Game theory and power theory: a critical comparison

Balzer, Wolfgang

Rethinking Power

EDITED BY

Thomas E. Wartenberg

STATE UNIVERSITY OF NEW YORK PRESS

SUNY Series in Radical Social and Political Theory
Roger S. Gottlieb, editor

and

SUNY Series in Feminist Political Theory
Marion Smiley, editor

Chapter 4

Game Theory and Power Theory:
A Critical Comparison

———————————— *Wolfgang Balzer*

The last decade has witnessed the application of game theory to social science. Created by abstracting from social games like chess and bridge, von Neumann's model was seen to be applicable to more "serious" social phenomena as well: to situations with two or more agents with conflicting goals and with the ability to influence each other. Starting essentially from prisoner's dilemma (see below), research focused on models capturing "socially rational" behavior. After Michael Taylor's demonstration that such models were possible in terms of super-games,' a veritable explosion took place in sociological research using game-theoretic models.[2] By now it seems fair to say that those branches of social science that deal with interaction and conflict in exact ways are dominated by the game-theoretic approach.

On the other hand, studies of power continued to develop with the same slow pace they had for centuries, articulated mainly by prac-titioners of politics and by philosophers, and more recently by more professional scientists.[3] In contrast to the flourishing of game theory, however, no comparable evolution took place in models of power. Only recently, an account by Thomas Wartenberg opened a broader perspective that may be said to combine more operational approaches (like Dahl's) and "internal" conceptual analysis with comprehensive views of social reality.[4] Wartenberg's account can be shown to com-bine most other approaches to power and to provide a frame in which all other special forms of power can be defined.[5] In addition, it is presented in a conceptually clear, analytic form. For these reasons I will take his approach as *the* representative of power theory.

My aim in this paper is to analyze and compare the two approaches—the game-theoretic and that of power—from the perspective of applications in social science. The result of comparison will be stated in the form of a somewhat provocative thesis of incommensurability, and the analyses performed will be used to justify that thesis.

My account of both approaches will concentrate on the *basic* models (or most general theoretical assumptions) used on each side. It is not possible to consider all the specializations that have developed. This will not affect my results, however, because the two approaches may be conceptualized as *theory-nets* in which more special models, assumptions, or "pictures" are all built on one fundamental, basic model such that most questions of comparison between the full theory-nets in a precise sense reduce to questions of how the basic models are related.[6]

Game Theory

The classical social application of game theory is the prisoner's dilemma. Two individuals, p, and p, are arrested, say for the illegal possession of guns. The district attorney believes that they robbed a bank the other night, but he has no evidence. Both are put into jail in isolation and accused of bank robbery. The district attorney proposes the following deal to each p_i individually. If p_i confesses the bank robbery and p_j does not confess, p_i will get the preferred status of a chief witness and go to jail for one week. If p_j does confess, p_i will go to jail for five years; if neither confesses, they will both go to jail for one month for possession of a gun. Thus, each prisoner p_i is faced with the following dilemma. If his friend (p_j, $j \neq i$) confesses, p_i is also inclined to confess, otherwise he alone will get five years while his friend gets by with one week. If the friend does not confess, then p_i is still more inclined to confess for this will get him the minimal punishment. So whatever his friend may do, p, is inclined to confess. The situation is completely symmetric, so p_j has the same inclinations. If both of them follow this line of reasoning dictated by individual rationality in the state of isolation, both of them will confess. This, however, leads both of them to the *worst* result they can achieve in that situation. So the dilemma is that applying the principle of individual rationality obtains the worst result, while a better result can only be obtained at the cost of "irrational" behavior.

In order to analyze this situation, game-theoretic apparatus uses the notions of *individuals,* of *strategies* or *alternatives,* and of *preference* or *utility.* We will not deal here with iteration of moves or with "mixed strategies." For each individual p, there is a set of alternatives (moves, actions, courses of actions, sequences of moves) comprising exactly those that are entertainable by p_i in the situation to be modeled. It is not necessary to spell out precisely what is meant by entertainable here. Usually, the set-is larger than the set of alternatives actually entertained, but smaller than the set of physically possible ones. It is usually restricted by the social and institutional frame in which the individual finds himself.[7] In the example above, p, and p, both have two alternatives: to confess or not to confess.

Each choice of an alternative by actor p_i amounts to some action of p_i that leads to subsequent events or states causally determined by this action. Out of the sequence of subsequent events, one distinguished event or state is picked out and made the subject of an evaluation by actor p_i; we call it the *resulting state.* So each choice of an alternative by p, leads to some resulting state that is then evaluated by p_i; p_i reflects, or is assumed to know anyway, how much the resulting state is worth to her. In other words, there is a utility function for each individual p_i that assigns some value to each state resulting from p_i's choice of any alternative and the corresponding action.

The crucial point is that p_i's choice and action do not *fully* cause the resulting state; p_i's action is only a partial cause, that is, one event among others, which only *together* cause the resulting state. Because of the interactive nature of the situation, different moves of other agents will lead to different resulting states for a given individual p_i even though p_i sticks to his choice. In other words, one alternative chosen by p_i will lead to different resulting states, depending on which alternatives are chosen by the other agents.

The example cited makes this very clear. If p, chooses the alternative to confess, the resulting state that he evaluates depends on the choice of p. If p, also chooses to confess, then the state resulting for p, from his choice is five years in jail. But if p, chooses not to confess, the state resulting for p, from p_1's action (to confess) is very different, namely one week in jail. The same holds for p,, of course.

As the resulting state is (assumed to be) uniquely determined by the sum of all the individual's actions, game theorists assign utilities to the sums of such actions rather than to the resulting states. Once each individual has made his choice and taken the corresponding action, the resulting state will evolve "automatically" (causally). If there

are n ($n \geq 2$) individuals, the "sum" or combination of their choices is represented by an n-tuple, $\langle a_1, \ldots, a_n \rangle$, each a_i representing the choice of individual p_i. Formally, then, the arguments of the utility functions are taken from the space of all such n-tuples. For each individual p_i and each combination $\langle a_1, \ldots, a_n \rangle$ of choices there is a utility value $U_i(a, \ldots, a_n)$ expressing the utility that individual p, has from the state resulting for him from combination $\langle a_1, \ldots, a_n \rangle$. In the cited example, we may take the lengths of the periods in jail as indicators of (inverse) utility values: the longer the period in jail, the smaller the utility of that resulting state. In general, however, the utilities of one common resulting state need not be the same for different individuals.

