

## Inequality in Collective Action Problems

*Katharina Holzinger*

### 1 Introduction

Rainer-Olaf Schultze was a second supervisor for my dissertation and he took this role very seriously. He showed me what to do and how, and more importantly: what not to do. I was very grateful for his advice, if not always in the very first moment, than surely later on. Looking back, I am still grateful because his level of engagement was not at all a matter of course. Although my ways led me elsewhere soon, geographically and academically, Rainer-Olaf further supported me in the early phases of my career and we never lost contact. There was one strong bond between us, namely our disagreement about the value of rational choice analysis. In what follows I do not strive to persuade Rainer-Olaf of the usefulness of this kind of reasoning; instead I want to demonstrate what I have learned at least: the rational actor is a fiction.

It was Mancur Olson's *The Logic of Collective Action* which made the notion of a collective action problem widely known in political and social science. Olson dealt with the provision of public goods by groups. He focused on the non-excludability property of public goods, and thus, with the term collective action problem he mainly referred to the free rider problem in the organization of interests (Olson 1965, chapter I.B). Thereafter, collective action problems have mostly been identified with social dilemmas, that is, situations where rational individual actors are trapped to produce collectively and individually sub-optimal results. For example in Elster's "strong" definition, a collective action problem is the same as the prisoners' dilemma (Elster 1985). Similarly, Hardin equates collective action problems with the prisoners' dilemma (1982: 25). Taylor and Sandler have shown that this it is surely too restrictive (Taylor 1987: 18; Sandler 1992, chapter 2). Taylor's own definition states "that a collective action problem exists where rational individual action can lead to a strictly Pareto-inferior outcome, that is, an outcome which is strictly less preferred by every individual than at least one other outcome" (Taylor, 1987, 19; cf. also Liebrand/Messick, 1996).

This definition, like most others, is only concerned with the welfare or Pareto-optimality aspects of collective action. Dilemmas between individual and collective rationality are, however, not the only problems to arise in collective

action. This becomes obvious if one looks at different sub-disciplines in social science, like sociology, positive political theory, or empirical political research. Many other kinds of problems have been identified by social scientists, which arise in collective action (see Sandler 1992; Scharpf 1997, chapter 4). To name a few: Pure conflict between the actors has been the subject of early game theory dealing with zero sum games (von Neumann/Morgenstern 1943). Problems of co-ordination have been distinguished from problems of co-operation (Runge 1984; Ochs 1995). Others have pointed to problems of finding agreement (Genschel/Plümper 1996) and to problems of inequality (Martin 1993; Zürn 1992), which can arise in the strategic interaction of individuals trying to achieve some collective outcome. Finally, the instability of collective outcomes has been identified as a problem (Ordeshook 1992, chapter 6; Morrow 1994: 87). Inequality, instability, co-ordination failure, and sub-optimal outcomes are widely discussed in political and social science. The authors quoted here have in common that they link these problems to game theory.

Against this background it seems justified to rely on a wider notion of collective action problems. In my definition the term “collective action” refers to the joint or independent actions of a number of individuals who have the potential to achieve and distribute some joint gain. The strategic constellations of actors can be such that they lead to difficulties in achieving these benefits. I call all problems arising in the pursuit of joint benefits which are a consequence of strategic interaction “collective action problems”. This definition includes inefficient outcomes, problems of co-ordination, inequality or instability. Closest to this definition is the term “problematic situation” (Zürn 1992: 154ff.).<sup>1</sup>

In the literature quoted above collective action problems are commonly represented by 2 x 2 matrix games. Matrix games have the merit of demonstrating a given strategic structure very clearly. They reveal a surprising amount of information on the problematic aspects of such a constellation. In fact, matrix games are the most parsimonious means of analyzing collective action problems.

In the following a typology of collective action problems is proposed which is built on a classification of matrix games. After a brief discussion of earlier taxonomies of matrix games (section 2), a classification is introduced, based on four attributes of games (section 3). Whereas three of these criteria are familiar, the fourth one is new in this context: it relates to the inequality of payoffs in the outcomes of a game. From this classification of games a typology of five distinct problems of collective action is derived (section 4). Finally, the value of intro-

---

<sup>1</sup> This definition does not include all problems of collective decision making. It excludes problems not being the result of strategic behavior, e.g., fundamental uncertainty about the consequences of alternative actions.

ducing the inequality criterion and its meaning for social science are discussed (section 5).

## 2 Taxonomies of matrix games

In their theoretical history, matrix games have been classified according to different attributes. In the beginning, the distinction between zero-sum and non-zero-sum games became important (von Neumann/Morgenstern 1943). Later the number of Nash equilibria (one, two or more) served as a criterion (Nash 1951). Also, basic types of collective action problem played a role from the very beginning: Games were classified as prisoners' dilemmas, chicken or battle of the sexes games (cf. Kreps 1990: 37; Taylor 1987), as volunteers' dilemmas (Rapoport 1988), or as co-ordination and dis-coordination games (cf. Rieck 1993: 35). However, attempts to produce systematic taxonomies have been rare. An exception are Rapoport/Guyer (1966) and Rapoport et al. (1976). They develop a taxonomy, which uses several solution concepts (dominant strategies, natural outcomes, Nash equilibria) and some factors that can make equilibria instable, such as threat-vulnerability. In this way they were able to rank matrix games according to the stability of their solutions.

