Reflections on Equilibrium

Ideal Rationality and Analytic Decomposition of Games

by

Siegfried Berninghaus*, Werner Güth** and Hartmut Kliemt***

Abstract: **Taking seriously the philosophical foundations of classical strategic theories of choice-making we scrutinize to what extent planning on equilibrium strategies can be justified "eductively" among rational players and how this can be utilized to analyze games by their "game-like" sub-structures, in particular by their sub-games in the extensive and by their cells in (agent) normal form. "Material" principles of rational choice and "formal" methodological requirements of consistent theory formation are considered and it is claimed that there *can* be consistent "conventions of rationality". Which of the possible conventions will prevail and *define* rationality may depend though on which of the theories of ideal rationality will be absorbed among rational agents. Once established "conventional rationality" can lead to unique solutions for strategic games.**

1 Introduction and overview

Its recent success in economics and social science notwithstanding the basic theory of "rational choice-making" – i.e. non-cooperative game theory – presently is in a state of reflective *dis*equilibrium (see Rawls 1951; Rawls 1971, Daniels 1979, and Hahn 1996; Hahn 2000 for the concept of a "reflective equilibrium"). On the one hand, working out the details of the original research program it became exceedingly implausible that real actors are rational choice-makers in the sense of non-cooperative game theory. The deeper our understanding of the theory the more problematic external or empirical validity and applicability of game theoretic results became. The more we think through the basic knowledge assumptions of classical game theory the more outrageous become attempts to interpret it as a behavioral theory of human choice-making. On the other hand, there were many attacks on the internal validity or coherence of non-cooperative game theory as well. In view of those it might seem that the purely conceptual exercise of a game theoretic explication of rationality (see on explication Siegwart 1997) encounters serious problems of internal consistency.

In fact, if we take the two objections together is not game theory like Lichtenberg's knife that after it lost its grip also lost its blade? Less loosely put, if on both central methodological accounts, external validity and internal consistency, classical non-cooperative game theory seems to fare badly should we not give it up completely as a theory of choice-making?

We think not. However, the empirical and the philosophical side of game theory must be kept apart carefully. The former should basically be developed in field studies or experimental game theory and

in the last resort presumably merge with economic psychology. The latter should be seen as a philosophical account of ideally rational choice-making.

A closer scrutiny of the foundations and implications of non-cooperative game theory is of great interest for "economic philosophy". It is also central for any rational individual who intends to reach a self-understanding of what it means to be rational and to interact with other rational beings. We all, at least in a way, conceive ourselves as members of a world of rational beings. As long as we do so we have an interest in what may be called the "ideal type" of rationality. Driving elementary notions of rationality and reasoning to their extremes as we intend in this paper is one way to approach such an ideal type of rationality. In doing so we will devote special attention to the fundamental notion of a strategic equilibrium (see Cournot 1838, Nash 1951) and to the question of how selecting equilibria can serve as a tool of analysis that allows for decomposing a larger problem of choice-making into smaller ones. We shall defend the view that analysis of "larger" problems in terms of "smaller" ones is possible only *if regardless of their place in a larger game context isomorphic game-like sub-structures of a game are solved the same way.*

Contrary to some recent criticisms solving games by equilibrium refinements or selection is not to be given up. Discriminating between equilibria is supported by arguments stronger than some purely formal consistency notions that have been proposed in recent years (see for instance Peleg and Tijs 1996, Norde, Potters et al. 1996, Van Heumen, Peleg et al. 1996).

We first sketch our own account of the underpinnings of a strict rational choice perspective in general (2.). We then discuss the use of equilibrium notions as solution concepts for games (3.) and reject purely formal consistency notions that would rule out the equilibrium refinements needed for our favorite decomposition process of game theoretic analysis (4.). We go on to show that convincing consistency notions are compatible with equilibrium refinements (5.). Some general philosophical reflections conclude the paper (6.).

2 On the rational choice perspective in general

On an elementary level all rational individuals – including merely boundedly rational ones – are characterized by their ability to distinguish between their own potential to exert causal influences on the world and the course of the world as emerging independently of their actions. They can and do sometimes make forward looking choices that serve their own ends, aims or values, as they perceive them. And they know that they interact with other such beings who themselves are equipped with the same abilities and knowledge to some extent.

Such a world of realistically rational beings is transformed into the world of ideally rational choice-makers by assuming that rational actors command unlimited resources to analyze models of their interaction situation and to draw inferences from those models. To this the so-called "common

knowledge of rationality" is added including the heroic assumption that all actors form their models of the world on the "rational expectation" that they as all others *do behave* according to the rational reasoning ascribed to them. The essential question of whether rational individuals would in fact behave and would make their choices as ideally rational reasoning suggests is assumed away for the purpose of modeling. As far as the rational players' theory of co-player behavior is concerned rational choice theory simply *ascribes* to the choice-makers the belief that rational choice theory – as envisioned by the theorist – is behaviorally valid. The theorist assumes that the objects of his theory share his theory, behave accordingly and therefore can predict behavior by the theory.[1]

The preceding assumptions are very heroic. But people have accepted them because with these idealizing assumptions in hand, the theoretically very rich traditional analysis of choice-making in terms of reasoning about choice-making unfolds. Since traditional analysis is clearly not externally valid – neither as a theory of mental processes nor of overt behavior – it must at least be internally valid in forming a consistent and in some material sense coherent endeavor. Checking on this let us start with the basic behavioral model.

2.1 Opportunistically rational behavior

We think that five assumptions about rational actors are essential to rational choice modeling:

- First, all actors base their choices on a model of their own action situation. Rational actors form mental representations of the world. Reasoning about these representations – thinking them through, so to say – is crucial for rational players' planning and acting.
- Second, in their modeling rational actors distinguish strictly between the causal effects of their own choices and those aspects of the action situation that are beyond the causal influence of the individual acts they consider.
- Third, they make (or at least plan) all their choices intentionally and purposefully in view of the anticipated causal consequences of each act of choice taken separately.
- Fourth, choices are evaluated and made according to individual preferences that can be represented by utility functions that take into account *all* relevant evaluative considerations.
- Fifth, individuals who reason in the presence of other individuals who can themselves reason are aware of that presence. It is part of their own rationality that they take into account that other individuals are endowed with the same rationality and thus theories of rational plan and play are also part of the reality to which they apply.

Behavior complying with these five requirements we shall call "opportunistically rational" or, more succinctly, "rational", subsequently. Of course, one might legitimately try to propose another concept of rationality. For instance, a "moralist" might feel that answering the question "what if everybody did the same?" provides "rational" reasons for acting (see paradigmatically on the implications of such views for human practice Kant 1991). But if individuals take each act separately (third premise)

---

[1] In short, via "rational expectations" rational choice modeling projects itself on reality and avoids most serious questions of external validity.

neither what *would* happen if *all* (or some) would act the same way nor what *would* happen if the actor *would* always (or sometimes) act the same way is relevant for a rational actor when deciding on how to act. Likewise a "sociologist" might argue that actions brought about by a general disposition to act in a certain way should be called "rational" provided that having this general disposition rather than deciding each case on its own merits is in the rational interest of the actor.

It is an essential characteristic of rational actors that they can decide each case on its own merits, however. Relying on games in extensive form may serve as a constant reminder of this. Moreover, an extensive form game with its time structure illustrates best how larger games can be decomposed into smaller game-like sub-structures like sub-games that can be solved independently of the larger context. For purposes of illustration the reader of the following discussion might want to think in terms of sub-games even when we speak of other game-like sub-structures like cells. Sub-games of a game in extensive form representation naturally translate into cells in the agent-normal form representation that we use as a reference or standard representation of games. One can perform all relevant analyses in terms of cells but for purposes of illustration the timing and causal structure of the extensive game representation is a most natural way to illustrate the effects of assuming opportunistic rationality throughout.

Those who want to take seriously what differentiates a "rational" choice approach from its scientific neighbors better stick to the five basic (counterfactual) assumptions. Being based on those five "idealizations" the rational choice approach to choice-making becomes "normative" in the peculiar sense of being untrue to the facts of actual behavior. Nevertheless theorists of rational choice insist that their normative models, though based on counterfactual assumptions, have explanatory value. But as we shall argue there is no way to make good on the promise of providing "true explanations" by rational choice modeling. Relying on predictions without an explanation of why they work is very doubtful as well.

2.2 Rational choice explanations?

Many adherents of a rational choice approach to human behavior are fond of using terms like "explanation" and "prediction". This is acceptable only if these terms merely indicate that certain models have certain implications. Of course, any model allowing some kind of logical inference can "predict" or "explain" what is implied by the premises of the model. For example, in this sense the assumption of the “hidden hand” (as opposed to “invisible hand processes” of selective adaptation) of an omniscient, benevolent, almighty planner could "explain" why biological organisms are well adapted to their environment. However, as far as we accept that a supra-human planner does not exist the implications of counterfactually assuming that such a planner exists do not amount to (empirically) valid or true explanations or predictions. They are merely "potential explanations" of doubtful credentials (see on the concept and use of "potential explanations" the introductory chapters

of Nozick 1974). They are explanatory tales that might conceivably be true in theory but are untrue in fact.[2]

More often than not the claim that rational choice modeling per se provides true or valid explanations of empirical phenomena is not even approximately correct but rather off the mark (for a survey of experimental evidence see Kagel and Roth 1995). As a matter of fact social phenomena do not emerge from the *rational* choice-making of individuals. Even if phenomena seem *as if* resulting from rational choices we do not have good reason to assume that rational choice explains what we observe. The relation between what is to be explained and what explains is quite the other way round: Whenever in the real world we encounter phenomena that seem in line with the implications of some rational choice model or other then this observation itself forms an expla*nandum* rather than an expla*nans*. It raises a question of explanation rather than providing an answer (see for a more extensive methodological treatment Kliemt 1996). The fact that the phrase "as if" is used acknowledges this at least in a way. The question why it is so that individuals behave "as if" they were fully rational and which causal influences – in particular of an institutional kind – induce them to behave that way is raised rather than answered.

In their typical answers to such queries social theorists have in fact eliminated rational choice assumptions from their explanata. For instance, according to Armen Alchian's seminal argument, even if firms would myopically follow fixed behavioral programs only those that behave *as if* they were opportunistically rational would survive competition (see Alchian 1950). In the same vein, the fact that market equilibria that cannot conceivably be improved by fully rational individuals might emerge from the choices of "zero-intelligence traders" (see Gode and Sunder 1993) is a very interesting result about the effects of market *institutions*. But it does not corroborate the theory of fully rational individual choice since in such an evolutionistic approach behavior complying with rational choice standards of rational decision-making is explained without assuming that individuals are in fact rational in the relevant sense. People act *as if* an invisible hand led them. They behave *as if* they were non-myopic, perfectly rational actors but they do so without being fully rational. In short, apparent rationality is explained by factors other than rationality.

Setting aside the fact that evolutionary dynamics do not necessarily imply the survival of material payoff maximization (see on this and for further references Güth and Peleg 2001) we do not deny that

---

[2] As David Hume already observed when he required that in politics everybody should be supposed a knave, "it appears somewhat strange, that a maxim should be true in *politics* which is false in *fact*." Hume, D. (1985). Essays. Moral, Political and Literary. Indianapolis, Liberty Fund. Essay VI, pp. 42-43. This is indeed strange within an empirical approach. Neither Hume nor modern economists ever came up with a convinicing answer to the querie. Yet counterfactual analyses remain appealing to most of us.

the evolutionary perspective offers valuable insights into the mechanisms of adaptation and the emergence of adaptive social practices (see in particular Young 1998, Sugden 1986). It may also be helpful in reformulating the "dynamics of rational deliberation" (see Skyrms 1990). However, only a strict focus on reasoning can help us to understand what separates the specifically rational, opportunity taking forms of behavior from adaptive “trial and error” in choice-making.

Accordingly we should take to rational choice modeling proper rather than to "adapted" versions of it. In its most idealized form it studies a world of fully rational beings engaged in strategic interaction. These beings command unlimited faculties to reason and a full knowledge of the rational choice theory itself. They are themselves guided by rational choice models or mental representations of their own interaction situation and at the same time endorse the symmetry assumption that all others are guided by such models, know the relevant theories and can draw inferences from them.

Those who engage the task of unfolding the implications of full rationality should be aware that their object of study is fictitious. Real world processes like "trial and error" or "adaptive behavior" are alien to the "reasoning about knowledge" (see Fagin, Halpern et al. 1995) that characterizes the mental processes of players envisioned by “normative” game theory. Such game theory is "eductive" (see Binmore 1987/88) in that it models the inferences that fully rational individuals *would* draw. It focuses on conclusions of perfect inference machines in a world populated exclusively by such beings among whom their rationality is common knowledge. It thereby “explicates” what “(ideal) rationality” means but it does not "explain" choice-making in any nomological sense.

We believe that it is of some value to engage the task of analyzing models based on assumptions of fully rational choice-making even if such models are non-explanatory but only explicatory. Being of the species homo sapiens we share a proclivity to reflect on our own rational capacities of reasoning and reflection. This is a purely philosophical interest in one sense. Yet in another it is also, if indirectly, influencing our self-image as humans and thereby our practical relations towards other humans whom we perceive as *participants* of rational interaction rather than as objects of action and manipulation. The implications of this "participant's attitude" (see Strawson 1962) towards social interaction have traditionally been modeled by non-cooperative game theory to which we therefore turn next.

