Games Real Actors Could Play: The Problem of Complete Information

Fritz W. Scharpf

89/9

Max-Planck-Institut für Gesellschaftsforschung
Lothringer Str. 78
D-5000 Köln 1
Federal Republic of Germany
Telephone 0221 / 336050
Fax 0221 / 3360555

MPIFG Discussion Paper 89/9
ISSN 0933-5668
August 1989

Abstract

Game theory is a powerful tool for the disciplined analysis of interacting choices. Nevertheless, its use in empirical research is considered questionable since the standard assumptions of mathematical game theory seem to place exceedingly high information costs on real-life actors as well as on researchers. The paper tries to show that these misgivings are largely unjustified. If players were in fact ignorant about each other's strategies and payoffs, they could either resort to generalized caution, or they could endogenously create preconditions for trustworthy communication in an iterated "truth game" that is embedded in ongoing interactions. Furthermore, actors and researchers alike are able to use standardized expectations derived from institutional rules and social norms for their orientation. As a consequence, the information costs of interactions are sufficiently reduced for players as well as for researchers to make game-theoretical explanations a feasible and promising proposition.

* * * * *

Trotz ihrer hohen Leistungsfähigkeit bei der Analyse interdependenter Entscheidungen gilt die Spieltheorie weithin als empirisch unanwendbar, weil ihre informationellen Anforderungen anscheinend weder von realen "Spielern" noch von der empirischen Forschung erfüllt werden können. Der Aufsatz versucht diese Bedenken zu entkräften. Bei Ungewißheit über die Optionen und Präferenzen der Partner könnten Spieler sich entweder auf risikominimierende Strategien beschränken, oder sie könnten die Voraussetzungen für glaubhafte Kommunikationen in einem (iterierten und vernetzten) "Wahrheitsspiel" endogenisieren. Darüber hinaus können sowohl die Spieler als auch die empirische Forschung sich an standardisierten Erwartungen orientieren, die durch Institutionen und soziale Normen begründet werden. Wenn diese Möglichkeiten vorausgesetzt und genutzt werden, vermindern sich die Informationskosten spieltheoretischer Analysen so weit, daß empirische Anwendungen möglich und aussichtsreich werden.

Contents*

---

* The present paper (which is to be followed by a second one on "The Problem of Connectedness") draws on an earlier draft ("Core Games, Connected Games and Networks of Interaction") that was written jointly with Andreas Ryll. It was presented at the workshop on "Connected Games: Theory, Methodology, and Applications" at the Max-Planck-Institut für Gesellschaftsforschung, Cologne. November 17-18, 1988. The present version owes much to discussions at the workshop and to helpful comments by Jens Alber, Jürgen Häusler, Hartmut Esser, Renate Mayntz, Andreas Ryll, David Soskice, Uwe Schimank and Volker Schneider.

I. INTRODUCTION

In the social sciences as much as in the natural sciences, our aspiration is to explain much with little. Yet all too often we are forced to choose between parsimonious theoretical models that do not seem to explain much at all, and more complete explanations that are barely less information-rich than historiographic accounts - and equally useless for prediction (Peterson 1970). In this predicament, many social scientists have turned to the rational-choice paradigm in the hope that it might do for sociology and political science what it apparently has done for economics - to generate an axiomatic-deductive theoretical system whose propositions are capable of producing parsimonious and nonobvious explanations and predictions of real-world phenomena (Lindenberg 1983; 1989).

1. The Challenge of Rational Choice

What we have not always realized, however, is the vast distance between the generic form of the rational-choice paradigm (Elster 1986) and its use in standard economics. In order to explain behavioral regularities as rational choices, it is not enough to know the decision environment and its reaction functions, but we must also know the subjective perceptions and preferences of actors. And as the generic form of the paradigm has nothing to say about the formation of subjectivities, these must be treated as exogenous data in each individual instance - with the conse-

quence that rational-choice explanations must have truly discouraging information requirements.

By contrast, the ability of economics to employ its impressive apparatus of applied mathematics for the production of nonobvious conclusions rests entirely on the reduction of these information requirements through a battery of simplifying and standardizing "auxiliary factual assumptions" (Simon 1986: S212; see also Simon 1978). Foremost among these are the postulates that actors have unbounded rationality, that their perceptions are true representations of objective reality, that they will maximize an integrated utility function, and that this utility increases monotonically (but at a decreasing marginal rate) with the quantities of material goods, services and money obtained (Lindenberg 1989: 181). From an empirical perspective, these are heroic assumptions. And yet, as Ronald Heiner (1983) has argued, even they could not assure the predictive usefulness of micro-economic theory. If omniscient rational optimizers should in fact respond without friction to all environmental perturbations, ex-post explanations would still require a complete account of the total decision environment, and prediction would be out of the question. Thus, Heiner concludes, any claim for the predictive power of economic theory must also presuppose behavioral regularities reflecting bounded, rather than perfect, rationality and rule-based behavior.

Yet if that is so in the economic heartland of the rational-choice paradigm, why should the rush to colonize the territories of adjacent social-science disci-

plines continue? And why should social scientists who are fully aware of all Simonian objections nevertheless defend the rational-choice paradigm on the grounds that "one can't beat something with nothing?"[1] One reason must surely be the macho appeal of being able to work with complex mathematical models. Another seems to be the lack of generality of alternative explanations. Even if it is granted that complete rationality is impossible, one cannot assert that any specific instance of bounded rationality could never be overcome by additional search efforts and learning.[2] Even more important, from my own point of view, is the experience that, in empirical research, we will often be able to obtain a reasonably good understanding of the options, world views and preferences of particular actors, or even of classes of actors in given types of

---

[1] "Why has satisficing theory not replaced neoclassical economics? The answer lies in a disturbing element of ad-hoc-ness in the notion of satisficing. The theory does not offer an answer to the crucial question of why people have the aspiration or satisfaction levels they have... These levels must simply be taken as given, which means that the theory offers little more than 'thick description'. Neoclassical theory will be dethroned if and when satisficing theory and psychology join forces to produce a simple and robust explanation of aspiration levels, or sociological theory comes up with a simple and robust theory of the relation between social norms and instrumental rationality. Until this happens, the continued dominance of neoclassical theory is ensured by the fact that one can't beat something with nothing." (Elster 1986: 26-27). For similar responses to the challenge of bounded rationality, see Przeworski (1986) and Tsebelis (1988a: Chap. 2).

[2] By the same token, bounded-rationality explanations will often turn into self-defeating prophecies when they are communicated to the actors (Binmore 1987: 33).

situations. When that is so, explanations based on the demonstration that a particular course of action was the best one available to actors under given circumstances have a degree of persuasiveness and conclusiveness unmatched by alternative theoretical models.

While institutionalist or functionalist explanations can always be challenged by asking why actors should have obeyed the assumed imperatives, it may be possible to question the factual premises, but not the conclusions, of rational-choice explanations. In their explanatory logic at least, they seem to fully satisfy our desire for Weberian Verstehen. And once we have reached that point, it is indeed tempting to restate this understanding in an analytical model that not only will be available as a heuristic for the explanation of choices in similar situations, but that will also facilitate the discovery of nonobvious implications of the present interaction.

But in order to be useful for empirical research, such models must in fact allow for valid representations of the understanding that we have gained. Thus, economic theory with its reliance on anonymous price mechanisms for the explanation and aggregation of individual choices may provide useful models for many types of large-number phenomena (e.g. migrations) even outside of the economy, while game theory seems to be more promising for the analysis of interactions among small numbers of individual or corporate actors who are aware of the interdependence of their choices. In my own work, which is mainly concerned with the influence of institutional arrangements on public policy, I have

used such models for explaining failures of macroeconomic policy coordination between national governments, central banks and labor unions (Scharpf 1987), and the emergence of inefficient constitutional arrangements in German federalism and the European Community (Scharpf 1988; 1989a). In the present article, I will explore in a more abstract fashion the conditions under which game-theoretical analyses could be useful in institution-oriented empirical research.

2. The Elusive Promise of Game Theory

Like all other types of formal modelling, the game-theoretical representation of interactions has the obvious advantage that

> "the critical role of assumptions is laid bare. In the ideal, all such assumptions must be made explicit and then carefully examined from the viewpoint of their realism in terms of "real-world" phenomena the model is intended to represent" (Blalock 1989: 450).

In my own experience, the importance of this advantage cannot be stressed enough, and game theory in particular is invaluable for the discipline it imposes on speculative hypothesizing about actors' interests, their available options, and the likely consequences of alternative strategies.

More specifically, what game theory adds to the usual agenda of institutional research is, first, a more systematic focus on the interaction among multiple actors with interdependent choices (rather than on the strategies of an idealized unitary "policy maker"). It is thus able to deal with the fact that many outcomes

are not under the full control of a single actor, and that the actual consequences of human action are often not desired by their authors. Rather than treating all of these as failures of rationality, a game-theoretical approach suggests that many of them could be explained as the anticipated (and hence intended even if undesired) outcomes of choices under conditions of mutual or multiple interdependence.

The second important contribution to institutional research is the concept of an equilibrium solution. By contrast to equilibrium analysis in economics, however, there are no implications of social optimality here. Game-theoretical equilibria are exclusively defined by the criterion that no player should be able to improve her present position by unilaterally changing her present strategy. In fact, one of the important uses of game analysis is precisely the identification of equilibria that are "social traps" with suboptimal outcomes for all participants.

For empirical research, it is also important that there are games that have either no equilibrium solution at all, or that have more (sometimes many more) than one equilibrium. Thus game-theoretical analyses may lead to three kinds of empirically relevant conclusions: In games with a single equilibrium, it will be possible to explain, as well as to predict, the outcome that rational actors are likely to converge upon. In games with multiple equilibria, the set of possible outcomes can be identified, while the choice of a specific solution will be path-dependent (permitting only historical explanations). Nevertheless, game

analysis will be able to explain the stability of the outcome that was in fact chosen (David 1985). If there is no equilibrium solution, finally, game theory will only be able to identify the reasons for the indeterminacy and instability of outcomes. But even that is, of course, a most useful contribution to our understanding of real-world situations.

Less obvious, but even more important, is the ease with which available social science knowledge can be integrated into game-theoretical models. The opening is provided by the concept of "rules of the game" which is used to construct the payoff matrices that can then be subjected to game-theoretical analysis. In mathematical game theory, of course, these rules are stipulated rather than empirically determined - which explains the sense of unrealism and naivety with which many empirically oriented social scientists respond to game-theoretical exercises. But once the need for their empirical determination is acknowledged (as it must be in empirical applications), the perspective is reversed. Unlike neoclassical economics, which has developed its own "auxiliary factual assumptions" that are for practical purposes impervious to empirical information generated by psychological, sociological or even economic research, game-theoretical applications are completely dependent on exogenous information. Thus, the empirically empty but formally precise concept of "rules of the game" becomes a perfect template for translating available social-science knowledge about real-world actors, institutions and power relationships into parameters of a game-theoretical model that facilitates analytical solutions.