In summary, a basic model of game theory is made up of the following items:

1. a set of individuals $p, \ldots, p_n$
2. n corresponding sets $A, \ldots, A_n$ of alternatives, one for each individual
3. n utility functions $U, \ldots, U_n$, which are all defined on the set of all n-tuples $\langle a_1, \ldots, a_n \rangle$ of elements of $A, \ldots, A$, and which all take real numbers as values.

This material has to have a special form in order to pass muster as a proper model. In game theory the basic assumption of form is that the individuals choose rationally, in some sense. There is no unique explication of what a rational choice for p_i is in a frame given by items 1–3. The general idea is that the individuals choose such that the resulting states (or rather the corresponding n-tuples $\langle a_1, \ldots, a_n \rangle$ of actions, choices, or alternatives, as just explained) have some kind of "equilibirum property." The notion of equilibrium, in turn, may be defined along different lines, which in general are not equivalent. The most commonly used notion is that of Nash equilibrium. A combination of choices $\langle a_1, \ldots, a_n \rangle$ is a point of Nash equilibrium if and only if for any $i \leq n$, any deviation $\langle a_1, \ldots, a_i^*, \ldots, a_n \rangle$ with $a_i^* \neq a_i$ would result in some decrease of utility for p,:

$$U_i(a, \ldots, a_i^*, \ldots, a_n) < U,(a, \ldots, an)$$

One way to formulate the axiom of rationality is to say that each individual should choose, or will choose, an alternative that belongs to some point of Nash equilibrium, provided the game has such a point. In order to turn this into a more descriptive form we might add

to the model, for each individual p_i, one alternative $a_i o \in A_i$ which is interpreted as that alternative which p_i actually chooses when the game is played. Collecting all these distinguished alternatives we obtain an n-tuple $\langle a_1 o, \ldots, a_n o \rangle$, and the axion of rationality basic in game theory says that this n-tuple is a point of Nash equilibrium. It should be noted that game theorists usually do not make explicit this move from the rule of behavior to a descriptive formulation.

Instead of going into details of refinements, we have to concentrate on a more fundamental feature inherent explicitly or implicitly in the basic models described above, namely the status of the utility functions. Utilities are difficult to determine. It seems fair to say that up to now there is not a single method for a practical determination of utilities in "real-life" situations. There are some experimental determinations in laboratory situations, but it is not clear in what way these can be made to function in real-life situations without substantially changing such situations.[8] Moreover, it is common sense that utilities are not stable over time: people may change their taste, and thus their utility.

In game theory utilities are assumed to be given and stable. In the light of the previous considerations these assumptions are rather strong and unrealistic. We may rephrase the assumptions in terms of the notion of the rules of the game. Each game essentially is given by its rules; the only possible variations of a game are who participates and how many participate. In the conceptual frame of a simple game as described above there are two components that together make up the rules of such a game.

Let us look at these components first in the special case of social games. Here, we first have the different alternatives open to each individual as determined by the rules of the game. If a player had some alternative not allowed by the rules, we would not say that she plays according to these rules or that she plays the game given by these rules. Second, there are the utility functions telling how much the players get out of the game. In the case of social games, game theorists often speak of *payoff functions* instead of utilities. Each player's payoff is determined by the rules of the game, so the payoff functions are part of, or constitutive of, the rules of the game. Since there are no other components in the basic models, we may say that the sets of alternatives together with the utility (or payoff) functions in social games make up the rules of the game. This analysis is by no means forced. It makes precise our intuitive notion of the rules of the game in a natural way.

If we pass from social games to games in general, this definition of the rules of the game no longer seems self-evident, and thus needs further explanation. To see that the rules of the game may be identified with the sets of alternatives and utility functions also in the general case, we have to further reflect on the notion of the rules of the game. What are the rules of the game on a more informal level? They are *rules,* of course, but this is a difficult concept, and we have nothing to say about it. Second, they determine the procedure of the game. This does not mean that once we know the rules we know precisely how the game will proceed. Certain choices can be made at every step. But the set of possible alternatives from which one has to choose is determined by the rules of the game. Moreover, the rules tell who wins and how much. Again, this is not determined a priori but depends on the actual course of the game. For a given course, however, the rules determine the gains and losses in a strict sense. Now if, in a general model of game theory, we look for the components that determine the course of the game, we find those already mentioned: alternatives and utilities. These two components work together to determine the course of the game, and there is no further component in the model with such an effect. This justifies identifying the two components as the rules of the game even in those general cases in which ordinary language would not apply the term. In the prisoner's dilemma, the rules of the game are given externally. They consist of the possible choices (to confess or not to confess) and the different lengths of the periods in jail. Sometimes one normally speaks of rules of the game also in such contexts, but this is metaphorical. Note that the prisoners do not participate in the game voluntarily.

For the justification of our final claim it will be important that game theory assures that the rules of the game are given and must not be changed by the players while the game is continuing. This has to be stressed because game theorists, when confronted with this line of reasoning, tend to belittle the assumption that the rules are given and to give an image according to which the rules of the game have a weaker and more flexible status. We have to be aware of the kind of evidence relevant for deciding such an issue. The evidence here cannot consist of a statement of one single game theorist who imagines a new theory having some notions in common with game theory, and who for whatever reason wants to call it "game theory." The evidence here can come only from historical and metascientific studies of the literature and self-representation of the whole group of game theorists. On such a basis we may claim that the rules of the game in fact

have the status of being given externally, or in advance. Game theory presupposes these rules; it does not aim at introducing, defining, or giving meaning to that notion. There is no space to argue for this in metascientific detail; we will only cite some authorities: "The rules of the game, however, are absolute commands. If they are ever infringed, then the whole transaction by definition ceases to be the game described by those rules."[9] "A game is distinguished from a real situation of conflict by being performed according to completely determined *rules. . . .* In order to make the game accessible to mathematical analysis the *rules of the game* have to be formulated exactly."[10] "The general concept of a game therefore comprises the following three elements: (1) the sequence of steps decided on by persons or by chance; (2) the level of information of the players; and (3) a payoff function."[11] These citations show that the rules of the game are essential to the game's identity: to change the rules means to change the game, that is, to play a *different* game.