3 Conventional rationality in eductive approaches

3.1 On non-co-operative game theoretic modeling

3.1.1 The basic characteristics of non-co-operative modeling

In the last resort the distinction between co-operative and non-co-operative game theoretic modeling is a very simple one: In non-co-operative modeling it is assumed that all strategic options of each decision maker, party or player involved in a strategic interaction have been explicitly modeled before analysis starts. If it is understood that some strategic options which do not show up in the game model itself are relevant for analysis then we have left the domain of non-co-operative game theoretic modeling.

Co-operative modeling may be extremely useful if we want or need to by-pass very subtle strategic aspects like the order of moves or detailed informational assumptions. In contexts where the latter aspects either do not matter much or would raise the level of complexity beyond what can be handled explicitly in a model it may often be good policy to take such short cuts. But if our focus is on what ideally rational players know and infer from their knowledge before they must play games we better assume that all strategic options have been explicitly modeled before analysis begins. Otherwise we would not have an explicit model of what the ideally rational players know and thus could not analyze theoretically what they should infer from their knowledge. In such eductive analyses all strategic options, all evaluations, the fact that all players are rational and the theory of rational planning itself are assumed to be common knowledge. For convenience we refer to these conditions jointly as *common knowledge of rationality* (CKR) and formulate the following adequacy condition for eductive non-co-operative game theory: *Any theoretical suggestions for rational plan and play must be acceptable for rational players under CKR*.

3.1.2 Expectations in a world of fully rational beings

3.1.2.1 Common knowledge assumptions

A game theoretic analysis of what rational players might have on their minds when analyzing a game must start from what they know. The analysis assumes that like the external game theorist, players will analyze the game beforehand. Players make their plans before the game is actually played. Therefore what they know *initially* is of central importance. Since the eductive theory of rational play is all about the conclusions players might draw from their information about the rules of the game, analysis cannot even start unless we make definite assumptions about what players know about the game and each other. Actually we must start at a level of analysis on which common knowledge prevails.

To see this let us begin with a simple game.

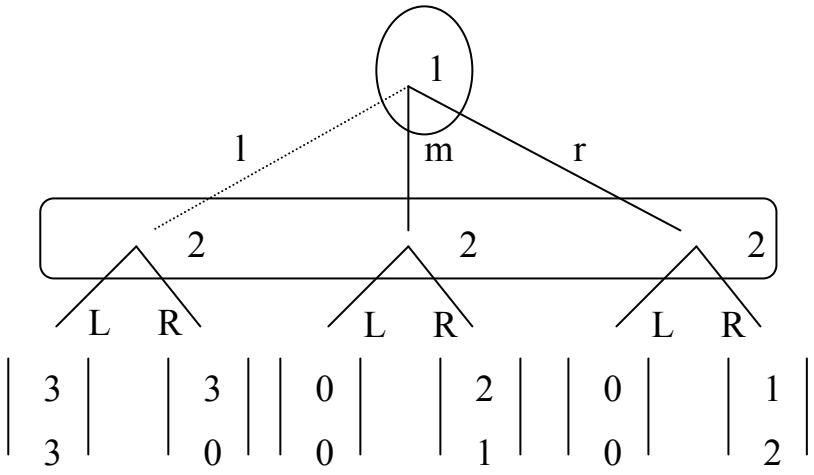

Figure 1:

In the game of Figure 1 player 1 is assumed to know that the game tree describes all possible plays and their resulting payoffs. The dotted lines are meant to indicate that player 2 does not know whether or not player 1 knows that strategy l is available.

Clearly, player 2 should use R if she thinks that 1 will not choose l simply because he is not aware of this option. On the other hand, if player 2 knew that player 1 were aware of strategy l she would expect that player 1 would use l. Given this behavioral expectation player 2 would plan to play L rather than R. Expecting L would not deter player 1 from playing l if he were aware of this option. Thus player 2 knows that she should play L if player 1 were aware of l and play R against an ignorant co-player. Still, since she does not know whether player 1 is aware of l player 2 cannot make up her mind. She does not know which game player 1 analyzes.

Obviously, basically the same problem emerges if the dotted lines in the game of figure 1 are interpreted as indicating that player 2 has the suspicion that player 1 will never use l for some reason or other (for instance an ethical one). This amounts to the assumption that player 2 does not know the other player's utility function (his "type"). According to the basic premise of non-co-operative modeling all relevant features of the game are assumed to be explicit. Therefore any lack of knowledge should be explicitly modeled. And, in fact it can be so modeled. For, ever since Harsanyi made his ingeniously simple proposals to that end (see Harsanyi 1967-8), models of interactions like those characterized by figure 1 can be "closed" by the assumption that "nature" moves first and in a fictitious chance move chooses whether the game with or the game without the dotted options is

played (or in the other case of ethical restraint nature chooses a utility function for one of the players). Afterwards it is revealed to player 1 which game is played while player 2 cannot observe the outcome of nature's move.

After closing the model by an initial move a well-specified game emerges. However, in principle the same questions may be raised with respect to the knowledge of the rules of the latter game, too. In particular it may be asked again whether each player knows what the other player knows as far as the "new" game is concerned. Conceivably this might lead to a setting in which it is again not obvious what the problem under consideration actually is because the dotted lines "come in again" so to say on higher levels of knowledge. If we want to avoid such problems in a general and compelling way it seems natural to require that in eductive analyses, non-co-operative modeling be always pushed to a level on which *common knowledge of the rules of the game* applies, i.e. on which the players not only know the game tree but also know that they know, that they know that they know that they know, ... in arbitrary progression.[3]

Comprising this requirement the assumption of CKR provides a definite format for "normative" game theoretic analysis. Without the assumption of CKR rational individuals could not proceed to arbitrary levels of rational expectation formation on the assumption that all others form their expectations rationally on all levels of reflection according to the same normative theory. If they could only progress finitely many steps there would be a finite stopping point. At that point analysis could not be based on rational expectations but would be cut off since there would be no well-defined problem to which the underlying normative theory would apply. Again, for many issues it may be interesting and useful to define such a finite cut off point. But if we take seriously the enterprise of analyzing what goes on in strategic interaction of ideally rational decision makers who are mutually aware of their own rationality, cutting off their rational reflection at some arbitrary level would seem rather arbitrary and certainly not in line with plausible notions of *ideal* rationality (see on the qualitative distinction between finite and indefinite repetition Rubinstein 1989).

In view of their foundational role it seems clear that in eductive analyses we ought to refrain from such arbitrary measures. Still, ought presupposes can. Since there are quite some claims around that CKR is in itself an inconsistent notion it seems doubtful whether we *can* in fact consistently pursue the project of developing a convincing normative theory of rational choice within an eductive game theoretic framework. Before we turn to this issue let us, however, explicitly state what goes almost

---

[3] See for a more detailed list of what must be commonly known Binmore, K. and P. Dasagupta, Eds. (1989). Economic Organisations as Games. Oxford, Basil Blackwell.

Binmore, K. (1990). Essays on the Foundations of Game Theory. Oxford, Blackwell.

without saying: *Except for extremely simple cases in which a very simple game is presented in a public event and individual preferences are very carefully induced by inter-subjectively recognizable incentives in public, too, CKR rules out any realistic interpretation as literally true*.[4]

3.1.2.2 "Localizing" analysis

3.1.2.2.1 A standard objection against CKR

Some of the strongest, at least most popular objections against the assumption of CKR have been based on variants of the centipede game (see Rosenthal 1981). Fortunately we need not go into the details of a game with a hundred legs. It is sufficient to consider the game tree of figure 2.

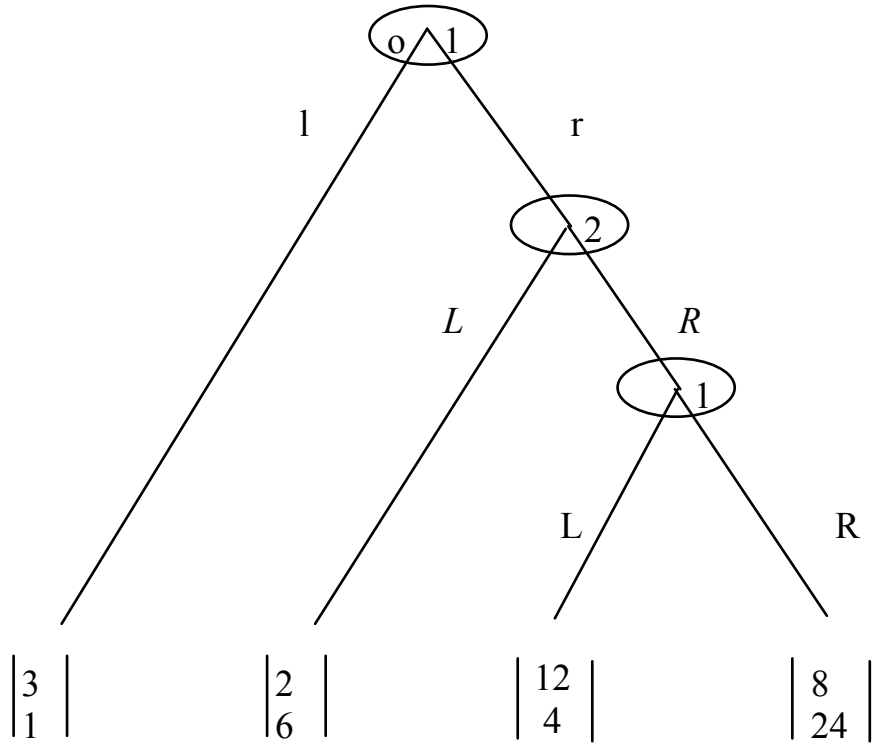

Figure 2

[4] We deliberately put aside arguments that cast doubts on the consistency of theories of fully rational behavior based on self-referential structures or discuss the computability of complex decision problems relying on Gödel's theorem since the underlying approach is quite different from our perspective; yet see, on this for instance Ibid., Canning, D. (1992). "Rationality, Computability and Nash Equilibrium." Econometrica **60**: 877-888., Rustem, B. and K. Velupillai (1990). "Rationality, Computability and Complexity." Journal of Economic Dynamics and Control **14**: 419-432.

The game of figure 2 may be used to illustrate some of the difficulties that seem to stem from the assumption that all players are rational and know that they are up to arbitrary levels. Clearly, if 1 is ever going to choose between L and R we must infer that he is going to choose L (first payoff is that of 1). Anticipating 1's choice of L player 2 will decide on *L* if her information set is ever reached. Finally, player 1 will anticipate these considerations and thus should initially choose l.

According to this process of repeated elimination of weakly dominated strategies players should plan to play according to (l, *L*, L). This vector of plans is self-enforcing in the sense that no player *before* the game is actually played has any reason to change the plan if she expects other players to be rational and to play according to their plans. But players do not only plan to play they actually do play. We may wonder therefore how anticipations of the information generated during the course of playing the game can affect the plans of the players and their expectations about each other.

As long as they play according to their self-enforcing plans observed behavior enforces sticking to the plans. But what if behavioral deviations from the plans occur? Since the plans were formed beforehand under the presumption that all decision makers plan rationally on the basis of expected rational behavior should players then infer from the fact that some deviation occurred that the rationality assumptions underlying their analysis do not hold good?

More specifically, if player 2 in the play of the game of figure 2 has actually to decide between *L* and *R* should she then conclude that player 1 is not rational? After all, according to the preceding analysis rationality prescribes that player 1 should use l and therefore rational play would exclude 2 from making a choice at all. However, now she does have to choose. Should she make her choice on the presumption that player 1 will choose rationally at his next decision node even though she could have reached her decision node only because player 1 did deviate from the rational plan?

In answering such questions we should ask ourselves what kind of a sub-game player 2 should expect to play if she has in fact to decide between *L* and *R*. If after reaching her decision node she cannot anymore know how 1 would choose between L and R then this amounts to assuming that she does not know the rules of the (sub-)game. As in the discussion of the example of figure 1 we then must conclude that the problem is not well specified for the purposes of an eductive non-co-operative game theoretic analysis. To put it slightly otherwise, how could we at all reach the conclusion that the last mover in the game should use L? Either the last sub-game is as depicted in the tree then it seems to follow that L is chosen or it is not, then it does, of course, not necessarily follow anymore that L will be chosen. In the latter case the correct tree should be written down. Once it is written down and depicts the relevant sub-game correctly why should analysis change once the sub-game is actually reached?

In its standard form eductive analysis is based on the premise that the utility functions represent individuals' choices. If player 1's utility function comprises all considerations relevant to his choice at his second decision node and if we assume that player 2 knowing the rules of the game also knows

player 1's utility function then player 2 must expect that player 1 will choose L at his last decision node. If we assume that the model is correctly specified before the game is actually played – that is when the game theoretic analysis of players' planning takes place – no other conclusion seems possible.

If it is assumed that utility functions are not a short hand for action or choice (and are, thus, also not necessarily revealed in choices) but represent preferences (or the order of states of the world resulting after evaluating them) then individuals can conceivably deviate from their preferences by intended actions. Preferences do not directly imply actions. There is an independent layer of choice-making with outcomes that are partly independent of preferences (this is close to a line of argument pursued in Dufwenberg and Lindèn 1996). More specifically, in the tree of figure 2 players know their preferences beforehand. Playing the game and finding herself in the position of making a choice herself the player then does not conclude that the other player has different preferences but rather that the player is not acting according to those preferences.

