of knowledge in different strategic interaction situations may manifest itself in various ways. The following table summarizes some important possibilities:

LACK OF KNOWLEDGE CONCERNING:

	OWN	OTHERS
1. Players	1	2
2. Action alternatives	3	4
3. Preferences	5	6
4. Utilities	7	8
5. Number of moves	9	10
6. Timing of moves	11	12
7. Recall of moves	13	14
8. Outcome function	15	16

Lack of knowledge concerning the four first issues is usually dealt with in the relevant literature, so we do not want to take up these issues for further discussion. The issues 6, 7 and 8 are more interesting here.

By crosstabulating the alternatives 5 and 6 we get the following matrix:

DIFFERENT GAME TYPES:			TIMING OF MOVES	
			Simultaneously	Successively
NUMBER OF MOVES	One shot		1	2
	Repeated	Finitely	3	4
		Infinitely	5	6

Many strategic interaction situations in real world politics are of types 2 and 4. But note that, for instance, voting usually is of type 1, 3 or 4. In some cases the players do not know the actual number of moves. They may know that the game they are playing must (for some reasons) have only a finite number of moves, but they do not know how many moves are left for themselves, the other player(s) or both. This empirical phenomenon can also be taken into consideration in terms of game theory by introducing the number of the subsequent moves in a probabilistic fashion.

What about the recall of previous moves? Consider the following 2 × 2-table:

RECALL OF PREVIOUS MOVES:

		THEY	
		Perfect	Imperfect
WE	Perfect	1	2
	Imperfect	3	4

There are, of course, various determinants of these factors. One thing that should be taken into consideration is the *timespan* between the actual moves: people usually forget things. One should take this into consideration in exact terms, too, because it is so obviously true in many human interactions.

Another determinant of the recall is, again quite obviously, the importance of the previous move(s). *Ceteris paribus* we remember important things better. Technically, this is not difficult to take into consideration in game theory, since all we need is to put weights on the different moves: the more important the move is to some actor, the heavier the weight should be assigned to the move.

A further complication, which we will not deal with but is still worth mentioning, is that there may be *information lags* of various kinds. It is not always the case that the other player(s) receive information immediately concerning the other player(s)' move(s) or the consequences of his own or other players' move(s). Because this may be of some importance for the further development of the game, the possibility of information lags should also be taken into consideration.

Games with incomplete information give rise to an infinite regress of reciprocal expectations on the part of the players. This can manifest itself

in many different ways. Players could, for instance be unaware of each other's payoff functions, but could know their own payoff functions. In such a game each player's choice will depend on what he expects or believes to be the other players' payoff functions as these are of crucial importance for their behavior in the game. Furthermore, each player's strategy choice will also depend on what he expects to be the other players' expectations about his own payoff function. However, there is much work in game theory that completely avoids the difficulties associated with sequences of higher and higher-order reciprocal expectations. Harsanyi's classical treatises (1967–1968) started this path of research by reducing the analysis of games with incomplete information to the analysis of certain games with complete information so that the problem of sequences of higher and higher-order reciprocal expectations will simply not arise. We could, of course, in principle use the sequential-expectation model for any given game with incomplete information by considering the relevant infinite sequences of higher and higher-order subjective probability distributions, i.e. subjective probability distributions over subjective probability distributions. But this would make the analysis very complex indeed. In Harsanyi's model, it is possible to analyze any game with incomplete information in terms of *one* unique probability distribution of the game. Thus, his approach amounts to replacing a game involving incomplete information by a new game which involves complete but imperfect information. Harsanyi's analysis rest on a Bayesian hypothesis, that is, in dealing with incomplete information, every player assigns a *subjective* joint probability distribution to all variables unknown to him – or at least to all unknown independent variables, i.e. to all variables not depending on the players' own strategy choices.

The Harsanyi analysis can be extended to games in which the players do not have consistent subjective probabilities, so that they might have different probabilities attached to the same event (see Aumann 1974; Shubik 1983, 280), but this does not concern us in the present context. Instead of that, let us look at some further important features of use of information in strategic interactions.

There is one special case of sequential expectations or beliefs which is of some importance for the present paper. The concepts of *mutual beliefs* and of *common knowledge* have recently received much attention from economists, philosophers and statisticians (cf. Lewis 1969; Aumann 1976; Heal 1978, Milgrom 1981). Mutual beliefs are usually characterized by saying that in a certain context C it is mutually believed that p if and only if (1) everyone in C believes that p; (2) everyone in C believes that everyone in C believes that p; (3) everyone in C believes that everyone in C believes that everyone in C believes that p; and so on *ad infinitum*. The last clause should be taken quite seriously: the cases of mutual beliefs and of common

knowledge must be separated from all those cases where we have only some shorter sequence of iterated believing or knowing. (Note that games with complete information means that the strategies available to all players and the resulting payoffs and how they result from strategy n-tuples for all actors are common knowledge among the actors.)