These are attractive promises for theory-oriented empirical research. Yet before we get carried away, we must remind ourselves that game theory, even more so than standard economics, has developed into a branch of applied mathematics that works with analytical procedures whose empirical referents are of no concern to the theoretician. In their quest for identifying (unique) equilibrium solutions for all types of game constellations, game theorists have in fact developed high-powered solution algorithms that are hardly accessible to actors in real-world interactions. But since all such solutions must ultimately be imputed to the rational calculations of the players themselves,[3] this creates a problem for empirical applications that does not arise in quite the same way in economics.

---

[3] That is not true of evolutionary game theory as it has been applied in biology and was then retransferred into the social sciences. It dispenses with the assumption of actor rationality by treating "strategies" as genetically fixed behavioral traits of biological species or as otherwise "hardwired" (Maynard Smith 1982; Axelrod 1984). If it is assumed that a rigorous selection mechanism will permit only the survival of traits with superior "fitness", the outcome, at the level of populations, will be identical with what could have been predicted on the basis of rational-choice assumptions.

But the success of evolutionary game theory in biology depends on hardwired strategies - which are difficult to square with assumptions of purposeful human action (however boundedly rational) and of learning, even if routines and "standard operating procedures" are important in many areas of (mainly organizational) decision making (Hannan/ Freeman 1977; 1984; Nelson/ Winter 1982). It also depends on the presence of a rigorous selection mechanism at the level of populations. Unless precise social analogues to both conditions can be identified, the specific explanatory potential of evolutionary game theory cannot be legitimately invoked in empirical social-science research.

There, individuals are assumed to play a one-person game against nature where the opponent has no objectives and no known strategies (Latsis 1972: 210-11), while the powerful analytical algorithms are of concern only to the analyst who is trying to model the "invisible hand" that aggregates individual choices (Ullmann-Margalit 1978). In empirical applications of game theory, by contrast, the direct correspondence between the mental operations that are used by the analyst, and those that are imputed to the actors, forces us to work with only a subset of the analytical procedures developed by mathematical game theorists. We must, in other words, disable all elements of the analytical machinery that could not plausibly be used by real-life players if we intend to create valid explanations of empirical interactions.[4]

---

[4] Foremost among these, in my view, are solution concepts relying on "mixed strategies" in single-shot games. Analytically, the use of mixed strategies is attractive because it assures that all zero-sum and finite non-zero-sum games will have at least one equilibrium solution. But the notion that players should resort to a randomized mixture of their pure strategies is highly counterintuitive, or even meaningless, in non-iterated games (Neisser 1952), and it is in fact not used by subjects, even in long iterations of experimental games where its application would have been profitable (Colman 1982: 77).

Of course, by ignoring mixed strategies in analyses of empirical interactions, we also increase the number of situations for which game theory will be unable to identify a unique, or any, equilibrium solution. By the same token, empirical work will perhaps not profit much from recent advances toward a "general theory of equilibrium selection" that uses extremely sophisticated solution algorithms to identify unique equilibria in constellations where hitherto only multiple equilibrium points could be identified (Harsanyi/Selten 1988). But what we lose in analytical power, we probably gain in greater realism.

Yet even if that necessity is accepted, it is unclear under what conditions actors could have the information that they are supposed to have in game-theoretical analyses, and whether the information costs for researchers could be sufficiently reduced to make empirical applications practicable. These are difficult problems. Yet contrary to a widespread prejudice among empirically oriented social scientists as well as among mathematical game theorists, it is by no means a foregone conclusion that adequate solutions cannot be found, if not generally, then at least for certain kinds of situations. And given the potential explanatory power of game models, it seems highly worthwhile to explore the conditions under which one might have good reasons to use them in social-science research.

Of the particularly vexing problems that arise in empirical applications of rational-choice (as distinguished from evolutionary)[5] game theory, the present

---

[5] See above, footnote 4. What seems more promising than evolutionary models are quasi-evolutionary or "game-learning theories" (Selten 1985: 83; Witt 1986) that relax, but do not eliminate, the assumptions of actor rationality. One might assume, for instance, that even if strategies are selected by trial and error, actors will be able to distinguish between more and less satisfactory outcomes. Thus, actors could reach at least local optima (Hernes 1976), and one could explain "lock-ins" (David 1985) or "joint-decision traps" (Scharpf 1988; 1989a) as game-theoretical equilibria even if actors did not, or could not, calculate their optimal responses in advance. But such explanations presuppose ongoing interactions in environments whose rate of change is slow compared to the actors' rate of learning. By contrast, the present paper will explore the realism of game-theoretical analysis based on more stringent rationality requirements.

paper will examine conditions under which players might have sufficient information to anticipate each other's choices of strategy. In a subsequent paper, I intend to discuss the connectedness among games and the conditions under which players might be able to cope with the potentially overwhelming complexity of multi-actor networks of interaction.

II. THE PROBLEM OF COMPLETE INFORMATION

At the most basic level, game-theoretical solutions are derived from two types of postulates which John Harsanyi (1977: 116-118) has labeled "payoff dominance" and "rational expectations." While the first set specifies that players will prefer outcomes maximizing their own payoffs, postulates of the second type explicate the notion that players should choose their moves in anticipation of the moves of other players who are also rationally maximizing their own payoffs. It is this second set of postulates that constitutes the specific analytical power of game theory, as distinguished from single-actor decision theory. It implies "empathy" - i.e. the ability to correctly analyze the interaction situation from the other players' point of view as well as from one's own.

1. The implications of information deficits

In classical game theory, the ability to analyze the interaction situation from other players' points of view is assured by the postulate of complete information, which implies that the strategy options and payoffs of every player should be "common knowledge" among all players. In empirical applications, that means that actors would need to know each other's action resources and constraints in order to arrive at a mutual understanding of their respective sets of strategies. They would also need to estimate the causal consequences of all combinations of strategies in order to determine the physical outcomes obtainable. And they would need to know each other's perceptions and subjective valuations of these outcomes in order to construct the "payoff matrices" which could finally be subjected to game-theoretical analysis. For real-life actors, these are demanding requirements indeed, whose lack of realism is - perhaps too readily - granted by leading game theorists:

> "Classical game theory cannot handle games with incomplete information at all... This obviously poses a very serious limitation since virtually all real-life game situations involve incomplete information. In particular, it very rarely happens that the participants of any real-life social situation have full information about each other's payoff functions. Uncertainty about the strategies available to the other players is also quite common." (Harsanyi/ Selten 1988: 10)

In other words, "classical game theory" would be inapplicable in practically all "real-life social situations." In order to overcome these limitations, John

Harsanyi (1967/68) has proposed a "Bayesian" solution concept that replaces uncertainty with players' subjective probability estimates of all elements of the game about which they are ignorant. In effect, this solution concept will transform games with incomplete information into games with "imperfect information"[6] that can be analytically handled by classical game theory (Selten 1982). But while the analytical validity of this ingenious solution is generally accepted, its applicability to "real-life social situations" remains very much in doubt. In fact, from an empirical point of view, the cure may be worse than the disease. It replaces unrealistic assumptions about actors' access to information with equally unrealistic demands on their computational capabilities (Eichhorn 1982; Colman 1982: 23-24; Friedman 1986: 116), and on their willingness and ability to use probabilistic solution concepts. Thus it seems fair to conclude that

> "Bayesian rationality has its place in normative theory but a more realistic approach should be developed for descriptive purposes" (Selten 1985: 82).

It seems unlikely, however, that a "more realistic approach" could arise entirely, or even primarily, from innovations in mathematical game theory. We would gain more, I think, from an examination of the type of real-life situations to which (a subset of) the avail-

---

[6] While "completeness" refers to information about payoffs and strategies, "perfectness" characterizes information available at any point in a game about prior moves of players (including chance moves attributed to "nature"). In fact, Harsanyi's solution postulates that unknown features of the game should be interpreted as having been fixed by a chance move.

able tools of game-theoretical analysis could plausibly be applied. In this spirit, I suggest that exploration could be assisted by a thought experiment in which it is assumed that actors have no knowledge of each others' strategy options and payoff functions[7] - but they do know that they are in an interdependent decision situation, and they know their own strategies and payoffs.

If Bayesian solution concepts are ruled out, and if no additional sources of information are available, players would then need to resort to pragmatic rules for dealing with uncertainty which are likely to resemble the prudential "rules of evidence" that have always permitted legal systems to reach firm conclusions in the face of conflicting assertions.[8] Under the assumed conditions, rational actors would do well to concentrate their intelligence efforts on obtaining objective indicators for the action resources available to their opposite numbers. And if inferences derived from these indicators are associated with margins of error, they would do well to err on the side of greater cau-

---

[7] Harsanyi (1967/68: 167) has shown that incomplete information about strategy spaces can be analytically reduced to incomplete information about payoff functions. But as the practical difficulties of obtaining valid information are quite different in the two cases, they need to be distinguished in empirically oriented discussion.

[8] Courts typically use a combination of prima-facie proof, controvertible and incontrovertible presumptions, and rules for allocating the burden of proof (i.e. the burden of non-persuasion), in order to resolve disputed questions of fact.

tion by applying appropriately skeptical rules of evidence.

But while it will be generally possible to obtain pragmatically sufficient (conservative) estimates of strategic capabilities, expectations about other players' likely choices among these strategies are another matter. They imply knowledge of payoff functions which depend not only on objective states of the world, but also on their interpretation and evaluation by other players. Unlike capabilities, these subjective decision premises cannot be estimated by a combination of objective indicators and general cause-and-effect rules. Definitions of the situation are of crucial importance, and they may change. There is a world of difference between friendship and mortal enmity, yet one can turn into the other without prior changes in observable circumstances (Flam 1989). And when such redefinitions of the situation occur, innocuous resources may turn into deadly weapons - and deadly weapons may again change their character from an essential means of self-protection to an unbearable burden on the economy. And even then, one may fall victim to strategic dissimulation and deception.

However, if rational players have less reason to trust their estimates of other players' payoff functions, they have even more reason to resort to prudential rules in drawing their conclusions. At least in situations where the other side is capable of inflicting serious damage, they would do well to discount uncertain estimates of others' payoffs, and to focus instead on the consequences which others' potential

strategies would have for their own position. This is, of course, the "capabilities-not-intentions" logic, based on worst-case scenarios, that is considered "rational generalship" (Colman 1982: 51) and that was in fact characteristic of strategic thinking in the context of the cold-war arms race (Snyder 1971: 77; Jervis 1978).

More generally, when players are ignorant about each other's payoffs, they cannot apply the central game-theoretical "rationality principle", which would have directed them to anticipate the optimal choices of other players in determining their own best response (Hamburger 1979: 44). Of Harsanyi's (1977: 116-118) two sets of rationality postulates, "payoff dominance" and "rational expectations", they are thus able to use only the former in order to select a strategy that is entirely derived from an analysis of their own payoffs. In other words, their interaction turns into a game against nature - except for the fact that they know that human opponents, unlike nature, are capable of guile and malevolence.

In practice that means that players who are ignorant of each others' payoffs should resort to cautious solution concepts. Thus, they should first eliminate strategies from their own repertoires whose outcomes are dominated by those of another strategy. Next, they ought to eliminate strategies which do not meet the maximin, or perhaps the maximin regret criterion (Colman 1982: 22-30). If that should still leave them with more than one eligible strategy, they might plausibly

pick the one with the highest average payoff (Hamburger 1979: Chap. 3).