If player 2 by observing a deviation from the anticipated plan learns that the other player did not act rationally then player 1 either must have made a mistake or must have deviated for a systematic reason. The first alternative is fully compatible with the rational choice analysis of the game including backward induction. Mistakes are events of nature occurring with some (typically low) probability. If they lead into sub-games that would remain unreached according to optimal plans this does not imply the necessity for a revision of plans. For if we take seriously the five assumptions introduced above then the strategies of the players must contain a plan for the contingency of ending off the equilibrium path. After it occurred the mistake is an event of the past. Planning is like the tortoise in the fable of "Achilles and the tortoise"; it has anticipated any possible play and is always already "there" with a plan when the relevant stage of the play is reached. Wherever the play leads, a plan has already been formed and if a full sub-game is reached the plan must have been made independently of the past entirely in terms of that sub-game and forward looking rational choice in that sub-game. In sum, as long as rationality is forward looking and treats bygones as bygones an analysis of games by sub-games is meaningful. It may be necessary, though, to perform all rational planning in anticipation of the possibility of mistakes or "trembles" (see on this Selten 1975).

The second alternative of systematic deviations – as opposed to deviations by mistake – from choices as dictated by preferences does not leave much room for eductive analysis. As long as we engage the task of forming a rational choice model of action at all we must assume that purposefully rational pursuit of aims, ends or values does play a role. But if that is true then it must be possible for fully rational beings to form a rank order among states of the world as emerging from their evaluations. Even if only the ends, aims or values rather than the preferences derived from those evaluations are "exogenously given" eventually preferences would result after taking all those evaluative concerns into account. Even if construed in a complicated process preferences simply represent the order that is

emerging from multi-criterial evaluations after *all* things have been considered. Virtually everything that matters (all value criteria) in ranking expected results of choice-making at any decision node must have been taken into account when writing down the preference order. Since the rank order simply sums up the *complete* evaluation it seems almost analytical that only mistakes can lead to choices deviating from the preference order (unless we are willing to change the nature of rationality itself such that it goes beyond means ends relationships).

In sum, there may be deviations from rational plans. But in a fully specified model they must have the status of mistakes. In the game of figure 2, player 2 anticipating such possibilities as mistakes or trembles must form her (now more inclusive) plan before the game is played. Taking this into account it seems obvious that she for the sake of consistency should plan her reactions to anticipated deviations from rational play in view of the same model that she uses in analyzing the game as a whole in the first place. But then, under common knowledge of utilities – even if the latter are representing preferences expressing evaluations rather than choice – analysis can decompose the game such that player 2 will still use the same utilities when she is to move as when analyzing the game before it is actually played. As a fully rational being she will anticipate that – and the above result from backward induction immediately emerges. Otherwise, contrary to what many critics of backward induction arguments seem to assume, not some new form of analysis based on some new form of rationality would result. Rather the game could not be analyzed at all by decomposition. Since we doubt that there is another convincing method of analysis this seems to imply that there would be no game theoretic analysis at all.

Pointing out here that eductive kinds of modeling under CKR are "unrealistic" is beside the point. Of course, from an empirical point of view the possibility of mis-specification of a model can never be excluded. However, the analysis of a given model necessarily proceeds on the assumption that the model is as specified – which may include updating of beliefs about utility functions in an incomplete information model. But if we assume an incomplete information model then in non-co-operative game theory we must model the latter assumption explicitly. On the other hand, if a game theoretic model is as specified, then the utility functions of the players are as specified and this is the end of the story. Whether the underlying preferences represented by utility functions are interpreted in terms of choice or in terms of evaluative rankings between states of affairs does not matter much. As long as we stick to the premise that in a well specified model at any instance of choice-making we can anticipate *all* "things" that choice-makers will consider in their rankings of outcomes at that instance rational planning seems to imply backward induction (see Aumann 1995).

3.1.2.2.2 Utility and local behavior

The rules of the game – including the utility functions – are not subject to strategic choices of the players otherwise they would not be the rules. To put it slightly otherwise playing the game cannot

change the rules of the game. What non-technically speaking is regarded as a change of rules amounts in general to choosing different structures, e.g. sub-games, of a game. In the last resort such a choice must be based on a set of given rules that do not change when the game is actually played. (Otherwise – according to what has been said before – no analysis of what rational players have on their minds when rationally planning their play could be possible.)

Again one might well say that eductive game theory is of no real interest because its underlying assumptions are so outrageously unrealistic. It is suitable only for an unrealistic world of ideally rational planners and actors. But even if we grant that, it still must be said that if somebody as a matter of fact takes some interest in analyzing interaction between ideally rational decision makers from their points of views he should better play by the rules of *that* game. The latter theoretical enterprise presently incorporates utility representations as a short hand describing individual choice behavior. If we take as seriously as needs be the fact that the utility functions at each decision are "complete" in the sense of including *all* considerations relevant for that decision then nothing that conceivably can happen along any path through a game tree can have an effect on the utilities associated with the decision nodes on and off that path. The representative utilities relevant for any decision node are formed only after all "things" have been considered. After the utility function describing choice (or complete evaluations) at a certain decision node is formed and known to the players, choice at that node is separated from all other choices in so far as the utility function at the decision node reaches over all possible subsequent plays and contains all relevant information (including the anticipation of how it affects our feelings to have actually reached a specific node unexpectedly).

The concept of a “satiated” or “complete utility function” almost amounts to the same thing as accepting the concept of a *local* player throughout; i.e. to associate a separate decision maker with every information set.[5] For, if the rules of the game are supposed to be common knowledge players know all the utility functions characterizing the game. Unless all adopt the same point of view to all decision makers in the game the rules could not be common knowledge. In particular, for every player knowledge of the rules must comprise the utility functions associated with her as a personal player. When analyzing a perfect information game beforehand she must adopt the same point of view as every other player as far as her own decision making in the game is concerned. At each instance of choice or information set the behavioral expectations at this juncture – including the expectations of her own behavior except for strategic considerations – are fully described by the utility function associated with the end nodes originating from the information set. In an imperfect information game

---

[5] Obviously there are some links with the famous Lockean notion of personal identity based on what a present self is aware of.

with private type information what is private and what is common knowledge must be explicitly modeled and in that sense again be commonly known.

Modeling on the basis of representative utility functions thus seems to imply that for the purposes of non-co-operative game theoretic analysis even the Harsanyi-Selten standard form (see Harsanyi and Selten 1988, chap. 2) goes too far. It seems that employing the agent normal form based on the concept of local players is already implied by the very notion of a classical utility index and the concept of analyzing a game by analyzing its parts. From this we may conclude that in non-co-operative game theory the agent normal form is not "a" but rather "the" adequate or canonical representation of strategic interaction in non-extensive contexts (see on this also Güth and Kliemt 1995).[6] Likewise the corresponding model extensive of games with one utility function for each decision-maker at each information set is the canonical extensive form.

3.1.2.2.3 Local decision making and CKR

If the preceding argument is correct the introduction of the agent normal form – and the corresponding extensive form with one decision-maker for each decision – is in no way ad hoc. Reliance on the notion of representative utility already dictates that non-co-operative game theoretic analysis should be based on the agent normal form. This has further implications since common objections against CKR lose some of their force then. In particular the view that CKR may lead to self-contradictory conclusions with respect to decisions at decision nodes that could never be reached under rational play would seem much less compelling. If we accept that introducing representative utility already implies that strictly speaking there are only local players then preceding and subsequent decisions in a game tree are never made by the same decision maker and thus – as far as the model is concerned – the decisions of a former cannot cast any doubts on the rationality of a later agent.

---

[6] It may be worth noting again that the preceding argument merely takes seriously the completeness or the all-inclusive character of the modern notion of utility. Unless utility functions represent all relevant considerations at any particular instance of choice the rules of the game cannot be common knowledge. Only if the latter holds good the strategic problem can be minimally well defined. Only then all players in their strategic planning focus on the same problem. Since their own expected choices are characterized by utility functions too their perspectives actually coincide with each other and that of an external analyst. This, in turn, shows that relying on a notion of representative utility to a large extent drives out what has been called the "internal point of view" of the choice-maker but not completely so since analysis is still based on the assumption that every agent in the game knows the rules and anticipates all potential consequences of his decision making against this background of common knowledge.

More specifically, consider figure 3. This figure is derived from figure 2 by listing players' payoffs in the order of sequential play. Since now modeling is based on the concept of local players the game involves three rather than only two players. Player 1 is split into two agents one – 1a – choosing between l and r and one – 1b – deciding between L and R. The two agents are linked or associated with each other by the fact that the utility functions characterizing their choices are both identical with that of the personal player 1 in figure 2. One should note that due to the sequence of moves the payoffs, though identical for the agents, are not of the same relevance to them. For instance the payoff "3" for 1a after choosing left is irrelevant for a forward looking rational agent 1b.

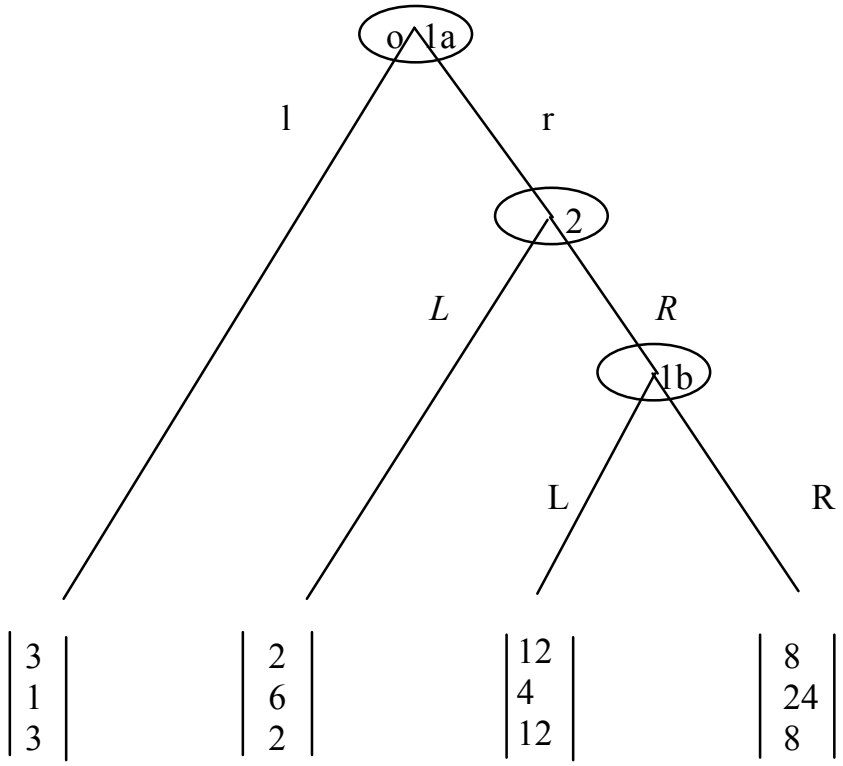

Figure 3

Whatever agent 1a in figure 3 does it will not reveal information about the rationality of agent 1b since the latter is viewed as an independent choice-maker. This way of modeling may seem artificial and provoke the obvious objection: Why should we split up personal players in this way? Or more loosely speaking: Why should anybody buy this? But, as we have argued before, those who accept the concept of (satiated) representative utility of completely specified preferences have already bought it. If anybody assumes that individuals' choices are characterized completely – all things considered – by a utility function then he has already accepted something that amounts almost to the same thing as using local players and the agent normal form as building blocks of eductive non-co-operative game theoretic modeling.

Again, we do not deny that analyses of interactions that are based on game forms or, for that matter, on "material" (money, pleasure, offspring etc.) rather than on utility payoffs might be useful. Quite to the contrary, we think that such game theoretically inspired forms of analysis can be most fruitful. But it should be noted, too, that these kinds of analyses belong to a different theoretical enterprise that has some links to but is not identical with an eductive analysis of rational players' planning under common knowledge of completely specified rules.

In any case, as long as we stick to the standard game theoretical interpretation of the utility function as an all comprising short hand those problems of CKR that are related to assumptions of personal identity would seem to be of very minor importance. But the problems commonly associated with CKR are not only rooted in our concept of personal identity. Similar problems can be construed for the agent normal form case and its corresponding extensive form in which every player makes exactly one choice.

So, for the sake of the argument, consider only games in which all personal players can make no more than one decision from the outset. In specifying his plan of rational play each decision maker must specify a decision and form expectations for all contingencies. But the contingencies off the path of rational play prescribed by the normative theory require that expectations for other individuals' play are formed on the basis of a theory that seems to be falsified by the very fact that the contingency arises. Doubts are raised about the theory rather than about the players now. Is it not necessary to revise the theory if for instance an information set and in particular a sub-game was reached even though it should not have been reached according to the theory?

If the theory of rational play of an embedded sub-game off the solution path is to be different from the theory of the stand-alone sub-game that emerges after the preceding game is cut off or did not exist in the first place then "analysis by parts" is rendered impossible. But this is implausible since according to the five assumptions of game theoretic modeling of choice-making introduced above, the future directedness of rationality clearly implies that "analysis by sub-game" – off and on any solution path – should be possible. That the sub-game contains the "complete future" is a very strong philosophical reason why a theory of "sub-game play" should not be revised after the sub-game has been reached off the equilibrium path. Moreover, claiming that the theory of rational play has been falsified if games off the solution path specified by the theory itself are reached, one must be quite precise about what exactly is falsified if anything at all. A prescription as such cannot be falsified by the fact that its addressee does not comply with it.