Information is costly and processing capacity is limited. After a certain limit, it does not pay anymore to buy or search and/or process more of it. Before determining whether or not it is worth gathering and processing more information, which is usually not completely decisive, it is first desirable to ask 'What is *perfect* information worth?'. There are several determinants of information usefulness. The value of an information system is the increase in expected utility resulting from utilization of the system. The demand value for an information system is the maximum amount, measured in the same units as those in which the outcome of the decision is measured, that the decision maker would be willing to exchange for the system, given that he now employs the null system. The following four determinants of *information value* for a decision maker are of crucial importance (see Hilton 1981 and the literature mentioned there).

(i) The decision maker's action set, i.e. the decision maker's flexibility.
(ii) The payoff function for the decision maker, i.e. the decision maker's technology and environment and his relative preferences for outcomes.
(iii) The decision maker's initial uncertainty about aspects of the technology or environment.
(iv) The nature of the information system, e.g. such attributes of information as timeliness and accuracy.

It has been analytically proven (see Hilton 1981 for references) that there is no general monotonic relationship between (1) action flexibility and information value, nor between (2) the degree of absolute or relative risk aversion and the information value, nor between (3) wealth and information value (see also Arrow 1984 esp. ch. 9 and Ratchford 1982).

There are also other informational activities beside that of information *acquisition*. Strategic interaction involves information transmission. The *dissemination* of information to other political agents can be very important. Sometimes it may pay to disseminate gratuitously, or even to incur cost to 'push' information to others (cf. advertising). There is also a choice between disseminating publicly ('publishing'), or else privately to a select audience. As a question of authenticity might arise in all such cases, the receiver of information may devote effort to the process of *evaluation*, possibly assisted by *authentication* activities (or hampered by *deception* activities) on the part of the disseminator. There is also the possibility of unintended dissemination (espionage or monitoring) on the part of the information seekers – possibly leading to countermeasures in the form of security

(secrecy-maintaining) activities by the possessors of information (Hirschleifer & Riley 1979, 1398).

A political actor may profit from not disclosing all information he possesses. And an agent may even better attain his goals by deceiving the other actor(s) in some respects. Deception means the conscious misrepresentation of one's preferences to induce other actor(s) to behave in such a way that the outcome induced is better for the deceiver than one that could have been realized without misrepresentation (Brams 1977). A deceiver as well as an agent that just does not disclose information he possesses usually confronts the information *leakage* problem. Whether an agent confronts the leakage problem depends, of course, on what the other actors believe about the first mentioned agents' information. A political agent confronts the information leakage problem only if at least one of the other political actors recognize (or believe) that the agent in question has an informational advantage. Stated technically, this amounts to the fact that at least one actor has information that is not common knowledge among all the relevant actors. This can still be the case, even if they all have the same immediate perception (Milgrom & Roberts 1982). Direct information transmission from one actor to another is especially susceptible to the hazard of deception if the actors' motivations differ or are not known for sure to be identical. In the 'economics of information' this is dealt with in terms of *signalling*, which concentrates on sending and receiving information messages of various kinds (see Hirschleifer & Riley 1979, 1406–1409). After these general points let us now turn more precisely to the concept of reputation.

What Is Reputation?

Reputation accounts for strong intertemporal linkages along a sequence of otherwise independent situations. Wilson (1985) provides a good starting point for a general account of the concept of reputation. The following points are his. Reputation is a characteristic or attribute ascribed to one actor by another (e.g. 'A has a reputation for courtesy'). Operationally, this is usually represented as a prediction about likely future behavior (e.g. 'A is likely to be courteous'). It is, however, primarily an empirical statement (e.g. 'A has been observed in the past to be courteous'). Its predictive power depends on the supposition that past behavior is indicative of future behavior. In a narrow sense, the actor's reputation is the history of his previously observed actions.

To be optimal, the actor's strategy must take into consideration the following chain of reasoning. First, his current reputation affects others'

predictions of his current behavior and thereby affects their current actions; so he must take account of his own current reputation to anticipate their current actions and therefore to determine his best response. Second, if he is likely to have choices to make in the future, then he must realize that whatever are the immediate consequences of his current decision, there will also be longer-term consequences due to the effects of his current decision on his future reputation, and others' anticipation that he will take these longer-term consequences into account affects their current actions as well (Wilson 1985, 28).