There is another allied assumption of game theory that usually is also ignored. To see this assumption, let us look at the typical case in a social game in which all the players accept the rules of the game before they start playing. This is a voluntary act on the part of each player. If a player is forced to participate, the payoff values derived from the rules of the game may be quite different from her "real" utilities. For instance, it may be of highest utility for her to lose the game if she knows her opponent has executed other winners in the past. To accept the rules of the game voluntarily is an indication of independence and of being roughly at the same level as the other players. As soon as one player is dependent on, or much inferior to, the other player, there is some probability that her utilities are not adequately represented by the payoff function determined by the rules.

These considerations suggest the following broad corollary. The rules of the game will be respected in greater degree the more equal and independent of each other the participants are. We can express this in the form of a slogan: Given utilities indicate equality and independence. This is admittedly a very vague formulation. There is no space here to argue in detail for the connection between given rules and equality and independence. We have to leave the issue with the status of intuitive plausibility. To give it more force, the point may be illustrated by the flexibility of individual utilities. A person's utility function may change for various reasons, among which we certainly must include emotions. A player recognizing that the rules of the game are very (un)favorable to her may change her utility function just for

emotional reasons. Such emotions need not, but may, indicate strong dependencies or strong inequalities. Conversely, the presence of inequalities or dependencies may indicate that utilities are not given and stable, but may quickly change.

If this analysis is correct, the assumption of givenness of the rules of the game indicates another, more implicit, assumption: that the different players in a game are independent of each other and are roughly of equal status with respect to the game. Of course, the notions of independence and equality of status have here a large degree of variability and are further complicated by being relative to the particular game. A master may play a game of chess with his slave. However, we think that these assumptions have an important role in delineating the intended applications of game theory from other, contrived applications. For example, it seems beside the point to analyze a fight for life or death as a game (though this is conceptually possible and is even done by game theorists).

Theory of Power

In order to describe Wartenberg's theory of power,[12] let us also begin with an example. Consider some individual p_1, a politician, whose history includes a dark side. Another individual, p_2, coming to know about this talks to p_1, threatens to tell the story to the local newspaper and in this way manages to get a job in the city administration. Clearly, this is an instance of p_2 exercising power over p_1—though perhaps not of the most brute kind.

As in game theory, each of the two individuals has a set of possible alternatives before him, any one of which he may choose to the exclusion of the others. It is difficult to characterize such a set in abstracto. Certainly, any alternative has to be physically feasible; given the individual's means, we may imagine all possible causal consequences following from any action made possible by these means. However, the difficulty then is pushed back to determining what are an individual's means. To allow for any physically possible means would be unnecessarily general. In the example, p_1 might be much stronger and thus able to beat p_1. However, because of their social context this would not increase his probability of getting a job and therefore such an alternative need not be considered from the beginning. In general, a characterization of the alternatives open to an individual will be strongly theoretical and will require consideration of the social system and the individual's special manner of socializa-

tion.[13] In the example the politician's space of actions may be taken to consist of four alternatives. First, he might do nothing and act as if no threat had been made. Second, he might wait but start preparing some campaign for survival in case the thing gets published. Third, he might behave "cooperatively," and use his influence to get p, a job in the city administration. Fourth, he might try to retaliate and threaten p, in some way. It is clear that many other possibilities might be considered, but the level of detail and of relevance depends on the concrete case.

As a second theoretical component there is an *assessment* of the respective situation by that individual. Assessment has two components. First, it consists of an *understanding* of the situation and of the different alternatives on the individual's side. This is a basic term from the hermeneutic tradition, and a difficult one. It reflects at least two major human properties: intentionality and the ability of interpretation. Any understanding of a situation depends on the interpreter's intentions. If I am in a hurry to reach a plane, my understanding of getting stuck in the subway is likely to focus on the incompetence of the workers, while in another situation I might see the incident as caused by a technical mistake, such as a short circuit. Similar differences of understanding result from differences in interpretation. As a scientific interpretation my understanding of a rain dance is one of an exotic, esthetic cultural event, whereas as a magical interpretation, the event becomes a major social and political issue. Interpretation may depend on intention, and vice versa. Usually, however, interpretation further varies with other parameters, in particular with knowledge and special forms of socialization. Whether the same holds for intentions is less clear, and in our opinion very doubtful. Technically, the effect of understanding is to filter out a subset of alternatives from the set of all possible ones as those that are relevant and seriously considered by the individual. In the example, we may assume that the politician's intention is not so friendly that he would get p, a job without the threat. Positively, his intention by and large may be to have an efficient administration in which performance is the major criterion for getting a job or losing it. His understanding of the situation is also influenced by this intention. He interprets p_2's approach as a threat, which he might not if p, were a close relative or if the whole administration were known for being completely corrupt. In such an environment, the publication in the newspaper probably would not create any problem for him. We will not attempt to analyze more formally the notion of understanding in the following.

The second component of assessment consists of a *valuation* of the given, understood alternatives. Formally, a valuation may be represented by means of a utility function. Each alternative gets assigned some numerical value. On the basis of these numbers, comparisons can be made of alternatives that by themselves are hard to compare, and rational decisions may be made by applying one or the other formal criteria of rationality. As was pointed out above, numerical representation of values or utilities is difficult to achieve in practice. If two numbers have been assigned, then the difficulty must have been overcome in the way in which these numbers were assigned. Since no real-life methods exist here, the real difficulties cannot be said to have been solved. In the example, the politician's valuation is hard to tell because this amounts to determining which of the four alternatives he prefers. If he is an active, aggressive character, alternative four (retaliation) might be his preferred reaction; if he is lazy, alternative one (do nothing) might be preferred; and so on.

In the context of the present stage of power analysis, there is no need for quantitative analysis. However, in order to facilitate comparison we take each individual $p_i's$ valuation as represented by a function V_i that maps alternatives into real numbers. We require V, to be a partial function only, so that some alternatives may pass unevaluated.

Following Wartenberg, the three components, alternatives, understanding, and valuation, may be said to form an individual's *action-environment.* An action-environment then has the form

$$<p,A,UN,V>$$

where *A* is individual *p's* set of alternatives, UN *p's* understanding, and *V p's* valuation of a given situation.

