Analyse & Kritik 26/2004 (© Lucius & Lucius, Stuttgart) p. 364–381

*Werner Güth/Hartmut Kliemt**

Perfect or Bounded Rationality?

Some Facts, Speculations and Proposals

Abstract: Simple game experiments of the reward allocation, dictator and ultimatum type are used to demonstrate that true explanations of social phenomena cannot conceivably be derived in terms of the perfect rationality concept underlying neo-classical economics. We explore in some depth, if speculatively, how experimental game theory might bring us closer to a new synthesis or at least the nucleus of a general theory of 'games and boundedly rational economic behavior' with enhanced explanatory power.

1. Introduction and Overview

Explanations in terms of bounded rationality are typically based on strongly domain dependent theories of quite short range. Using them leads to a fragmented view of the world. As opposed to that, decision and game theoretic language convey the impression of a unified view of the world that renders diverse phenomena more intelligible by putting them in (the same) perspective. Yet if we look more closely at the way the general language of decision and game theory has to be adapted to specific domains mostly by ad hoc assumptions, it is obvious that the rational choice theorists' claims to universality and unity of their theories are based on a kind of collective self-deception. They have a common language—not a small accomplishment—but no common theory with general empirical content and explanatory power.

It may well be that in the end there are no convincing truly general theories of human behavior at all. But if that is so we should take pains to expose as openly as we can where and why our theories of human behavior in social contexts fail and what their limitations are. Nothing can be won by concealing the failure to formulate general theories with empirical content behind a veil of deceptively general optimization models. If the theories that carry empirical weight are in the end quite commonsensical, so be it. If we can only give some advice as to how domain specific theories can be formulated without being in a position to contribute substantial general information, why not accept that as well?

In a first step of our programmatic discussion we will discuss the merits of generality in social theory formation (2.). We then turn to claims of generality as contained in neo-classical theories of fully rational behavior (3.). The deficiencies of explaining the results of reward allocation, dictator and ultimatum games in a perfect rationality framework are discussed then (4.). A speculative

* The authors would like to express their gratitude for helpful comments of two anonymous referees. Of course, the conventional disclaimer applies.

sketch of some attempts to explain experimental results in a bounded rationality framework follows (5.). Final remarks conclude the paper (6.).

2. Generality of Social Theories?

From its beginnings to at least the middle of the 20th century, neo-classical economics saw itself as a domain specific theory confined to market behavior, to exchange and commercial relationships narrowly defined. Only in the second half of the 20th century did neo-classical economics venture into new territory beyond the realm of the 'commercial'. The theory of games was developed, public choice theory was established as a theory of non-market decision making (the journal *Public Choice* started out as 'papers on non-market decision making'), the law and economics movement and new institutional (and constitutional) economics took off (with contributions ranging from Buchanan/Tullock 1962, over Milgrom/Roberts 1993 to Young 1998).

The so-called 'new world of economics' (see McKenzie/Tullock 1978) is a theoretical image of the social and institutional world that promises a unified view of all social phenomena. It is a world of rational decision makers, of utility maximizers who act on the basis of given preferences (see for instance Becker 1976). In all contexts, from the family to politics, the same *r*ational, *e*xpecting, *m*aximizing *m*en (REMM, see Meckling 1976) are acting. The corresponding theory of rational decision-making forms the core of the economic approach to human behavior. It promises to connect diverse phenomena in one unified body of mathematical social theory. Social science thereby reaches a stage of universality and mathematical precision comparable to that of, say, Newtonian mechanics—or so it seems.

If it were in fact true that the behavioral assumptions underlying 'rational economic man' formed a universally applicable model of human behavior leading to valid explanations in social realms as diverse as competition among political parties at the polls and among firms for customers (phenomena as diverse as the fall of apples and the movement of planets) the world of social science would be fundamentally altered. However, regardless of the charms of the economic approach to human behavior the uniformity of its language and of its descriptions is not matched by unifying behavioral laws applicable to several realms.

To say that in all realms individuals act 'as if' they maximize is not good enough. First, it is, to put it mildly, very unlikely that 'as if' maximization would with any plausibility apply universally. Second, even if all individual behavior could be described 'as if' it were maximizing, we would have an explanandum (a—very astonishing—phenomenon to be explained) rather than an explanans (an explaining set of hypotheses and conditions for their application). Conscious and intentional maximization of a given objective function is not at the root of human choice making. Rational choice models cannot describe the true causes of observed behavior. But what is it then that brings about behavior that is as if the rational choice model were true? Third, human behavior is sometimes so-well adapted to circumstances that an omniscient maximizer could not do better.

Yet, again, well-adapted behavior is *not* the result of conscious maximization. It is in all likelihood the outgrowth of (adaptive or selective) processes other than intentional maximization that somehow lead to the observable optimal result. Merely to identify overt behavior as optimally adapted—or at least as stable in the sense of, say evolutionary stability—does not provide an explanation. Rather optimality as well as stability requires an explanation in terms other than optimization (the origin of predictable behavior is discussed from a more or less neo-classical perspective in several studies, e.g. Heiner 1983).

