

Perfect or Bounded Rationality?

Some Facts, Speculations and Proposals

Werner Güth and Hartmut Kliemt

Abstract: The paper focuses on empirical studies of a class of very simple games of the reward allocation, dictator and ultimatum type. It is demonstrated that true explanations of social phenomena cannot conceivably be derived in terms of the perfect rationality concept underlying neo-classical economics. Analyzing the simple paradigm games it is explored in some depth, if speculatively, how experimental game theory might bring us closer to a new synthesis or the nucleus of a general theory of “games and boundedly rational economic behavior” with more explanatory power.

1. Introduction and overview

Evaluated in entrepreneurial terms neo-classical economics is a success among the social sciences. The numbers of professionals is ever increasing and economics is well respected in adjacent fields of social inquiry. Some of its core techniques are widely used by other social scientists while economics is so self-confidently tackling all sorts of social science issues within its “optimization” or “rational choice” paradigm that one may indeed speak of “economic imperialism”. Political science departments at least in the US lean more and more towards “economic theories of politics” while journals like “Public Choice”, “Theory and Decision”, “Economics and Philosophy” or “Rationality and Society” are flourishing. Game and decision theory in particular provide a modern “lingua franca” for social science (reaching even into biology). The prisoner’s dilemma, the battle of the sexes and the chicken game which all originated from elementary rational choice modeling of social phenomena have taken hold on the minds of social scientists across the board. We see the world as full of “games and decisions” ((Luce, D. and H. Raiffa 1957)), study the “economic games that people play” ((Maital, S. and S. L. Maital 1984)) and

generally use “games as models of social phenomena” ((Hamburger, Henry 1979)).

The existence of the common decision and game theoretic *language* along with common paradigm examples of elementary social problems is extremely valuable in facilitating communication about social research.¹ However,

¹ For instance the project “game theory in the behavioral sciences” organized in the academic year 1987/1988 at the ZiF – center for interdisciplinary research – of the university of Bielefeld brought together an interdisciplinary group from fields as economics, philosophy, mathematics, political science, sociology and very markedly biology. Due to their training in modern decision and game theory people in the group could actually talk to each other quite easily across the borders of distinct disciplines as traditionally defined. It may be noted in passing that the two authors of the present paper (an economist and a political philosopher) met the first time at the ZiF and started to find out how much common ground there was between them. It may also be instructive to note that we later on – in the academic year 1999/2000 – co-organized jointly with Joachim Frohn and Reinhard Selten a project “making choices” which focused on bounded rationality as opposed to the game equilibrium models that were the focus of the earlier project. It indicates, we think, part of the dynamics of social science research. The role of the ZiF in our research and also preparing the ground for writing this paper is gratefully acknowledged; for further documentation of the work in the two projects see the volumes edited by Selten, Reinhard ed. 1991a. *Game Equilibrium Models I - Evolution and Game Dynamics*. Berlin / Heidelberg / New York / London / Paris etc.: Springer Verlag.

Selten, Reinhard ed. 1991b. *Game Equilibrium Models II - Methods, Morals, and Markets*: Springer Verlag.

Selten, Reinhard ed. 1991c. *Game Equilibrium Models III - Strategic Bargaining*. Berlin / Heidelberg / London etc.: Springer-Verlag.

Selten, Reinhard ed. 1991d. *Game Equilibrium Models IV - Social and Political Interaction*. Berlin / Heidelberg / New York etc.: Springer-Verlag., Frohn, Joachim, Werner Güth, Hartmut Kliemt, and Reinhard Selten eds. 2001. *Making Choices I & II*. München: Accedo.,

interpreting decision and game theoretic language literally implies that human individuals do optimize and choose intentionally on the basis of given preferences. These implications are widely off the mark and should not be accepted.² They are at root of a neo-classical orthodoxy that quite naturally leans towards the view that the true art of the economist consists in fitting everything to a description in decision theoretic terms.