The kind of distinctions that can be found in political science literature are usually not concerned with stability but with different types of collective action (for example Schelling 1960: 88; Taylor 1987: 34-60; Tsebelis 1997: 61-68; Scharpf, 1997: 69-79; Martin 1993). Prisoners' dilemmas, chicken, and assurance games are analyzed with respect to the social problems of co-ordination, co-operation or inequality (Ullmann-Margalit 1977), and they are used to model empirical problems of the respective types.

In political science, the most comprehensive attempt to classify 2 x 2 matrix games has been undertaken by Zürn (1992). Zürn distinguishes four types of games that represent what he calls "problematic situations": co-ordination games without distributional conflict, co-ordination games with distributional conflict, dilemma games, and rambo games. He uses three criteria for defining these types: the number of Nash equilibria, the Pareto-optimality of Nash equilibria, and the notion of a qualified Pareto-optimum, which serves to distinguish the rambo games (1992: 161-165). In introducing rambo games as a class, Zürn was the first one to point at inequality in the outcome as a problematic aspect in strategic interaction.

The classification presented here uses somewhat different criteria. In contrast to Rapoport et al. (1976) I am not only interested in stability but also in co-operation, coordination and inequality. In contrast to Zürn, I present a *complete*

classification of all  $2 \times 2$  games. Two aspects of Zürn's taxonomy are contested: I argue that pure conflict games also represent problematic situations and that games without a Nash equilibrium are a class in their own right, with specific problems.

### 3 A classification of strategic form games

The classification of matrix games developed here allows *each*  $2 \times 2$  game to be identified as belonging to a certain type of collective action problem. Payoffs are given in the usual general form; with two players and two strategies there are four different outcomes for each player. I use the letters a, b, c, and d to denote the payoffs. Using  $2 \times 2$  matrix games implies five more or less obvious assumptions:

1. there are only two strategies,  $s = 1, 2$
2. there are only two players,  $i = A, B$
3. the game is played "simultaneously", that is, there is imperfect information
4. there is complete information on payoffs and strategies
5. the game is played only once

In addition, I apply three assumptions that are not routinely made for matrix games. First, the payoffs are interpreted as observable material units, for example, units of some good, or as money. Second, I assume that identical letters imply identical payoffs for the players. Here I definitely leave what is common ground in game theory. The assumption implies that interpersonal comparison of payoffs is possible. The payoffs of the players are "tied to each other"; a transformation of player A's payoff has to be accompanied by a transformation of B's payoff if the strategic structure of the game is to be kept. This underlines that payoffs are not interpreted as personal utilities here. The assumption introduces comparability among the players into the matrix game. This runs *contra* the usual assumption in game theory textbooks; however, comparability is often implied in applications using monetary or material payoffs and it is also implied in experimental game theory. Third, I assume an ordinal scale in assumption (8).

6. payoffs  $\pi_i$  are interpreted as observable "material" payoffs, for example, units of some good or money,  $\pi_i = a_i, b_i, c_i, d_i$
7. identical letters imply identical payoffs for both players,  $\pi_A = \pi_B$
8. payoffs satisfy the condition  $a_i > b_i > c_i > d_i$

The aim of the classification is to distinguish all important types of collective action problems relevant for political science. For this purpose it is necessary to classify all strategically distinct  $2 \times 2$  matrix games according to certain properties, which reveal collective action problems. The number of different matrix games is infinite, if cardinal payoffs are used. This is true, even if the games are restricted to two players and two strategies. Therefore it is impossible to tell how many individual games belong to one class of games. How can one be sure to include all relevant classes of games if there are an infinite number of games?

Most importantly, the ordinal formulation of payoffs is sufficient to capture differences in the strategic structure, which is implied in my third additional assumption. Games with cardinal payoffs can be re-formulated into ordinal ones without any loss of information about the strategic structure.<sup>2</sup> In order to construct a complete classification it is thus sufficient to classify all games with ordinal payoffs that are strategically different. In a  $2 \times 2$  game with strict ordinal preferences the four outcomes are ranked by the players such that they satisfy the condition  $a_i > b_i > c_i > d_i$  (assumption 8).

Each player can place the letters into the four cells of the matrix in  $4! = 24$  ways. Since there are two players, there are a total of  $24 \times 24 = 576$  possible combinations of strictly ordinal payoffs. Not all of them represent different strategic structures, however. The game matrices are strategically equivalent whenever only the rows, the columns, both rows and columns, or, in symmetric games, the players, are interchanged. Taking this into account, only 78 strategically distinct games remain (Rapoport *et al.* 1976: 14-17). If indifference about the outcomes or strategies is permitted, however, the number of strategically different games rises to 732.

For the purpose pursued here it seems sufficient to classify the 78 strictly ordinal games and to treat them as representatives of all  $2 \times 2$  games. However, some interesting "borderline" games result if the players are allowed to be indifferent about the outcomes. Therefore I include some prominent games that are based on indifference, as well as some prominent others. I would like to emphasize that the classification provided is a classification for *all*  $2 \times 2$  games, cardinal and ordinal ones.