3.1.2.2.4 Sub-game consistency

If one can solve sub-games of a larger game completely independently of their embeddedness this has far reaching consequences in particular in the case of so called super-games; i. e. games emerging from an indefinite repetition of an identical, so-called "normal" or "base" game. Every indefinitely

repeated game with perfect information is isomorphic to each of its proper sub-games after cutting off at most finitely many initial rounds of play. Therefore the game as well as its isomorphic sub-games should be solved isomorphically by an adequate solution theory.[7]

The criterion of "sub-game consistency" captures the requirement of solving isomorphic games isomorphically (cf. for the preceding argument and a more rigorous definition of isomorphic games also Güth, Leininger et al. 1991).[8] Intuitively sub-game consistency seems a very appealing requirement in super-games of the preceding kind. For, after cutting off the past, every sub-game of a game that is repeated indefinitely looks identical and contains the whole future form its initial node. It is isomorphic to the whole game. Thus different plays up to any decision node cannot have an influence on the rational plan of future play in such a game. Bygones are bygones and therefore identical futures regardless of how they are reached must induce identical strategic plans as endorsed by fully rational players who analyze that future.

Of course the argument applies to finitely repeated games as well as to infinitely repeated games. The crucial point in both cases is the *consistency requirement that identical sub-games on and off the solution path must be solved in identical ways regardless of the history of play in a larger embedding game.*[9]

Structurally identical sub-games must be solved in identical ways regardless of embeddedness. Yet to know that the solution will be invariant in this respect does not amount to a specification of the solution. Something must be said on what a solution should be like. And, within our approach the obvious answer is it must be an equilibrium *and* it must be unique. Since the latter implies that the

---

[7] Isomorphisms allow for positive affine linear transformations of the pay-offs and some other variations that are structurally irrelevant. We shall not go into the specification of what is and what is not structurally irrelevant here.

[8] The requirement of sub-game consistency also makes sense in view of the above mentioned background assumption that rational actors plan all their choices in view of the causal consequences of those very choices. Players plan and play in strictly future directed ways. If two "futures" do look exactly the same plans for both futures should coincide. What players did in the past should not influence their future directed choices unless leading to different "futures". Making different plans for isomorphic futures depending on the past -- like in so-called Folk Theorems -- does not make sense under the idealized rationality assumptions of eductive analysis.

[9] Conditional strategies that select an equilibrium in a sub-game in a way that makes that choice dependent on the play before entering that sub-game would not have the relevant property of invariance of solutions with respect to preceding play. This rules out even sub-game perfect equilibria of the folk-theorem kind as well as any forward induction argument.

equilibrium selection problem must be solved and it has been claimed that this is impossible some real problems seem to be in our way here if we insist that there can be a reflective equilibrium on the explication of fully rational behavior in interactive situations emerging in a world of ideally rational beings.

3.2 Equilibrium

3.2.1 The notion of equilibrium

Initially Cournot gave a justification of equilibrium behavior in terms of an adaptive process (though converging to equilibrium only for special classes of games, see Cournot 1838). His notion is about playing rather than about planning. Or to put it slightly otherwise, Cournot's vision of adaptation is behavioral rather than deliberational. Players are assumed to act and to adapt myopically in a process of trial and error.

Although "behavioral adaptation" is nowadays often substituted by "learning" or "evolution" current adaptive justifications like their predecessor still refer to a process that is so to say "pushed rather than pulled". It is "driven" by past experience rather than "drawn" by plans for the future. On each round of adaptation the preceding stage(s) of the adaptive process determine its next stage. Forward looking anticipation is basically left out of account. In this regard adaptive or evolutionary justifications for the central role of the equilibrium notion differ fundamentally from those that are cast in terms of self-stabilizing expectations.

Adaptive justifications can be very instructive but since we are presently interested only in eductive analyses of choice-making we shall leave aside the adaptive line of argument. Focusing on the eductive approach consider an agent-normal form game $\Gamma = (S_1, S_2, ..., S_n; u_1(.), u_2(.), ..., u_n(.))$ with i=1, 2,..., n decision makers, their strategy sets S_i and their payoff functions u_i. As indicated above we assume that each "personal player" is split into several agents with identical preferences.[10] Since there is a separate agent for each decision to be made all decision makers (agents) move at most once and thus have to consider but one contingency in their planning. An n-tuple, vector or profile of individual strategies, $s_i \in S_i$, is denoted by $s=(s_1, s_2, ..., s_n)$. For all $\underline{s}$ and any s_i the complementary n-1 tuple of strategies $s_{-i} = (s_1, s_2, ..., s_{i-1}, s_{i+1} ..., s_n)$ can be formed. Evidently for all s we have $s=(s_i, s_{-i})$;

---

[10] The assumption that only the causal consequences of a praticular decision matter, at least in a way, isolates each decision from the other decisions. This provides another reason for relying on the agent normal form. It may be noted incidentally that this way of modeling naturally allows for incomplete information about the future preferences of a personal player composed of several agents and thus for reasonable ways to analyze intra-personal conflict.

analogously, for any subset M of the set of players N we can form –M:=N\M and pairs of complementary strategy tuples: $s=(s_M, s_{-M})$.

The equilibrium concept for any game with N≥1 players then can be formulated as follows:

A profile $s^* \in S$ forms an equilibrium in Γ iff for all individuals $i \in N$ and for all strategies $s_i \in S_i$ the best reply condition $u_i(s^*_i, s^*_{-i}) \geq u_i(s_i, s^*_{-i})$ is met or more formally:

$[(\forall\ i)\ (\forall\ s_i)]\ [u_i\ (s^*_i, s^*_{-i}) \geq u_i(s_i, s^*_{-i})]$.

On the basic level of analysis only the question of *unilateral* deviation from theoretically predicted behavior is meaningful. For, in non-co-operative game theory, by assumption – namely that of complete specification of models – players are "all on their own". Unless we have pushed our eductive analysis to the point where all players can meaningfully be assumed to be informationally isolated we have not yet formulated a completely specified non-co-operative game model. In a completely specified non-co-operative game model it is assumed that *all* possible information flows are explicitly modeled (for instance a correlated equilibrium introduces by assumption privately observable signals, see Aumann 1987, that should be explicitly modeled). After *all* possibilities have been modeled there are, of course, no further possibilities to exchange information, to send signals etc.

In short, on the ultimate level of analysis or in a *complete* game model players are informationally isolated. In that sense decisions of single individuals and thus solutions for one agent or single player games can be regarded as fundamental for the whole game theoretic enterprise. So let us turn to those.

3.2.2 Solutions in equilibrium?

3.2.2.1 Solutions for games with one player

It seems quite obvious what a theory of opportunistically rational plan and play requires from a single actor in a one-player game Γ [11]: A single player should plan on making a utility maximizing choice.[12]

---

[11] Since by assumption all games are presented in agent normal form decision problems that conventionally would be modeled as involving a single personal player making several decisions must be modeled as games with more than one player (agent). The same applies to the corresponding extensive representations.

[12] Of course, in executing the plan she could make mistakes with a certain probability. This probability is determined by "nature". It is not the outcome of a (strategic) choice. Therefore it cannot itself be planned. But the likelihood of making mistakes can be taken into account in the theory of rational plan and play provided that it or at least some distribution of it is known. Since we assume that a fully specified non-co-operative game model exists the likelihood of making mistakes in executing plans by assumption must be known to the single

Only a utility maximizing strategy can qualify as a (rational choice) solution of that game. This is expressed by

> Axiom 0: If a game Γ has exactly one player (agent), then a strategy of the single player (agent) qualifies as a solution of Γ only if it is utility maximizing.

Axiom 0 secures individual rationality in the material sense of choices that are undominated in the preference order of the choosing individual. There can be several undominated alternatives, though. In that case a non-singleton solution set that comprises them all can be formed as expressed by

> Axiom 0*: The solution set $E:=\{s \mid (\forall\ s')\ (u(s)\geq u(s')\}$ of a one player (agent) game Γ is the set of all strategies maximizing the only player's utility.

It is obvious that all elements from E are equilibria of the one player game Γ and all equilibria of that game are also elements of E. Thus, in the one player case rational choice solutions and equilibria coincide.

Individual optimization in a world of interdependent choice-makers means that the Axiom 0* must be reformulated to yield

> Axiom 0**: For all plans s the solution set E_i of all players i of an n-player (n-agent) game Γ with n>1 in which all choices s_{-i} of players $j\neq i$ are fixed is the set of all strategies maximizing player i's utility
>
> $\forall i \in N_n: E_i:=\{s_i \mid (\forall\ s_i')\ (u(s_i, s_{-i})\geq u(s'_i, s_{-i})\}$
>
> given the choices or plans s_{-i}.

Here the behavior of other individuals is "parameterized" to yield s_{-i}. As long as we can assume for each player i that his beliefs may be summed up in some probability distribution or other over the set of complementary strategies s_{-i} the presence of other players would induce a complication but not a qualitative change of the analysis. All individuals i could conceivably treat all others as parts of nature and then optimize against subjective probability distributions over the sets of s_{-i} describing their co-players possible behavior. Whether one adopts such a parametric view of games with n>1 strategic players or subscribes to a stronger "eductive" concept of strategic interdependence a solution concept for interactive situations with more than one strategic player must be found. We focus here on the solution concept for eductive analyses.

---

agent or player as part of the rules of the game (see Selten, R. (1975). "Reexamination of the Perfectness Concept for Equilibrium in Extensive Games." International Journal of Game Theory **4**: 25-55.)

3.2.2.2 Solutions for games with more than one player

3.2.2.2.1 The philosophical difference between games against nature and strategic games

According to the assumptions underlying eductive analyses players conceive themselves and all other players as members of a world of fully rational beings. They also presume in their reasoning that all other players are reasoning as they do. The rules of the game being common knowledge they all know that they are drawing inferences from the same mental representation of the interaction situation. But if there are more than one player (agent) who all are thinking through an interaction situation under common knowledge that they do new problems emerge. If all reason about the reasoning of others then "predictions" of other players' choice-making seem to be more difficult or even be ruled out. For each and every *participant* of the game the ancient problem of "what I think that you think that I think" emerges. What “I” should rationally plan to do crucially depends on what “you” think that “I” expect “you” to plan etc. In short, rational utility maximizing choices can be made only under some presumption about the choices and plans of the other players that can be made only under the same presumption etc. When initially forming their rational plans the plans of all players do not exist yet. Since non-existent facts cannot be "known" in the strict sense of that term the *initial* formation of plans encounters an obvious and seemingly insurmountable difficulty if we stick to the basic game theoretic premise that ultimately we have to start from commonly known facts.

In their efforts to overcome the difficulty of formulating initial beliefs many theorists regard it as obvious that planning should start with some kind of initial probability distribution over the plans of other players. This probability distribution is seen as a prediction of what the other players shall do. As a "prediction" in the standard sense of that term it must be based on some (probabilistic) psychological or behavioral *laws of nature*. Since these are commonly known, there are "common priors". Players who according to the initial assumptions discriminate between what they can and what they cannot influence will regard priors as beyond their strategic influence and thus treat them as part of the rules of the game or like a move of chance of "mother nature" whose probabilistic laws are commonly known among the strategic actors.[13]

For instance, to refer to but one important recent example, Aumann and Brandenburger in their analysis of "epistemic conditions for Nash equilibrium" use the term "theory" for such a distribution derived from some common prior (see Aumann and Brandenburger 1995). However, this is not only a somewhat unusual use of the term “theory” it also implies that there is no categorical distinction between games against nature and strategic games proper. For, according to this view strategic multi-

---

[13] If priors are treated as being determined by the rules of the game the common prior assumption seems a "natural" consequence of the assumption that the game model is commonly known.

player games are just more complicated than one-player games but the same in kind. All games are ultimately games against nature.

We do not deny that game theory could conceivably be restricted to analyses of games against nature. But we think that the original spirit of game theory was clearly different. It was much more "Austrian", "subjectivist" or, for that matter, "eductive". And from an eductive point of view, the planning and choices of players cannot be subject to a standard prediction since we cannot regard other players' planning as an event of nature *emerging independently of the theory of rational planning itself*. We, as well as the players, must rather base inferences about the planning and playing of players on inferences from the *theory of rational plan and play*. Other than a theory in the Aumann/Brandenburger sense that is based on or describes information that is given independently of the theory, the eductive theory of rational plan is *constitutive* of what players shall plan rather than describing some theory-independent facts.

In an eductive approach a player must treat all other players as *theorists* who *draw inferences* from the same *theory* as she herself. If there is a genuinely eductive approach at all – which one might want to deny[14] – then that theory cannot simply amount to a descriptive theory of what is likely to happen (or a probability distribution summing up all such information about nature's "propensities"). The theory must rather be normative in the sense that it itself lays down what rational planning and playing amounts to (typically under the additional premise that all plan accordingly).

In short, the eductive theory does not and cannot *describe* the rational plan and play of others as a fact emergent independently of itself. Expectations cannot be formed on the basis of common priors that are based on laws of nature but must be based on the theory of rational plan.