Assuming that players are risk averse,[9] these rules of generalized caution would generate highly predictable outcomes even in single-shot encounters under conditions of incomplete information. But how would such outcomes differ from solutions of the same game played under conditions of complete information? It is easy to see that opportunities for optimization must be lost if players cannot anticipate each other's moves. Yet whether these would have been opportunities for better cooperation or for more precise exploitation seems to depend entirely upon the character of the game that is being played (Diagram 1).

Take the "Chicken" game, for instance, where mutual uncertainty about the payoffs of the other side would lead either player to adopt a maximin strategy that would assure a cooperative outcome. Or take Reinhard Selten's (1978) famous "Chain-Store Paradox", where each one of twenty small independent stores must separately decide whether to challenge a monopolist by entering its market, while the monopolist must then choose between acquiescence and a costly price war.

---

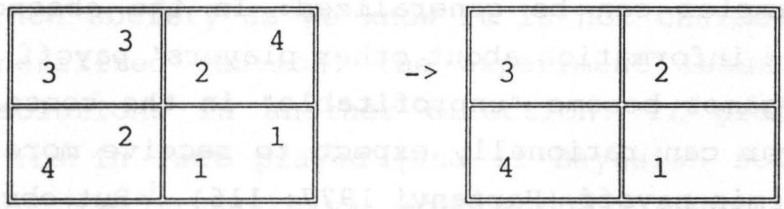
[9] Under conditions of complete information, maximin would only be rational when players' interests are strictly opposed, but not otherwise. With payoff uncertainty, however, that is precisely the aspect of the interaction which is unknown. Logically, optimistic and pessimistic presumptions might then be equally valid or invalid. But psychological research has shown that potential losses seem to weigh more heavily than potential gains (Kahneman/ Tversky 1984; Tversky/ Kahneman 1986).

Within the present context, it is obvious that a small store (for whom the game is a single-shot encounter) should not be willing to risk disaster if it were uncertain about the true payoffs of the monopolist (Kreps/ Wilson 1982). In short: In the absence of reliable information about each other's payoffs, players are unable to exploit the weakness of their partner's position.

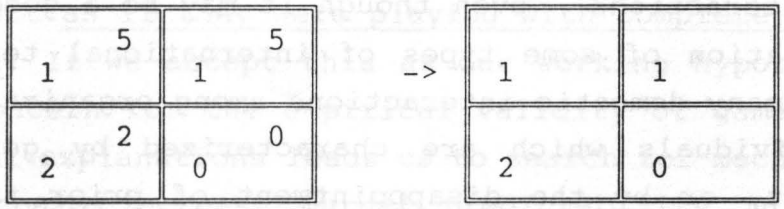
By the same token, however, players are also unable to exploit opportunities for cooperation arising from an underlying harmony of their interests. Even under the benign conditions of the "Assurance" game, uncertainty about the payoffs of the other side would render defection the preferred choice. The same is true in "Battle of the Sexes" and, a fortiori, in the "Prisoner's Dilemma", where cooperation could at best arise out of a sophisticated calculus of long-run mutual interest when the game is iterated.[10]

---

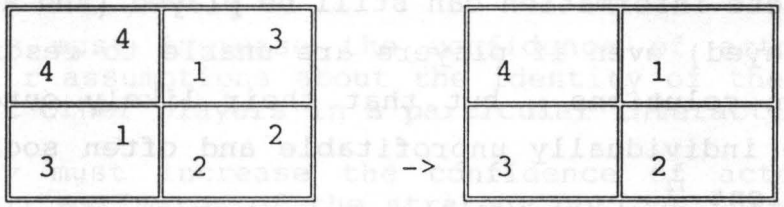
[10] These analytical conclusions are reinforced by experimental research showing that even in long iterations of the games of "Leader" and "Hero" (which is equivalent to Battle of the Sexes), cooperation is significantly lower when the players are not informed about the payoff matrices of their partners (Guyer/ Rapoport 1969). Similar findings have been reported for the Prisoner's Dilemma (Rapoport/ Chammah 1965; Hamburger 1979: 238-242). Thus we have reason to think that the "evolution of cooperation among egoists" discovered in computer simulations of long iterations of the game (Axelrod 1984) depends critically on the empathy of players with the other side's calculus of self-interest. For that reason, and not merely because of greater monitoring difficulties (Hechter 1987: 75), cooperation is less likely to evolve, and more likely to break down, when more players are involved (Fox/ Guyer 1977).



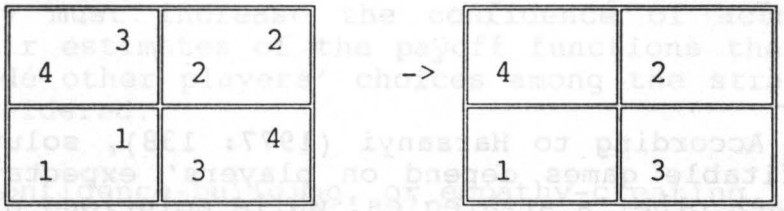
CHICKEN



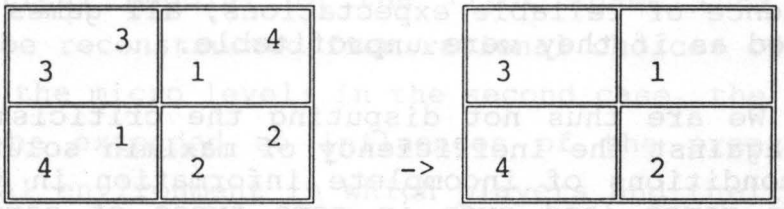
CHAIN-STORE PARADOX



ASSURANCE



BATTLE OF THE SEXES



PRISONER'S DILEMMA

Diagram 1: Strategies under Incomplete Information

The examples can be generalized: In the absence of reliable information about other players' payoff functions, games become "unprofitable" in the sense that no player can rationally expect to receive more than her maximin payoff (Harsanyi 1977: 116).[11] But obviously, our thought experiment is not a model of all real-life interactions - even though it may be a good representation of some types of international tension and of many domestic interactions among organizations or individuals which are characterized by general distrust, or by the disappointment of prior trust. What is important is the demonstration that games with incomplete information can still be played (and are in fact played) even if players are unable to resort to Bayesian solutions - but that their likely outcomes will be individually unprofitable and often socially inefficient.[12]

---

[11] According to Harsanyi (1977: 138), solutions of profitable games depend on players' expectations about each other's strategies, while solutions of unprofitable games are independent of such expectations. The present argument reverses this relationship: In the absence of reliable expectations, all games must be played as if they were unprofitable.

[12] We are thus not disputing the criticism directed against the inefficiency of maximin solutions under conditions of incomplete information in variable-sum games (and even in some types of zero-sum games: Kaysen 1952) with which the Bayesian solution is justified (Selten 1982). Our point is that such inefficient outcomes will in fact occur unless the players find ways to overcome their uncertainty about each other's payoffs - and that among these ways, the Bayesian solution is empirically less plausible than the mechanisms discussed below.

Yet since society as we know it is not characterized by generalized caution, the experiment leads us to seek solutions in another direction: If profitable games are in fact played (and if Bayesian solutions are empirically unlikely), actors must somehow be able to create or exploit conditions that allow them to interact as if they were playing with complete information. If we accept this as our working hypothesis, our concern for the empirical validity of game-theoretical explanations leads us to search for mechanisms that could increase mutual predictability in three ways:

- they must increase the confidence of actors in their assumptions about the identity of the relevant other players in a particular interaction;
- they must increase the confidence of actors in their estimates of the strategy options that other players will in fact consider; and
- they must increase the confidence of actors in their estimates of the payoff functions that will guide other players' choices among the strategies considered.

Such confidence-building, or empathy-creating, mechanisms could be either endogenous or exogenous to game-theoretical models. In the first case, they would could be reconstructed from rational choices of players at the micro level; in the second case, the search would be extended to influences of the preexisting societal environment in which players and their games are embedded. In the following sections, both directions will be explored for the second and third of these mechanisms. Questions relating to the determination of relevant players will be taken up in the second part of this article.

2. The individualistic construction of empathy[13]

The possibility that the informational preconditions for profitable games might be endogenously created through the interaction of rational individuals should be theoretically interesting even if we acknowledge the practical importance of socially constructed coordination mechanisms (institutions and internalized social norms). For one thing, individualistic constructs might help to throw light on the evolution of norms and institutions (Ullmann-Margalit 1977; Kliemt 1986; Shepsle 1989). Secondly, they might be necessary for explaining international and other types of "lawless" interactions. And even when institutions are in place, complex, dynamic and pluralistic modern societies could not exist without wide margins of choice and discretionary action. The question is how actors could develop the converging expectations that are needed to structure these choices.

There is of course an easy individualistic answer to the problem of incomplete information: If players have an interest in coordinating their expectations, they should agree on their permissible strategies and communicate their own payoffs to each other - in other words, they should play a cooperative game. But then, how could they trust each other? If truthful communication and binding agreements were dependent on exogenous enforcement, we would have left the realm of

---

[13] The following section owes much to a discussion with David Soskice who is, however, in no way responsible for my tentative answers to his incisive questions.

individualistic construction - unless one could show how such enforcement mechanisms could themselves arise in interactions that do not depend on exogenous enforcement. In game theory, that condition is expressed by the maxim that "any equilibrium, cooperatively reached or not, must be a noncooperative equilibrium" (Ordeshook 1986: 303).

This seems a reasonable requirement in a world in which the existence of institutions is not presupposed. But solutions are made unnecessarily difficult if "noncooperative" games are defined by the absence of both communication and the exogenous enforcement of agreements, as is the usual convention in game-theoretical literature. Instead, I will here follow Harsanyi's proposal to base the distinction exclusively on the availability of exogenous enforcement, so that "noncooperative" solutions could be derived from assumptions that allow for the possibility of communication, and for nonbinding commitments among players. I will also adopt Harsanyi's further suggestion (1977: 115) that every game may be thought to be preceded by a (possibly tacit) bargaining game in which players will try to achieve an agreed-upon definition of their allowable strategies and relevant payoffs - just as children's games or parlor games are often preceded by discussion and agreement on the applicable rules (Shepsle 1989).

By itself, of course, that would not yet overcome the crucial problem of how players could be able to trust each other's communications and commitments in the absence of exogenous enforcement, nor does it help

much to know that some agreements may be self-enforcing (Harsanyi 1977: 110). It is true, of course, that there are certain classes of games (e.g. "Assurance"), where the players have an interest in truthful communication and where agreements reached in pre-play bargaining would define a strong equilibrium from which no one has an interest to depart. But there are also other types of games (e.g. Chicken) where players would profit from misrepresenting their preferences and/or from defaulting on their commitments.[14] And since, given their initial ignorance of each other's payoffs, players could not know which type of game they are in fact playing, they cannot distinguish one type of communication from the other. They would remain locked in mutual distrust in both types of situations.