Wherever results corresponding to the predictions of full rationality should be observed—and we are very reluctant to concede that in complex decision tasks they are ever observed—an explanation in terms of behavioral laws is required. Rather than feeling reassured in our economic views of the world we should ask: Why is it so that the anomaly of seeming full rationality is observed? What are the circumstances that led to such an astonishing phenomenon under general behavioral laws? If optimality in the sense of 'as if' rationality is widespread, how can we explain this extraordinary fact? What are the general behavioral laws that can explain empirical observations?

We will not be in a position to offer fully convincing answers to these questions. Turning to the task of outlining what can and to our opinion should be done, let us start with a more specific, if brief, look at the traditional theories of perfect rationality that were involved in making stronger claims to generality. We then will try to move away from the more traditional view and gradually bring in a more realistic perspective on choice making to which we will refer quite generally as a concept of 'bounded rationality' without thereby identifying with any specific approach (as for instance that of Herbert Simon or Reinhard Selten).

3. (Commonly Known) Perfect Rationality and its Difficulties

Consider a standard *battle of the sexes* game:

2 1	s_2^1	s_2^2
s_1^1	\$10, \$20	0,0
s_1^2	0,0	\$20, \$10

Table 1

An omniscient outside observer might conceivably be able to predict the choices of both individuals and to feed each of them the correct information on what the other will do. But even for an omniscient observer this task is not an easy one because as an outside observer she must take into account how individuals will respond to the information she provides. For instance if she sends the signal that player 1 will choose s_1^1 to player 2, she has to anticipate how player 2 will respond

to that information and feed that anticipation back to player 1. This will lead to valid predictions only if player 1 does not alter his plans upon receiving the information about player 2.[1]

More specifically, in an experiment with monetary payoffs two individuals could be brought into a room in which the matrix of the game is presented to them. They are jointly present and know that they are. A theory of rational play is given to both players. It suggests that they choose (s_1^1, s_2^1). Then they are taken into separate rooms to actually play the game without further communication. In this situation it is safe to assume that the matrix of table 1 along with the theory of rational play is common knowledge among the players. Co-ordination on (s_1^1, s_2^1) can be predicted according to the instructions of the players. Due to their common knowledge a rational choice explanation of equilibrium play is presumably very close to the truth. However, once we leave this simple case, severe problems emerge.

For instance, the theory of rational play in this game is still dependent on the presence of a third person. It was specifically tailored to the situation at hand by the instructor. Without such an individual a general theory would have to specify something as weird as the prescription that each player in battle of the sexes games should choose the strategy *numbered* as strategy 1.

More generally speaking, there are cases in which rational choice explanations in the narrow sense of that term apply in strategic contexts involving multiple equilibria (as for instance in Table 1). This holds good for specific situations with contingent commonly known signals tailored to the situation (which also could be features of prominence) but it is very unlikely that a theory of equilibrium selection per se could provide the true behavioral explanations.[2] To assume that it would, it would be necessary to claim that the theory of equilibrium selection be present in the mental processes of the players. Interpreted literally, as is necessary if we aim at true explanations based on the true behavioral laws, this is absurd.

We all act frequently the 'fast and frugal way' (see Gigerenzer 1996; 1997; Gigerenzer/Czerlinski/Martignon 1998). We constantly face constraints inducing boundedly rational decision-making throughout. These constraints are even more marked in strategic interaction than in single actor contexts. But even in non-strategic contexts models of bounded rather than perfect rationality must be used to approach the reality of choice making in the real world of homo sapiens as opposed to the economic world of *homo oeconomicus*.

[1] The relationships of this line of argument to the problems of 'theory absorption' and the self-supporting or self-destroying character of the absorption process should be obvious. One might add here that this problem was the center of Oskar Morgenstern's thinking about 'equilibrium' in economics and thereby more decisive for the development of game theory than is generally acknowledged today.

[2] Related criticisms are offered in Sugden 1991.

4. Two Paradigm Examples from Experimental Economics

Even though much of experimental economics is still pursued within the traditional neo-classical framework its most basic results are not in line with the behavioral model of perfectly rational choice making of *homo oeconomicus*. Experimentalists with a background in economics have not, however, striven to drive home this point. They have been reluctant to abandon the maximization framework completely and have rather tried to adapt experimental results to the established corpus of neo-classical economic theories—typically by incorporating them into the utility functions (see as a particularly instructive recent example Bolton/Ockenfels 2000). However, introducing new factors into utility functions is a rather incoherent effort to include psychological findings in a framework meant to avoid psychology. For neo-classical economists the main attraction of relying on utility maximization has been that their analyses could start 'after psychology', so to speak, since by assumption all psychological aspects have already been taken into account by utility functions representing individual preferences that were 'given all things considered'.