In a repeated strategic interaction situation with incomplete information the actions of early rounds can provide information about the actors' preferences. Thus the actions in early rounds may be overwhelmingly influenced not by immediate payoffs but by considerations of what information is being transmitted. The amount of uncertainty that is needed to obtain the reputation effect is quite small, as long as the actor who is developing the reputation has many opportunities to give evidence about his preferences (Kreps & Wilson 1980, 69).

Note that the notion of reputation in repeated games appear in every specific context where one actor can threaten others (Kreps & Wilson 1980, 70–71).

At least four ingredients are necessary to enable a role for reputations (Wilson 1985, 29):

(1) There must be several actors in the game.
(2) At least one player has some private information that persists over time.
(3) This actor with private information is likely to take several actions in sequence.
(4) This actor is unable to commit himself in advance to the sequence of actions he will take.

The key ingredient in all reputation considerations is that an actor can adopt actions that sustain the probability assessments made by other actors that yield favorable longterm consequences (Wilson 1985, 33). Differences in the information available to actors make their strategies acutely sensitive to their beliefs and expectations. This in turn affects the behavior not only of the uninformed actors, but also of the informed one, who realizes that his current actions affect others' later beliefs, their expectations about his subsequent behavior, and ultimately their choice of actions (Wilson 1985, 59).

What does then follow from taking reputation considerations seriously? Stigler in his classical treatise on 'The Economics of Information' (1961, 224) was the first to note that reputation of good quality has a huge effect on the decision making for customers, because it economizes on search.

Alt, Calvert & Humes (1986, 8–9) emphasize another aspect. If an actor is weak, then to build a reputation requires the *cultivation* of the opponent's uncertainty, rather than its alleviation. It may be advantageous for an actor to change some things just for the sake of change, in order to gain a temporary reputation-based advantage. Thus reputation, if established, economizes on coercion or enforcement.

We have thus far seen that reputation is always *relational*: a certain agent has a certain type of reputation with respect to some other agent on certain grounds and at a certain time. After these general points, let us now turn to a few specific games which nicely illustrate the effects of reputation.

Reputation in Cabinet Coalitions

As our first example, let us consider a process of cabinet coalition formation. Party A may form a cabinet alone or together with party B, which is politically close to it. By this A receives zero utility and B gets $b > 1$. If A doesn't accept B, then B may choose between low profile politics (R), and opposition politics (L). Strategy R gives B the utility 0, and A gets a, which is greater than zero but smaller than one. If B chooses to use strategy L, it gets -1 while A receives $a - 1$ utilities. Stated nontechnically, A prefers to rule alone, unless B then chooses opposition politics. If it does, it is better for A to form a coalition with B. On the other hand, B prefers a coalition above all other outcomes, and prefers to pursue a low profile rather than opposition if the coalition is not formed. This simplified situation is represented in Figure 1.

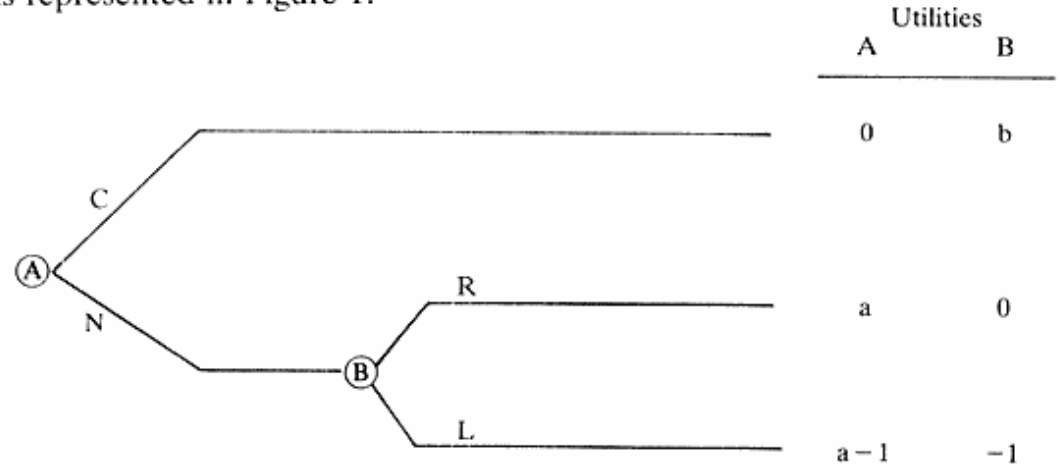

Fig. 1. The Game Tree with Probability $1 - p$.