In short, players cannot create the conditions of mutual credibility when the individual game is considered in isolation. However, they may be able to do so if the game is placed within a wider context. Among game theorists, it is a well known "folk theorem" that infinite[15] iteration may change the outcome of other-

---

[14] The temptation to misrepresent one's payoffs exists whenever players' most preferred outcomes differ from one another, and it is irrelevant when the other player has a dominant strategy. The temptation to default on an agreement exists when the agreed-upon (Pareto-efficient) solution is not a noncooperative equilibrium. Thus both types of temptation will exist in Chicken, and none of them in Assurance, while only the temptation to misrepresent exists in Battle of the Sexes, and only the temptation to default in the Prisoner's Dilemma.

[15] In finite iterations, rational players would defect on the last play, and hence also on the second-

wise dilemmatic games because it permits players to make their next move contingent on their opponent's previous choice. Thus it has been shown for the iterated Prisoner's Dilemma that cooperation will emerge as a self-enforcing equilibrium when players follow a "tit-for-tat" strategy that responds in kind to each instance of the opponents cooperation of defection (Axelrod 1984; Taylor 1987). The same logic, it is suggested, may also help to solve the problem of truthful communications and credible commitments under conditions of incomplete information.

However, since not all real-life games have the form of a Prisoner's Dilemma, and since infinite or even indefinite iterations of identical games are rare in any case, the logic must be modified in order to become more generally applicable. This is achieved with the help of Harsanyi's distinction between the primary game that is in fact being played, and the pre-play bargaining game in which payoffs may be communicated and strategies agreed upon. While all primary games may differ from one another, they may still be embedded in an open-ended series of interactions among the same players. When that is the case, the corresponding instances of pre-play bargaining may be conceptualized as a more abstract and indefinitely iterated "truth game"[16] in which each player must choose

---

to-last play, and so on. Whether the backward induction of this "endgame effect" can be avoided in games with incomplete information is disputed: Fudenberg/ Maskin (1986); Güth et al., (1988).

[16] The idea of a "truth game" was introduced by Brams (1985: 119-126) in the context of his analysis of the arms race among superpowers. What is different

each time whether to tell the truth about her own payoffs, and whether to keep agreements about her strategies. Correspondingly, each player must also choose each time whether to trust other's communications and commitments.

Now if all instances of default could be detected after the fact, each player would be able to respond in the next round to the other's previous performance in the truth game. If we further assume that players do in fact have a common interest in being able to play profitable games based on trustworthy communications and commitments, the game-theoretical "folk theorem" suggests that they should be able to reach that outcome through individually rational strategies (Fudenberg/ Maskin 1986). But that conclusion is based on two critical assumptions that need to be discussed.

First, the folk theorem will not help if defaults cannot be detected before the other player has to make her next move - which may be so in arms-control negotiations. But even when the actual choice of strategies can be verified after the fact, there will often be no possibility of direct verification for the payoffs that have been communicated. While their truth may sometimes be indirectly confirmed or disconfirmed by the actual choice of strategies (if these can be observed), there are also cases where even that is not possible (Diagram 2).

---

here are the assumptions of iteration and of the diversity of the underlying primary games.

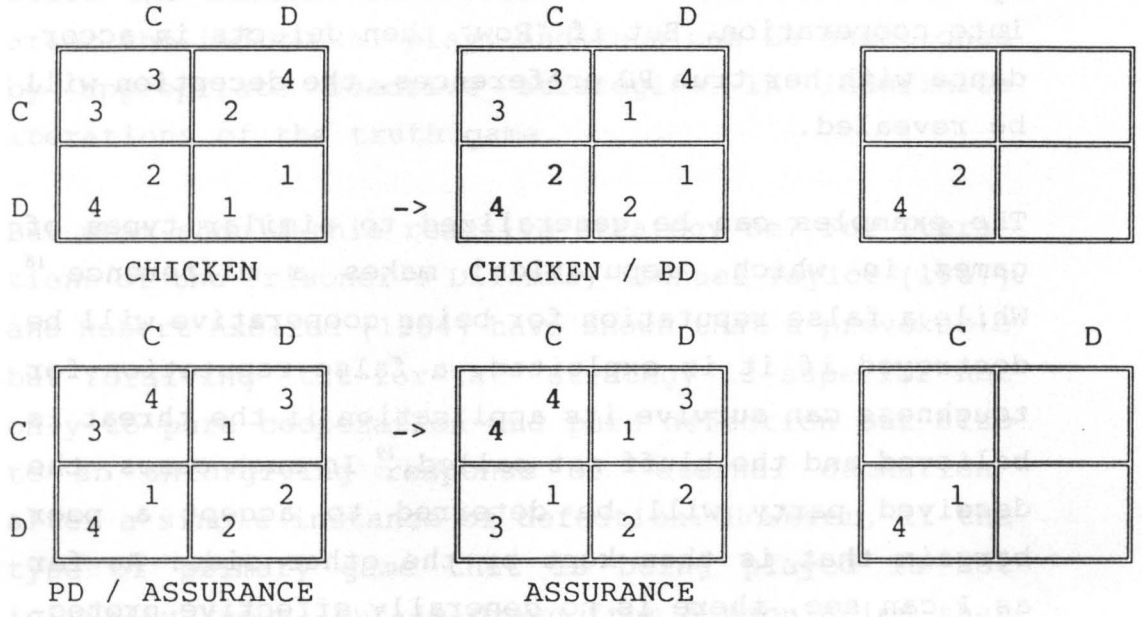


Diagram 2: True Payoffs, Misrepresented Payoffs and Outcomes

Take the example of a Chicken game in which the row player misrepresents her own payoffs as being of the Prisoner's-Dilemma variety. If she is believed, "Column" will expect "Row" to pick D as her dominant strategy, and will choose C herself in order to avoid her own worst-case outcome. As this will then allow "Row" to in fact play D in accordance with her true (Chicken) preferences, the deception cannot be discovered.[17] A counterexample is provided by a game where "Column" has Assurance-type and "Row" Prisoner's-Dilemma-type preferences. If "Row" pretends to also

[17] Of course, if both should misrepresent their preferences in this fashion, and if each should believe the other, the deception would become visible - and both would be surprised to find themselves with a cooperative outcome in the C/C cell.

have Assurance preferences, "Column" will be deceived into cooperation. But if "Row" then defects in accordance with her true PD preferences, the deception will be revealed.

The examples can be generalized to similar types of games in which "reputation" makes a difference.[18] While a false reputation for being cooperative will be destroyed if it is exploited, a false reputation for toughness can survive its application if the threat is believed and the bluff not called.[19] In such cases, the deceived party will be deterred to accept a poor bargain that is then kept by the other side. As far as I can see, there is no generally effective protection against the misrepresentation of payoffs in threat games - except of course the fallback strategy of generalized caution. But it is important to realize that the victim of a false threat is in fact no worse off than she would have been if she had disbelieved all communications and stayed with her maximin strategy to begin with. So the real dangers that players must guard against are not false threats but false promises - which may in fact lead to outcomes that are

---

[18] On the general importance of reputation for the explanation of economic outcomes, see Akerlof (1980).

[19] In the game-theoretical literature, the use (and the risk) of a reputation for toughness has been discussed in the contexts of international relations and of legislative leadership (Ward 1987; Calvert 1987; Alt/ Calvert/ Humes 1988). It has also been shown that under conditions of incomplete (Kreps/ Wilson 1982) or imperfect information (Trockel 1986), the reputation for toughness can be used to construct a solution of the "Chain-Store Paradox".

below the maximin threshold. But these will show up after the individual play, and thus can be sanctioned by appropriate reactive strategies in indefinite iterations of the truth game.

But what should this reactive strategy be? For iterations of the Prisoner's Dilemma, Michael Taylor (1987) and Robert Axelrod (1984) have shown that a provokable but forgiving "tit-for-tat" strategy is superior not only to pure cooperation and pure defection but also to an unforgiving response of "eternal damnation" after a single instance of defection. However, if the type of primary game that is being played is not initially known, but is drawn from a population that includes not only the PD, but also Assurance, Battle of the Sexes, Chicken and other games, tit-for-tat is not a plausible strategy for the truth game. In fact, a player who was known to be committed to tit-for-tat would become the perfect victim of exploitation - deceived whenever the primary game happened to be Chicken, and willing to forgive after each play of Assurance (where telling the truth is in the other player's self-interest). By contrast, an unforgiving response ("once a liar, always a liar") to a single instance of deception would seem to be a superior strategy in the truth game.[20]

---

[20] The strategy makes sense on the assumption that a player who cheats in the truth game reveals herself to have a high discount of the future that justifies the expectation that she will cheat again when that is profitable in a primary game.

Alternatively, one might consider a strategy that begins with generalized caution in first encounters and that incrementally increases the level of one's trust in opponents' communications with each subsequent instance of demonstrated truthfulness (in games

An opponent confronted with the unforgiving strategy would then face the following choice in the truth game: By being truthful, she would maintain access to future games in which her own communications are trusted - and which, therefore, could be more profitable than encounters governed by generalized caution. A single case of (detected) deception, however, would thereafter restrict her to interactions in which her own communications would be distrusted, and in which her opponent would revert to generalized caution in all primary games. Thus the temptation to deceive in an individual case would have to be weighed against the lesser profitability of all future interactions.[21] The same logic does of course apply a fortiori to all situations where different games are simultaneously ongoing among the same parties (McGinnis 1986).

There are two reasons, however, why the deterrent effect of an unforgiving strategy cannot generally assure the truthfulness of communications and the credibility of commitments. First, if punishment must be applied, it is often costly (Molm 1989). More specifically, the player who must in fact revert to generalized caution after a single instance of decep-

---

where deception would have been advantageous). This accumulated "credit rating" could be depleted by a single instance of detected deception, but might be rebuilt again (perhaps under more difficult conditions).

[21] If the opponent applies generalized caution thereafter, there would be no more opportunities to profit from cheating in games with conflicting interests, and both would not be able to profit in games with compatible interests.

tion must also deprive herself of future profitable interactions even in games with common interests like Assurance. Thus the temptation to forgive in later games may be too great to maintain the deterrent effect.[22] Moreover, even a credible threat to be unforgiving would not offer protection against a "hit-and-run" or "confidence-game" player who would build up trust in the truth game in order to facilitate one big "killing" after which interactions are terminated.

Since breaches of trust do in fact occur, we have no reason to construct a theory that would exclude them altogether. Nevertheless, it seems worthwhile to seek endogenous mechanisms that could explain why they do not occur much more often. Such mechanisms can be found by again extending the truth game to include a wider context. In the real world, ongoing relationships are not restricted to dyadic interactions, but often will take place within a population of actors who are aware of, and able to communicate with, each other. Thus other players could also respond with an unforgiving strategy of generalized caution when it is their turn to play against a former defector. But as that form of sanctioning also entails costs for these later players, what still needs to be explained is why they should be motivated to punish.[23]

---

[22] That presupposes that the "interaction orientation" (which will be discussed below) remains constant. But if the deceived party should be sufficiently angry or humiliated, she might switch from an "individualistic" to an "aggressive" orientation in which her own costs would no longer matter.