If economists are confronted with the argument that evidence about real cognitive processes clearly speaks against any assumption of the individually rational pursuit of given preferences, they normally respond that it is sufficient that individuals behave 'as if' maximizing a utility function. If asked how this miracle comes about they tend to admit that

> "... (t)he assumption that conduct is prompt and rational is in all cases a fiction. But it proves sufficiently near to reality, if things have time to hammer logic into men. Where this has happened, and within the limits it has happened, one may rest content with this fiction and build theories upon it ... and we can depend upon it that the peasant sells his calf just as cunningly and egoistically as the stock exchange member his portfolio of shares. But this holds good only where precedents without number have formed conduct through decades and, in fundamentals, through hundreds and thousands of years, and have eliminated unadapted behavior. Outside of these limits our fiction loses its closeness to reality." (Schumpeter 1959, 80)

What Schumpeter said (originally in the early 20th century) has been repeated in some form or other in defence of *homo oeconomicus* again and again (see for instance Alchian 1950; Gode/Sunder 1993 and Young 1998). It should be noted, however, that the argument, though ingenious in itself, is absolutely hopeless as a defence for the explanatory role of rational choice models. For, firstly, within the proper 'adaptive or (social-)evolutionary' context, utility maximization (or, for that matter, *homo oeconomicus* behavior) is part of an explanandum rather than being a component of an explanans. Secondly, outside the proper context it not only fails to explain anything but the "fiction loses its closeness to reality" completely.—As the following illustrations show, outside the context of

long term behavioral adaptation, processes other than rational forward looking choices must be invoked in adequate behavioral explanations.

4.1. Reward Allocation Games

A simple class of experiments originally introduced by social psychologists as 'reward allocation games' (see for instance, Shapiro 1975 or Mikula 1973) illustrates that straightforward maximization of monetary rewards can hardly be typical of human behavior.[3] As performed by social psychologists the experiments adopt the following general form:

- two subjects X and Y must jointly perform some work;
- as a result of their joint effort they jointly earn some monetary reward or 'pie', p>0,
- they are informed that actor X contributes a fraction c∈(0, 1) of the total effort of '1 unit' that the two invested in joint production,
- individual X allocates shares of the pie by 'dictating' that individual X receives x while individual Y receives y=p-x; where x, y≥0,
- the two individuals will not meet after the game and are aware that there will be no subsequent interaction.

Were monetary rewards the dominant motive of individuals the outcome should be (x*=p, y*=0). However this is rarely the outcome of such experiments. Typically participants act as good Aristotelians—as described in book V. of the *Nicomachean Ethics*—and allocate proportionally to the contribution c (see for a standard philosophical account Frankena 1966); i.e. X chooses (x=c p, y=(1-c) p). If the work effort is rather trivial then, since p more or less descends like manna from heaven, we may also find the allocation $(\frac{p}{2}, \frac{p}{2})$ as the most frequently chosen alternative.[4]

In view of such falsifying evidence as presented by the results of reward allocation games, either maximization of 'utility' must be given up altogether or motives other than monetary ones must be included in the utility function (i.e. be included as determinants of preferences in the broader picture of human motivation). Only the latter alternative is open to the neo-classical approach. Moreover, it is quite respectable to start with an austere model of human motivation like the classical *homo oeconomicus* model and gradually add motivational factors so as to develop a fuller account of human motivation. If the modifications are intended to apply not only to reward allocation games but rather to several social situations as specified by antecedent clauses of the theory, we are in the realm of potentially sound theorizing—a criterion for avoiding ad hoc

[3] A variant of such games became more widely known as 'dictator games' among economists.

[4] Letting manna descend from heaven is, as may be noted in passing, the way economists normally frame their somewhat bowdlerized version of the reward allocation problem in which the phases of earning claims by some work is left out (see for exceptions Gantner/Güth/Königstein 2002; Königstein 2000).

repairs being that suggested modifications are aiming beyond the problem that they are supposed to fix.[5]

For the sake of illustration let us look merely at the minimum modification by one factor with an application to another, slightly more complicated, type of social interaction. The additional factor is so-called 'inequality aversion' (see Fehr/Schmidt 1999) and the interaction situation the 'ultimatum game' (see conceptually Güth 1976 and experimentally Güth/Schmittberger/Schwarze 1982). Let us first sketch the game and then discuss how inequality aversion may 'explain' observed results.

4.2. The Ultimatum Game and Additional Motives

There is a pie p to be allocated among two players X, Y. One, X, is in the role of the 'proposer' while the other, Y, adopts the role of the 'responder'.

- First X can 'propose' some allocation (x, y) of p with $x, y \geq 0$ and $x+y=p$.
- The responder Y can accept the proposal (x, y) or reject it.
- If the responder rejects the proposal both players receive nothing (0, 0).
- If the responder accepts the proposal the pie will be allocated according to the proposal (x, y).
- The players do not—and do not expect to—meet each other after the experiment.[6]

If the ultimatum game is analyzed in terms of maximizing monetary rewards as the only motive then the recipient should accept all proposals $y>0$ (and be indifferent as towards acceptance or rejection of the proposal $y=0$). The proposer should anticipate this if 'rationality is common knowledge' and propose the minimum monetary unit such that $y>0$. Again, observations tell a different story:

- Responders Y frequently reject offers y from the range $0 \leq y \leq p/3$
- Proposers X in general offer $y>p/3$, mostly an equal split $y=p/2$ of the pie, which nearly all responders accept.

These observations[7] are clearly not in line with the classical motivational assumptions of rational economic behavior. At least the responder behavior cannot coincide with the model. Moreover, proposers' theories about human motivation must be such that they predict non-opportunistic rejections on the side of

[5] In case of the reward allocation game it would be purely ad hoc if the modification amounted to no more than that the theory holds except in case of reward allocation games.

[6] The ultimatum game has also been used under conditions in which subjects knew that it was played 'double blind' meaning that they knew that even the experimenter would be ignorant of the identity of the players see Bolton/Zwick 1995.