Strategy C means that A forms the cabinet coalition with B, and N means that the coalition is not formed. There are two Nash equilibria in this game: (C, L) and (N, R). The Nash equilibrium is the most important solution concept in game theory. In a Nash equilibrium no player, regarding the

other actors as committed to their choices, can improve his lot. If A chooses N, then B may freely threaten to choose the opposition policy. If B threatens to choose the opposition policy, then it would be profitable for A to form the coalition with B (because zero is greater than a − 1, and if C is chosen by A, B has no opportunity to execute its threat policy). But B has no incentive to execute its threat even when the coalition is not formed. This is true, because if the coalition is not formed, B really has to choose between the two policy alternatives (low profile or opposition). In this case B chooses low profile politics, because this guarantees it greater utility than would come from the opposition policy. Anticipating this, A does not form the coalition with B. Although both (C, L) and (N, R) are Nash equilibria, only the latter is plausible. This is true, because the only thing which makes (C, L) an equilibrium is the empty threat by B to choose opposition politics if A does not form the cabinet coalition with it.

Stated technically (N, R) is the only *subgame perfect equilibrium* in this game. An (Nash) equilibrium is subgame perfect, if and only if it induces an equilibrium in every subgame. The only (proper) subgame here is the game starting from node 'B'. It was noticed earlier that B here has to choose R. Hence, (N, R) is the only subgame perfect equilibrium. It is still the only subgame perfect equilibrium, when this game is repeated finitely many times and the actors evaluate their performance by summing their utilities from the component games. This is easily seen by backward induction: in the last component game, the players choose (N, R), independently of the history of the game. In the next to last component game, players anticipate this, and they therefore choose (N, R) again, and so on.

This example is mathematically equivalent to Selten's (1978) 'Chain Store Paradox', but our interpretation is a political one. If the game is repeated, and A has even the slightest uncertainty about how B values its opposition policy, then A forms the coalition with B in almost all periods.

Suppose that A believes that instead of having the preferences depicted in Fig. 1, party B prefers the opposition policy in the case that the coalition is not formed. To be more specific, we assume that A believes with probability p, $0 < p < 1$, that B's preferences are as in Fig. 2, and with probability $1 - p$, that Fig. 1 is the correct model.

Kreps & Wilson (1982a) show that even for very low values of p, if the game is repeated sufficiently often, A always forms the coalition with B except possibly during a few last periods. We don't reproduce their argument here, because it is very similar to the analysis of the next example.

Hiring Political Trustees

We now turn to another political case, where reputation plays a central role. Assume that a board of in relevant respects identical individuals must

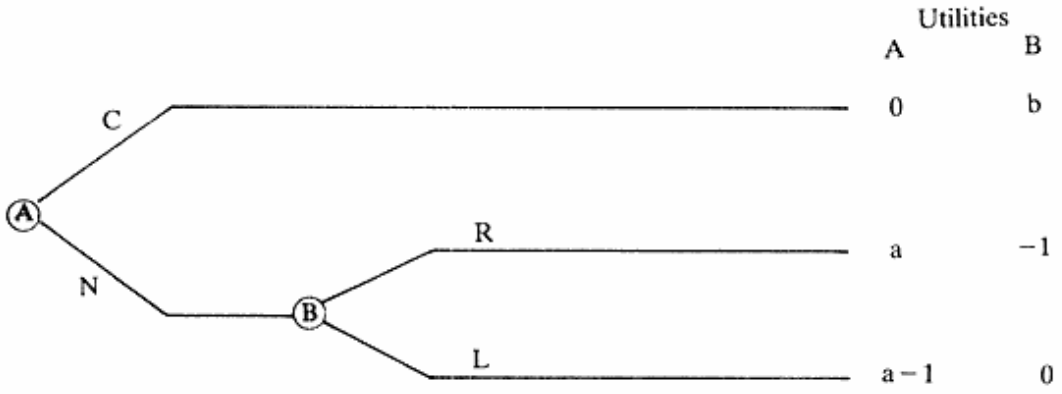

Fig. 2. The Game Tree with Probability p.

choose a person for some task for one period. The task in question could be for instance the chairmanship of a trade union club or the like. Let there be only two candidates, 1 and 2. Person 2 is the sure thing, which gives the choosers 0.5 utility with certainty. Person 1, the efficient one, may be of type R or of type L. He has two policies available: L and R. If 1 is chosen, and he employs the policy R, the choosers receive the utility 1, while the policy L gives them zero. If 1 is of type L, and he is chosen, he gets 2 from R and 3 from L. If he is not chosen, he gets zero from R and one unit of utility from L. These preferences exhibit the situation, where type L likes the policy L more than the policy R, and he gets a better salary, if he is chosen. The preferences of type R are the diametrical opposite of L's preferences. The situation is depicted in Fig. 3.

1 = L	The Board 1	2
Player 1 R	(2,1)	(0,1/2)
Player 1 L	(3,0)	(1,1/2)

1 = R	The Board 1	2
R	(3,1)	(1,1/2)
L	(2,0)	(0,1/2)

Fig. 3. The Game Matrices.