In general, the action-environments of different individuals may be rather different; it would be too unrealistic to require them to be very similar or even identical for similar individuals. However, some weak form of similarity constraint seems realistic and appropriate. If individuals have been brought up in the same social group under very similar conditions and have acquired similar social positions and roles, there is some plausibility to the idea that their action-environments will have some similarity, just as in microeconomics the use of such "stability assumptions" is important because it yields strong empirical claims (at the risk, of course, of being too idealized or "false").

We now may describe the basic model of power theory. It captures situations in which one person exercises power over another person. The fundamental axiom characterizing such a model says that one of the individuals exercises power over the other. This informal requirement may be made precise in the conceptual frame outlined in the following way:

> p exercises power over p* if and only if p intentionally changes $p^{*'}s$ action-environment in a fundamental manner.[14]

Basically, therefore, to exercise power means to change some action-environment. Any change consists of a transition from one state to another, so we arrive at a quasistatic representation in which change of an action-environment is described by two succeeding action environments $E(b)$ and $E(a)$ (b and a as in before and after) such that the latter is different from the former. For given individual p, let us write $E_i(b)$ and $E_i(a)$ to denote $p_i's$ action-environments before and after some change. We may identify the two action-environments with that change itself, provided they are different from each other. So we define a *change of action-environment* to be a pair $<E(b),E(a)>$ of two different action-environments of one individual such that $E(a)$ follows after $E(b)$ in time. Using this ontology, a model of power theory consists of a set $\{p,p^*\}$ of two individuals and four action-environments $E(b),E(a),E^*(b),E^*(a)$:

$$<\{p,p^*\},E(b),E(a),E^*(b),E^*(a)>$$

where $E(b)$, $E(a)$ are the two succeeding action-environments of individual p, $E^*(b)$, $E^*(a)$ are those of individual p'', and $E(b)$, $E^*(b)$ are simultaneous. In order to be a proper model, such an entity has to satisfy the above-stated axiom of change, i.e., one of the individuals exercises power over the other. If, for instance, p exercises power over p* this means that p intentionally changes $p^{*'}s$ action-environment in a fundamental manner, the change being represented by $<E^*(b),E^*(a)>$.

The requirement of the change being fundamental can be read in two ways. It may be read as a threshold for degrees of change to be overcome in order to constitute an exertion of power, or it may be read as making the definition fuzzy. We do not prefer either of these options. The question may be a matter of further development of the theory. "Intentional" is necessary to exclude unintended changes from counting as exercises of power. If I injure somebody by crashing into her car unintentionally, we do not say that I exercised power over her

though the result may be a very fundamental change of her action-environment.

On the basis of the defined notion of exercising power it is easy to introduce the notion of *having power* by means of counterfactuals. *p has power* over *p** if and only if there exist situations in which *p* could exercise power over *p** by acting accordingly.

"Change" may occur in one of three types or combinations, depending upon the three components making up an action-environment. First, the set of alternatives may be changed, existing alternatives may be taken away, new alternatives may be added, or both possibilities may occur in the same step. Second, the understanding of the situation may be changed. And third, the valuation may be changed. Each of these changes may occur in isolation, but mixed cases also are possible: if a new alternative is added, it may be valuated in the same process. This amounts to a simultaneous change of the valuation function. In the above example, the politician's set of alternatives may be seen as being enlarged. Though this sounds counterintuitive at first sight, we think it is the correct way to see the situation. The "new" alternatives coming into play are those different from the first one, those in which he reacts to the threat in some way. Though these alternatives were physically possible to him even before the threat was made, it is unlikely that he would have thought of them or would have taken them into account in his actions. Of course, the situation may also be described by including those alternatives in the set of alternatives from the beginning and assuming that the politician did not understand them or, if he did, did not evaluate them. The most natural description, however, seems to be one in which they do not occur originally.

The models introduced capture the most basic form of exercise of power. Other interesting forms are obtained by adding further, special requirements, for instance on the particular way in which the action-environments are changed. In this way, the most important *forms of power,* force, coercion, influence, manipulation, may be characterized.[15]

A final remark about the model is that the axiom of change essentially is antisymmetrical. Only one of the two individuals, the so called *superordinate agent* (formally it does not matter which one) exercises power, and in this case the action-environmeht of the other individual, who is called the *subordinate agent,* is changed. The model does not specify the superordinate agent's action by which she causes the subordinate agent's action-environment to change. The superordinate agent's action-environment does not play any role in this formulation.

In formal terms, however, symmetry is not excluded. In a model of the theory of power it may well be the case that one individual p exercises power over the other individual p*, and simultaneously p* exercises power over p. Such kinds of symmetry in real applications usually reveal different forms or kinds of power exercised in the two directions. For instance, p may exercise power in the form of coercion over p*, while p* at the same time manipulates p (stupid master p and clever slave p*). Inclusion of the superordinate agent's action-environment in the models provides a frame for insertion of the items just mentioned.

Comparison

There is not enough space here for accounts of both theories' historical developments, which certainly would further the comparison. We will confine ourselves to the systematic level on which there are different criteria of comparison: formal structure, method of application, nature of objects, problems, problem-solving capacity, empirical content, success. This list is certainly not complete, and it must be admitted that the theory of power described in the previous section does not yet have the status of a generally acknowledged theory. This, together with the fact that it is not completely formalized, suggests relaxing the standards of comparison as far as formal structure is concerned. At the present stage, results of comparison that can be logically proved would not be ultimately convincing because one always might try (and succeed) to escape by changing the power theory. Nevertheless, formal comparison is the most important dimension of comparison we have, so it will be discussed first.