[7] For results of a newspaper experiment with more than 1000 participants, see Güth/Schmidt/Sutter 2002; 2003, while older evidence is discussed in Roth 1995.

responders or, if not so, the proposers must be intrinsically motivated not to allocate rewards (too) unevenly.[8]

One way of dealing with the observations in a rational choice framework relies on the aforementioned motive of 'inequality aversion'. Forming utility functions $u_x(x, |x\text{-}y|)$, $u_y(y, |x\text{-}y|)$ increasing in the first and decreasing in the second argument[9] has the clear advantage that it, in principle, applies to actors in both roles. It explains behavior in reward allocation and ultimatum games (and beyond) by 'inequality aversion'. Both selfish and other regarding motivations are included in a systematic way specifying their relative weight. Finally the qualitative results derived are in line with observational data.

In the preceding regards the argument from inequality aversion seems to be impervious to the more obvious methodological criticisms. It is clearly not pure ad hocery. But is it a plausible explanation in terms of real motivational processes?

It is true that the results of many resource allocation processes pertaining to a wide class of game experiments are less unevenly distributed than they would be if expedient choice were operative alone. Still, even though there is less inequality and less opportunity taking behavior than maximization of monetary income would dictate, this need not suggest that avoiding inequality is among the motives of action. In fact, without independent evidence that a corresponding motive is present, the assumption that inequality is avoided *because* actors are motivated by an aversion against inequality resembles explanations of the sleepiness of opium smokers by appeal to the 'dormitory power' of opium (where dormitory power means no more than that opium makes you sleepy).

Inequality aversion applies to both the ultimatum and the dictator game (and beyond) and to both player roles. This is a virtue. On the other hand, such generality per se is a secondary rather than a primary virtue of theories. It may well be that certain forms of motivation apply only to the role of a second mover who reacts while different motives may apply in situations in which somebody acts without being in the position of a responder. Identifying the true motives rather than their generality and power in explaining some set of data is the primary aim.

Independent evidence for the presence of the motive of inequality aversion rather than merely the effects of such a motive is necessary also because there are many other competing motivational hypotheses that could explain the emergence of the observed results. For instance, the proclivity of the responder to reject offers that are 'too low' could quite naturally be explained in the traditional way by the presence of retributive emotions.[10] However, retributive emotions would

---

[8] In a fuller account the comparison between dictator and ultimatum games on the one hand and on the other hand reward allocation games and ultimatum games with a preceding joint effort—as in the reward allocation case—would be appropriate.

[9] Such effects need not be continuous, though, see for experimental evidence Güth/Huck/Müller 2001.

[10] Philosophers like John Mackie believe that retributive emotions are fundamental to ethics in general, see Mackie 1982.

not explain the observations in dictator games as typically used in experimental economics.[11]

It is one of the great merits of experimental game theory that it raises such questions in a systematic manner thereby inducing the emergence of a dynamic research agenda. But we do not think that it will in the end help to rescue neo-classical economics as an explanatory enterprise. Even though much of experimental economics still sees itself within the neo-classical framework, experimental evidence like the preceding should rather be interpreted as anomalous. The results of such economic experiments lend themselves to a (re-)interpretation in categories of boundedly rational choice making. More often than not this interpretation seems much more natural than an interpretation in categories of perfectly rational (maximizing) behavior under additional non-monetary motives like inequality aversion (or whatever).[12] So let us turn to an account of behavior in dictator and ultimatum games in categories of bounded rationality—using the latter term again very broadly for all sorts of deviations from the neo-classical maximization of utility framework.

5. Bounded Rationality in Dictator and Ultimatum Games

Modern economists tend to refer to themselves as methodological individualists. Accordingly explanations of social phenomena should in the last resort be 'reduced' to hypotheses about individuals. However, strangely enough such hypotheses hardly play any role in approaches based on perfect rationality in which individuals are represented only by the stenographic device of utility functions. According to a common misconception of economists, starting from such given preferences as represented by utility functions eliminates any need to rely on psychology. If we reject the axioms that guarantee the existence of the utility function we also eliminate what economics has left of the individual. No wonder that a kind of 'horror vacui' takes hold of many economists if utility maximization is taken away.

Our fears of the void or the vacuum notwithstanding the optimization paradigm as based on individual utility maximization should be eliminated from economics without further ado. If real progress is to be made economists must put aside their illusions of having a general theory of human behavior first. 'Optimization under constraints' must be abolished as the fundamental explanatory concept of economics. We think that this should be done even if there is nothing yet that can be put in its place and even though we believe that something must be put in its place eventually. Waiting for a full fledged substitute for the 'optimization under constraints' paradigm before experimenting with new forms of theory formation would be unwise since it would block investment in alternative

[11] The desire not to disappoint others could be useful in that regard (Rabin 1993) since it would explain first mover behavior in the dictator game as well.

[12] At the risk of flogging this to death let us reiterate that *homo oeconomicus* cannot be rescued by appealing to 'as if' arguments.

forms of research and thereby perpetuate a dismal status quo of theorizing. As we shall indicate in the final, more constructive but extremely tentative step of our present discussion, economists can and should explore new approaches immediately.