[23] Axelrod (1986) demonstrated in computer simulations that a meta-norm which makes not punishing a

In game-theoretical analyses, it is usually assumed that players do not have the choice of refusing to play a game. That maximizes the costs of punishment. The cost-benefit balance is much improved, however, when it is possible to relax this assumption (Tullock 1985). In many real-life situations, actors are indeed able to choose whether and with whom they are willing to play. Thus, when confronted with a former defector, a player would have the following choices: (1) playing with the defector and trusting her communications; (2) playing with the defector and applying generalized caution; (3) playing with someone else; and (4) not playing at all. If options (3) or (4) are at all attractive, the cost of punishment associated with avoiding option (1) could be much reduced, while the severity of punishment, and hence its likely deterrent effect on potential defectors, would increase.

Under such conditions, the motive of self-protection against the possibility of being cheated may be all that is needed to assure the ostracism of players that are known to have defaulted on previous occasions. It may even be sufficient to provide a commercial market for reliable information about the "credit-worthiness" of other players whose past defaults one could not have observed directly (Milgrom/ North/ Weingast 1988). And if the discounted disadvantage of being

---

defector a punishable offense could generate a stable equilibrium of universal punishment. But the emergence of this meta-norm is not itself explained as the outcome of rational choices. Similarly, Witt (1986) combines social learning and evolutionary hypotheses to show how the tendency to cooperate, and to punish defectors, might be stabilized after it has, somehow, become dominant in a population.

excluded from future profitable games is greater than the present gain from cheating in a single game, rational players will have reason to maintain their reputation for keeping promises and telling the truth even in single-shot encounters with strangers whom they will not meet again.[24]

To conclude this exercise in speculative exploration: Even in the absence of exogenous enforcement and of exogenous motivations to punish, it seems possible to construct "incentive-compatible" mechanisms (Hurwicz 1972) that would allow rational, self-interested players to create environments where complete-information conditions can be approximated. And to the extent that this enlarged truth game is in fact being played, it should create favorable conditions for the credible communication of payoffs, and for reliable agreements on strategy choices,[25] that will greatly increase the relevance and usefulness of "classical game theory" for the analysis of real-life interactions.

It must be seen, however, that the notion of pre-play bargaining is no more than an analytically useful

---

[24] It is, of course, possible that sanctions will be applied only when all players involved are members of an in-group, while cheating against outsiders is permissible. Uwe Schimank has pointed out the similarity of a such a pattern to the "amoral familism" in backward societies described by Banfield (1958).

[25] In zero-sum games, truthful pre-play bargaining would of course serve no useful purpose. Thus, "no comment" would be an appropriate communication in the truth game, while cheating would still need to be ostracized in order to maintain a context that assures credible communications.

fiction,[26] and that the existence of a truth game also needs to be empirically verified before it can be used for explaining real interactions and outcomes. Strictly speaking, we have only shown that it is possible that actors might achieve common knowledge of payoffs and strategies under conditions whose existence is not ruled out by what we know. Yet in an empirical world of complex interactions, where undisciplined theoretical speculation is not self-correcting, while barefoot empiricism cannot find the forest behind the trees, we must have an interest in internally consistent analytical models derived from plausible assumptions. Thus, the construct of a truth game could at least serve as a useful guide to what is worth searching for in empirical research (Abel 1948).

But even if that is granted, pre-play bargaining by itself cannot generally assure the empirical applicability of game-theoretical models. First, the postulated preconditions for the enlarged truth game, low-cost choice among partners and low-cost information about past defaults of potential partners, will often not exist.[27] Second and more important, the assumptions

---

[26] Plausibility is not increased by Harsanyi's (1977: 112) "principle of tacit bargaining", which assumes that players of sufficient intelligence can reach any agreement on payoff distribution by tacit understanding if they could have reached it by explicit bargaining. This presupposes that payoffs have become common knowledge in some other fashion.

[27] For instance, many corporate actors such as nation-states, government organizations, political parties, labor unions and employer associations, do not have the option of avoiding each other, and the same is true of individual members of organizations or families. On the other hand, these captive players may

introduced could only assure the credibility of communications and agreements, but they do nothing to reduce the information costs associated with idiosyncratic perceptions and preferences. To construct common-knowledge conditions on this foundation alone would impose enormous burdens on the communicative capacities of real-life actors who would have to recollect, specify, transmit, and interpret, all decision premises that are potentially relevant in characteristically ambiguous interaction situations. And even if such feats of communication are conceivable among parties with a long history of close association, they would create serious problems for empirical research. Game-theoretical explanations might thus be entirely valid, and at the same time pragmatically useless, if they should in every case depend on a precise reconstruction of the idiosyncratic (and mostly tacit) understandings among specific actors.

3. The social construction of empathy

Thus we have reason to extend our search for solutions to the common-knowledge problem to include mechanisms that have lower information costs for actors as well as for researchers. If such mechanisms cannot be found within the strict confines of rational-choice premises, they nevertheless seem to be available just across the paradigmatic boundary. In trying to clarify the preconditions of social order, social scientists working in the functionalist, institutionalist or

---

have excellent opportunities for monitoring defaults and applying varieties of sanctions.

symbolic-interactionist traditions have always focused on exactly those factors that are able to reduce information costs by limiting the range of human choices and world views. From a rational-choice perspective, these approaches are often attacked for their lack of generality, their inadequate conceptualization of purposeful action, and their implicit determinism. In response, there are comparable attacks by social scientists on the imperialism of the economic paradigm and the incredible naivety of its underlying assumptions. More promising than such confrontations, however, is an emergent perspective, on both sides of the paradigmatic fence,[28] that considers traditional social-science and rational-choice approaches not as competitors, but as useful complements to each other. While this paper is not intended as a substantive contribution to this unifying perspective, it may be useful to indicate how it could help to facilitate the use of game-theoretical explanations in empirical research.[29]

---

[28] Among examples that come to mind are the works of Bruno Frey (1980), Nelson and Winter (1982), Raymond Boudon (1983; 1988), Mark Granovetter (1985), Hartmut Kliemt (1985; 1986), Elinor Ostrom (1986), James S. Coleman (1986) Amitai Etzioni (1988) or March and Olsen (1989).

[29] By this I do not intend to take sides in the philosophical debate between "monists" (e.g. Lindenberg 1983; 1989), who believe that "ultimately" all social phenomena must be reducible to self-interested choices, and "dualists" (e.g. Kliemt 1985; Etzioni 1988), who insist that institutions, social norms and moral laws cannot be so reduced. My reasons for using game theory as the dominant frame of reference are pragmatic: while I do not see how choice-oriented knowledge could be integrated in, say, a functionalist perspective (Scharpf 1989b), game theory seems able to

Regardless of all divisions among competing schools, sociologists and political scientists have always been in broad agreement about the importance of two sets of factors that are directly pertinent to the information problems associated with game-theoretical explanations, namely

- power structures and institutionalized rule systems that will exclude a wide range of physically feasible behavior as being infeasible or impermissible in a given interaction situation, and
- socially or culturally constructed systems of meaning that will exclude a wide range of subjectively conceivable preferences and perceptions as being irrelevant in a given interaction situation.

Both sets of factors define a social context that does not have to be created de novo by individual actors. When games are "embedded" in this context, strategies and payoff matrices become "socially constructed" - and hence transparent for players who must choose their moves within a well-defined set of options and who know that their partners must do so as well. As Norton Long (1961: 140-1) put it in his discussion of "The Local Community as an Ecology of Games":

> "Here we deal with the essence of predictability in social affairs. If we know the game being played is baseball and that X is a third baseman, by knowing his position and the game being played we can tell more about X's activities on the field than we could if we examined X as a psychologist or psychiatrist. If such were not the case, X would belong in the mental ward rather than in a ball park. The behavior of X is not some disembodied rationality

---

use social-science knowledge about actors, and the institutions and cultural orientations influencing their choices.

> but, rather, behavior within an organized group activity that has goals, norms, strategies, and roles that give the very field and ground for rationality."

In short, "baseball structures the situation" - and so does the "political game", the "banking game", the "contracting game", the "newspaper game", the "civic organization game", or the "ecclesiastical game," all of which are simultaneously ongoing and interacting with each other in Long's perception of the local community. Similarly, when Michel Crozier (1976) and Crozier and Friedberg (1979: 66-73) speak of the "games" that constitute the reality of the organizations they have studied, they refer to empirically discoverable (formal and informal) "rules of the game" that define the strategies among which actors must choose, and the gains and losses about which they must be concerned if they wish to continue playing within the organization.

Thus, even though neither Long nor Crozier and Friedberg were much interested in the information requirements of game-theoretical models, or in formal models at all,[30] they have specified conditions under which the contingency and ambiguity of the world is sufficiently reduced to make complete information games a plausible assumption. Moreover, if relevant strategies

---

[30] Crozier does in fact consider the possibility that the empirically discovered rules of the game "could eventually be formalized according to rough game theory models" (1976: 196, 203). But in his view, such formalization can only "be done ex post - i.e. when one knows the outcomes. The strategies and the payoffs can be interesting only for comparative purposes" (1976: 206, note 3).

and payoffs are in fact defined by reference to existing power relations, formal and informal rules and shared meanings, their discovery should also become a much more manageable task for empirical research.

In some fields of highly structured and highly public interaction - e.g. in the legislative arena - it may be entirely sufficient to draw on legal and customary rules and publicly documented outcomes in order to determine both the set of available strategies and the (proximate) payoffs that players are trying to maximize. That explains the persuasiveness of game-theoretical explanations of coalition formation and roll-call voting in Congress and its committees (Shepsle 1979; Shepsle/ Weingast 1981; 1987; Denzau/ Riker/ Shepsle 1985). In many other areas, however, power relations and (legal and organizational) systems of rules that are backed by institutionalized sanctions will not fully determine the value of payoffs, but will nevertheless constrain the range of strategy choices that actors and observers need to take into consideration (Ostrom 1986; Burns/ Flam 1987).

- Thus boundary rules may effectively restrict access to decision arenas to legitimate participants and may permit actors and observers to ignore "outsiders" even if they are also affected by the outcome of interactions in the arena. In game-theoretical terms, they define who the players are likely to be. This point will be taken up again in the second part of the article.

- Constitutional and procedural decision rules will restrict the issues that may be dealt with in an arena, and the types of action that are permissible. They also define which actors are allowed to participate in what role in which type of decisions; and they prescribe the rule by which the contributions of individual participants will be aggregated (unanimity, qualified majority, plurality, hierarchy). In game-theoretical terms, decision rules restrict the range of strategies that players can be expected to use, and they may also restrict the range of outcomes that can be achieved in interactions - and hence the payoffs obtainable by players.

While actors in control of potent power resources (e.g. the means of physical violence) may sometimes be able to disregard formal rules, institutional legitimacy is itself an action resource that is difficult and often impossible to substitute by other means. Nevertheless, it is rare that power constraints and institutionalized rules will completely determine the choice among feasible strategies. Thus actors' perceptions of the situation and their preferences among outcomes - i.e. their perceived payoff matrices - will generally be important for any attempt to predict their choices. But, again, these perceptions and preferences need not be empirically determined de novo in every case. Often it will be possible to refer to plausible "bridge assumptions" instead.