That individuals are theory guided does neither imply any strong philosophical assumptions about a separate mental reality nor about acts of choice or free will as "roots of new causal chains" etc. Automata could be programmed and in that sense be guided by theories as well. Essential is the assumption that plans and play are guided by the theory of rational choice itself and that the content of the theory is *not* going to be predicted or to be derived from our knowledge of natural laws alone but always needs to rely on the theory itself.

"Initially" the theory that is assumed to be present in the plans of others does not "yet" exist and thus cannot be known before it is formed. This confronts us with a puzzle: On the most fundamental level of eductive analysis of rational plan and play we need the theory for its own formation. In formulating that theory we must proceed on the assumption that plan and play of the players are guided by the theory and thus that the theory already exists when it is formulated. – In theory-based eductive

---

[14] We have nothing to say on this issue since it is our working hypothesis that there should be such a thing as eductive game theory and we try to take this hypothesis seriously and to its logical conclusions.

analysis there seems no way to avoid the precedingly sketched circularity. One must try to cope with it in ways that prevent the circularity from becoming vicious. Next we shall explore some such ways.

3.2.2.2.2 Unique rationality

As already stated, in non-co-operative theory games must be analyzed by (and "for") "informationally isolated" players. Moreover, since players cannot commit themselves beforehand all choices must be made when the game is actually played. In other words, when the players are planning their play their choices do not yet exist. Therefore, strictly speaking, at the planning stage choices cannot be (commonly) known. What could conceivably be commonly known is the (normative) *theory* of rational plan and play. Common knowledge of that theory can induce common knowledge of (sets of) rational *plans* among players who follow the prescriptions of the theory.

For the time being let us set aside the issue of relying on weaker conditions than common knowledge (see again Aumann and Brandenburger 1995) and assume that for a class G of games Γ a commonly known theory of rational plan and play exists. If the theory implies a *unique prescription* for every player in any given game Γ from a class of games G we refer to it as *definite on G* or a *G-definite* theory. A theory that does not yield a unique prescription for every player for at least one game $\Gamma \in G$ may be called G-non-definite or non-definite on G.

Now, consider a G-definite normative theory and a game $\Gamma \in G$. Assume that the G-definite theory uniquely prescribes a profile of individual plans $s^* \in S$ such that there are an i an s_i with $u_i(s^*_i, s^*_{-i}) < u_i(s_i, s^*_{-i})$. Assume further that players know that player i is rational in the sense of axiom 0**. Then, since by assumption the theory is commonly known among the players, all know that player i would plan to play s_i rather than s^*_i. The profile $s^* \in S$ can not be a solution. The argument applies to any i and thus to all or $\forall i \in N_n$. We have thus shown

> Proposition 1
>
> If it is commonly known that a specific G̃definite theory is applied to $\Gamma \in G$ by all players in forming their plans and expectations of rational play before $\Gamma \in G$ is actually played they will simultaneously plan according to the theory's prescriptions only if the prescriptions entail an equilibrium.[15]

In other words, unless a commonly known G̃definite theory of rational play meets the test of leading to an equilibrium in the game $\Gamma \in G$, it cannot lead to an *absorbing state* in the reflective or intellectual search process in which fully rational players who know of each other that they are

---

[15] For a parallel argument see Binmore, K. and P. Dasgupta (1989). Game Theory a Survey. Economic Organizations as Games. K. Binmore and P. Dasgupta. Oxford, Basil Blackwell., 5

rational seek to answer the question of how they should play (see on theory absorption Morgenstern 1972; Morgenstern and Schwödiauer 1976, Dacey 1976; Dacey 1981, Güth and Kliemt 2000). At least according to a purely instrumental or "means-ends" view of rationality, in the preceding case player i would have good reason to reject the theory itself: Accepting the theory would lead to sub-optimal results for him on which he could unilaterally improve. Therefore a G-definite theory not leading to an equilibrium could not conceivably be accepted as an explication of the notion of strategic rationality among actors who are rational in the minimal sense of 0**. In other words: *If* (normative) game theory proposes a unique rational solution for a game, then it must be an equilibrium (cf. on that also Sugden 1991, Pearce 1984).

Since we know that at least one equilibrium must exist for any finite game and since we assumed that all $\Gamma \in G$ are finite we know that G-definite theories can always be formed in principle. However, though the argument supporting the preceding proposition is simple, it is burdened by the uniqueness assumptions on which it rests. First, the theory must yield a unique rational plan for every $\Gamma \in G$, that is, it must yield at least one and not more than one plan. Second, the G-definite theory itself must be uniquely selected. We shall first discuss the possibility of theories that not always select a unique plan and then turn to cases in which the problem of selecting among (G-definite) theories emerges.

3.2.2.2.3 Choosing among plans and theories

3.2.2.2.3.1 Theories that not always select a unique plan

There could be less or more than one plan for some games. As far as the first possibility is concerned we shall assume – as seems plausible for a world of ideally rational beings – that for each player there is always at least one plan. After all, rational planners facing the game know that they must play the game and thus eventually must make one choice. Given that knowledge rational players will form a plan and at the same time will not plan on something they deem irrational. In that weak sense there should be at least one rational plan for every player. As far as the second possibility is concerned the argument of those who deny that solution functions must select equilibria then hinges on the fact that solution theories might select more than one solution. We doubt, though, that the development of non-definite theories is a "life-option" within an *eductive* analysis of players' planning. For it could always plausibly be argued that an eductive analysis that does not lead to a theory that selects a definite solution is simply incomplete.

To see why that is so, consider a non-definite theory on a class of games G. There would exist at least one game Γ and at least one player i in Γ for whom at least two different strategies were considered rational plans. Now, eductive game theoretic analysis is about plans not about choice. Choice is based on plans. If there are random choices then at the theoretical level of analyzing rational players' plans beforehand players would have to plan on using the natural mechanisms leading to such "spontaneous

choices" as well. In particular, if – due to some natural factor or other – there are in fact "mixed" strategies then those "mixed" strategies must be taken into account by rational planners as separate pure strategies. There should be a random mechanism that can be chosen as a separate strategic option.

To guarantee the existence of equilibria it needs to be assumed that fully rational players can in fact choose a random mechanism such as to reach any mixed strategy in the convex closure of strategies. If such a strategic choice is to be made, then in a full analysis this separate decision must be assigned to some agent in the agent normal form of the game. We make the assumption here that fully rational players – as part of their rationality – have access to a pseudo random number generator of arbitrary complexity that will generate probability distributions of their choosing – if they plan on this.

In principle, we could go on in the precedingly sketched way until – in the theory of solving games of class G – unique plans emerge. If, as we assumed for the sake of the present argument, all players apply the same commonly known theory of solving games of class G then the unique plan for every player would be common knowledge as well and could persist only if leading to an equilibrium. Let us therefore turn to the second uniqueness assumption and ask to what extent relying on the equilibrium concept can be justified even if the uniqueness assumption on the level of theories cannot be justified.

3.2.2.2.3.2 Selecting among definite theories

If the set of G-definite theories contains more than one G-definite theory that is in line with the minimum requirements of rationality then what is deemed rational in strategic interaction over G can in principle itself be *chosen*. To put it slightly otherwise, the G-definite theory chosen to some extent defines the notion of rationality over G.[16] If we assume that there is no unique commonly known G-definite theory and thus no unique commonly shared and known concept of rationality for strategic interactions as modeled by games from class G we must account for the possibility that different players might endorse different concepts or (G-definite) theories of rationality.

Take the simplest case of a two person bi-matrix game played by two players (cf. for the following line of argument Jacobsen 1996). More specifically, let $G = ((A, B), N, S_1, S_2)$ be a class of bi-matrix games with pay-off matrices A, B, player set N={1, 2} and strategy sets S_1, S_2. Let T^1 be the theory of rational play endorsed by player 1 and T^2 the theory endorsed by player 2. It is assumed that the game is common knowledge. But it is not assumed that the theories of rational play are commonly

---

[16] We neglect here the possibility that the option of choosing a theory may amount to choosing something akin to a pseudo-random choice generator. Using a complicated theory may be almost equivalent to throwing dices for an outside onlooker and therefore close to the choice of a random device in the preceding case.

known. Moreover, both players being engaged in eductive analyses must "predict" their own as well as their co-player's choice according to the (normative) theory of rational plan and play itself. They must use the theory to emulate their co-player. But which theory should they use?

In the subjective interpretation of the "symmetry assumption" according to which players ascribe the same type of rationality to each other, player 1 *thinks* that (ultimately) player 2 uses the *same* theory of rational play player 1 uses. Likewise player 2 thinks (believes) that player 1 uses the *same* theory as player 2 does. Both players know their own theory of rational plan and play. Therefore, within an eductive approach two players who by assumption conceive themselves as rational in the same way (after choosing the G-definite theory) should simply use their own concept of rationality – as defined by their *own* theory of rational play – in emulating the thoughts of their co-player. For each player, subjectively, there is only his concept of rationality and if he must ascribe not only rationality but the *same* rationality to the other player he is stuck with the theory of rational play that he in fact does himself employ.

In analyzing a specific game $\Gamma' = ((A, B), N, S_1, S_2)$ each player knows that he eventually will play the game either as player 1 or player 2. However, within an eductive analysis each player must put himself into the shoes of both players. Applying his own theory of rational play to both roles for the purpose of predicting the plan of each of the players yields a complete plan comprising both roles. The first player's analysis yields $(\mathrm{T}^{1,1}(\Gamma'), \mathrm{T}^{1,2}(\Gamma'))$ while the second player's application of his theory of rational plan and play yields the complete plan $(\mathrm{T}^{2,1}(\Gamma'), \mathrm{T}^{2,2}(\Gamma'))$. Since both players know the G-definite theory applied in both roles and certainly know that they know it etc. Proposition 1, obviously applies and rational players would endorse only such G-definite theories that select an equilibrium.

The preceding analysis does not exclude the possibility of a "false consensus", however. Though each player assumed that he and his co-player both used the same G-definite theory each may have made that assumption for a different theory. Without a convention guiding choices among G-definite theories it may well be that as a matter of fact $T^1 \neq T^2$ and no equilibrium will be implemented even though both players subjectively intend to implement an equilibrium by their choices. Only if the game Γ' has exactly one equilibrium we can infer from the preceding argument that the subjective theory choices of the two players must lead to the choice of this equilibrium.

In the case of a single equilibrium of Γ' the possibility of a "false consensus" does not matter since the theories – even though they may differ in other respects – must lead to an equilibrium choice. In case of several equilibria this argument does not apply. Objectively the subjective equilibrium choices can differ. What is intended subjectively to be equilibrium choices can in fact or objectively lead to the implementation of non-equilibrium results. If players are aware of this, then, when playing a game $\Gamma \in G$, they do not believe to know which G-definite theory their co-player endorses. That

each follows a unique prescription is commonly known but it is not common knowledge among the players which theory each of them uses. They cannot anymore “naively” ascribe their own "favorite" G-definite theory to each other but must "push up" their analysis up to the level where G-definite theories are chosen or selected.

On this level the possibility space will be narrowed down somewhat by principles of theory formation. A priori considerations of several kinds may also influence which theory may prevail. Still, there is no guarantee that all players will be guided by the same theory.

For example, imagine a simple symmetric two by two battle of the sexes game. This game has two pure and one mixed equilibrium. Mixings would be modeled as additional strategies. Neglecting the additional strategies representing "mixing" it is a priori open which strategies should be chosen. The theory must somehow arbitrarily select one of the equilibria. Which one that will be cannot be deduced by any a priori reason. How could a theory deal with this problem? Trivially a theory could specify that always the main diagonal equilibrium with the lowest strategy numbers should be chosen. But is this a meaningful way to go about the problem of equilibrium selection in eductive analysis? At least it would not be invariant with respect to isomorphic renumbering. Such invariance would be possible only if mixed strategy equilibria would be included.

If we cannot deduce from first principles which G-definite theory is appropriate this does not imply that eductive analysis does not make sense at all. Quite trivially, eductive arguments *relative* to a commonly known G-definite theory do make sense once such a theory exists. Moreover, a reflective equilibrium can legitimately refer to conventions that actually prevail while allowing for their critical assessment in "rational" argument. All the discussions about the "pros and cons" of alternative solution concepts should in fact be regarded as meta-theoretical efforts towards establishing or altering *a convention of rationality*. Rational argument can be treated as one of the "forces" that bring about changes in the search for reflective equilibrium. But in the end all depends on which of the theories will be absorbed by all rational players.

From the external point of view of an outside onlooker the absorption process may be regarded as "adaptive". Nevertheless, subjectively or from their internal point of view the participants in the collective search process for a reflective equilibrium are engaged in argument. They use evaluative standards of what should be the case and not simply adapt to what is in fact the case. Though what is deemed rational is *created* rather than discovered what emerges is not (completely) arbitrary. It is constrained by certain "a priori" evaluative standards that we apply in our discourse about what is rational. Before we go on let us sum up our discussion up to this point.

3.2.2.2.4 Intermediate result

According to the preceding an eductive theory of fully rational plan and play in strategic interaction needs to focus on theories that select unique equilibria. These theories themselves define what

rationality in such interactions does mean. The only a priori criterion for the theories is derived from Axiom 0** which implies that definite theories must lead to the selection of equilibria. However, the problem which definite theory is to be used as defining the meaning of rationality may itself be contested. Reasoning from a priori principles alone cannot guarantee which theory will prevail in a process of theory absorption among fully rational individuals. However, once the process of theory absorption has reached an absorbing state in which a unique theory prevails isomorphic game like sub-structures like sub-games or cells of any game must and can be solved uniquely according to that conventional rationality independently of any larger game context. The unique solution can then be substituted for the whole sub-game or cell to yield some kind of truncated game and analysis of games as an iterative process becomes viable. In this process the most elementary "strategically closed" problems are solved first, then the problems of the next level etc.