In the left hand matrix, person 1 is of the type L, and the only Nash equilibrium is (L, 2), i.e., person 2 is chosen by the board, and 1 employs

his most preferred policy L. In the right hand matrix, person 1 is of type R, and the only Nash equilibrium is (R, 1), where the board chooses 1, and he chooses his preferred policy. Suppose that the board believes with probability 1/3 that person 1 is of type R, and that he is L with the complementary probability. In the one shot game person 2 is chosen, because this maximizes the expected utility of the board. If this game is repeated, say five times, things change radically. In order to see this, we must introduce the notion of *sequential equilibrium*. By a sequential equilibrium is meant a pair of strategies and beliefs such that given the beliefs, the strategies are in a Nash equilibrium. And on the other hand the beliefs are consistent, that is, new beliefs do not contradict the old ones (see Kreps & Wilson 1982b). We will show that there exists a sequential equilibrium, where person 1 is chosen during the first four periods, and he plays the strategy R during the first three periods, *independent of his type.* During the fourth period, the type L uses a *mixed strategy* between L and R, and in the last period he chooses L. In the last period the board either mixes between 1 and 2, or chooses 2 with certainty, depending on which of the strategies L or R was observed in the fourth period. Thus, it is possible that the type L is chosen in all five periods. This result depends on the ability of the L type to exploit the uncertainty of the board and build a reputation as being of type R. We index the time backwards, so that the first period is 5 and the last 1. We begin by giving the description of the beliefs and strategies, and show that these form a sequential equilibrium. (We would like to stress that our analysis is almost equivalent to one of Kreps' and Wilson's, and that our primary interest here is to show by an example how the reputation effect functions.)

Beliefs. Player 1 knows which type he is, so the beliefs are needed only for the board. Because person 2 plays no active role in the game, his beliefs and strategies may be omitted. The beliefs in the period t are given by probability p(t), which is the probability by which the board believes person 1 to be of type R. Clearly $p(5) = 1/3$. If R is observed during the period t, then $p(t-1) = \max(p(t), 0.5^{t-1})$, if $p(t) > 0$. If L is observed during the period t, or if $p(t) = 0$, then $p(t-1) = 0$. According to these beliefs, it is sufficient evidence that person 1 is of type L, if policy L has been observed once in the history, even if 1 was not chosen during that period. (In this example, the board observes the policy chosen by 1 also when person 2 has been chosen.)

Strategies. If 1 is of type R, then he always chooses R. If he is of type L, then he chooses L during the last period. If $t > 1$, and $p(t) \geq 0.5^{t-1}$, then 1 chooses R. If $p(t) < 0.5^{t-1}$, then 1 chooses R with probability $p(R: 1 = L) = p(t)(1 - 0.5^{t-1})/(1 - p(t))(0.5^{t-1})$, and L with the complementary probability. The board chooses 1, if $p(t) > 0.5^t$. If the inequality is reversed,

the board chooses 2, and if it is equality, then 1 and 2 are chosen with equal probabilities.

Proposition. The beliefs and strategies given above constitute a sequential equilibrium.

Proof. The beliefs are consistent: If $p(t) \geqslant 0.5^{t-1}$, then 1 chooses R. If $p(t) = 0$, then 1 chooses L. In both cases the Bayes' rule implies $p(t-1) = p(t)$. If $0 < p(t) < 0.5^{t-1}$, then $p(t-1) = p(t)/(p(t) + (1 - p(t))p(R: 1 = L)) = 0.5^{t-1} = \max(p(t), 0.5^{t-1})$, that is, if R is observed in the period t, and 1 has chosen the strategy $p(R: 1 = L)$, then the belief that $p(t-1) = 0.5^{t-1}$ is consistent. Bayes' rule doesn't apply if a zero probability event happens (given the strategies). Only zero probability events in this case are the following: $p(t) = 0.5^{t-1}$ and 1 chooses L, and $p(t) = 0$, and 1 chooses R. In these cases new beliefs may be chosen freely. In particular, it is perfectly legitimate (and not totally ad hoc) to assume that in these cases $p(t-1) = 0$, that is, all deviations from the equilibrium strategies imply that 1 is L. If the board deviates from the equilibrium, this doesn't change its beliefs, of course.