Consider two models of the respective theories, a model of game theory of the form $\langle \{p_1, \ldots ,p_n\}, A_1, \ldots ,A_n, U_1, \ldots ,U_n \rangle$, and a model $\langle \{p,p^*\}, E(b),E(a),E^*(b),E^*(a) \rangle$ of power theory. In the light of the interpretations of the different components of these models, some formal identifications are possible. First of all, the individuals in both models can be the same. The possible difference in number of individuals is not essential; for the sake of comparison we may restrict ourselves to models consisting of just two individuals on either side. Next, let us look at the sets of alternatives attached to the individuals. The alternatives themselves may be compared without difficulty. Any alternative open to an individual in game theory may be taken as an alternative for the same individual in a model of power theory, and vice versa. However, in the model of power theory each individual

has two sets of alternatives, one before and one after power is exercised. The game-theoretic model, on the other hand, does not contain a distinction of before and after. Starting from the game-theoretic model, we may try to build up a power-theoretic model by taking over the individuals and by taking the game-theoretic sets of alternatives to be the power-theoretic sets of alternatives present *before* power is exercised. The problem then is how to fix the sets of alternatives *afterward.* Of course, simple identification would not work, for in power theory, the two sets of a person's alternatives before and after may be different.

However, formal comparison is not restricted to a term-by-term identification. It may involve further constructions using the full theoretical pictures of either theory to be compared. In the present case, it seems possible to construct reasonable sets of alternatives before and after out of one game-theoretical set of alternatives by using the additional structure of the game-theoretical model. Consider first the case in which individual p_i's set of alternatives after power is exercised is smaller than that before; i.e., the effect of exercising power was to eliminate one or more of p_i's alternatives (by force, for example). In this case the following kind of identification suggests itself. We may look for a game-theoretic model in which the sets of alternatives are just those occurring in the power-theoretic model "before." Now for any alternative to be eliminated and for any state resulting from that alternative in the sense of the section on game theory, p_i's utility should be smaller than that for any alternative not to be eliminated and for any state resulting from the latter. In other words, the states resulting from those alternatives to be eliminated have minimal utility for p_i irrespective of what the other individual chooses to do. Taking such a model of game theory (which can be easily defined), we may identify the sets of alternatives "after" in the power-theoretic model with the sets obtained by taking away the minimal alternatives just described in the game-theoretic model. If we manage to find a game-theoretic model in which those minimal alternatives are uniquely determined (which we always can find), we can construct the power-theoretic sets of alternatives out of the sets of game-theoretic alternatives in the way just described. A similar kind of construction may be performed in the second possible case in which exercise of power leads to an extension of the set of alternatives present before.

These constructions show that different power-theoretic sets of alternatives can be obtained out of one set of game-theoretic alternatives if we choose appropriate models possessing further properties

on the game-theoretic side. Of course, the construction cannot succeed for any given pair of models. Nevertheless, the method provides rather strong identifications: For any given game-theoretic model (with two individuals), we may trivially find some power-theoretic model such that the sets of alternatives "before" in the latter are identical with the sets of alternatives in the former. Conversely, for any given power-theoretic model we may find a game-theoretic model with additional special properties such that the sets of alternatives of the first model before and after can be constructed out of those of the second model as described above.

A third component of the two models that might be seen as a candidate for formal comparison consists of the utility functions and valuation functions. Here things become difficult. Purely formally, the problem is this. If alternatives are identified along the lines just considered, then the arguments of a game-theoretic utility function and a power-theoretic valuation function are different. Whereas the valuation function takes a single alternative as argument, the utility function needs a resulting state in the technical sense, i.e., an n-tuple of alternatives. If the alternatives are identified one by one in two corresponding models, such identification is impossible for utility and valuation functions. Still, we may try to construct one from the other as we previously did for the alternatives themselves. Starting from a power-theoretic model, we might construct a corresponding game-theoretic model as follows. Individuals and alternatives (before, say) are taken over as above. We define each p_i's utility function, U, as being independent of other individuals' choices; i.e., if $<a_1, \ldots, a_n>$ and $<a_1^*, \ldots, a_n^*>$ are resulting states such that $a_1 = a_i^*$, then $U_i(a, \ldots, a_n) = U_i(a_1^*, \ldots, a_n^*)$. A simple way to achieve this is to define $U_i(a_1, \ldots, a_n) = V_i(a_i)$. In the game theoretic model thus constructed, the utility functions cannot be strictly identified with those of the power-theoretic model. They can be regarded, however, as inessential formal variants of the latter. Intuitively, the utility p_i gets from a list of choices $<a_1, \ldots, a_n>$ is just the value V_i that p, attaches to the alternative a_i from her own set of alternatives. For this value the choices of other individuals do not matter. It seems possible to obtain game-theoretic "descriptions" of every power-theoretic model in this way. Conversely, if we start with a game-theoretic model, we can get a power-theoretic "image" by giving up a one-by-one identification of alternatives and simply taking the power-theoretic alternatives to be the n-tuples of alternatives from the model of game theory. This yields a straightforward identification of utilities and values by defining the valuation functions to be the same as the utility functions.

In summary, there are quite substantial possibilities of formal comparison and identification between models of game theory and power theory. They indicate that both theories have the same or very similar applications and objects. The meanings of the terms "individual" and "alternative" may be kept unchanged, and the meanings of "utility" and "valuation" may be identified to a large extent. These facts agree with the informal observation that both theories deal with similar kinds of phenomena—at least in a large area of overlap. This may be expressed by saying that there are many real situations or systems that can be captured simultaneously by models of game and power theory. The situation of the prisoner's dilemma gives rise to a power-theoretic model if we concentrate on the district attorney's exercise of power over the two criminals, and the politician's blackmail may easily be analyzed to yield a model of game theory.

On the other hand, it is difficult to obtain a satisfactory, complete comparison that would show that one theory can be completely "reproduced" in the other. The previous discussion was "local," i.e., two corresponding terms were considered without looking at the impact of their comparison on the other parts of the models. We did not take into account whether the identifications considered are all compatible with each other and did not ask whether they are compatible with the basic axioms characterizing the models on each side. A full intertheoretical relation would imply that not only one theory's terms can be matched with, or constructed out of, those of the other theory, but also that the basic axioms are related by implication, at least in addition to some translation or identification of the terms involved. It is obvious that neither theory is a specialization or a theoretization of the other. It is less clear whether one of them can be reduced in a precise sense to the other.[16] This would require, among other things, a relation of derivability of the axioms. It is far from clear whether the basic axiom of rationality characteristic of game theory as formulated above implies, or is implied by, the basic axiom of change in power theory, even if appropriate constructions of the kind discussed earlier are inserted. Intuitively, this does not seem possible as long as natural kinds of translations of the terms are used. However, it is not easy to say which kinds of translations of one theory's terms into terms of the other are natural.