Since there is no unified, general theory of boundedly rational behavior and space is limited we will focus in our discussion of reward allocation (dictator) and ultimatum games on the fairly standard and prominent category of satisficing rather than optimizing behavior (see Simon 1957; 1986). According to this view, humans do not go for the best but rather for results that satisfy their aspiration levels. Aspirations of, say a car manufacturer, may be, 'to stay in business', 'avoid losses', 'avoid cutting back the work force in any dramatic manner', 'keep share prices from falling', 'try to be ahead of the average performance', 'try to meet a profit target' etc. The car manufacturer tries to see to it that the aspirations are met but is not restlessly striving to find the best alternative (neither is he directly 'jumping' to the best alternative all the time since information processing is neither costless nor perfect).

To refer to satisfying aspirations or to 'satisficing' rather than to the single-minded desire to 'maximize' utility has a realistic ring to it. However, from a theoretical point of view it does not say much unless we can specify to some extent how certain motives will be operative in a boundedly rational manner.[13] We think that again—as can be learnt from the difficulties of rational choice approaches—perceptions and the framing of situations (see, of course, Kahneman/Tversky 1984) do play a crucial role in triggering certain responses and cognitive cum motivational processes.

5.1. The Formation of Aspiration Levels

With respect to the manner in which aspirations become operative, considerable evidence suggests that human subjects often have aspiration levels along several dimensions. If their aspirations are not met even after extended effort, modification of at least one of the levels will be pending. This modification can involve several dimensions and follow some lexicographic pattern.[14]

How aspirations are generated is a difficult question. For the sake of specificity let us again take a look at the simple ultimatum game. In such a setting according to a first plausible hypothesis a reacting individual or responder will not aspire to get more than p/2. From the results of many experiments we know that offers of p/2 are practically always accepted and offers greater than p/2 are practically never made by the proposer. There will presumably be types of individuals who in the responder role will reject any offer $y<p/2$. Their aspiration level is $y^*=p/2$. Others will tend to accept some offer $y<p/2$. Of special

[13] To apply the theory of aspiration satisfaction to theory formation itself let us note that we should start from an aspiration level of theory formation that actually can be fulfilled. The aspiration to present a theory of human aspiration level setting that would specify aspirations across contexts in a general manner clearly cannot be fulfilled. But we may hope to say something more specific about the manner in which aspirations become operative and also about the process in which they are formed or modified through time (starting from some initial set up) in some exemplary contexts.

[14] Lexicographic modes of thinking are discussed in the case studies in Ahlert/Kliemt 2001.

interest is the acceptance threshold or the aspiration level y^* such that they will not accept any y with $y<y^*<p/2$.[15]

It is impossible to say something about the absolute value of individual aspiration levels in general. There is also heterogeneity in any population of individuals. Different individuals will be endowed with different aspirations y^* that must be satisfied should they not become inclined in the responder role to reject the offer of the proposer. Given y^* the following classification of response behavior to offers y emerges:

$y=p/2$	acceptance will be forthcoming;
$y^*\leq y<p/2$	acceptance will be forthcoming, if grudgingly;
$y<y^*<p/2$	rejection will be triggered since the aspiration level is not satisfied.

Forming aspirations in the role of the proposer is somewhat more complicated. Being confronted with such a decision task as playing the ultimatum game in the proposer role, individuals might reason along the following lines which we present in dialogic 'query and answer' mode to give just one example of conceivable cognitive processes of deliberation:

1. Q What am I trying to achieve?

   A To get a large x is desirable but my proposal must be accepted (there is a 'maximization under constraints'-ring to this consideration).

2. Q Is there a conflict between my desire for x and the acceptance by Y and when does it emerge?

   A $y<p/2$ may upset the responder while $x\leq p/2$ will bring me on the safe side as far as this is concerned and in particular $x=p/2$ would also satisfy my own feelings of self-esteem.[16]

3. Q If I go for $x>p/2$ where is the critical threshold which will trigger Y's retributive response?

   A There is no definite answer to that query. Either the risk must be taken or not. But a kind of prominent fraction like $y^*=p/3$ is the most likely rejection threshold and the corresponding rejection rule guiding Y's behavior is 'reject if $x/y>2$'.

4. Q What should I do in the light of the foregoing?

   A Either offer $y=p/2$ or take the risk and offer $y=p/3$.

The preceding reasoning is not very complicated. We do not have much more to offer in its favor than its plausibility. But we readily admit that the reasoning

[15] Such acceptance thresholds were 'elicited' by questionnaire, for instance, in the KRESKO experiments; see Tietz/Weber 1972.

[16] See for an inroad into the territory of the 'economy of esteem', Brennan/Pettit 2004.

could be otherwise as well. In particular it could have stopped earlier in the process. An individual searching for a plausible demand that might end her own uncertainty about what her own aspirations should be, could for instance stop after step 2. She might say after the answer to the query in 2 that the alternative p/2 has much in its favor and should therefore be chosen. For rather small p it may not appear worthwhile to go into such a matter too deeply. So somebody who is economizing on decision effort might be content after step 2 to let things rest at that. Somebody else might of course go on and, at a deeper level, come to the same conclusion. For instance someone might reason that if p seems small to the responder he may become more inclined to reject the offer since the opportunity costs of expressing retributive emotions are low. Again the same result may be reached from going down the list to 4.