We think that this is as should be in a world of fully rational beings. Once coordination on a theory of rational choice-making has emerged rational choice-makers who conceive of themselves as members of a world of purely rational beings can solve any problem of choice by reasoning according to the theory.

So much for the limits and capabilities of pure reason. But are there perhaps not some other criteria of an a priori nature that could be used? It has been argued, that consistency requirements going beyond sub-game consistency are desirable. These consistency requirements rely on the assumption that arbitrary fixings should be possible. They are, however, inconsistent with equilibrium refinements or selection. It may seem that there is a deeper paradox here that could subvert classical game theoretic analyses of choice-making. As we shall argue next this is not true, however. Classical game theory can and should plausibly reject consistency notions that require consistency under arbitrary fixings. The requirement of arbitrary fixings is itself arbitrary in that there is no systematical philosophical reason in its favor.

4 Non-arbitrary and arbitrary fixings

In every theoretical endeavor beyond logical consistency some general principles of theoretical consistency of a more or less formal kind must be met. In particular we require that (relevantly) isomorphic problems or "structures" be treated isomorphically by our theories. If we aim at methods of analysis that can decompose larger into smaller problems such invariants must apply independently of whether or not a problem is embedded in a larger one. Unless invariant in that sense we could not conceivably analyze a problem by parts (such methodological insights are clearly embodied in parts 2 and 3 of Harsanyi and Selten 1988). The aforementioned game-like sub-structures of games that, like in particular cells in normal and sub-games in extensive game representation, are in one sense or other strategically closed are of particular importance here. We again start with the intuitively

plausible example of sub-games and only afterwards turn to somewhat more general considerations about cells.

4.1 Sub-game consistent fixings of player choices

As we have argued before, in case of conventional non-co-operative game theory the solution of sub-games is assumed to be invariant with respect to the precise location of such games in any larger game. In view of the crucial role of the assumption of future directedness in rational choice theory there is a strong substantial rather than merely formal reason to go for the specific requirement of sub-game consistency. Philosophically speaking a sub-game contains the "full future" of the game as is open once the initial node of the sub-game under consideration has been reached. A sub-game is closed with respect to all causal influences still possible when the past up to its initial node has been fixed. Such causal closure suggests that fully rational individuals who act in a forward looking manner and treat bygones as bygones would analyze games by their sub-games and would reach the same conclusions concerning a sub-game independently of where in a game tree a sub-game is located. Solutions of sub-games should therefore be invariant with respect to the games of which a sub-game is part. One should not overlook though, that this rules out all kinds of forward induction. In particular, sub-game perfect solutions that dependent on the history of previous play prescribe selecting different strategies for otherwise isomorphic sub-games are ruled out by the future directedness of opportunistic rationality.

The first three background assumptions about opportunism seem to imply that it is legitimate to treat choices in sub-games as fixed in the sense of not depending on other anticipations of future effects than already contained in the models. In the context of sub-games we may substitute unique solutions of a sub-game whenever they exist in any analysis of the full game without changing the solution. This means that the expectations of what happens in the sub-game as formed by a G-definite theory may be treated as if derived from natural laws or as if exogenous to the theory when the theory is applied to the larger game of which the sub-game is a part.[17]

Analyzing by sub-game amounts to counterfactually fixing the choices of a subset of the agents in the game in two senses. First, when the sub-game that commences with a specific information set is

---

[17] Though, as in the reasoning leading to the introduction of the axiom 0** one might have second thoughts about leaving out the information that the result is the outcome of strategic interaction we think that it is justified to neglect that information since by construction a sub-game is causally closed from the starting point and only expected causal effects of choice matter for a rational choice-maker. Therefore, in the case of sub-games what amounts to "truncation consistency" in the Harsanyi and Selten approach should clearly apply, see Harsanyi, J. C. and R. Selten (1988). A general theory of equilibrium selection in games. Cambridge, Mass., MIT Press., 101.

analyzed independently of the preceding game this is based on the assumption that the sub-game has been reached. This way of planning for the contingency that the sub-game will be reached implicitly fixes the choices in the preceding game to a specific play leading to the sub-game. The requirement that the analysis of isomorphic sub-games must lead to the same solution regardless of where a sub-game is located in a game tree amounts to imposing a plausible form of consistency. Second, when the solution of a sub-game – the result of the analysis – is substituted for the initial node of the sub-game then this implicitly fixes the choices of the agents in the sub-game for the further solution process. Here assuming that the solution for that part of the full embedding game that corresponds to the reduced game (which emerges after fixing the sub-game play at the solution of the sub-game) does not change again amounts to a plausible consistency requirement.

From a slightly different point of view solution by sub-games implies a specific stepwise reduction of the set of strategic players. The reduction follows the order of moves in plays of a game. It always focuses on the possibility of strategic influence and respects the constraints of causal closure. Formally this kind of reduction process raises the question whether arbitrary reductions of the player set of a game in agent normal form would have similar properties (and if not which reductions would have those properties). We do not believe that there is a convincing systematic argument or a philosophical reason to the effect that arbitrary fixings be possible without affecting solutions. However, from a purely formal point of view it would seem convenient if that property were fulfilled. The condition would be stronger than systematically or philosophically desired but it would imply the philosophically and systematically desired property. So let us consider reduced games that are created by fixing the choices of arbitrarily chosen players while considering only the interaction among the remaining players as strategic.

4.2 Consistent arbitrary fixings of player choices

Let us consider now games in agent normal form representation. The basic purely formal "general consistency" notion for such games is that any game with $n \geq 2$ players after fixing the choice of $m \geq 1$, $m < n$, player(s) and making the choice(s) known to the other player(s) is a game with n-m player(s). Therefore, if the fixed choice(s) coincide with the solution for the game of n players then the solution for the "reduced" game should coincide with the corresponding part of the solution vector for the original game, too. To put it slightly otherwise, if the solution theory is directly applied to a "reduced game" then the solution should coincide with the solution that is induced on the reduced game by solving any larger game in which the reduced game is embedded. It is asked whether a solution theory's prescriptions for the interaction within a subset of players would change if it were commonly

known that players from the complementary subset of players (agents) had already complied with the theory's prescriptions and the theory were then applied to the remaining reduced game.[18]

More precisely, consider a non-empty proper subset H of the set of players N, or H≠∅, $H \subseteq N$, H≠N. Assume that the players in H are already "committed" to plans s_j with $s_j = s^*_j$, j∈H, that is, they are not anymore in a position to make opportunistically rational choices. Let M:=N\H denote the set of uncommitted players who can still choose opportunistically rational, while $\Gamma_M^{s^*}$ refers to the reduced game that emerges among the "active" players j∈M after restricting the choices of all j∉M – that is of all j∈H – to $s_j = s^*_j$. A solution s* of Γ induces a "solution candidate" $((s^*_j)_{j \in M}) =: s^*_M$ for the reduced game $\Gamma_M^{s^*}$ played by the active players.

Now, let G be a class of games. For any $\Gamma \in G$ let R(Γ) denote the set of reduced games $\Gamma_M^{s^*}$ that can be derived from Γ in the way indicated before. Let us call any set of games G *closed* iff for every $\Gamma \in G$ we have $R(\Gamma) \subseteq G$.

Let L be a solution concept for the closed class G. L is basically a function that assigns a non-empty subset of its strategy set S to any $\Gamma \in G$.

For $\Gamma \in G$ with the set S of strategy vectors s

$L^*(\Gamma) := \{s^* \in S | s^*_M \in L(\Gamma_M^{s^*})$ for all $M \subseteq N$ with M≠∅,M≠N$\}$

is the *set of generally consistent solution candidates* for Γ (under L). The *consistency axiom* for solution concepts requires that a solution concept L for a class of games G assigns only solutions that are generally consistent under L.

> *Axiom of consistency C*: Let G be a closed class of games. For all games $\Gamma \in G$ the solution concept L fulfills $L(\Gamma) \subseteq L^*(\Gamma)$.

Consistency means that L assigns the same strategy choices in L($\Gamma_M^{s^*}$) as in L(Γ) to each player in M if it is assumed that the choices of the players of the complementary set N\M are fixed at $s^*_{N \setminus M}$. This is "cell-consistency" extended to arbitrary sets $M \subseteq N$ rather than merely "cells" $C \subseteq N$.[19] There is, however, as we indicated already and shall argue in some detail below no good reason for assuming general consistency under arbitrary fixings. If general consistency has any merit at all then

---

[18] In other words, it should not matter whether the expectations concern aspects of nature or result from the application of a G-definite theory.

[19] On cell-consistency see Harsanyi, J. C. and R. Selten (1988). A general theory of equilibrium selection in games. Cambridge, Mass., MIT Press., 101.

because it shows up as one of the axioms in a full characterization of the equilibrium notion. Before we can state the relevant characterization theorem we need to introduce:

Axiom of converse consistency C': Let G be a closed class of games. For all games $\Gamma \in G$ with |N|≥2 the relation $L^*(\Gamma) \subseteq L(\Gamma)$ holds.

According to converse consistency all generally consistent solution candidates must qualify as solution candidates. None may be excluded. Here is the characterization of equilibria by

Theorem (Van Heumen, Peleg et al. 1996): Let E refer to a solution function such that for all Γ∈G the set valued function E assigns the set E(Γ) of equilibria as solution candidates. For any closed set G of games Γ and solution function L: 0**&C&C' <=> L=E.

That is, L will fulfill all three axioms

(i) decision rationality

(ii) consistency

(iii) converse consistency

if and only if L=E and thus for all Γ∈G we have L(Γ)=E(Γ).

The theorem seems to imply that all equilibrium refinements have to be excluded if the solution theory is to be consistent in the preceding sense. So which of the two consistency notions, cell- respectively sub-game consistency or general consistency should go?

4.3 Against general consistency

If general consistency and definiteness were both part of the core of full rationality then an inconsistency would emerge. Now, in our view, definiteness of solution theories must not be given up since for any class of games G our very concept of ideal rationality seems to require that a G-definite theory of fully rational choice-making be presented (ruling out set valued functions (as in Dufwenberg, Norde et al. 2001)).[20] If definiteness must not go, then general consistency must. And we can let it go without losing anything substantial since we believe that general consistency is a purely formal requirement that lacks substantial support and even plausibility.

A simple example suffices to show why this is so. Consider a game Γ∈G where G is a closed class of simultaneous move games in strategic representation. In all games from G each player (agent) i∈{1, 2, 3} has to move exactly once by making a choice from her strategy set $S_i = \{X_i, Y_i\}$. The payoffs are utilities listed in the natural order of the players 1, 2, 3 in the following table that represents our game example(s) (see Güth 2003)

---

[20] Of course, insisting on definiteness is characteristic of the Harsanyi and Selten approach as presented in Ibid.

	X_3		Y_3	
	X_2	Y_2	X_2	Y_2
X_1	4,4,z	2,2,0	0,0,0	4,0,0
Y_1	2,2,0	6,6,0	0,4,1	5,5,5

Table 1

For reasons that will become obvious immediately assume z>5/4. With this specification the game has two strict equilibria namely $X = \{X_1, X_2, X_3\}$ and $Y = \{Y_1, Y_2, Y_3\}$. Definiteness of a solution theory requires choosing one of these equilibria. Assume for the sake of specificity that the conventional solution theory selects equilibria in cases as the class of games G according the criterion of "unilateral deviation stability". Unilateral deviation stability is ranking equilibria according to the size of the products of all losses ensuing from unilateral deviations from the equilibrium under consideration. Since z>5/4 has been assumed $2 \cdot 2 \cdot z > 1 \cdot 1 \cdot 5$ and the strict equilibrium X would be chosen.

Note carefully that the criterion of equilibrium selection takes into account the effects of unilateral deviations from equilibrium as experienced by *all* actors. It is not by accident that all actors are included in the criterion as applied in the present case. Since equilibrium selection always involves a collective coordination problem the criterion should incorporate all *concerned*. For instance, in the case of sub-games as discussed before only those agents who are active in the sub-game under consideration are concerned because the sub-game is "causally-closed". But in the example at hand we do not have such a substantial reason to exclude any of the players from consideration when solving the collective coordination problem of equilibrium selection. If such a reason is lacking a reduction of the coordination problem to a subset of players can change its nature. Again our specific example and the specific criterion of "unilateral deviation stability" can serve as an illustration.

Consider the reduced game that emerges among the subset of players (agents) M={1,2} if the choice of the third agent is fixed at X_3:

	X_2	Y_2
X_1	4,4	2,2
Y_1	2,2	6,6

Table 2

If a solution theory based on unilateral deviation stability would be generally consistent the theory should still fix the solution at (X_1, X_2). Applying a solution theory that incorporates unilateral deviation stability to the reduced game will lead to the choice of (Y_1, Y_2), however, since $2 \cdot 2 < 4 \cdot 4$. General consistency is violated by the exemplary conventional solution theory incorporating unilateral deviation stability. But this does not provide convincing evidence against incorporating equilibrium refinements into a solution theory. It seems obvious that the coordination between two players is of a different nature than that between three. Why should it be otherwise? After all, the

problem is one of co-ordination in a group of agents. And, clearly we would never assume that the necessity to co-ordinate with an additional independent actor would not affect the co-ordination reached between previous actors. Vice versa, if at least one actor need not be taken into account anymore as a strategic actor because his choice has been fixed this opens up new possibilities of coordination among the remaining actors. This holds good even if the choice of the exiting actor has been fixed exactly as he would have chosen in a co-ordination process involving him as a strategic choice-maker. But after reduction he is not anymore part of the strategic game. The remaining players do not need to take him into account anymore as a strategic actor and therefore may have good reason to solve their restricted co-ordination problem otherwise than in his presence.