The term refers to empirically based assumptions about regularities in the orientation of certain classes of

actors in certain situations. While some of these bridge assumptions may be of near-universal validity, most of them will be time and space specific in their applicability. In either case, however, they may be (provisionally) used as substitutes for specifically collected data on individual preferences and perceptions in the formulation of empirically testable hypotheses (Homans 1961; Lindenberg 1983; 1989). I find it useful to distinguish three levels of such bridge assumptions - definitions of basic self-interest, social norms, and interaction orientations.

At the first level, we are interested in definitions of self-interest that are so general and so basic that they may plausibly be imputed to most actors. While Adam Smith had tried to get by with two types of universal definitions of utility, "social approval" and "physical well-being", Siegwart Lindenberg (1989: 190) found it necessary to add the "minimization of loss" to that short list, and other authors have provided much longer catalogues of "values" (Lasswell/ Kaplan 1950)[31] or of "basic human needs" (Maslow 1970). For corporate actors, a similar list might include organizational survival, autonomy and growth. To the extent that these fundamental interests are seriously at stake in an interaction, they may be all that is needed to explain actors' choices in what has been

[31] It is interesting to note, however, that the eight Lasswellian values are grouped into two larger categories corresponding closely to Smith's original dichotomy, namely "welfare values" (including well-being, wealth, skill and enlightenment) and "deference values" (power, respect, rectitude and affection) (Lasswell/ Kaplan 195; 55-56).

called "single-exit" or "high-cost situations" (Latsis 1972; Zintl 1989).

In many other situations, however, basic self-interest by itself will be less directly useful as a predictor. Even if it should still be the sole motivating force of human action (as would be claimed from a radical "monistic" perspective), its pursuit would be mediated by social norms in the widest sense. They define, in Lindenberg's apt metaphor, the "production function" through which an actor's choices contribute to the attainment of her ultimate goals (Lindenberg 1989: 190-192).[32] If a doctor seeks social approval, she must fulfill the role expectations of a good doctor, and if political parties seek to maximize votes, they must present candidates and platforms that believably pursue the public interest. In this sense, the abstract calculus of self-interest is less instructive for explanations and predictions even of purposeful choices than a "logic of appropriateness" (March/Olsen 1989), which defines specific criteria of relevance and rightness that actors use in judging their own and others' choices.[33]

---

[32] Lindenberg (1989: 187-189) also introduces the theoretically promising notion that - as a consequence of bounded rationality - actors will at any one time only be able to focus on a single goal (among many potentially relevant ones) that then "frames" the choice situation. This provides a theoretical opening for treating social norms as "situationally prescribed goals" which may, however, lose their predictive power if another value gains in situational salience.

[33] March and Olson (1989) oppose the "logic of appropriateness" to the "logic of consequentiality", rather than to a "logic of self-interest". This seems problematical if one assumes that norm-oriented behav-

Among these criteria we may distinguish norms defined either in terms of universalistic justice, morality or fairness (Etzioni 1988; March/ Olsen 1989), or in terms of the specific (individual and collective) identities to which actors must refer for their "self-interested" evaluation of outcomes.[34] At least the latter kind of group-oriented social norms are often sanctioned by the highly effective threat of collective disapproval and ostracism. Of even greater practical importance are functionally specific orientations defined by the specific codes of relevance and excellence associated with different societal subsystems (Luhmann 1986), with socially coherent professions, or with the development of large technical systems (Mayntz/ Hughes 1988). Outsiders will have learned to take these specific criteria for granted, and insiders will have internalized them through professional socialization supported by the incentives

---

ior is not necessarily purely expressive, but may also be concerned with expected consequences for the values that the norm serves. A more systematic treatment would thus need to distinguish at least two dimensions, "instrumental vs. expressive" and "self-interested vs. norm-oriented."

[34] As Jon Elster (1983: 35-36:) has pointed out, certain arguments simply cannot be stated publicly in a particular setting. "In a political discussion it is pragmatically impossible to assert that a given solution be chosen simply because it favors oneself or the group to which one belongs." The reason, I submit, is a definition of the collective identity that is appropriate for the particular debate. If the same issues were discussed by the same individuals within an interest organization, or within their own family, different criteria of relevance would exclude different types of arguments.

and disincentives associated with professional reputation and ostracism.

In addition to universalistic, group-oriented and function-oriented social norms, whose relevance is generally acknowledged in the social sciences, I find it useful to discuss a third level of bridge assumptions which, for want of a better label, I will here describe as "interaction orientations".[35] It is of particular relevance in the context of game-theoretical analyses which focus on interactions among the goals or interests of two or more actors. Like standard economics, game theory usually assumes an "individualistic" orientation in which all actors are only concerned with their own utility (however its definition may be mediated through social norms). But that is not always true. Interactions between political parties or sports teams are clearly governed by a "competitive" orientation in which one's own success or failure depends very much on payoffs received by other players. In a different way, these also matter in interactions within a sports team, a government coalition, a joint business venture, or perhaps a happy marriage - all of which depend on "solidaristic" orientations that bring members to care not only about their own payoff, but about the aggregate payoff received by all partners jointly. Other possibilities

---

[35] The phenomenon has been discussed under a variety of labels, including "social motives" (MacCrimmon/ Messick 1976), "styles of decision making" (Scharpf 1988; 1989), "social orientations" (Schulz/ May 1989), or "social payoffs" (Burns 1989). The term proposed here has the advantage of emphasizing interaction effects and differing more clearly from "social norms".

include "egalitarian", "altruistic" or "aggressive" orientations.[36]

All of these non-"individualistic" orientations have the effect of relating the player's utility in specific ways to the payoffs received by other players, and they often derive their motivating force from underlying emotions (love, communal feelings, envy, fear, anger or hate) that make people care about others in one way or another (Flam 1989). Given the potential instability of the underlying emotions, interaction orientations may also be quite volatile. Theoretically more interesting, in the present context, is the possibility that particular types of mixed-motive games may generate endogenous shifts of dominant orientations in response to past moves of other players.[37] When such shifts occur, the "effective matrix" of the game[38] may radically change its character. While a solidaristic orientation will transform all dilemmas into games of pure coordination, a shift from individualistic to competitive orientations will transform

---

[36] For the complete list of experimentally identifiable "social orientations", see Schulz and May 1989).

[37] A good example is the characteristic cycle of political coalitions moving from a "honeymoon" of solidaristic orientations through self-interested individualism to competition and, perhaps, outright hostility. Such cycles are to be expected in game constellations resembling Battle of the Sexes.

[38] The "effective matrix" which governs players' choices is the transformation of an objectively "given matrix" by a transformation rule representing what I have labelled the interaction orientation (Kelley/ Thibaut 1978: 14-17).

any mixed-motive game into zero-sum conflict (Scharpf 1989).

That is not meant to deny that volatility is not everywhere, and that these interaction orientations are also regulated by social norms.[39] Buyers and sellers in the market are expected to be guided by individualistic "interests", rather than by their "passions" (Hirschman 1977); sports teams and political parties should compete with each other, but their internal interactions should be solidaristic; and guardians as well as doctors, psychotherapists and other members of the helping professions are supposed to have an altruistic concern for the interests of their wards and clients (Barber 1983). Nevertheless, interaction orientations should probably not be considered as just another type of social norm. Norms are typically addressed to individual actors. Yet when the very definition of a relationship is concerned, "it takes two to tango" - meaning that it may be difficult or impossible to unilaterally maintain a solidaristic, altruistic, egalitarian or even individualistic orientation if other players are competitive or hostile. As a consequence, the binding force of social norms is likely to be attenuated by the mutual dependence of interaction orientations - which also means that empirical game research must be sensitive to the fact that actors will retain a high degree of mutual con-

[39] As Helena Flam has pointed out, social norms may either define what is appropriate to feel in a situation, or they may prescribe rule-oriented action that substitutes for emotion-based behavior.

trol over the effective definition of their relationship.[40]

4. Discussion

But where do all these possibilities of socially constructed orientations leave us in our quest for parsimonious explanations? Even if they help us in formulating empirically testable hypotheses, it is important to realize that the bridge assumptions discussed above can, at best, provide conditional simplifications. There are essentially three reasons why predictions or explanations based on them might fail in empirical applications.

First, since human actors have acquired the potential of thinking and acting on their own, there is no way in which socially constructed interests, rules, norms, or interaction orientations can completely determine their perceptions and valuations. As individuals are capable of ignoring their own survival interests, they are also capable of violating institutionalized rules and social norms, even if they have to pay a price for it. Even more important, for practical purposes, are the inventiveness and determination of individuals in shaping, bending and manipulating institutions and

---

[40] From a research point of view, it is nevertheless encouraging that these interaction orientations seem to fall into a limited number of patterns that can be operationally defined. It is not yet clear, however, whether it will be possible to develop information-efficient methods for identifying these orientations in the type of materials accessible to empirical (as distinguished from experimental) research.

norms to serve their own purposes. There is, in other words, reason to be on guard against "oversocialized conceptions of human nature" (Granovetter 1985), and to allow for the possibility of idiosyncratic action that does not conform to the regularities predicted by bridge assumptions.

Second, and theoretically more interesting, the socially constructed mechanisms that standardize and coordinate behavior are only able to support time and space specific empirical regularities or "quasi-laws", rather than general laws of universal validity. The reason is straightforward: Coordinating mechanisms are the more advantageous, the more they are in fact followed. In technical language, they have positive scale effects. Thus, behavioral regularities that already exist have a competitive advantage over new proposals, even if these would have been superior when considered in the abstract. When that is so, their evolution becomes path-dependent, and historical accidents make a difference in the selection of one pattern of behavioral regularities over its potential competitors (David 1985).

As a consequence, there is no theoretical reason to assume that history should be "efficient" (March/Olsen 1989: Chap. 4) or that institutional arrangements, social norms or social orientations should converge toward unique and hence universal equilibria. In practice that means that in our empirical work we cannot rely on bridge assumptions whose empirical validity was established elsewhere or at another time. The norms and interaction orientations guiding Swedish

unions in the 1970s were different from those prevailing in Britain, but they also differed from those that had been documented in Sweden during the 1960s. For the same reason, the assumption - which is still characteristic of most economic research - that business firms or banks should have identical utility functions in the United States, in Europe and in Japan rests on weak theoretical foundations. Explanations and predictions based on bridge assumptions may thus fail because we have not taken care to determine the specific content of the quasi-laws that apply in the time-space region that we are studying.

They may also fail for a third reason that relates back to the individualistic construction of empathy discussed above. Bridge assumptions, whether derived from institutionalized rules, basic interests, social norms or socially regulated interaction orientations, can only define standardized expectations. They are useful tools for empirical research within a game-theoretical frame of reference because we can assume that actors must rely on them for their expectations when they do not know very much about their partners in a given interaction situation.