I analyzed these games with respect to a number of game theoretic properties: symmetry or asymmetry, the number of Nash equilibria, the number of dominant strategies, the number of Pareto optimal outcomes, Pareto-optimality at equilibria, conflict over outcomes, inequality of payoffs in Pareto-optimal outcomes and in equilibria, and the Kaldor criterion. The classification was devel-

---

2 This is true as long as we assume strictly rational players who care only about their own payoffs. If players care about differences in cardinal payoffs among the players, as I will assume later on, the proposition still holds as a result of assumption (7).

oped on the basis of the results of this tentative analysis. Only four of these formal properties have proven to be crucial for a distinction between basic collective action problems. These four properties of games and their equilibria are used firstly to classify matrix games and secondly to develop a typology of collective action problems. The properties can be easily identified for each game. The classification is empirically complete insofar as all 2 x 2 games can be subsumed under the disjunctive classes as defined by the following four criteria.

1) *Number of Nash equilibria in pure strategies (none, one, two or more)*. I have chosen to use the number of Nash equilibria in *pure* strategies as a criterion because it is the number of pure strategy equilibria that determines whether problems of co-ordination or disco-ordination arise. If I had chosen to count the number of pure and mixed strategy equilibria, the distinct classes created would have remained the same anyway. The only difference is that the categories of this factor would have to be re-labeled as *one mixed, one pure, three or more pure or mixed* strategy equilibria.

2) *Pareto-optimality of the equilibria (optimal, sub-optimal)*. The Pareto-optimality determines whether an outcome of the game entails conflict between individual and collective rationality. In cases of multiple pure strategy equilibria, it is distinguished whether all equilibria are Pareto-optimal or whether there are optimal and sub-optimal equilibria. This criterion cannot be applied to games that do not have an equilibrium in pure strategies.

3) *Conflict over the valuation of the outcomes (no, irrelevant, partial, pure)*. The conflict factor refers to the players' valuation of the four possible outcomes. *No conflict* means that the players rank all outcomes equally. Conflict over outcomes is *irrelevant* if both players rank the same outcome first, and different valuation occurs consequently only with respect to the second, third or fourth preference. *Pure conflict* means that the players' ranking of the outcomes is completely oppositional. There are no outcomes off the Pareto-frontier. The group of games that can be classified as pure conflict games, however, is not restricted to constant- or zero-sum games, as the cardinal formulation allows for non-linearity. For all other games, conflict is *partial*. Usually games in the last group are called "mixed motive" games.

4) *Equal payoffs in Pareto-optimal outcomes or equilibria (yes/yes, yes/no, no)*. The previous three factors constitute common ground in matrix game analysis. The strategic relevance of these attributes is obvious. The factor of inequality introduces a distributional dimension into the typology. Including inequality goes beyond standard rational choice theory, and thus has to be defended more extensively. In order not to interrupt the flux of the argument this will be done in section 5, in which the inequality property is discussed.

For the classification, I use the equality criterion in the following way: there is equality if both players rank an outcome the same; there is inequality if they rank an outcome differently. Since I am concerned with what makes a game problematic, I am interested in the equality or inequality of those outcomes that form an equilibrium or those that are Pareto-optimal. Thus, the inequality of non-equilibrium outcomes or of sub-optimal outcomes is of no relevance.

If at least one player has a strong preference for equal payoffs, a Pareto-optimal outcome may not be achieved, or an equilibrium may even be rejected by one of the players, because it is characterized by inequality. Games are not problematic with respect to equality if there is a Pareto-optimal equilibrium, which is associated with equal payoffs. They may become problematic, however, if the Pareto-optimal equilibrium is characterized by inequality, or if they have no equal Pareto-optimal outcomes at all (implying that there are no equal Pareto-optimal equilibria). Thus, two questions have to be asked:

1. Is there at least one Pareto-optimal outcome with equal payoffs (yes or no)?
2. If yes, is one of these outcomes a Nash equilibrium (yes or no)?

If the answer to the first question is no, there cannot be a Pareto-optimal and equal Nash equilibrium. The two questions are equivalent to applying the three criteria of Pareto-optimality, equality, and Nash equilibrium to a potential solution, or, respectively, to applying the two normative criteria Pareto-optimality and equality to the Nash equilibrium. The two questions create three groups of games:

- those with (at least one) Pareto-optimal and equal Nash equilibria (*yes/yes*); that is, both normative criteria are fulfilled in equilibrium;
- those with equal Pareto-optimal outcomes which are no equilibria<sup>3</sup> or those with Pareto-optimal equilibria characterized by inequality<sup>4</sup>, or those which meet both conditions (*yes/no*); that is, only one normative criterion is fulfilled in equilibrium;
- those without any equal Pareto-optimal outcomes – and thus also without equal Pareto-optimal Nash equilibria (*no*); that is, no normative criterion is fulfilled in equilibrium.

There is no distributional problem in the first group. The distributional problem is most severe in the third group, as there are simply no optimal and equal solutions. In the second group, at least optimal and equal outcomes exist, which the players could try to achieve, if they were able and willing to co-operate.

---

3 The only example is the standard prisoner's dilemma.

4 The only example is a chicken game, which has a third and equal Pareto optimal outcome.

The classification presented in table 1 is the product of the four factors discussed above. The complete product would include 72 cells. However, even now not all categories of each factor are applicable to all categories of all the other factors. For example, the two equilibrium-related criteria cannot be applied to games without equilibrium in pure strategies. Other sub-classes are empirically empty; for example, in “no conflict” or “pure conflict” games equality does not vary. Thus, several rows and columns can be omitted and finally only 30 cells remain. In this way, the classification becomes more informative and remains at least as fine.