What this all shows is, of course, that our theories of cognitive processes leading to the formation of aspiration levels even in such simple cases as ultimatum games are rather underdeveloped. We readily admit that. At the same time we think that experimentation in economics should aim to develop such a theory. Rather than repairing a neo-classical maximization approach which quite clearly is not present in the cognitive processes underlying choice making, we should try to lift the veil of 'utility maximization' and form a model of the individual as a choice maker.

Starting with examples like the ultimatum game may be good policy since it keeps things reasonably simple. Complications immediately emerge if we try to generalize. For instance, ultimatum games are sometimes discussed under the heading of 'ultimatum bargaining games'. Posing an ultimatum is just the limiting case of bargaining. If we allowed for several rounds of offering and responding between two actors X and Y, a real negotiation process about the distribution of the pie p could unfold. What both actors demand in this process may not be their true aspiration level. Fixing an initial demand d_{i0} for actor i=X, Y may rather be the result of some kind of strategic act in which both are 'testing the water'. The initial demands d_{i0} will therefore typically exceed p; i.e. $p < d_{X0} + d_{Y0}$. A theory of aspiration adaptation will involve steps which lower the initially incompatible demands such that individuals descend to the next level below. One meaningful hypothesis here might be that individuals tend to reduce their demands by stepping down one aspiration level at a time. Simultaneous or alternating concessions proceeding to the next lower aspiration level may then lead to an aspiration balancing equilibrium which is reached after an equal number of concessions (see Tietz/Weber 1972 and for an effort to present the process of aspiration adaptation and balancing in bilateral bargaining Ahlert 2003).

5.2. Avenues of Research on Boundedly Rational Reasoning

If we want to know more about boundedly rational behavior it will be necessary to learn more about actual reasoning processes of human individuals. Of course, we are entering the field of cognitive psychology here. Acknowledging this we nevertheless think that economists cannot stay out of the 'psychological

fire'. Being well aware of our own deficiencies in these regards as well as of the immature state of the field of cognitive sciences, we shall confine ourselves to a speculative discussion of the basic examples of reward allocation and ultimatum games.

5.2.1. Thinking Aloud

Thinking aloud procedures are fairly well-established in psychology. They are a way to get some handles on actual human reasoning processes by letting the reasoners 'think loud'. For the sake of specificity imagine that a reward allocation game including the problem of fixing an allocation is played under the 'think aloud' regime. Individuals in the role of the 'dictator' are asked to report what ever comes into their mind. Whatever they say is recorded, transcribed, analyzed and then classified in broader categories according to content.

The difficulties that such studies encounter are obvious. There are introspection-biases created in reporting itself. The validity and reliability problems of content analyses are well known from other branches of social and psychological theory, e.g., media research, and cannot be neglected in thinking aloud studies as well. Still, there are at least some remedies for the deficiencies, e.g., the technique of letting at least two experts look through the protocols independently and to let them classify reasons according to pre-specified catch-words or phrases.

Regardless of their deficiencies, 'think aloud' studies can clearly assist us in the formation of theories. But it should also be noted that they are no substitute for theorizing. Neither is it a good idea to hope for the inductive emergence of a general theory by simple repetitions of 'think aloud' studies. It can be hoped though that conducting several 'think aloud' studies may provide a good initial grasp of the several factors involved. In a next step, theory formation may and should ensue. Moreover, introspectively created 'common sense' theories may be 'tested' or at least critically assessed.

5.2.2. Artificial Agents

To provoke specific reactions in a controlled way artificial agents may be used. For instance, imagine again a reward allocation game r. If the individual X in the dictator role were asked by an artificial agent (typically a computer program that is possibly endowed with some kind of artificial intelligence) assuming the role of a partner, interviewer, adviser, consultant etc. whether she thinks that fairness is important in that situation, this might give us some clue on how she sees the situation.

To use artificial agents is in all likelihood not neutral with respect to results but it can create valuable information in a controlled way (see for an early application of such ideas to social science in the present *Journal*, Baurmann/Mans 1984). After all, the artificial agent, unlike a human being, will always respond in exactly the same manner to identical inputs. Technically speaking, its reactions are a function of the responses of participants in an experiment. The next question asked by the program is triggered by the response of the natural or

personal agent. It seems that the potential of research based on artificial agents has not yet been sufficiently explored. In particular if it is used in re-runs of former experiments it might create some useful insights. Since we come back to the suggestion of 're-runs' in the next section, in which personal actors are teamed with another personal actor rather than an agent, let us move on.

5.2.3. Teams as Unitary Decision-Makers

Repeating some of the experimental games with teams whose internal discussions can be observed and recorded rather than with persons as players may, on the one hand, control for biases and, on the other, produce new insights. If the distribution and patterns of results of team play are broadly the same as those of individual play of the same games, one can be fairly sure that team- and explicitness-biases are not too strong. Letting teams of players play the same games as individual players, we may hope to learn something about the truth of the hypotheses that were formed about the reasoning of persons. There will be evidence about what individual team members regard as relevant considerations for the team. Since many of our decisions and, for that matter, many of those most important in business life are made in inter-personal (discussion) processes, such ways of experimenting may not seem too strange to participants. We may therefore generally hope that the techniques of research may not bias results too much. Finally it seems quite obvious that decentralized decision-making as known from firms or companies is mirrored in individual decision processes and vice versa. There seem to be interesting analogies between, say, 'mental accounting' in the case of an individual and setting budgets to sub-divisions in case of a firm.