It is not our aim here to show that one of the theories is fully reducible to the other. This would show that the reduced theory in some sense is more restricted and "poorer" than the other, and in this sense could be replaced by the other one. Rather, the aim of the previous formal comparison was to show that both theories have very much

in common: terms, meanings of terms, and a large overlapping set of applications. If this is so, if both theories' models are sufficiently different and if neither theory can be reduced to the other, they are rivals, at least in their domain of overlap. We state this result for later reference:

(1) *Game theory and power theory are rival theories in a large domain of common applications. There are many ways to establish or construct identities between models of the two theories.*

Having stressed these identities, we may now turn to the differences. Clearly, the basic axioms of both theories are rather different from each other, not only in formulation but alsq in spirit. The game-theoretic axiom of rationality requires that all individuals choose alternatives that make up a point of equilibrium, i.e., alternatives that under the constraints of the game for each individual yield utilities that in some sense can be called maximal. In Nash equilibrium, for instance, they are maximal because any deviation from the chosen alternative would yield a decrease in utility for the individual who deviates. This axiom requires behavior disciplined by the rules of the game: reflection on what the other individuals' alternatives and utilities are and a kind of calculation of one's own best (equilibrium) choice from a rather complicated range of choices. The power-theoretic axiom of change is very different. It cannot be regarded as a rule of behavior; it is purely descriptive. It just states that one individual changes the action-environment of the other. It does not have any implications of rationality or **maximization** of utility. An exercise of power can be entirely irrational in the sense of yielding utilities that, for both the subordinate and the superordinate agent, are much lower than those resulting from a choice of other alternatives. The axiom of change implies neither strategic consideration of the possibilities of the other agent or agents nor evaluation of one possibility against another. It does not rule out such behavior, of course, but it also does not require it.

It seems difficult to draw some natural link between those two requirements. No translation of either axiom that would turn the translated statement into one derivable from, or implying, the other axiom seems possible. On the other hand, in the light of statement (1) above, there has to be some connection. We may get a better understanding by trying to add each axiom to the frame or surrounding of the other one. First, what would it mean to add the axiom of rationality to that of change? In the model of power theory, there are two places where

the axiom of rationality could be added. First, it could be added on the superordinate agent's side, so that her action that causes the change of the other agent's action-environment is rationally chosen, i.e., in a way to maximize her own valuation. Second, the subordinate agent might be required to change her action-environment in a rational way, namely such that the changed environment yields better courses of action than were previously available given the superordinate agent's action. Although it is not clear whether the second possibility still falls under the established paradigm of rational behavior, it is still a genuine possibility. Both these extra assumptions would amount to an additional feature not present in the original models. *So* power theory might be enriched by assumptions of rationality.

Conversely, there is no clear way to add the assumption of a change of action-environment to those of game theory. If we try to construct something like an individual's action-environment in a given game-theoretic model, we are bound to fail. On the one hand, in a given model there are no means to identify an individuals' valuation of his alternatives. Utilities are given only for resulting states, and from these there is no general way to obtain a valuation of single alternatives. On the other hand, the idea of change is alien to the game-theoretic model, as we argued earlier. Game theory assumes that the rules of the game, including utilities and alternatives, are given and stable.

The two possibilities of adding the assumptions of one theory to those of the other show a certain asymmetry. It is easily possible to add assumptions of rationality to those of power theory, but it is difficult to add assumptions of changes of action-environments to those of game theory. This indicates that power theory starts from a more fundamental level. By adding rationality, we ascend to the higher level of game theory. Conversely, to ascend from game theory to power theory in this way is difficult, as just discussed, and there is no way to descend from game theory to power theory by omitting the assumption of rationality.

This leads to a fundamental difficulty: by adding an assumption of change to game theory, we obtain a set of assumptions that contradicts the basic *presupposition* of game theory, namely that the rules of the game are given and stable. This is a situation whose significance can be seen by reflection on other episodes in science. Feyerabend proposed that two theories related in this way are *incommensurable.* According to his proposal, two theories are called incommensurable if and only if the meaning of their essential descriptive terms rests on contradictory principles.[17] Now principles on which the meaning of a

theory's essential descriptive terms rest have traditionally been called presuppositions. So Feyerabend's characterization amounts to saying that the theories' presuppositions contradict each other. By a slight liberalization — allowing for presuppositions as well as for axioms — we obtain the case before us. Two theories are incommensurable if and only if their axioms or presuppositions contradict each other.[18] Of course, this stipulation makes sense only in the presence of a strong overlap of applications of the theories in question as guaranteed by (1) above.

So we have finally arrived at the thesis mentioned in the introduction: game theory and power theory are incommensurable. They are rival approaches to a large array of common phenomena, they have many features (terms and meanings) in common, but their models and basic assumptions are distinct to an extent that makes full comparison difficult if not impossible. Further, the axioms of power theory contradict the presuppositions of game theory. The first three points have emerged in this section. The contradiction of basic axioms and presuppositions is justified by our earlier elaboration of those axioms and the game-theoretic presupposition.

Some objections that rest mainly on misunderstanding may be dealt with right away. A first objection consists in pointing out that game theory can describe exertions of power. This is acknowledged. The point, however, is that it can do so only *in the given frame* of the rules of the game. Since the essence of power consists in changing these rules, game theory fails to grasp the essential feature of power. A more subtle version of this objection refers to forms of influencing each other's choices, in particular in connection with supergames. The tit-for-tat strategy mentioned above, for instance, may be seen as a means by which player p tries to influence her opponent to play cooperatively. Again, the objection does not take into account the presupposition of game theory that the rules of the game are given. In the present example this means that every player's *supergame* strategies are fixed before the game begins and are known to each player. The opponent p^*, therefore, will *not* see tit-for-tat as a means to coax her to play cooperatively. She chooses her strategy in the light of all given strategies of p, one of which is tit-for-tat. If p^* in fact chooses some strategy that takes into account p's playing tit-for-tat, we cannot say that p changed p^*'s action-environment by playing tit-for-tat. p^* chooses her alternative *before p's* move, not knowing which strategy p actually will play. p^*'s only "reaction" (the term is not really appropriate) is a reaction to the given set of p's strategies. Still, it may be said that the

existence of suitable strategies or alternatives in a game changes the players' action-environment, for if they were not present, the players would choose differently. This is correct, but in this form the statement is no longer an objection. For now it is not the other player that induces a change, but the existence of certain strategies, i.e., the rules of the game. The initial objection confuses behavior caused by the rules of the game with behavior caused by the other player's behavior. In game theory the latter *cannot* occur.