In table 1, factors (1) and (2) are given in the vertical dimension and factors (3) and (4) in the horizontal dimension. The table shows first how the 78 games with strict preferences are distributed over the 30 cells of the classification. The numbers given in the table refer to the number of strategically different games in each class. Some prominent examples of each important class of games are mentioned in the cells in quotation marks. Second, four additional games are given in brackets that are based on indifference to outcomes or strategies. These less well-known games are the “pure co-ordination” game, the “degenerate co-ordination” game, “triangle co-ordination”, and “matching pennies”. Finally, a zero sum and a constant sum game are also included.

In the following section, table 1 will be interpreted and the typology of the collective action problems involved will be derived. Although the classification of matrix games along the four criteria may look complicated at first sight, it yields only five types of collective actions problems, as we will see below. Like the taxonomy provided by Rapoport/Guyer/Gordon (1976), it is a comprehensive classification of all games, however, a much less complex one. Like the taxonomy of Zürn (1992), it ends up with only a small number of empirically relevant problem types; however, it is based on a systematic and comprehensive analysis of all possible matrix game structures.

Nash equilibria in pure strategies: Number	Conflict		No Yes/Yes 3 3	Irrelevant Yes/Yes 18 18	Yes/Yes 7	Partial (mixed motive games)		Pure 3 No 2
	Equality	Pareto-optimality				Yes/No 7	No 40	
			1	2	3	4	5	6
Unique 57	Pareto-optimal 53	1	2 "pure harmony"	13 "impure harmony"	7 "weak harmony"	3 "Rambo"	26 "Rambo"	"pure conflict" ("zero/ constant sum)
	Sub-optimal 4	2		HARMONY		MERE DISTRIBUTION PROBLEMS		
	Pareto-optimal 6	3	("pure coordination")	("triangle coordination")		1 "prisoners' dilemma"	3 asymmetric dilemmas	
Multiple 12			CO-ORDINATION PROBLEMS	LEMS		1 "chicken"	5 "battle of the sexes"	
	Both optimal and sub-optimal 6	4	1	5 "assurance" ("degenerate coordination")		DISAGREEMENT PROBLEMS		
None 9	(not applicable)	5				2	6	1 constant sum games ("matching pennies")
						INSTABILITY PROBLEMS		

#### 4 Typology of collective action problems

The classification of matrix games allows distinguishing basic types of collective action problem. Each type is defined by a certain combination of the four matrix game properties analyzed. There are only six distinct groups of games which are interesting for collective action and only five of them pose a collective action problem. First, this is due to the fact that some of the cells are still empty (cf. table 1). Second, not all factors produce crucial distinctions in each of the sub-classes constituted by the other factors. Thus, several cells can be subsumed under one type of collective action problem.

The types are marked in table 1 by bold frames. They have been labeled in accord with the dominant problem of a certain class of games. Sometimes classes of games combine several elementary aspects of collective action problems – for example a co-ordination plus a distribution problem – and thus form a new characteristic type. In the following, definitions and descriptions of the six groups of games are given. The number of games belonging to each type and sub-type, is given in brackets.

1) *Harmony*. The first group to be distinguished are harmony games. They have a unique Nash equilibrium and thus a stable solution, which is Pareto-optimal and maximizes joint gains. There is no conflict concerning the valuation of outcomes (or conflict is irrelevant), and there is no inequality in the players' payoffs in the equilibrium. From a collective action perspective, the distinction between "pure" harmony games (2), where all outcomes are ranked equally by both players, and "impure" harmony games (13), where outcomes off the equilibrium are ranked differently, is irrelevant. "Weak" harmony games (7) involve partial conflict and have two Pareto-optimal outcomes. However, these games are still harmony games: that is, there is no inequality in the Pareto-optimal equilibrium and no other problem. This group of games does not constitute a collective action problem.

2) *Mere distribution problems*. On the right hand side of the harmony games there is a group of games that have a unique and Pareto-optimal Nash equilibrium. They differ from harmony games in that they possess a second Pareto-optimal outcome which is not an equilibrium, whereas the equilibrium is characterized by inequality. There is partial conflict over the valuation of outcomes. In the first group there are three games, in which there is inequality in the Pareto-optimal Nash equilibrium but equality in another Pareto-optimal outcome. In the second and larger group (26) each of the two Pareto-optimal outcomes (equilibrium or not) is characterized by inequality. Both groups pose a distributional problem. It is *merely* distributional since, from the perspective of co-ordination, stability, and welfare effects, there are no problems with these games.

They have been labeled “rambo” games by Zürn (1992: 209ff.) and suasion games by Martin (1993). According to Martin (1993: 104) “suasion problems have equilibrium outcomes that leave one actor dissatisfied”. So far, rambo games have not received much attention in the literature. There are two reasons why this presents a significant deficit in the political science literature: First, from a normative perspective, the inequality may be judged to be unfair. Second, from a positive perspective, there may be an “irrational” incentive to defect, for example driven by the players’ considerations of fairness, or an incentive to re-negotiate if communication is possible. Rambo games are probably very common and should not be neglected.