The strategies are in equilibrium, given the beliefs: In the period 2 happens for the first time that $p(t) < 0.5^{t-1}$, namely, $1/3 < 0.5$. Then 1 chooses R with probability $p(R: 1 = L) = 0.5$. It is still true that $1/3 > 0.5^2 = 0.25$, which implies that the board chooses 1 in this period. Because 1 chooses the mixed strategy above, it is possible that his choice at this stage is R. We saw above that this generates $p(1) = 0.5$. Given this belief, the board is indifferent between 1 and 2, in particular; choosing 1 and 2 with equal probabilities in this (last) period is among its best replies, given that the type L chooses L and the type R chooses R with certainty. Clearly, these strategies are optimal for these types in the last period. Now it is straightforward to verify that it is indeed optimal for type L to choose a mixed strategy in period 2, if the board chooses him, and vice versa. Further, it is easy to check that it is profitable for all players to choose their above described strategies during the first three periods. One may wonder if it is not profitable for the board at some stage to deviate from the equilibrium and choose 2, in the hope that this would give the type L incentives to reveal his true type in the next period. This is not the case. In the next to last period it does not pay for the board to choose 2, because L chooses L in the last period anyway. In the third period it is not profitable for the board to choose 2, because it is still profitable for L to continue to play the equilibrium, if it assumes that the board follows the equilibrium in the rest of the game, and so on. So, the players are wise when following the equilibrium under all circumstances, also if something unexpected happens.

The sequential equilibrium above is not unique, as we have chosen the beliefs off the equilibrium path somewhat arbitrarily: we assume that the

type R can never deviate from his strategy R. On the other hand, R seems to be the only rational choice for the type R, and therefore this assumption is not totally *ad hoc*. Of course, different equilibria would arise if the utility numbers in the matrices in Fig. 3 were different, or if the board had a prior difference from 1/3. Notice that in this example, the board is very happy indeed by *not* knowing with certainty which is the true type of player 1, because nicely behaving player 1 is much more valuable for the board than the boring middle of the road man, 2. If it is known with certainty that player 1 is L, then the board and L have no chances to 'cooperate', a very sad situation for both players. It is also interesting to note that type L chooses his dominated strategy (in the one shot game) R during the first three periods in equilibrium. Hence, *maintaining reputation is costly, but profitable.*

The above model also gives one possible explanation as to why making a job sufficiently attractive (paying high salaries etc.) may keep (potentially) revolutionary but effective types silent for a very long time, although their ideology remains unchanged. If important posts are manned by middle of the road types, it may indicate that the board has not behaved completely rationally, or, that it is too costly to make the jobs attractive enough for effective types. However, if the time horizon is of sufficient duration, then it may be possible to make the post attractive at a quite moderate cost. Every man has his price, and the board should be happy to pay it.

Bargaining over Cabinet Coalitions

Now we turn to *bargaining* over cabinet coalitions. Let us assume that Party A may form a cabinet coalition with party B, if they agree how to divide the seats. If no such agreement is reached, A receives zero utility and B x units of utility. There are only two seats 1 and 2, which are valued similarly by both parties: 1 gives both u(1) and 2 u(2) utilities, $0 < u(1) < u(2)$. A situation where one of the parties gets both seats is equally valuable (or invaluable) as the conflict outcome for both parties. We assume that there are only two bargaining rounds, and party A makes one offer in both rounds, which may be accepted or rejected by B. If A's offer is accepted in the first period, the game is over, and the distribution of the seats is as proposed by A. If B rejects A's offer, then in the next and last period A makes a new proposal, which may, again, be accepted or rejected by B. If accepted, the players receive their conflict payoffs and the situation is over. Both players discount their second period payoffs by a factor d, $0 < d < 1$; hence both players prefer early agreements.

We assume that x, the conflict payoff for B, can only take values u(1) or u(2). If A knows with certainty that $x = u(1)$, then there exists a unique

subgame perfect equilibrium, where A proposes in both periods that A gets the seat 2 and B the seat 1, and B accepts these offers in both periods. Hence the game ends in the first period, and the distribution of seats is as proposed by A. This can be seen by the backwards induction argument: in the last period B is indifferent between accepting and rejecting A's proposal, since both alternatives give him u(1). In the first period, B accepts A's proposal, because $u(1) > du(1)$. Similarly, if $x = u(2)$, there is a unique subgame perfect equilibrium, and the game ends in the first period, A gets the seat 1 and B the seat 2. It is clear that A is better off if he has the opportunity to play against the type of B who has values $x = u(1)$.

Suppose now that A doesn't know what is the true type of B, and assumes that it is equally probable that $x = u(1)$ and $x = u(2)$. In this case, A is in a worse bargaining position, because B may successfully imitate the behavior of type $x = u(2)$, even if it is in fact of the weaker type. On the other hand it is possible that an agreement is never reached. We analyze here only the case $2u(1) > u(2)$. In this case there is no possibility for disagreement. The case $2u(1) < u(2)$ is more complex (see Fudenberg & Tirole 1983, where this kind of model was first analyzed. In their model the player who makes (price) offers has a continuum of pure strategies).