More specifically, think of teaming up pairs of individuals in an ultimatum game. They now earn the same payoffs jointly that personal players earned before. To control for the effect of splitting the pie by four rather than two the monetary pie p might be transformed into 2p. A 50-50 within-team payoff split should be imposed by the experimenter. The players could be kept incommunicado except for a communication channel that allows for the exchange of messages over the computer key-board. The individuals forming the team in the proposer role might even get some advice concerning the structure and direction of their opinion formation and likewise the individuals forming the team at the receiving end. Different designs or treatments are conceivable here. Except for such variations the rules should be the same as in a former ultimatum experiment and in all likelihood results will be qualitatively the same (if not, this would suggest that there be a team influence as such and the observations made might be less relevant as evidence about individual cognitive processes).

Recording and analyzing what individuals say etc., we may hope to gain some insights on what is in fact driving their decisions. It seems likely, for instance, that one can learn whether motives such as 'inequality aversion' are present. Intentions in the proposer role might become more transparent. Likewise for the responder role—we may find out whether an aversion to inequality drives subjects or whether retributive emotions induce individuals to reject offers.

5.2.4. Changing the Lab Environment

Fancy research techniques like connecting individuals to brain-scanners may increasingly be used in the future. But the workings of the human brain can be better understood, too, by employing less fancy strategies. One particularly interesting method emerges if we let individuals retrieve information in a more formalized manner from a computer on which a program like 'mouse lab' is installed.

'Mouse lab' has been aptly called 'the poor man's brain and eye scanner'. Basically it records the movements of a mouse or pointer on the computer screen. Thereby we know when and how often individuals open certain boxes containing specific information relevant for a decision problem that they face. For instance if there is a decision process with several stages then backward induction arguments would apply among fully as opposed to boundedly rational players. Mouse lab allows for recording whether and how often information on later stages of the game is retrieved. Since information that allows for backward induction is often ignored, it is obvious that humans do not reason according to the precepts of full rationality. This is true even if the information is available easily.

6. Conclusion: *Homo Sapiens* as Bounded and Expecting Man

Human behavior is always both drawn by the future and pushed by the past. Humans can in exceptional instances act in a purely opportunistic manner. But more often they act in a way that fails to exploit the full potential of opportunities. Their cognitive capacities as well as their ethical will impose constraints on them. They are cognitively or normatively bounded in their ability to engage in opportunity taking behavior. Sometimes they can overcome boundedness. Consulting experts, engaging in an exceptional effort of situational analysis may be helpful in transcending cognitive routines. Likewise, those who intend to abandon old habits and allegiance to some ethical norm of conduct or other may do so as well by some special effort or by simply giving in to temptation.

We have indicated very tentatively some possibilities for future research. This research must take into account the fact that human behavior cannot be adequately understood unless *both* future directedness and adaptation to past experience are taken into account. To find the fundamental mechanisms of human coordination we must look at repeated situations under several influences including frequent external shocks of minor and sometimes major proportions. Even though we reject the typical neo-classical focus on optimization as brought about by selective adaptation in repeated situations, we do not deny that repetition, learning and adaptation in gradual processes of trial and error are of the essence of the economic process and therefore must be studied in detail.

The case-by-case maximization assumption of the standard perfect rationality model in neo-classical economics is so far off the mark that it becomes hard to find real world examples for it. It would be much better if economists ceased

to acclaim those most who manage to 'explain' everything in terms of 'rational choice'. What can be won by economic story telling of that kind?

On the other hand, the deficiencies of the bounded rationality approach, its lack of specificity, sometimes even its lack of empirical content and certainly of general applicability are obvious. But let us not forget that utility maximization, though its mathematical precision nurtures all sorts of illusions, is no better in that regard. Worse, it is a dead-end since it does not support efforts to form realistic models of decision processes. The bounded rationality approach is at least a step towards modeling real decision processes. Since it does not insist that everything be cast into the maximization under constraints mold, it can be more open with respect to all sorts of theorizing.

However, a new synthesis of several strands of research on human decision making is needed if real progress is to be made. Cognitive psychology can be utilized in the move towards theories of boundedly rational behavior (reaching from Festinger 1957 to Gigerenzer 1996; 1997). Cognitive science in the broader sense of that term as inspired by Simon and then pushed on in several directions may bear promise too (see for some non-standard contributions opening new perspectives of research Clark 1997; Hutchins 1995). Even though one should be careful not to be carried away by the intellectual fashions of the day there is some hope out there.

It seems striking how complex the discussion of even such simple game structures as reward allocation, dictator and ultimatum games can get. Complexity is not a good but rather a bad thing for boundedly rational theoreticians like us. But if it unfolds from going over every nook and cranny of such simple structures as we studied we may speculate that we might be on the right track towards the exemplary understanding of the delicate relationship between norm orientation and opportunism in human behavior that characterizes homo sapiens as we know him in practice and should get to know in theory.