Further to the right in table 1 there are pure conflict games. In pure strategies they have just one equilibrium, which, as all other outcomes, is by definition Pareto-optimal. In strictly ordinal terms there are two of them. Zero-sum games are well-known examples. They pose distributional problems because their equilibria inevitably entail inequality between the players. Thus, these games belong to the mere distributional problems. Pure conflict games have traditionally been considered a very special type of game. From the perspective of the four problematic aspects in collective action, however, pure conflict games are not much different from those rambo games, which possess no equal outcomes at all. The only remaining difference is that rambo games have sub-optimal outcomes, while this is not the case in pure conflict games.

3) *Defection problems (dilemmas)*. The class of games below the rambo games entail a problem of defection (4). They have a unique Nash equilibrium. However, this is a sub-optimal equilibrium. There is partial conflict, and equilibria may result in either equality or inequality of players’ payoffs. The collective action problem associated with these games is a conflict between individual and collective rationality. The Pareto-optimal outcome is not achieved by the players, as there is an individual incentive to deviate should the other player conform. The most well-known example is the prisoners’ dilemma. Both the Pareto-optimal outcome and the Nash equilibrium are equality outcomes in the prisoners’ dilemma.

However, there are also asymmetric dilemmas (3) in which the Pareto-optimal (non-equilibrium) outcome is characterized by inequality. The distributional problems that go along with these games make the situation worse: Even if the players could conclude a contract and enforce it, it would be difficult for them to agree to the Pareto-optimal outcome because that outcome is associated with inequality. Moreover, the Nash equilibria in these games may be equality outcomes. If the actors strongly value equality, they have one more reason to choose the sub-optimal equilibrium. While in the symmetric prisoners’ dilemma

the equal Pareto-optimal outcome is a “focal point” for agreement, there is no such “focal point” in asymmetric dilemmas.

4) *Co-ordination problems*. Below the harmony games there is a class of games that pose a co-ordination problem. The reason for this is the fact that these games have multiple equilibria in pure strategies. Some of them are no conflict games (1), and some are irrelevant conflict games (5). Both groups may have two Pareto-optimal equilibria. The “impure” co-ordination games have an optimal and a sub-optimal equilibrium, such as the assurance game or the “degenerate co-ordination” game. There is no inequality in their equilibria, and they maximize joint gains in the optimal equilibria. The main problem for the players is to co-ordinate their strategies such that at least one of the equilibria, and hopefully the optimal equilibrium, will result.

5) *Disagreement problems*. The next class, to the right of the co-ordination games and below the defection games, combines several problems. Games belonging to this class have two equilibria in pure strategies, and both are Pareto-optimal. However, both entail inequality – one disadvantages Player 1 and the other disadvantages Player 2. Thus players disagree about which equilibrium is to be selected. Examples are the battle of the sexes or the chicken game. Usually, these games are treated under the heading of co-ordination games. However, they have to be treated as a type of collective action problem in its own right, because of the distributional problem they imply.

In its standard ordinal formulation the chicken game does not only have two Pareto-optimal Nash equilibria – both of which are inequality outcomes – it also has a third Pareto-optimal outcome, which is no equilibrium but shows equality. The other five games in this class, including the battle of the sexes, possess only two Pareto-optimal outcomes. Thus, as with the rambo and the defection games, there are two sub-groups: the first group has a Pareto-optimal outcome characterized by equality, while the second group has no such outcome.

There is no problem with collective welfare in these games, but there is both a co-ordination and a distribution problem. Given this combination, the players have difficulties finding an agreement. Both players prefer a different equilibrium, and there is thus a good chance that they will miss the equilibria. This is even true if both players act altruistically and pursue the other player’s preferred equilibrium. If the players can communicate and thus eliminate the co-ordination problem, they will still have difficulties agreeing on one of the equilibria.

6) *Instability problems*. The last group is formed by the so-called dis-coordination games (Rieck 1993: 53). These games have no pure strategy Nash equilibria. Therefore the criterion of Pareto-optimality of the Nash equilibrium cannot be applied. The main problem of these games is instability. In these games, rational players generally want to prevent their strategies from intersecting – they

dis-coordinate strategies. There is always a unilateral incentive to deviate from Pareto-optimal or joint gain-maximizing outcomes. Most of these games are partial conflict games (8); however, they may appear in pure conflict situations as well (1).

Instability games imply a distributional problem. Some of them possess an equal Pareto-optimal outcome (2), which could serve as a “focal point” in a cooperative solution. Most others (7) do not, however. A classical example that allows for indifference is “matching pennies”. Although instability games have been discussed in game theory textbooks (Morrow 1994: 87; Ordeshook 1992, chapter 6), they have not found much attention in the wider political science literature.

## 5 Discussion of Inequality

In this section I justify why I introduced the inequality criterion into the classification and discuss the salience of the distributional aspect in collective action problems. The inequality factor is interesting from a political science perspective for two reasons. First, equality (and in the absence of context information equality may serve as a proxy for justice) may be used as an outside normative standard for judging the outcomes of social interactions – similar to Pareto-optimality. Second, and more importantly, the distributional inequality of the outcome of a game may produce incentives to negotiate or to defect, even if the solution is rational, i.e., an equilibrium.