A second objection is that the theory of power presented here is inadequate because it completely neglects rationality. A "good" theory of power, it is held, should incorporate rational agents from the beginning. To this there are two replies. First, as mentioned already, we easily may add features of rational behavior to the model of power theory. So the present theory may serve as a basis on which a fuller theory of "rational power" may be erected. Second, by not incorporating those features in the basic models, we have a much more general point of departure. Humans may *be* rational, but we certainly do not always *behave* rationally. So the theory of "rational power" envisaged is not a theory more adequate than the one considered here, but only a special case of the latter, or more technically, a theoretization.

A third objection holds that the phenomena of exercising power and of playing games are disjoint, so that the two theories presented here are as well. In other words, the phenomena both theories deal with are different, so the theories have no common applications. If this were so, a claim of incommensurability would make no sense, for incommensurability presupposes that the two theories are rivals and have many common applications. We have to be careful here to make clear what we mean by the phenomena studied by a theory. This term may be used with two different meanings. It may refer to the brute facts, the real systems as given completely independent of the theory in question. But it also may refer to the facts or systems as seen from the point of view of that theory. The brute fact of a man knocking down another man under ordinary conditions is one phenomenon, the situation seen as an exercise of power is another phenomenon, and the situation seen as a scene played realistically for the camera installed further away is a third phenomenon. We admit that a real situation as seen from the point of view of power theory and a real situation as seen from the point of view of game theory may be different. Nevertheless, there is some real situation giving rise to the two interpretations, though we cannot say much about it. But the claim that power theory and game theory have overlapping domains of application is

meant in the sense that there are common brute facts or real systems in the first sense just mentioned. By "application" we mean a process starting at a level independent of the theory in question, a process including conceptualization in a first step. A real system may be conceptualized in different ways and thus may give rise to the application of two different theories. In these meanings of "phenomenon" and "application" our assumption of an overlap between game and power theory is rather trivial. Examples of systems to which either theory can be applied have been mentioned already, and the reader can easily come up with others.

The theme of incommensurability has been much discussed recently, and several definition-have been proposed. Kuhn has linked it to more comprehensive entities like Gestalt, world-views, and sociopsychological features. Cases of incommensurable theories can be identified at the sociopsychological level by continuous unproductive discussions by proponents on both sides in which no real arguments are advanced and propagandistic elements are substantial. The transition from one such theory to the other, or from one theory's model to a model of the other theory, involves a Gestalt-switch, a radical and deep reinterpretation or reorientation. Finally, the two theories are closely associated with different comprehensive views about the world or substantial parts of it.[19]

These further characterizations of incommensurability may be used to further justify our hypothesis. All three features are present in our case. First, there is little communication between scholars of the two camps. I have myself experienced rather emotional and bitter discussions when the comparison was brought up. Second, the transition from one model to the other involves something like a Gestalt-switch. Game-theoretic individuals are independent of each other as far as the game is concerned, they are equal to each other as far as the game is concerned, they are free to play the game, and they stick to the rules of the game. The picture usually associated with the notion of power is one of everyone being everyone else's enemy. Individuals are neither independent nor equal. Exercise of power is the open expression of dependence and inequality. The subordinate agents are not free to leave the situation captured by a model of power theory. There are no rules of the game, there is only the rule of the stronger. Third, these strong contrasts point to more comprehensive views about humans and society. Game theory is in line with Hobbesian ideas of social contract as a means to establish social order, while power theory looks at social order essentially as a means to stabilize social stratification.

I do not want to close with passages that give the impression of a political fight rather than of a scientific study. If my analysis is correct, then both theories are rivals insofar as they have a large array of common applications and their models are quite different from each other (perhaps irreducibly). Furthermore, the basic axiom of power theory contradicts the basic presupposition of game theory, which according to Feyerabend means that both theories are incommensurable. However, in contrast to the situation in the natural sciences, incommensurability here does not mean that both theories or their proponents have to fight until one approach is eliminated and replaced. This pattern may be typical for physics, but it is not typical for social science. In physics, incommensurabilities are rather small, and the victorious theory usually incorporates all or most of the achievements of the losing theory. In our case, this is quite different. If one of the two accounts were to replace and eliminate the other one, this would mean a substantial loss. If power theory were simply replaced by game theory, the axiom of change would get lost. In game theory, there is nothing to indicate that individuals try to change each other's **action-environ**ments. Conversely, if game theory were given up in favor of power theory, the assumption of rationality inherent in game theory would no longer be made. In this case, one might enrich power theory by such an assumption, but the point is that mere replacement (without enrichment) would not save the basic assumptions of the replaced theory. This situation is not found only in the case of the two theories considered here, it also obtains for other pairs of theories in social science, such as microeconomics versus game theory, microeconomics versus power theory, or **Marxian** value theory versus any of the other three.

The significance of incommensurability in the social sciences must not be judged by means of its significance in the natural sciences, namely that one of the two rival theories will ultimately replace the other one. But then incommensurability loses its frightening aspect. If we imagine incommensurable theories in peaceful coexistence, another term seems appropriate: *complementarity.* Each theory focuses on one particular side or dimension or surface of the common real systems, and though the picture obtained is in some sense a complete description of the systems, it is incomplete insofar as there are other sides or dimensions or surfaces of the same systems that, when described in terms of another theory, make them look quite different. Game theory and power theory, in fact, are complementary in this sense. Game theory focuses on the rational, calculating aspect of human beings,

while power theory concentrates on the inherited Wille *zur* Macht. These are two sides of human existence that perhaps cannot be united without residue, and so each of them becomes important as a branch of social scientific research.