Yet there are in fact many interactions where the players do know a good deal more about each other because, as Mark Granovetter (1985) has emphasized, specific encounters are often "embedded" not only in the general context of institutional rules and social norms, but also in ongoing personal relations and networks of such relations. When such ongoing relations do exist, the reliability of actors' expecta-

tions, and their trust in each other's commitments, are raised far above the level that would be reasonable among even well-socialized strangers. As was suggested above, and as Granovetter himself points out, this increase in empathy is not an unmixed blessing. It creates opportunities not only for mutually advantageous cooperation but also for exploitation: "The more complete the trust, the greater the potential gain from malfeasance" (1985: 491).[41] But that is only the reverse side of the same coin. What is more important is the fact that these personal relations and networks of personal relations will widen the range of motives that can be pursued in an interaction.

"Profitable games", we have seen, can only be played if players' payoff functions are common knowledge. By reverse implication, that also restricts the range of motives, which even cooperative players are able to pursue, to a subset of which one can safely presume mutual awareness. Institutionalized rules, social norms and other bridge assumptions provide a general basis for such mutual awareness. But - as Georg Simmel (1917/1984: 33-39) pointed out long ago - the very generality of such orientations systematically ex-

---

[41] That explains why the mere communication of idiosyncratic concerns is facilitated by the absence of opportunities for consequential interaction (e.g. among strangers meeting in a railroad compartment or on a vacation). But where interaction is possible or inevitable, higher levels of voluntarily granted empathy depend on trust. In its absence, any attempt to increase one's understanding of another's payoffs beyond the limits of conventional concerns and motives becomes a violation of "privacy" or a case of "espionage".

cludes those aspirations in which human beings tend to differ from one another. Even if we should not share Simmel's pessimism that this must always discriminate against "higher" aspirations,[42] it is clear that it must exclude wide ranges of concerns which are extremely valuable to actors, and in which they could enormously benefit from empathetic interaction and exchange.

Ongoing personal relations and networks of such relations are in fact able to create the preconditions for higher levels of empathy among actors. Where that is so, they permit more idiosyncratic concerns to be communicated and understood and to become the object of interdependent action. Thus even if the importance of standardized bridge assumptions is fully recognized, the "individualistic construction of empathy" discussed above retains its practical significance, since it allows actors to achieve greater scope and precision in their coordination of expectations than can be attained through the "social construction of empathy."

---

[42] For Simmel, this constitutes the "sociological tragedy":

"Der Einzelne mag noch so feine, hochentwickelte, durchgebildete Qualitäten besitzen - gerade je mehr das der Fall ist, desto unwahrscheinlicher wird die Gleichheit und also die Einheitsbildung gerade dieser mit den Qualitäten anderer, desto mehr strecken sie sich nach der Dimension der Unvergleichbarkeit hin, auf desto niedrigere, primitiv sinnlichere Schichten reduziert sich das, womit er sich mit Sicherheit den anderen angleichen und mit ihnen eine einheitlich charakterisierte Masse formen kann" (Simmel 1984: 38).

What we seem to have, then, are different layers of structuration of social interaction,[43] all of which are amenable to game-theoretical interpretation. At the lowest level, there is the "minimal social order" of our thought experiment: In the absence of trustworthy expectations, rational actors will follow strategies based on generalized caution that minimize the risks, as well as the gains, of interaction. At the next level, mutual expectations are structured by relatively standardized bridge assumptions regarding actor interests and the impact of institutions and social norms. These will reduce the need for generalized caution by limiting the range of permissible strategy choices, and they will permit actors (and researchers) to refer to standardized payoff matrices for the game-theoretical analysis of specific types of interactions. At the third level, finally, ongoing personal relations and networks of such relations may so increase empathy and trust among actors that the range of interests and motives that can be processed in an interaction is significantly enlarged.

But even though all three layers of structuration seem amenable to game-theoretical interpretation, opportunities for empirically valid game-theoretical explanations and predictions seem to differ significantly

---

[43] A somewhat similar idea was expressed by Uwe Schimank who, in his comment on Milgrom/ North /Weingast (1988), distinguished between small-scale social order (based on ongoing personal interactions), medium-scale social order (based on reputations depending on third-party communication) and large-scale social order (depending on mass media). That does not, however, seem to exhaust the order-creating potential of institutions and social norms.

among these levels. Least problematical, perhaps, are the conditions of generalized caution. Actors' own knowledge is often reduced to (worst-case) estimates of each other's capabilities; and given their lack of mutual trust, they are unlikely to interact unless some salient interest of their own is in fact at stake. Thus their perceptions of each other, their own payoffs, and their solution concepts should be fairly easy to reconstruct from data that, in principle, are accessible to empirical or historical research. This explains the relative success of game-theoretical explanations under conditions of international tension and crisis (Snyder/ Diesing 1977).

More research effort is required in situations of the second type where socially constructed bridge assumptions are dominant. Nevertheless, the institutional structuring of interactions is likely to be quite compelling and relatively easy to ascertain at the level of permissible strategies. At the level of payoffs, however, researchers will have to cope with local variations, and with the coexistence of the different types of criteria discussed above: individual or organizational self-interest, ethical or function-related norms of appropriateness, and specific interaction orientations. Yet it is likely that one of these concerns will be dominant in a particular situation, and in any case the criteria are socially defined, so that their meaning will often be common knowledge within a wider community that includes the researcher as well.

Even more difficult to obtain are empirically-grounded explanations of interactions embedded in ongoing personal relations or in "relational contracts" among firms and other corporate actors (Macneil 1980). The degree of mutual trust that is facilitated by a common history, and an expected common future, reduces information costs among actors and facilitates a degree of openness and empathy that is unattainable in single-shot encounters and arms-length relationships. Their higher degree of empathy permits actors to introduce idiosyncratic concerns, and to fine-tune their mutual expectations in such a way that games of potentially much higher "profitability" can be played.

By the same token, however, information costs for the researcher must sharply increase since perceptions and valuations of actors engaged in ongoing relations are much more difficult to capture, and are easily misconstrued when standard bridge assumptions are applied. Ironically, therefore, while high-empathy interactions will most closely approximate the theoretical ideal of complete-information games, they are the ones whose game-theoretical explanation will require the greatest empirical effort, and is most likely to fail for lack of adequate empirical data. But that, it should be noted, is a problem of all social-science research dealing with highly individualized social interactions. It becomes more painfully visible here only because game theory could, in principle, provide more powerful theoretical tools for the analysis of such interactions.

On the whole, however, this essay ends on an optimistic note: Game theory is clearly a powerful tool for the disciplined theoretical analysis of interacting choices.[44] However, its analytical power is dependent on seemingly unrealistic assumptions about actors' common knowledge of strategies and payoffs. What I have tried to show is that actors are, in principle, able to cope with these informational demands by either resorting to strategies of generalized caution, or by using institutional rules and social norms for their orientation, or finally by creating the conditions for an individualistic construction of common knowledge through credible communications which are embedded in networks of ongoing relationships. I also have tried to show that game-theoretical research may profit from bridge assumptions in the reconstruction of norm-oriented choices, while research on interactions embedded in ongoing relationships will inevitably face very high information costs - which are, however, no higher than those encountered by other approaches. What still remains to be discussed are the problems of over-complexity arising from the connectedness of games, which will be the subject of a subsequent paper.

---

[44] Competitors that come to mind are transactional analysis in psychology and symbolic interactionism in sociology.

Bibliography

Abel, Theodore, 1948: The Operation Called Verstehen. In: American Journal of Sociology 54, reprinted in Hans Albert, ed., Theorie und Realität. Ausgewählte Aufsätze zur Wissenschaftslehre der Sozialwissenschaften. Tübingen: Mohr (1964), 177-188.

Akerlof, George A., 1980: A Theory of Social Custom, of Which Unemployment May Be One Consequence. In: Quarterly Journal of Economics 94, 749-775.

Alt, James E./ Randall L. Calvert/ Brian D. Humes, 1988: Reputation and Hegemonic Stability: A Game-Theoretic Analysis. In: American Political Science Review 82, 445-466.

Axelrod, Robert, 1984: The Evolution of Cooperation. New York: Basic Books.

Axelrod, Robert, 1986: An Evolutionary Approach to Norms. In: American Political Science Review 80, 1095-1111.

Banfield, Edward C., 1958: The Moral Basis of a Backward Society. New York: Free Press.

Barber, Bernard, 1983: The Logic and Limits of Trust. New Brunswick: Rutgers University Press.

Binmore, K.G., 1987: Why Game Theory "Doesn't Work". In: P.G. Bennett (ed.), Analysing Conflict and its Resolution. Oxford: Clarendon Press, 23-42.

Blalock, Hubert M., Jr., 1989: The Real and Unrealized Contributions of Quantitative Sociology. In: American Sociological Review 54, 447-460.

Boudon, Raymond, 1983: Individual Action and Social Change: a No-theory of Social Change. In: British Journal of Sociology 34, 1-18.

Brams, Steven J., 1985: Superpower Games. New Haven: Yale University Press.

Burns, Tom R., 1989: Notes: Toward a Sociological Theory of Games and Social Behavior. Ms. Uppsala University.

Burns, Tom R./ Helena Flam, 1987: The Shaping of Social Organizations. Social Rule System Theory with Applications. London: Sage.

Calvert, Randall L., 1987: Reputation and Legislative Leadership. In: Public Choice 55, 81-99.

Coleman, James S., 1986: Micro Foundations and Macrosocial Theory. In: Siegwart Lindenberg/ James S. Coleman/ Stefan Nowak, eds., Approaches to Social Theory. New York: Russell Sage Foundation, 345-367.

Colman, Andrew M., 1982: Game Theory and Experimental Games. Oxford: Pergamon Press.

Crozier, Michel, 1976: Comparing Structures and Comparing Games. In: Geert Hofstede/ M. Sami Kassem (eds.), European Contributions to Organization Theory. Assen: Van Gorkum, 193-207.

Crozier, Michel/ Erhard Friedberg, 1979: Macht und Organisation. Die Zwänge kollektiven Handelns. Königstein: Athenäum. First published as "L'Acteur et le Système". Paris: Edition du Seuil, 1977.

David, Paul A., 1985: Clio and the Economics of QUERTY. In: American Economic History 75, 332-337.

Denzau, Arthur/ William Riker/ Kenneth Shepsle, 1985: Farquharson and Fenno: Sophisticated Voting and the Home Style. In: American Political Science Review 79, 1117-1134.

Eichhorn, Wolfgang, 1982: Diskussionsbeitrag zu Reinhard Selten: "Einführung in die Theorie der Spiele mit unvollständiger Information". In: Erich Streißler (ed.), Information in der Wirtschaft. Berlin: Duncker & Humblot, 149-150.

Elster, Jon, 1983: Sour Grapes. Studies in the Subversion of Rationality. Cambridge: Cambridge University Press.

Elster, Jon, 1986: Rational Choice. Oxford: Basil Blackwell.

Etzioni, Amitai, 1988: The Moral Dimension. Toward a New Economics. New York: Free Press.

Flam, Helena, 1989: Emotional Man: A Third Perspective on Collective and Corporate Action. MPIFG Discussion Paper 89/7. Köln: Max-Planck-Institut für Gesellschaftsforschung.