An every day example of equilibrium rejection is the following: Couples living together in an apartment often have a problem of cleaning if they have different preferences with respect to how clean and orderly the apartment should be. Player 1, having the stronger preference for a clean apartment will eventually do the cleaning, although this is not his or her first preference, which would be that the partner does the cleaning. Player 2 gets his first preference of having a clean apartment and not doing the cleaning, just because his preference for a clean apartment is not that strong. This constitutes an equilibrium within a rambo game. Even if it is rational for player 1 to do the cleaning, he or she will be dissatisfied with this situation and might reject the equilibrium, thus sanctioning both with the sub-optimal situation of a disorderly apartment.

The anecdote corresponds to systematic empirical evidence. Perfectly rational equilibria are often rejected if inequality is high, whatever the exact motives are. This has convincingly been shown by experimental research on “ultimatum bargaining” (Güth/Schmittberger/Schwarze 1982; Falk/Fehr/Fischbacher, 2003; Güth *et al.* 2006). There is now a huge body of research in behavioral

economics showing that individuals care about fairness (Falk/Fehr/Fischbacher, 2008; Güth/Berninghaus/Kirstein 2006), equity and reciprocity (Falk/Fischbacher 2002), that they exhibit altruism (Fehr/Fischbacher 2003; Fehr et al. 2003) and inequality aversion (Güth/Huck/Müller 2001; Fehr/Naef/Schmidt 2006). Another strand of behavioral research shows that there are also motivations such as envy, revenge and punishing (Fehr/Rockenbach 2002; Fehr/Gaechter 2002) or competitive orientation. For example, higher income is often not valued as such but because it is higher than the income of someone else (Fehr/Götte 2008). In this whole body of research monetary payoffs are used and are the basis of equilibrium rejection and of outcome valuation by the players.

It is now common ground to accept that considerations of fairness or sensitivity to justice play a role in the selection of social outcomes. This became obvious in a recent discussion on Binmore's *Natural Justice* (Binmore 2006). The conflict is about whether it is necessary to revert to "other-regarding preferences" in explaining morality, or if modeling norms is sufficient (Binmore 2006; Ross 2006). For Binmore morality serves to coordinate behavior among possible equilibria which are based on "self-regarding preferences". The behavioral literature quoted above, however, shows that people "irrationally" defect from equilibria, that is, they choose an outcome not maximizing their payoff in terms of "self-regarding preferences" (Fehr 2005; Seabright/Gintis 2006). This may be a consequence of considerations of justice, or a consequence of envy, or some other kind of other-regarding preferences.

For the argument presented here, it is important that the distributional dimension and thus equality or inequality of payoffs play a role not only in equilibrium selection but also in equilibrium rejection. Both represent real world problems in collective action and I therefore take inequality into account as a criterion for classifying collective action problems.

The rejection of equilibria by players is usually based on the inequality of cardinal monetary payoffs. In ordinal games, however, outcomes that have the same rank in the preference order of both players, may well be associated with very different cardinal payoffs. If, in a battle of the sexes game, for example, the female gets her preferred equilibrium, it may well be that the male's overall cardinal utility is higher than the female's, as he values both spending an evening together and the female's preferred entertainment more than she herself does. So he should not be dissatisfied with the situation.<sup>1</sup> Assumptions (7) and (8) serve to avoid this problem. Payoffs having the same rank in the preference order are

---

<sup>1</sup> On the other hand, it is reasonable to assume that players also judge an outcome as unfair in which player 1 gets her first preference whereas player 2 gets only his second preference. Nida Rümelin (1991) uses this interpretation, for example, when he compares the outcomes of matrix games with ordinal payoffs in respect to justice.

equal in cardinal terms. The implication of this assumption is that we look only at those games in which equality or inequality of the numbers representing the payoffs have in fact a meaning for the players. Payoffs will actually be perceived as comparable (equal or unequal) by the players. The analysis is thus limited to situations of comparability of payoffs.

This is not to deny the impossibility of interpersonal utility comparison. It is surely true that individuals may derive different utility from the same amount of money (the beggar and the millionaire) or the consumption of the same amount of a certain good. Personal utility, however, is neither observable nor directly measurable and therefore also not inter-individually comparable. Other-regarding preferences, be it a sense of fairness or justice, altruism, competitive interaction orientation, or envy, seem to relate to observable, measurable and comparable entities, rather than to non-observable utilities. It may well be that the same absolute amount of money is valued differently by two players in terms of the personal utility they derive from it; but that they are within a social interaction eager to achieve more of the material payoff than their counterpart (or: not less, or: not more, ...). That is, it seems to be material, observable payoffs and not only personal utility, as well as the comparison of the own payoff with the other's and not only the own payoff, which motivate social interactions.

In my view it makes sense to look at collective action from this perspective and therefore to include comparability into the models to be able to judge equality. Comparison to others seems to play a great role in social interaction, as the empirical research has shown. Depending on context and personal orientations, strictly rational solutions are not always preferred: People sometimes prefer outcomes that give them individually more than another outcome over Pareto-optimal and over equal outcomes (strict rationality); sometimes they prefer equality outcomes over those that give them individually more (fairness?); sometimes they avoid outcomes that give them less than the other player (envy?); and sometimes they particularly enjoy outcomes that give them more than the other player (competitive orientation).