Conflict in Industrial Society (Stanford, Calif.: Stanford University Press, 1959), **175–79;** I have criticized Benton on this in my "On Benton's 'Objective Interests and the Sociology of Power': A Critique," *Sociology* **16,** no. **3** (August, 1982), **440–44.**

65. On this problem within classical Marxism, see Svetozar Stojanovic's excellent *In Search of Democracy in Socialism: History and Party Consciousness* (New York: Prometheus Books, **1981);** also Norman Geras, "The Controversy About Marx and Justice," New *Left Review* **150** (March–April, 1985), **47–88.**

66. See Robert A. Dahl, *Polyarchy* (New Haven, Conn.: Yale University Press, **1971);** *Dilemmas of Pluralist Democracy* (New Haven, Conn.: Yale University Press, **1982);** *Who Governs?* (New Haven, Conn.: Yale University Press, **1959).** For the argument that Dahl does not employ his own methodology in his actual research, see Peter Morriss, "Power in New Haven: A Reassessment of 'Who Governs?' " *British Journal of Political Science* **2 (1972).**

67. See Michele Barret's synthetic discussion in *Woman's Oppression Today* (London: Verso, **1981).**

68. Karl Marx and Freidrich Engels, "The Communist Manifesto," in *The Marx-Engels Reader,* ed. Robert C. Tucker, 1st ed. (New York: Norton, 1969), **347.**

69. See my *Power and Marxist Theory,* Part II.

70. C. Wright Mills, *The Power Elite* (London: Oxford, 1956), **4–5.**

71. Harold Lasswell and Abraham Kaplan, *Power and Society* (New Haven, Conn.: Yale University Press, 1950), x–xv.

72. David Easton, *The Political System* (New York: Knopf, **1953).**

73. David Easton, "The Political System Besieged by the State," *Political Theory* **9,** no. **3** (August, 1981):316.

74. Bob Jessop, *The Capitalist State* (New York: New York University Press, 1982), **221.** For a good summary of recent arguments about the state, see Martin Carnoy's *The State and Political Theory* (Princeton, N.J.: Princeton University Press, **1984).**

Notes to Chapter 4

1. M. Taylor, *Anarchy and Cooperation* (London: Wiley, **1976).**

2. See J. von Neumann and 0 . Morgenstern, *Theory of Games and Economic Behaviour* (Princeton: Princeton University Press, 1953), for the classical account of game theory, and J. W. Friedman, *Oligopoly and the Theory of Games*

(Amsterdam: North-Holland, 1977); R. Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge:Cambridge University Press, 1982); Kreps, et al., "Rational Cooperation," *Journal of Economic Theory* 27 (1982):245–52, for more recent developments.

3. See R. Dahl, "The Concept of Power," *Behavioural Science* 2 (1957):201–15; S. Lukes, *Power: A Radical View* (London:Macmillan, 1974); and J. W. Thibaut and H. H. Kelley, *The Social Psychology of Groups* (New York: Wiley, 1959), for recent accounts.

4. T. Wartenberg, *The Forms of Power: From Domination to Transformation* (Philadelphia:Temple University Press, 1990).

5. The first part of this claim cannot be defended here for reasons of space; for the second, see Wartenberg's book, in particular Chapter 5.

6. Compare W. Balzer, C. U. Moulines, and J. D. Sneed, *An Architectonic for Science* (Dorecht: Reidel, 1987), Chapter 4 for the notion of theory-nets, and Chapter 6 for the "reduction" claim.

7. This becomes more clear when the individuals are seen in the broader context of an institution. Compare W. Balzer, "A Basic Model for Social Institutions," *Journal of Mathematical Sociology* 16 (1990).

8. See F. Mosteller and P. Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy* 51 (1951):371–404, for an example.

9. Von Neumann and Morgenstern,*Theory of Games and Economic Behaviour,* 49.

10. J. S. Wentzel, *Elemente der Spieltheorie* (Leipzig:Teubner, 1964), 6. Translation is mine.

11. G. Owen, *Spieltheorie* (Berlin:Springer, 1971), 1. Translation is mine.

12. Wartenberg, *The Forms of Power.* The following definitions are taken from that book with minor variations.

13. Compare my paper on social institutions, "A Basic Model for Social Institutions."

14. Compare Wartenberg,*The Forms of Power.* I have merged his notions of control and change into that of intentional change in order to avoid a discussion of "control." We think that intentional change, though apparently weaker, covers all important aspects.

15. See Wartenberg, *The Forms of Power.* It may be noted that *bribery* also can be included in this list, though it is not treated by Wartenberg.

16. Compare Balzer, Moulines, Sneed, *An Architectonic for Science,* Chapter 6, for accounts of the intertheoretical relations mentioned.

17. P. Feyerabend, "Problems of Empiricism," in *Beyond the Edge of Certainty,* ed. *R.* G. Colodny (Englewood Cliffs, N.J.: Prentice-Hall), 145–260. See in particular 227, note 19.

18. This definition is a straightforward and natural generalization of my account in "On Incommensurability," in *Imre Lakatos and Theories of Scientific Change* eds. *K.* Gavroglu, et al. (Dordrecht: Kluwer), 287–304.

19. T. **Kuhn,** *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970).

Notes to Chapter 5

1. Brian Fay, *Critical Social Science* (Ithaca, N.Y.: **Cornell** University Press, 1987), 120. My emphasis.

2. I want to stress that both the example that I am using and the model that I am developing are necessarily abstract. In order to see how situated power works, both at a theoretical level and in terms of the example of grading, I shall abstract from various other features of the concrete situation in which teachers and students find themselves in order to focus my attention upon the manner in which a set of social relationships external to the teacher-student relationship constitutes that relationship as a power relationship.

3. There are institutions of higher education like Hampshire College that evaluate rather than grade their students. I leave out of consideration how such a divergence affects the power between students and teachers. The teacher-student relationship that I am discussing here is the standard one in American higher education.

4. Fay, *Critical Social Science,* 120.

5. Fay's use of "causal outcome" in his definition of power is also problematic. My telling you that today is a holiday may have as a causal outcome a change in your actions, but this does not mean that I have exercised power over you.

6. Previous social theorists have talked about "anticipatory reaction" in this context. See **Carl** J. Friedrich, *Man and His Government: An Empirical Theory of Politics* (New York: McGraw-Hill, 1963), Chapter 11, for an elaboration of this concept. The problem with this concept is that it describes the existence of such power as dependent solely upon the subordinate agent.