If $x = u(2)$, then B accepts only the seat 2. Because A cannot make any offer which is better for B than this, it is always accepted. It follows that if the first period offer is rejected, then A must necessarily believe that the probability of $x = u(2)$ is at least 0.5. Offering the seat 2 in the second period A receives u(1), and offering the seat 1 A receives *at most* 0.5u(2). Because $u(1) > 0.5u(2)$, we conclude that A always offers the seat 2 in the second period, which is always accepted. B knows this: he receives at least du(2) utility from the game. If $du(2) < u(1)$, then, if $x = u(1)$, B is better off by accepting the seat 1 immediately rather than waiting for the seat 2 in the second period. When d is sufficiently close to 1, it also happens that A offers the seat in the first period, which is accepted by B, if $x = u(1)$ and rejected otherwise. However, if $du(2) > u(1)$, then B prefers waiting, and indeed A in this case offers the seat 2 immediately, which is now accepted by both. The argument is in fact slightly more complicated, because we should take into account the possibility that A and B (if $x = u(1)$) use mixed strategies, but it still results in the same conclusion. Depending on the size of the discount factor, A may be fooled to offer the seat 2 already in the first period to B, which is very profitable for B, if $x = u(1)$.

Conclusion

We have now discussed some aspects of the reputation effects in politics. We have demonstrated by three concrete game theoretical

examples how the reputation effect works. Thus, it has been shown in a strict sense that reputation really plays a central role in political decision-making. These results could in principle easily be generalized.

We hope that our discussion has provided some fruitful insights. It is also hoped that our account has supported the view that game theory has much to give to concrete social research. With game theory, it is possible to substantiate ideas that without it would only be intuitive insights.

There is a huge research area to be covered in the years to come. Reputation is of central importance in almost all political contexts, and game theory may be the most fruitful approach to deal with these sometimes quite complex situations.

REFERENCES

Alt, J. E., Calvert, R. L. & Humes, B. D. 1986. 'Game Theory and Hegemonic Stability: The Role of Reputation and Uncertainty.' Paper prepared for delivery at the Midwest Political Science Association Meetings, April 10–12, 1986, Chicago.

Arrow, K. 1984. *The Economics of Information*, Volume 4 of Collected Papers of Kenneth Arrow Basil Blackwell: Oxford.

Aumann, R. J. 1974. *Values of Non-Atomic Games*. Princeton University Press: Princeton.

Aumann, R. J. 1976. 'Agreeing to Disagree', *The Annals of Statistics* 4, 1236–1239.

Brams, S. 1977. Deception in 2 × 2 Games, *Journal of Peace Science* 2, 171–203.

Calvert, R. L. 1986. 'Reputation and Legislative Leadership.' Paper prepared for delivery at the Carnegie Conference on Political Economy, Carnegie-Mellon University, May 2–3, 1986.

Crawford, V. P., Sobel, J. 1982. 'Strategic Information Transmission', *Econometrica* 50, 1431–1541.

Fudenberg, D. & Tirole, J. 1983. 'Sequential Bargaining with Incomplete Information', *Review of Economic Studies* 50, 221–247.

Harsanyi, J. C. 1967–1968. 'Games with Incomplete Information played by "Bayesian" players I–III', *Management Science* 14, 159–182, 320–334, 486–502.

Heal, J. 1978. 'Common Knowledge', *Philosophical Quarterly* 28, 116–131.

Hilton, R. W. 1981. 'The Determinants of Information Value: Synthesizing some General Results', *Management Science* 27, 57–64.

Hirschleifer, J. & Riley, J. G. 1979. 'The Analytics of Uncertainty and Information – An Expository Survey', *Journal of Economic Literature* 17, 1375–1421.

Kreps, D. M. & Wilson, R. 1980. 'On The Chain-Store Paradox and Predation: Reputation for Toughness', *Technical Report No 317*, The Economics Series Institute for Mathematical Studies in the Social Sciences, Stanford University.

Kreps, D. M. & Wilson, R. 1982a. 'Reputation and Imperfect Information,' *Journal of Economic Theory* 27, 253–279.

Kreps, D. M. & Wilson, R. 1982b. 'Sequential Equilibria', *Econometrica* 50, 863–894.

Lewis, D. 1969. *Convention*. Harvard University Press: Harvard.

Milgrom, P. 1981. 'An Axiomatic Characterization of Common Knowledge', *Econometrica* 49, 219–222.

Milgrom, P. & Roberts, J. D. 1982. 'Predation, Reputation, and Entry Deterrence', *Journal of Economic Theory* 27, 280–312.

Ordeshook, P. C. & Palfrey, T. 1986. 'Agendas, Strategic Voting, and Signalling with Incomplete Information', unpublished manuscript.