A conceptualization of payoffs as personal utilities misses these interesting phenomena. It is, of course, still valuable for the analysis of the "purely rational" collective action problems: defection, coordination failure and instability. However, it seems also interesting to systematically distinguish different elements in the utility function: the utility derived from payoffs and the utility derived from (in)equality in payoffs. This differentiation enables us to focus on the distributional aspect in collective action.

Turning to the frequency of inequality in collective action I shall first look at the distribution of the various problematic aspects throughout the classification. The distribution of the 78 strictly ordinal games shows that the

number of strategically different games in the problematic classes is significantly higher than the number of games in the harmonic class. There are altogether 22 pure, impure, and weak harmony games. There are three games that pose only moderate problems of inequality, and 26 more severe rambo games, four defection games, six co-ordination games, six disagreement games, eight dis-coordination games, and three games of pure conflict. Thus, 56 of 78 strategically different games cause some sort of collective action problem. However, only four different dilemmas have been found. Thus, in the strictest definition of the term, there would be not many “collective action problems”.

The inequality aspect is present in games of pure conflict, rambo games, disagreement games, dis-coordination games and asymmetric dilemmas, that is, in 49 out of 78 games. Whereas in 7 of these strategic structures an equal Pareto-optimal outcome does exist (and could thus be achieved by cooperative behavior or in the context of cooperative institutions), in 42 constellations there is no outcome that implies equality of payoffs. In sum, in almost two thirds of the classified strategically distinct  $2 \times 2$  games inequality is present in equilibrium payoffs and there is thus a risk of equilibrium rejection.

This result must be qualified, however. The empirical relevance need not be correlated with the theoretical distribution of strategically different game models over the types of collective action problem. As of yet, not much can be said about the empirical frequency of the collective action problems. It is true that representatives of most classes of collective action problems seem to have been frequently analyzed and applied – with the exception of rambo and instability games. The fact, however, that the prisoners’ dilemma has received a great deal of attention in social science literature, while rambo games have not, does not make it possible to draw conclusions about the empirical relevance of the two games. This fact is probably a mere result of researchers’ perceptions and interests, and of the greater logical attractiveness of the prisoners’ dilemma. More-over, systematic information on the empirical frequency of the different games is difficult to acquire. Real world situations usually exhibit several problematic aspects and it depends on the analyst’s interest how to exactly model the situation. Games are models of the world and cannot be found “as such” in the real world. Therefore, the universe of strategic constellations appearing in the empirical world is unknown.

It is important to note, however, that the typology itself is of high empirical relevance: All kinds of real world collective action problems can be analyzed in terms of the six problem types. The types make a difference not only for finding solutions but also for the explanation of empirical attempts to resolve those problems. This has been demonstrated by, for example, Genschel/Plümper (1996), Martin (1993), Runge (1984), Ostrom (1990), Sandler (1992, 1997),

Scharpf (1997), Holzinger (2001, 2003, 2005) and Zürn (1992) who apply this kind of analysis to many political problems.

The analysis presented here shows that matrix games can also be used to analyze problems of inequality and that a great deal of the conceivable strategic constellations exhibit the inequality property in the rational outcome. The problems of collective action seem to become more severe from top to bottom and from left to right in table 1. On the left hand side of the table and on the top, there is harmony. Closer to the bottom are the co-ordination problems which are in general easier to solve than the other four types. It is obvious that the most problematic games are to be found in the right hand side of the table. It is the dimension of conflict over outcomes and the dimension of distributional inequality of outcomes that makes collective action problematic.

## 6 Summary

This paper has argued that the distributional aspect of collective action can be made visible and accessible for strategic analysis by interpreting payoffs not as personal utilities but as *material and comparable* payoffs. A typology of collective action problems was developed on the basis of a classification of 2 x 2 matrix games according to four formal properties of the games. Five types of collective action problem were distinguished, which represent different combinations of four problematic aspects in collective action: problems of efficiency, of co-ordination, of distribution and of stability.

Almost two thirds of the strategically distinct structures classified in this paper exhibit inequality in equilibrium as a problematic aspect. We do not know how this translates into empirical appearances of problematic aspects in collective action; but everyday life and political experience suggests that distributional problems determine social and political action as much as defection and coordination problems and thus deserve attention and strategic analysis.