The requirement of general consistency does not allow for all considerations relevant for eductive theory. Within eductive theory the actors have to put themselves into the shoes of all participants. This includes the *reasoning* from any position rather than taking into account merely the result of such reasoning. Representing the results of reasoning suffices only if certain substantial conditions of strategic closure as in the case of cells (to which we shall turn below in more detail) or sub-games (in case of extensive games) are met.

But, one might want to object still that when solving the larger problem – before reduction – there was no communication, no strategic bargaining no cheap talk going on between the players. Since actors are by assumption conceived as being informationally isolated the theory alone has to determine choices. Without any further information flow the theory has to provide sufficient reasons for specific plans of the players. Why, so might the argument go on then, should such a theory be responsive to the reduction of the set of strategic agents by altering the theoretical result for the reduced set if the choices of exiting agents are fixed *according to the theory*? If the theory changes its advice after the choice of an exiting player is fixed according to the theory's own implied advice would that not amount to incoherence and is it therefore not desirable to require general consistency?

A brief rehearsal of what our reflections on equilibrium are all about may be helpful here. We are seeking a concept of fully rational choice that can co-ordinate the choices of fully rational choice-makers who conceive themselves as members of an idealized world of fully rational beings. Due to the crucial assumption of theory absorption the theory of rationality itself fully characterizes the expectations of other players' behavior in classes of games G that any of the players endorses. The meaning of full rationality is itself fixed by a convention or a conventional theory of rational plan and play (the criterion of "unilateral deviation stability" of the preceding example being part of such a theory).

Dealing with rationality itself as a convention solves the overall co-ordination problem of players who are aware of the presence of other rational players. Players commonly know that all players are reasoning according to a theory and due to the convention of rationality commonly know which theory that is. It is only this common knowledge of both the presence of other rational agents and of

the theory guiding their choices that allows reaching definite results. The awareness of the presence of other rational choice-makers as well as the awareness of the fact that these other rational choice-makers analyze the situation in the same terms and according to a commonly known theory is central to the reasoning of fully rational actors. It is of the essence of their reasoning that others reason too.

The whole point of eductive game theory as opposed to decision theory plain and simple is this awareness (and common knowledge of the awareness) of the presence of other rational beings. General consistency contrary to the very spirit of eductive game theory assumes that "parameterization" of other players is admissible across the board. However, without a special reason it is overwhelmingly implausible that such arbitrary fixing should not alter the character of games. For, once we fix the choice of one of the actors the remaining strategic choice-makers are aware of the fact that the "parameterized" choice-maker is not anymore a strategic actor. Then they do not need to put themselves into his shoes and they need not assume that others do or that he is putting himself in their shoes.

The difference between games against nature and other games is at stake here. If we assume that this difference crucially depends on our theorizing in view of the very same theorizing of others it makes a fundamental difference how many other theoreticians in which position are present. Specifically in our example there are three fully rational actors in the initial game $\Gamma \in G$ and two in the reduced game $R(\Gamma) \in G$. In an eductive context which of the theories can be absorbed should crucially depend on who reasons in which position. We can abstract from the reasons that the actors would have for their choice-making – their theory guided deliberation process – only if we have a special reason for that.

More generally speaking, it can be seen here why it may be quite misleading to analyze games parametrically. To work with assumptions like "suppose another player or all other players have chosen a profile s" implicitly neglects the reasons for the assumed choices. But in eductive game theory it is always the pair of theory and the collectivity of strategic actors using that theory that matters. The theory is all about reasoning. Therefore abstracting form the reasoning of actors amounts to abstracting away the very essence of eductive game theory unless there is a substantial reason admitting this abstraction. We have pointed out such reasons for sub-games of games in extensive form already and will now provide an analogue for games in agent-normal form.

4.4 From sub-games to cells

In this paper we relied on the concept of a sub-game and on the requirement of sub-game consistency rather than that of a "cell" and "cell-consistency" (see Harsanyi and Selten 1988). Basically the consistency notion introduced in the preceding axiom C is a generalization of cell-consistency to arbitrary game-like sub-structures. Cell-consistency restricts the consistency requirement of the axiom to specific game-like sub-structures. Accordingly the solution of a cell in a closed class of games that with any game contains all cells of that game must be the same for isomorphic cells.

Though we could have relied on cell-consistency throughout we chose to rely on sub-games for the sake of simplicity and because the time and causal structure of sub-games very transparently represents the substantial issues involved. On a somewhat more abstract and general level the concept of a cell drives at the same as that of a sub-game (and we referred to this analogy already in side remarks and footnotes).

Players who belong to a cell can never find a reason to revise their plans if individuals outside the cell change their plans. More precisely, consider again a game $\Gamma = (S_1, S_2, ..., S_n; u_1(.), u_2(.), ..., u_n(.))$ presented in agent normal form. A cell $M \subseteq N$ is a subset of the players and –M:=N\M is the complementary set. These players interact in a cell-game for which it is true that whatever happens outside the cell-game will not affect the best reply correspondence in the cell-game. In short, what choice-makers not belonging to the cell do does not matter at all for optimal choice behavior of members of the cell.[21]

More formally cell-games of a game Γ are sub-structures $\Gamma_M = ((S_i)_{i \in M}), (u_i(.))_{i \in M})$ of $\Gamma = (S_1, S_2, ..., S_n; u_1(.), u_2(.), ..., u_n(.))$ such that for any s'_M, s''_M:

$[\exists s^*_{-M} \in S_{-M}][(\forall\ i \in M)]\ [u_i\ (s'_M, s^*_{-M}) \geq u_i(s''_M, s^*_{-M})] \Rightarrow$

$[\forall s_{-M} \in S_{-M}]\ [(\forall\ i \in M)]\ [u_i\ (s'_M, s_{-M}) \geq u_i(s''_M, s_{-M})]$

An equilibrium s^* is called "cell-perfect" if for all cells M and all strategy sets S_{-M} the profile s^*_M is an equilibrium of the corresponding reduced cell-game $\Gamma_M^{s^*}$. Cell-consistency goes beyond cell-perfectness in that it is also required that the same equilibrium is selected for isomorphic cells no matter what.

The requirement of cell-consistency has substantial reasons in its favor whereas the general consistency requirement has not. Again a simple example suffices to show why this is so. Consider a game $\Gamma \in G$ where G is a closed class of simultaneous move games in strategic representation. In all games from G each player (agent) $i \in \{1, 2, 3\}$ has to move exactly once by making a choice from her strategy set $S_i = \{X_i, Y_i\}$. The payoffs are utilities listed in the natural order of the players 1, 2, 3 in the following table that represents our game example(s)

---

[21] In the special case of a sub-game, whatever those outside of the group of agents who act strategically in that sub-game do, it will not rationally affect the plans of fully rational forwardlooking choice-makers who act according to the standard teleological model of purposeful choice-making.

	X_3		Y_3	
	X_2	Y_2	X_2	Y_2
X_1	2,1,1	0,0,0	2,1,0	0,0,0
Y_1	0,0,0	1,2,0	0,0,0	1,2,1

Table 3

The best reply of player 1 to the choices of player 2 depends only the choices of player 2 and vice versa. Therefore the best reply correspondence does not depend on players other than 1, 2 and M={1, 2} is a cell and $\Gamma_M = ((S_i)_{i\in M}),(u_i(.))_{i\in M})$ a cell-game. However we fix the choices of the third player as a parameter this will not influence the preference orders of the remaining two players. The reduced (battle of the sexes like) cell game has two strict equilibria among which the cell-players will choose independently of the choices of player 3. In his planning player 3 can merely "adapt" to whatever equilibrium he foresees the cell players to choose since his own choices, cannot provide any substantial reason for the remaining players to revise their plans.

In view of the informational isolation of agents on the fundamental level of analysis player 3 must make his choice by means of a theoretical inference. This inference he can draw only if there is an established convention (a theory of rationality) solving the cell game (the battle of the sexes game being a particularly tough case for a closed a priori theory of equilibrium selection, see for instance Sugden 1991). This theory convention might not fulfill the requirement of invariance of solutions under, for instance, the renumbering of players but it clearly should comply with the requirement of cell-consistency. For, after all, in eductive analyses we are discussing strategic reasoning and the very notion of a cell shows which kind of strategic reasoning matters and, for that matter, which kind of consistency, namely cell- consistency rather than purely formal general consistency.

5 Concluding discussion

5.1 Adequate idealizations

Somebody who argues that prescriptions that can be derived from eductive game theory will not be obeyed since human individuals are only boundedly rather than fully rational is right on target. But he should resist the temptation to put eductive theory through some of the many obscure neo-classical repair shops to make it more realistic. Some have argued that conventional full rationality is an inherently incoherent concept and that therefore rather than for making models more realistic some of the idealizing assumptions must be removed. Moreover it has been also suggested that analysis proper is impossible since the embeddedness of a game in a larger game can never be neglected. Let us finally address these arguments.

To keep things simple let us restrict the argument to games with perfect information in which all information sets are singletons from which a sub-game starts. For the sake of specificity imagine a game like the one presented in figure 3. If we analyze that game we will quite naturally apply backward induction. It is clear then that the solution of the final sub-game originating at 1b's decision node is L. This sub-game cannot be reached if players in fact play according to the precepts of standard game theory based on backward induction arguments. Thus, if 1b is actually going to move, the expectations formed according to the commonly shared theory have not become reality. Should the player infer then that the theory of rational play itself is falsified?

Like in the preceding discussion we face a simple alternative here: Either it is true that theoretical analysis for any sub-game of a larger game is not affected by the history of play leading to the sub-game or any analysis of a game that is based on an analysis of its sub-games is illegitimate from the outset. If the first is true then there is no need for any decision-maker to adapt her expectations if a sub-game off the path of theoretically prescribed rational play is reached. The theory remains unaffected. Regardless of the history of play that leads to any sub-game identical sub-games should be solved in identical ways. This is also the only view fully in line with the future-directedness of purposefully rational behavior for which bygones are bygones. If the second holds good no analytical decomposition into smaller problems is viable. Only a kind of "holistic" theoretical discussion of games seems possible. Then it seems that each game would need its own theory of rational play. The latter would be the end of eductive game theory proper. For, whatever else we might associate with the concept of a theory a certain degree of generality must be among its constitutive features. The necessity to form a separate holistic account for each and every problem is certainly not ideal from a methodological point of view (though it might not be too far fetched to see some relations to Aristotelian notions of prudent thinking).

In any event, a game of perfect information can "naturally" be decomposed into sub-games. If it is nevertheless claimed that this cannot be done it should be possible to say something about general features that preclude analysis in the sense of decomposing a larger problem into smaller sub-games. Of course, every game is embedded (possibly but not necessarily in Granovetter's sense Granovetter 1985) in a larger context that in turn is embedded in a still larger context etc. But referring to this as a general feature precluding decomposition leads to absurdities. Unless the context of games may be neglected for analytical purposes there would be essentially only one single game, the "game of life".

Until somebody would come up with some proposal taking into account the phenomena of "embeddedness" in a general way it would seem doubtful to speak at all of "game theoretic analysis" if there were only holistic, tailor-made accounts. The latter would have to prescribe different solutions to otherwise identical sub-games depending on the length and structure of alternative preceding histories of play and possibly alternative courses of action that might occur if the sub-game were not reached. A truly general theory of that kind is somewhat hard to imagine at least as long as

we try to base it on convincing assumptions about individual rationality and do not altogether give up the premise of future directedness of rational choice. As long as we want to stick to the program of theoretical and general analysis of a hypothetical world of fully rational beings at all we need some invariants and cell-consistency is the best candidate we have for singling them out.

5.2 Rational choice vs. choosing as if being rational

General theoretical and methodological considerations of an a priori kind have been discussed so far. What about aposteriori arguments should they play a role in the process of explication as well?

We are aware that within the empiristic mind set of our age there is a strong feeling that a priori reasoning is dubious and a posteriori reasoning should take precedence whenever viable. We do not share this view entirely. On the other hand we are empiricists as well. Therefore some final remarks determining the proper place of a priori and a posteriori arguments may seem in order. To that end let us briefly recapitulate and evaluate some of the arguments concerning the relationship between eductive considerations and adaptive processes.

As compared to eductive analyses of cognitive processes going on in the minds of rational players best reply dynamics basically give up the assumption of common knowledge. Usually this move is defended by the observation that the common knowledge assumption is too "unrealistic" or too much of an idealization to be of any relevance as a model of real players' capacities and behavior. But approaches based on best reply dynamics still make use of quite daring assumptions of normative decision theory. Like the latter they allow for beliefs concerning other players' behavior and the assumption that players react optimally on these beliefs.