Bibliography

Ahlert, M. (2003), (unpublished manuscript), *Agreement by Aspiration Balancing – A New Axiomatic Approach to Bounded Rationality in Negotiations*, University of Halle

— /H. Kliemt (2001), A Lexicographic Decision Rule With Tolerances. The Example of Rule Choices in Organ Allocation, in: *Analyse & Kritik 23*, 191–204

Alchian, A. A. (1950), Uncertainty, Evolution, and Economic Theory, in: *Journal of Political Economy 58*, 211–21

Baurmann, M./D. Mans (1984), Künstliche Intelligenz in den Sozialwissenschaften. Expertensysteme als Instrumente der Einstellungsforschung, in: *Analyse & Kritik 6*, 103–59

Becker, G. (1976), *The Economic Approach to Human Behavior*, Chicago-London

Bolton, G./A. Ockenfels (2000), ERC: A Theory of Equity, Reciprocity and Competition, in: *American Economic Review 90*, 166–193

— /R. Zwick (1995), Anonymity versus Punishment in Ultimatum Bargaining, in: *Games and Economic Behavior 10*, 95–121

Brennan, G./P. Pettit (2004), *The Economy of Esteem*, Oxford

Buchanan, J. M./G. Tullock (1962), *The Calculus of Consent*, Ann Arbor

Clark, A. (1997), Being There. Putting Brain, Body, and World Together Again, Cambridge/MA

Fehr, E./K. Schmidt (1999), A Theory of Fairness, Competition, and Cooperation, in: *Quarterly Journal of Economics 114*, 817–868

Festinger, L. (1957), *Theory of Cognitive Dissonance*, Evanston

Frankena, W. K. (1966), Some Beliefs about Justice: Lindley Lecture, The University of Kansas, Lawrence

Gantner, A./W. Güth/M. Königstein (2002), Equitable Choices in Bargaining Games With Joint Production, in: *Journal of Economic Behavior and Organization 46*, 209–225

Gigerenzer, G. (1996), Reasoning the Fast and Frugal Way, in: *Psychological Review 103*, 650–669

— (1997), Bounded Rationality: Models of Bounded Rationality, in: *Swiss Journal of Economics and Statistics 133*, 201–218

— /J. Czerlinski/L. Martignon (1998), How Good Are Fast and Frugal Heuristics?, in: J. Shanteau/B. Mellers/D. Schum (eds.), *Decision Research from Bayesian Approaches to Normative Systems: Reflections on the Contribution of Ward Edwards Kluwer*, Dordrecht, 1–23

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

Güth, W. (1976), Towards a More General Analysis of von Stackelberg-Situations, in: *Zeitschrift für die gesamte Staatswissenschaft 132*, 592–608

— /S. Huck/W. Müller (2001), The Relevance of Equal Splits in Ultimatum Games, in: *Games and Economic Behavior 37*, 161–169

— /C. Schmidt/M. Sutter (2002), Bargaining Outside The Lab – A Newspaper Experiment Of a Three Person Ultimatum Game, Max Planck Institute for Research into Economic Systems, Jena. Discussion Paper Series

— / — / — (2003), Fairness in the Mail and Opportunism in the Internet – A Newspaper Experiment on Ultimatum Bargaining, in: *German Economic Review 4*, 243–265

— /R. Schmittberger/B. Schwarze (1982), An Experimental Analysis of Ultimatum Bargaining, in: *Journal of Economic Behavior and Organization 3*, 367–388

Heiner, R. (1983), The Origin of Predictable Behavior, in: *American Economic Review 73*, 560–83

Hutchins, E. (1995), *Cognition in the Wild*, Cambridge

Kahneman, D./A. Tversky (1984), Choices, Values and Frames, in: *American Psychologist 39*, 341–350

Königstein, M. (2000), *Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production: Lecture Notes in Economics and Mathematical Systems*, Berlin-Heidelberg-New york

Mackie, J. L. (1982), *Morality and the Retributive Emotions*, in: *Criminal Justice Ethics 1*, 3–10

McKenzie, R. B./G. Tullock (1978), The New World of Economics, Irwin

Meckling, W. (1976), Values and the Choice of the Model of the Individual in the Social Sciences, in: *Schweizerische Zeitschrift für Volkswirtschaft und Statistik 112*, 545–563

Mikula, G. (1973), Gewinnaufteilungsverhalten in Dyaden bei variiertem Leistungsverhältnis, in: *Zeitschrift für Sozialpsychologie 3*, 126–133

Milgrom, P./J. Roberts (1993), *Economics, Organization and Management*, Englewood Cliffs

Rabin, M. (1993), Incorporating Fairness Into Game Theory and Economics, in: *American Economic Review 83*, 1281–1302

Roth, A. E. (1995), Bargaining Experiments, in: J. H. Kagel/A. E. Roth (eds.), *The Handbook of Experimental Economics*, Princeton, 253–348

Schumpeter, J. A. (1959), *The Theory of Economic Development*, Cambridge/MA

Shapiro, E. G. (1975), Effects of Future Interaction on Reward Allocation in Dyads: Equity or Equality, in: *Journal of Personality and Social Psychology 32*, 873–880

Simon, H. A. (1957), *Models of Man*, New York

— (1985), *Models of Bounded Rationality (1&2)*, Cambridge/MA

Sugden, R. (1991), Rational Choice: A Survey of Contributions from Economics and Philosophy, in: *The Economic Journal 101*, 751–85

Tietz, R./H.-J. Weber (1972), On the Nature of the Bargaining Process in the KRESKO-Game, in: H. Sauermann (ed.), *Beiträge zur Experimentellen Wirtschaftsforschung*, Tübingen, 305–334

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