## Literaturverzeichnis

- Bates, Robert (1988): *Contra Contractarianism. Some Reflections on the New Institutionalism*. *Politics and Society*. 16. 387 401.
- Binmore, Ken (2006): *Why do people cooperate?* In: *Politics, Philosophy & Economics*. 5. 81 96.
- Elster, Jon (1985): *Rationality, Morality, and Collective Action*. In: *Ethics*. 96, 136 55.
- Falk, Armin/Fischbacher, Urs (2002): *The Economics of Reciprocity Theory and Evidence*. In: *Freeman, Richard (2002): 207 233.*
- /Fehr, Ernst/Fischbacher, Urs (2003): *Reasons for Conflict Lessons From Bargaining Experiments*. In: *Journal of Institutional and Theoretical Economics*. 159. 171 187.
- / Fehr, Ernst/Fischbacher, Urs (2008): *Testing Theories of Fairness Intentions Matter*. In: *Games and Economic Behavior*. 62. 287 303.
- Fehr, Ernst/ Schmidt, Klaus M. (1999): *A Theory of Fairness, Competition, and Cooperation*. In: *Quarterly Journal of Economics*. 114. 817 68.
- /Götte, Lorenz (2008): *Do Workers work More if Wages Are High? Evidence from a Randomized Field Experiment*. In: *American Economic Review*. 97. 298 317.
- /Fischbacher, Urs (2003): *The Nature of Human Altruism*. In: *Nature*. 425. 23 October 2003. 785 791.
- /Rockenbach, Bettina (2002): *Detrimental effects of sanctions on human altruism*. In: *Nature*. 422. 15 March 2002. 137 140.
- /Gächter, Simon (2002): *Altruistic Punishment in Humans*. In: *Nature*. 415. 10 January 2002. 137 140.
- /Gintis, Herbert/Bowles, Samuel/Boyd, Robert (2003): *Explaining altruistic behavior in humans*. In: *Evolution and Human Behavior*. 24. 153 172.
- /Naef, Michael/Schmidt, Klaus M. (2006): *Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments, Comment*. In: *American Economic Review*. 96. 1912 1917.
- Freeman, Richard (ed.) (2002): *Inequality Around the World*. New York: Palgrave MacMillan: 207 233.
- Fudenberg, Drew/Maskin, Eric (1986): *The Folk Theorem in Repeated Games with Discounting or with Incomplete Information*. In: *Econometrica* 54. 533 54.
- Genschel, Philipp/ Plümper, Thomas (1996): *Kommunikation und Kooperation in der internationalen Bankenregulierung*. In: *Zeitschrift für Internationale Beziehungen*. 3. 225 53.
- Gibbons, Robert (1992): *Game Theory for Applied Economists*. Princeton N.J.: Princeton University Press.
- Güth, Werner/Huck, Steffen/Müller, Wieland (2001): *The Relevance of Equal Splits in Ultimatum Games*. In: *Games and Economic Behavior*. 37. 161 169.
- /Schmittberger, Rolf/Schwarze, Bernd (1982): *An Experimental Analysis of Ultimatum Bargaining*. In: *Journal of Economic Behavior and Organization*. 3. 367 88.

- /Berninghaus, Siegfried/Kirstein, Annette (2008): Trading Goods versus Sharing Money An Experiment Testing Whether Fairness and Efficiency are Frame Dependent. In: *Journal of Neuroscience, Psychology, and Economics Research*. 33 48.
- /Fischer, Sven/ Stiehler, Andreas/Müller, Wieland (2006): From Ultimatum to Nash Bargaining, Theory and Experimental Evidence. In: *Experimental Economics*. 9. 17 33.
- Hardin, Russell (1982): *Collective Action*. Baltimore: Johns Hopkins University Press.
- Holzinger, Katharina (2001): Aggregation Technology of Common Goods and its Strategic Consequences. *Global Warming, Biodiversity and Siting Conflicts*. In: *European Journal of Political Research*. 40. 117 138.
- (2003): Common Goods, Matrix Games, and Institutional Solutions. In: *European Journal of International Relations*. 9. 173 212.
- 2005: Tax Competition and Tax Co operation in the EU. The Case of Savings Taxation. In: *Rationality and Society*. 17. 475 510.
- Kagel, John H./Roth, Alvin E. (eds.) (1995): *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press.
- Kreps, David M./Milgrom, Paul/Roberts, John/Wilson, Robert (1982): Rational Co operation in the Finitely Repeated Prisoners' Dilemma. In: *Journal of Economic Theory*. 27, 245 53.
- Kreps, David M. (1990): *Game Theory and Economic Modelling*. Oxford: Clarendon Press.
- Liebrand, Wim B.G. (1983): A Classification of Social Dilemma Games. In: *Simulation and Games*. 14. 123 38.
- / Messick, David M. (1996): *Frontiers in Social Dilemmas Research*, New York: Springer.
- Martin, Lisa L. 1993: The Rational State Choice of Multilateralism. In: Ruggie, John Gerard (1993): 91 122.
- Morrow, James D. (1994): *Game Theory for Political Scientists*. Princeton, N.J.: Princeton University Press.
- Nash, John (1951): Non Co operative Games. In: *Annals of Mathematics*. 54. 286 295.
- Neumann, John von/Morgenstern, Oskar (1943): *Theory of Games and Economic Behavior*, Princeton, N.J.: Princeton University Press.
- Nida Rümelin, Julian (1991): Practical Reason or Metapreferences? An Undogmatic Defense of Kantian Morality. In: *Theory and Decision*. 30, 133 62.
- Ochs, Jack 1995: Coordination Problems. In: Kagel, John H./Roth, Alvin E. (1995): 111 94.
- Olson, Mancur (1965): *The Logic of Collective Action, Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- Ordeshook, Peter C. (1992): *A Political Theory Primer*, New York London: Routledge.
- Ostrom, Elinor (1990): *Governing the Commons. The Evolution of Institutions for Collective Action*, New York: Cambridge University Press.
- Rapoport, Anatol (1988): Experiments with N Person Social Traps, I. Prisoner's Dilemma, Weak Prisoner's Dilemma, Volunteer's Dilemma and Largest Number. In: *Journal of Conflict Resolution*. 32. 457 72.