If it is pointed out to adherents of best reply dynamics that the assumptions about beliefs and response behavior may seem quite strange and unrealistic they will probably respond that it is sufficient for their aims that behavior is "as if" it were based on best reply dynamics. If this is their best reply we submit, though, that it is not good enough. For, even if it were true for certain games that individuals behave according to best reply dynamics and thus behave *as if* optimizing against expectations of other individuals' behavior in the way described before it is most unlikely that *optimization* on the side of individuals would form a true explanation. The fact that they behave that way would itself require an explanation rather than providing one.

To put it slightly otherwise, whenever we observe optimizing real world behavior that conforms with the implications of idealized models the adequate response is to regard optimization as an explanandum rather than an explanans. Almost always the real puzzle is how to account for optimal results without relying on the premise of optimal behavioral responses. The "art" of the empirically minded social scientist – with some very rare exceptional cases – basically consists in providing explanations for optimal behavior that do not rely on intentional best reply behavior on the side of the individuals.

If we apply this line of thought to the alleged justification of equilibrium behavior in terms of best reply dynamics it seems obvious that the latter justification does not accomplish its task. Behavior corresponding to best reply dynamics is itself as much in need of an empirical justification as the equilibrium behavior itself. The approaches based on best reply dynamics should be regarded as extensions of eductive analyses (and as almost as far away from reality as the latter). They characterize situations in which equilibrium behavior will emerge among *rational* players who lack common knowledge of rationality of all players. Such an accomplishment is not without merit. But it is complementary to outright normative approaches rather than a foundation for the idealized assumptions of normative analyses in general.

In approaches based on best reply dynamics beliefs about future behavior of co-players are formed according to some kind of past experience. The reply itself, however, is chosen according to maxims of rational forward looking choice. As opposed to that, approaches based on stimulus response learning stress the importance of past experiences in forming behavioral responses. There are no intermediate (or intervening) "variables" forming a cognitive model of the situation. For, at least in their more traditional variants stimulus response learning models view the decision maker as a kind of black box or automaton.

Now, even though there may be very sophisticated automata – there are even very sophisticated "chocolate dispensing machines" – it is obvious that human actors typically act upon models of their decision situation which in complexity and flexibility go far beyond what may be emulated by the capacities of such automata. It comes as no surprise then that studies based on stimulus response learning seem to be most valuable in situations in which players know very little about the environment in which they are acting, e.g. a stochastic environment where a player does neither know his pay off function nor that of his opponents nor even whether there are any opponents. But even in less extreme cases it may be quite interesting to study which kinds of intermediate rather than asymptotic behavior is implied by stimulus response learning. Again we do not deny that adaptive approaches based on stimulus response learning may be of great value. What we reject though is the thesis that such studies can serve as justifications for rational equilibrium behavior and the prominent role of the equilibrium concept in analyzing strategic interaction in general.

When focusing on evolutionary stability adaptive approaches disregard decision making processes altogether. If the evolution of behavior and evolutionary stability are center stage it does not really matter what the springs of action are. The evolutionary approach per se does not rely on any specific assumptions about rationality, learning or more generally, psychological or genetical mechanisms. Still evolutionarily stable behavior may look like rational equilibrium behavior or like behavior predicted by psychological theories. This "as if" character of evolved behavior has been noted ever since so called teleological arguments for the existence of God – mother nature looks as if planned by father God – have been related to evolutionary selection and adaptation. Suffice it here to repeat that

the observation of seemingly purposeful planning forms an explanandum rather than the explanans and thus cannot serve as an empirical foundation for idealized theory formation. Moreover, positive results where asymptotical stability justifies equilibrium behavior are still rare. They are limited to some special classes of games while for other classes the relationship between asymptotic stability of evolutionary dynamics and equilibrium behavior is not as tight as one might hope. In general there is little hope that real game playing behavior and its development over time can be captured by dynamics simple enough to allow for an analytical discussion of absorbing states. Adaptation dynamics that yield equilibrium behavior more likely than not will rest on rather questionable assumptions which raise more problems of justification than they solve. For more reasonable adaptation dynamics the results are most likely negative in that they converge to equilibrium only for very special cases or classes of games.

In any case adaptive approaches to game playing behavior should not focus on (as if-)rationality. The true empiricist might well claim that predictive power and empirical validity rather than consistency are crucial in assessing the value of a scientific approach. But then he should better take empiricism seriously and build his theories on empirically sound a posteriori premises rather than slipping in some a priori notions of optimizing behavior whenever convenient. From an empiricist point of view hypotheses about adaptation should be based on psychological knowledge of decision making – including "consistency" requirements as scrutinized by cognitive psychology – rather than simply stipulating optimal behavior or taking recourse to "as if notions of rationality".

It is not without irony that the focus on as if behavior does make sense only from the point of view of someone who is primarily concerned with problems of normative or pure rational choice analyses of an eductive kind but feels that they are lacking a proper foundation. Adaptive approaches – though of great interest in their own right – cannot provide such a foundation in general nor for the equilibrium notion in particular. If one cares primarily for explaining and predicting actual game playing behavior the equilibrium concept may indeed be dispensable altogether.

In sum, we should pursue a research agenda that comprises both eductive and adaptive analyses and at the same time keeps them apart. We should distinguish between analyses of idealized, fictitious mental processes of complete strategists and explanatory and predictive analyses of real world behavior. There is neither an explanatory nor a justificatory relationship between the two realms. It may be possible though to use equilibrium behavior developed in eductive models as a bench-mark or heuristic in empirically oriented studies. This is not to say, however, that eductive analyses are only of interest as a tool for those who work empirically. Besides explanation and prediction there are other legitimate aims of scientific discourse. The philosophical interest in what it means to be rational and whether we can find an explication of the very concept of rationality in strategic interaction under conditions of common knowledge of rationality which can be sustained in reflective equilibrium is a legitimate concern as well. Human as we are we all take an interest in understanding better what it

means to be rational by driving rationality to its extremes. Whether that will drive us out of any conceivable reflective equilibrium or not the reader may reconsider now.

6 References

Alchian, A. A. (1950). "Uncertainty, Evolution, and Economic Theory." Journal of Political Economy **Vol. 58**: 211-221.

Aumann, R. and A. Brandenburger (1995). "Epistemic Conditions for Nash Equilibrium." Econometrica **Vol. 63**(5): 1161-1180.

Aumann, R. J. (1987). "Correlated Equilibrium as an Expression of Bayesian Rationality." Econometrica **55**(1): 1-18.

Aumann, R. J. (1995). "Backward Induction and Common Knowledge of Rationality." Games and Economic Behavior **8**: 6-19.

Binmore, K. (1987/88). "Modeling rational players I&II." Economics and Philosophy **1987/88 (3 & 4)**: 179ff. & 9 ff.

Binmore, K. (1990). Essays on the Foundations of Game Theory. Oxford, Blackwell.

Binmore, K. and P. Dasagupta, Eds. (1989). Economic Organisations as Games. Oxford, Basil Blackwell.

Binmore, K. and P. Dasgupta (1989). Game Theory a Survey. Economic Organizations as Games. K. Binmore and P. Dasgupta. Oxford, Basil Blackwell.

Canning, D. (1992). "Rationality, Computability and Nash Equilibrium." Econometrica **60**: 877-888.

Cournot, A. (1838). Recherches sur les principes mathématiques de la théorie des richesses. Paris, L. Hachette.

Dacey, R. (1976). "Theory Absorption and the Testability of Economic Theory." Zeitschrift für Nationalökonomie **36**(3-4): 247-267.

Dacey, R. (1981). Some Implications of 'Theory Absorption' for Economic Theory and the Economics of Information. Philosophy in Economics. J. C. Pitt. Dordrecht, D. Reidel**:** 111-136.

Daniels, N. (1979). "Wide Reflective Equilibrium and Theory Acceptance in Ethics." The Journal of Philosophy **LXXVI**(1): 265-282.

Dufwenberg, M. and J. Lindèn (1996). "Inconsistencies in Extensive Games." Erkenntnis **45**: 103-114.

Dufwenberg, M., H. Norde, et al. (2001). "The Consistnecy Principle for Set Valued Solutions and a New Approach to Normative Game Theory." Mathematical Methods of Operations Research **54**: 119-131.

Fagin, R., J. Y. Halpern, et al. (1995). Reasoning about Knowledge. Cambridge, MA / London, MIT Press.

Gode, D. K. and S. Sunder (1993). "Allocative Efficiency of Markets With Zero Intelligence Traders: Markets as a Partial Substitute for Individual Rationality." Journal of Political Economy **101**: 119-137.

Granovetter, M. (1985). "Economic action and social structure: The problem of embeddedness." American Journal of Sociology **91**(3): 481-510.

Güth, W. (2003). "On the Inconsistency of Equilibrium Refinement." Theory and Decision **forthcoming**.

Güth, W. and H. Kliemt (1995). "Ist die Normalform die normale Form?" Homo oeconomicus **XII** 155-183.

Güth, W. and H. Kliemt (2000). From full to bounded rationality. The limits of unlimited rationality. Bielefeld, Center for Interdiciplinary Research (ZiF).

Güth, W., W. Leininger, et al. (1991). On Supergames and Folk Theorems: A Conceptual Analysis. Game Equilibrium Models. Morals, Methods, and Markets. R. Selten. Berlin et al., Springer. **II:** 56-70.

Güth, W. and B. Peleg (2001). "When Will Payoff Maximization Survive? -- An Indirect Evolutionary Analysis." Evolutionary Economics **11**: 479-499.

Hahn, S. (1996). Überlegungsgleichgewicht und rationale Kohärenz. Die eine Vernunft und die vielen Rationalitäten. K.-O. Apel and M. Kettner. Frankfurt a.M., Suhrkamp**:** 404-423.

Hahn, S. (2000). Überlegungsgleichgewicht(e). Prüfung einer Rechtfertigungsmetapher. Freibrug i.Br., Karl Alber.

Harsanyi, J. C. (1967-8). "Games with Incomplete Information Played by Bayesian Players." Management Science **14**: 159-82, 320-34, 486-502.

Harsanyi, J. C. and R. Selten (1988). A general theory of equilibrium selection in games. Cambridge, Mass., MIT Press.

Hume, D. (1985). Essays. Moral, Political and Literary. Indianapolis, Liberty Fund.

Jacobsen, H. J. (1996). "On the Foundations of Nash Equilibrium." Economics and Philosophy **12**(1): 67-88.

Kagel, J. H. and A. E. Roth, Eds. (1995). The Handbook of Experimental Economics. Princeton, Princeton University Press.

Kant, I. (1991). Political writings. The metaphysics of morals. Oxford et al., Oxford University Press.

Kliemt, H. (1996). Rational Choice-Erklärungen? Handlungs- und Entscheidungstheorie in der Politikwissenschaft: Eine Einführung in Konzepte und Forschungsstand. U. Druwe and V. Kunz. Opladen, Leske und Budrich**:** 83-105.

Morgenstern, O. (1972). "Descriptive, Predictive and Normative Theory." Kyklos **25**: 699-714.

Morgenstern, O. and G. Schwödiauer (1976). "Competition and Collusion in Bilateral Markets." Zeitschrift für Nationalökonomie **36**(3-4): 217-245.

Nash, J. (1951). "Non-Cooperative Games." Annals of Mathematics **52**(2): 286-295.

Norde, H., H. Potters, et al. (1996). "Equilibrium Selection and Consistency." Games and Economic Behavior **12**: 219-225.

Nozick, R. (1974). Anarchy, State, and Utopia. New York, Basic Books.

Pearce, D. G. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection." Econometrica **52**(4): 1029-1050.

Peleg, B. and S. Tijs (1996). "The consistency Principle for Games in Strategic Form." International Journal of Game Theory **25**: 13-34.

Rawls, J. (1951). "Outline of a Decision Procedure for Ethics." Philosophical Review **60**: 177-190.

Rawls, J. (1971). A Theory of Justice. Oxford, Oxford University Press.

Rosenthal, R. (1981). "Games of Perfect Information, Predatory Pricing, and the Chain Store Paradox." Journal of Economic Theory **25**: 92-100.

Rubinstein, A. (1989). "The Electronic Mail Game: Strategic Behavior Under
"Almost Common Knowledge"." American Economic Revue **79**(3): 385-391.

Rustem, B. and K. Velupillai (1990). "Rationality, Computability and Complexity." Journal of Economic Dynamics and Control **14**: 419-432.

Selten, R. (1975). "Reexamination of the Perfectness Concept for Equilibrium in Extensive Games." International Journal of Game Theory **4**: 25-55.

Siegwart, G. (1997). Explikation. Dialog und System. W. Löffler and E. Runggaldier. Sankt Augustin, Academia**:** 15-45.

Skyrms, B. (1990). The Dynamics of Rational Deliberation. Cambridge, Harvard University Press.

Strawson, P. F. (1962). "Freedom and Resentment." Proceedings of the British Academy: 187-211.

Sugden, R. (1986). The Economics of Rights, Co-operation and Welfare. Oxford, New York, Basil Blackwell.

Sugden, R. (1991). "Rational Choice: A Survey of Contributions from Economics and Philosophy." The Economic Journal **101**(July): 751-785.

Van Heumen, R., B. Peleg, et al. (1996). "Axiomatic Characterizations of Solutions for Bayesian Games." Theory and Decision **40**: 103-129.

Young, H. P. (1998). Individual Strategy and Social Structure. An Evolutionary Theory of Institutions. Princeton, Princeton University Press.