Ratchford, B. T., 'Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior', *Management Science* 28, 197–212.

Schlenker, B. R. & Bonoma, T. V. 1978. 'Fun and Games: The Validity of Games for the Study of Conflict', *Journal of Conflict Resolution* 22, 7–38.

examples how the reputation effect works. Thus, it has been shown in a strict sense that reputation really plays a central role in political decision-making. These results could in principle easily be generalized.

We hope that our discussion has provided some fruitful insights. It is also hoped that our account has supported the view that game theory has much to give to concrete social research. With game theory, it is possible to substantiate ideas that without it would only be intuitive insights.

There is a huge research area to be covered in the years to come. Reputation is of central importance in almost all political contexts, and game theory may be the most fruitful approach to deal with these sometimes quite complex situations.

REFERENCES

Alt, J. E., Calvert, R. L. & Humes, B. D. 1986. 'Game Theory and Hegemonic Stability: The Role of Reputation and Uncertainty.' Paper prepared for delivery at the Midwest Political Science Association Meetings, April 10–12, 1986, Chicago.

Arrow, K. 1984. *The Economics of Information*, Volume 4 of Collected Papers of Kenneth Arrow Basil Blackwell: Oxford.

Aumann, R. J. 1974. *Values of Non-Atomic Games*. Princeton University Press: Princeton.

Aumann, R. J. 1976. 'Agreeing to Disagree', *The Annals of Statistics* 4, 1236–1239.

Brams, S. 1977. Deception in 2 × 2 Games, *Journal of Peace Science* 2, 171–203.

Calvert, R. L. 1986. 'Reputation and Legislative Leadership.' Paper prepared for delivery at the Carnegie Conference on Political Economy, Carnegie-Mellon University, May 2–3, 1986.

Crawford, V. P., Sobel, J. 1982. 'Strategic Information Transmission', *Econometrica* 50, 1431–1541.

Fudenberg, D. & Tirole, J. 1983. 'Sequential Bargaining with Incomplete Information', *Review of Economic Studies* 50, 221–247.

Harsanyi, J. C. 1967–1968. 'Games with Incomplete Information played by "Bayesian" players I–III', *Management Science* 14, 159–182, 320–334, 486–502.

Heal, J. 1978. 'Common Knowledge', *Philosophical Quarterly* 28, 116–131.

Hilton, R. W. 1981. 'The Determinants of Information Value: Synthesizing some General Results', *Management Science* 27, 57–64.

Hirschleifer, J. & Riley, J. G. 1979. 'The Analytics of Uncertainty and Information – An Expository Survey', *Journal of Economic Literature* 17, 1375–1421.

Kreps, D. M. & Wilson, R. 1980. 'On The Chain-Store Paradox and Predation: Reputation for Toughness', *Technical Report No 317*, The Economics Series Institute for Mathematical Studies in the Social Sciences, Stanford University.

Kreps, D. M. & Wilson, R. 1982a. 'Reputation and Imperfect Information,' *Journal of Economic Theory* 27, 253–279.

Kreps, D. M. & Wilson, R. 1982b. 'Sequential Equilibria', *Econometrica* 50, 863–894.

Lewis, D. 1969. *Convention*. Harvard University Press: Harvard.

Milgrom, P. 1981. 'An Axiomatic Characterization of Common Knowledge', *Econometrica* 49, 219–222.

Milgrom, P. & Roberts, J. D. 1982. 'Predation, Reputation, and Entry Deterrence', *Journal of Economic Theory* 27, 280–312.

Ordeshook, P. C. & Palfrey, T. 1986. 'Agendas, Strategic Voting, and Signalling with Incomplete Information', unpublished manuscript.

Ratchford, B. T., 'Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior', *Management Science* 28, 197–212.

Schlenker, B. R. & Bonoma, T. V. 1978. 'Fun and Games: The Validity of Games for the Study of Conflict', *Journal of Conflict Resolution* 22, 7–38.

Selten, R. 1978. 'The Chain-Store Paradox', *Theory and Decision* 9, 127–159.
Shubik, M. 1983. *Game Theory in the Social Sciences: Concepts and Solutions*. The MIT Press: Cambridge, Mass.
Snidal, D. 1985. 'The Game *Theory* of International Politics', *World Politics* 38, 25–55.
Sobel, J. 1985. 'A Theory of Credibility', *Review of Economic Studies* 52, 557–573.
Stigler, G. J. 1961. 'The Economics of Information', *The Journal of Political Economy* 69, 213–225.
Wilson, R. 1985. 'Reputation in Games and Markets', pp. 27–62 in Roth, Alvin, ed., *Game-Theoretic Models of Bargaining*. Cambridge University Press: Cambridge.