Fox, John/ Melvin Guyer, 1977: Group Size and Other's Strategy in an N-Person Game. In: Journal of Conflict Resolution 21, 323-38.

Frey, Bruno S., 1980: Ökonomie als Verhaltenswissenschaft. Ansatz, Kritik und europäischer Beitrag. In: Jahrbuch für Sozialwissenschaft 31, 21-35.

Friedman, James W., 1986: Game Theory with Applications to Economics. Oxford: Oxford University Press.

Fudenberg, Drew/ Eric Maskin, 1986: The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. In: Econometrica 54, 533-554.

Granovetter, Mark, 1985: Economic Action and Social Structure: The Problem of Embeddedness. In: American Journal of Sociology 91, 481-510.

Güth, Werner/ Wolfgang Leininger/ Gunter Stephan, 1988: On Supergames and Folk Theorems: A Conceptual Discussion. Game Theory in the Behavioral Sciences 19. Universität Bielefeld: Zentrum für Interdisziplinäre Forschung.

Guyer, Melvin/ Anatol Rapoport, 1969: Information Effects in Two Mixed-Motive Games. In: Behavioral Science 14, 467-482.

Hamburger, Henry, 1979: Games as Models of Social Phenomena. San Francisco: W.H. Freeman.

Hannan, Michael T./ John Freeman, 1977: The Population Ecology of Organizations In: American Journal of Sociology 82, 929-964.

Hannan, Michael T./ John Freeman, 1984: Structural Inertia and Organizational Change. In: American Sociological Review 49, 149-164.

Harsanyi, John C., 1967-68: Games with Incomplete Information Played by Bayesian Players. In: Management Science 14: 159-182, 320-334, 486-502.

Harsanyi, John C., 1977: Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge: Cambridge University Press.

Harsanyi, John C./ Reinhard Selten, 1988: A General Theory of Equilibrium Selection in Games. Cambridge, MA: MIT Press.

Hechter, Michael, 1987: Principles of Group Solidarity. Berkeley: University of California Press.

Heiner, Ronald A., 1983: The Origin of Predictable Behavior. In: American Economic Review 73, 560-595.

Hernes, Gudmund, 1976: Structural Change in Social Processes. In: American Journal of Sociology 82, 513-547.

Hirschman, Albert O., 1977: The Passions and the Interests. Princeton: Princeton University Press.

Homans, George, 1961: Social Behavior. Its Elementary Forms. New York: Harcourt, Brace & World.

Hurwicz, Leonid, 1972: On Informationally Decentralized Systems. In: C.B. McGuire/ Roy Radner (eds.), Decision and Organization. A Volume in Honor of Jacob Marschak. Amsterdam: North-Holland, 297-336.

Jervis, Robert, 1978: Cooperation Under the Security Dilemma. In: World Politics 30, 167-214.

Kahneman, Daniel/ Amos Tversky, 1984: Choices, Values, and Frames. In: American Psychologist 39, 341-350.

Kaysen, Carl, 1952: The Minimax Rule of the Theory of Games, and the Choice of Strategies under Conditions of Uncertainty. In: Metroeconomica 4, 5-14.

Kelley, Harold H./ John W. Thibaut, 1978: Interpersonal Relations. A Theory of Interdependence. New York: John Wiley.

Kliemt, Hartmut, 1985: Moralische Institutionen. Empiristische Theorien ihrer Evolution. Freiburg: Karl Alber.

Kliemt, Hartmut, 1986: Antagonistische Kooperation. Elementare spieltheoretische Modelle spontaner Ordnungsentstehung. Freiburg: Karl Alber.

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

Lasswell, Harold D./ Abraham Kaplan, 1950: Power and Society. A Framework for Political Inquiry. New Haven: Yale University Press.

Latsis, Spiros, 1972: Situational Determinism in Economics. In: British Journal for the Philosophy of Science 23, 207-245.

Lindenberg, Siegwart, 1983: The New Political Economy: its potential and limitations for the social sciences in general and for sociology in particular. In: Walter Soudeur (ed.), Ökonomische Erklärungen sozialen Verhaltens. Duisburg: Verlag der sozialwissenschaftlichen Kooperative, 7-66.

Lindenberg, Siegwart, 1989: Choice and Culture: The Behavioral Basis of Cultural Impact on Transactions. In: Hans Haferkamp (ed.), Social Structure and Culture. Berlin: de Gruyter, 175-200.

Long, Norton, 1961: The Local Community as an Ecology of Games. In: Norton Long, The Polity. Chicago: Rand McNally, 139-155.

Luhmann, Niklas, 1986: Ökologische Kommunikation. Opladen: Westdeutscher Verlag.

MacCrimmon, Kenneth R./ David M. Messick, 1976: A Framework for Social Motives. In: Behavioral Science 21, 86-100.

Macneil, Ian R., 1980: The New Social Contract. An Inquiry into Modern Contractual Relations. New Haven: Yale University Press.

March, James G./ Johan P. Olsen, 1989: Rediscovering Institutions. The Organizational Basis of Politics. New York: Free Press, forthcoming.

Maslow, Abraham H., 1970: Motivation and Personality. Revised Edition. New York: Harper.

Maynard Smith, John, 1982: Evolution and the Theory of Games. Cambridge: Cambridge University Press.

Mayntz, Renate/ Thomas P. Hughes (eds.), 1988: The Development of Large Technical Systems. Frankfurt: Campus.

McGinnis, Michael D., 1986: Issue Linkage and the Evolution of International Cooperation. In: Journal of Conflict Resolution 30, 141-170.

Milgrom, Paul R./ Douglass C. North/ Barry R. Weingast, 1988: Third-Party Enforcement of Norms and Contracts: A Theoretical-Historical Analysis. Paper prepared for the "Workshop on Connected Games", Max-Planck-Institut für Gesellschaftsforschung, Cologne, November 17-18, 1988.

Molm, Linda D., 1989: Punishment Power: A Balancing Process in Power-Dependence Relations. In: American Journal of Sociology 94, 1392-1418.

Neisser, Hans, 1952: The Strategy of Expecting the Worst. In: Social Research 19, 346-363.

Nelson, Richard R./ Sidney G. Winter, 1982: An Evolutionary Theory of Economic Change. Cambridge, MA: Belknap Press.

Ordeshook, Peter C., 1986: Game Theory and Political Theory. An Introduction. Cambridge: Cambridge University Press.

Ostrom, Elinor, 1986: An Agenda for the Study of Institutions. In: Public Choice 48, 3-25.

Peterson, Richard A., 1970: Comment: Must the Quest After Variance End in History? In: Mayer N. Zald (ed.), Power in Organizations. Nashville: Vanderbilt University Press, 92-96.

Przeworski, Adam, 1986: Methodologischer Individualismus als Herausforderung der marxistischen Theorie. In: Prokla 62, 120-143.

Rapoport, Anatol/ Albert M. Chammah, 1965: Prisoner's Dilemma. Ann Arbor: University of Michigan Press.

Scharpf, Fritz W., 1987: A Game-Theoretical Explanation of Inflation and Unemployment in Western Europe. In: Journal of Public Policy 7, 227-258.

Scharpf, Fritz W., 1988: The Joint Decision Trap: Lessons from German Federalism and European Integration. In: Public Administration 66, 239-278.

Scharpf, Fritz W., 1989: Decision Rules, Decision Styles, and Policy Choices. In: Journal of Theoretical Politics 1, 149-176.

Scharpf, Fritz W., 1989a: Der Bundesrat und die Kooperation auf der "dritten Ebene". MPIFG Discussion Paper 89/4. Köln: Max-Planck-Institut für Gesellschaftsforschung.

Scharpf, Fritz W., 1989b: Politische Steuerung und politische Institutionen. In: Politische Vierteljahresschrift 30: 10-21.

Schulz, Ulrich/ Theo May, 1989: The Recording of Social Orientations with Ranking and Pair Comparison Procedures. In: European Journal of Social Psychology 19, 41-59.

Selten, Reinhard, 1978: The Chain Store Paradox. In: Theory and Decision 9, 127-159.

Selten, Reinhard, 1982: Einführung in die Theorie der Spiele mit unvollständiger Information. In: Erich Streißler (ed.), Information in der Wirtschaft. Berlin: Duncker & Humblot, 81-147.

Selten, Reinhard, 1985: Comment. In: Kenneth J. Arrow/ Seppo Honkapohja (eds.), Frontiers of Economics. Oxford: Basil Blackwell, 77-87.

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

Shepsle, Kenneth A., 1989: Studying Institutions: Some Lessons from the Rational Choice Approach. In: Journal of Theoretical Politics 1, 131-147.

Shepsle, Kenneth A./ Barry R. Weingast, 1981: Political Preferences for the Pork Barrel: A Generalization. In: American Journal of Political Science 25, 96-111.

Shepsle, Kenneth A./ Barry R. Weingast, 1987: The Institutional Foundations of Committee Power. In: American Political Science Review 81, 85-104.

Simmel, Georg, 1984: Grundfragen der Soziologie, Individuum und Gesellschaft. 4th edition. Berlin: de Gruyter (first edition 1917).

Simon, Herbert A., 1978: Rationality as Process and as Product of Thought. In: American Economic Review 68, Papers and Proceedings, 1-16.

Simon, Herbert A., 1986: Rationality in Psychology and Economics. In: Journal of Business 59, S209-S224.

Snyder, Glenn H., 1971: "Prisoners' Dilemma" and "Chicken" Models in International Politics. In: International Studies Quarterly 15, 66-103.

Snyder, Glenn H./ Paul Diesing, 1977: Conflict among Nations. Bargaining, Decision Making, and System Structure in International Crises. Princeton: Princeton University Press.

Taylor, Michael, 1987: The Possibility of Cooperation. Cambridge: Cambridge University Press.

Trockel, Walter, 1986: The Chain-Store Paradox Revisited. In: Theory and Decision 21, 163-79.

Tsebelis, George, 1988a: Nested Games: Political Context, Political Institutions, and Rationality. Book Manuscript. Los Angeles: Department of Political Science, UCLA.

Tullock, Gordon, 1985: Adam Smith and the Prisoners' Dilemma. In: Quarterly Journal of Economics C, Supplement, 1073-1081.

Tversky, Amos/ Daniel Kahneman, 1986: Rational Choice and the Framing of Decisions. In: Journal of Business 59, S251-S278.

Ullmann-Margalit, Edna, 1977: The Evolution of Norms. Oxford: Clarendon Press.

Ullmann-Margalit, Edna, 1978: Invisible-Hand Explanations. In: Synthese 39: 263-291.

Ward, Hugh, 1987: The Risks of a Reputation for Toughness: Strategy in Public Goods Provision Problems Modelled by Chicken Supergames. In: British Journal of Political Science 17, 23-52.

Witt, Ulrich, 1986: Evolution and Stability of Cooperation without Enforceable Contracts. In: Kyklos 39, 245-266.

Zintl, Reinhard, 1989: Der Homo Oeconomicus: Ausnahmeerscheinung in jeder Situation oder Jedermann in Ausnahmesituationen? In: Analyse & Kritik, forthcoming.