To describe human action as the result of “optimization under constraints” amounts to “economic story telling”.³ Evidence from psychology and in particular experimental economics and thus from “inside” economics itself (see (Kagel, John H. and Alvin E. Roth 1995)) seems to indicate that “explanations” of human behavior based on rational choice and optimization except for the simplest cases must be rejected. More often than not full rationality is a mistaken and misleading assumption. It may not even count as an idealization of the “true” behavioral laws if we assume that behavior is fully rational and optimizing. For, real behavior is in all likelihood guided by principles completely different from optimization and rational choice. It is “boundedly rational”.

If that be true then we obviously need a theory of boundedly rational behavior. Yet explanations in terms of bounded rationality have their disadvantages as well. They are typically based on strongly domain dependent theories of quite

Ahlert, Marlies and Hartmut Kliemt eds. 2001. *Making Choices in Organ Allocation*. Stuttgart: Lucius und Lucius.

² In the German speaking countries a fundamental critique of looking at the world from the point of view of decision theory rather than through the perspective of behavioral science has been offered in particular by Hans Albert since the early sixties, see the collection of papers Albert, Hans. 1967. *Marktsoziologie und Entscheidungslogik*. Neuwied/Berlin.: Luchterhand..

³ That the effort to describe things in decision theoretic terms imposes the discipline of making modeling assumptions explicit in our story telling is a great advantage over traditional verbal approaches that should not be underestimated.

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 by often 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 not a common theory with empirical content and explanatory power.

It may well be that in the end there are no convincing 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 common sensical, so be it. If we can only give some advice how domain specific theories can be formulated without being in a position to contribute substantial general information why not accept that?

It may be part of the human condition that human behavior cannot *generally* be foreseen or explained by human beings. Therefore the formation of social theories may remain an “art”. We may be able to train people in the art of forming short-range domain specific theories. In management science and the best of business administration studies this is done even though general theories of social behavior with explanatory content are lacking. And people well-trained in the techniques of formulating short-range (or “ad hoc”) theories carry high market value. If in view of the complexities of social life people have to form their theories in a case by case manner according to their good judgment,⁴ why

⁴ With out subscribing to it in full we would like to point out that this view is to some extent supported by insights reaching from Aristotle's arguments about the role of judgment to

not accept this and see how we can improve our ways of theory formation at least at the margin? Why not seek ways of coping with our ignorance as well as possible without engaging in self-deception?

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 the neo-classical theories of fully rational behavior (3.). The next section deals with some neo-classical repairs and the difficulties that they encounter (4.). The deficiencies of in explaining the results of reward allocation, dictator and ultimatum games in a perfect rationality framework are discussed then (5.). A speculative sketch of some attempts to explain experimental results in a bounded rationality framework follows (6.). Final remarks conclude the paper (7.).

2. Generality of social theories?

The integration of otherwise diverse experiences into one body of knowledge is among the chief purposes of theorizing. General theories shed light on relationships, bring to attention common aspects and build bridges between seemingly separate matters. Standard examples are found in natural science. For instance, with plain common sense hardly anybody would suspect that the movement of planets and the fall of an apple from a tree have something in common. But natural science tells us that there is a common aspect.

Insights of natural science into the common traits of apparently unrelated phenomena are essential to the success of that great enterprise to which we refer as (natural) science. In the realm of social phenomena it would obviously be desirable to have access to theories of comparable scope that could impose some unity on our diverse experience. Classics like most prominently David Hume thought that those who succeeded to develop a theory of “human nature” would

modern arguments about the limits of artificial intelligence as expressed in Dreyfuss. ???? *The limits of artificial intelligence*????.

hold the key to a universal theory of the social world (see [Hume, 1739/1978 #253]). In the famous opening passage of Hume's "treatise of human nature" he says: "Here then is the only expedient, from which we can hope for success in our philosophical researches, to leave the tedious lingering method, which we have hitherto followed, and instead of taking now and then a castle or village on the frontier, to march up directly to the capital or center of these sciences, to human nature itself; which being once masters of, we may every where else hope for an easy victory."

But social science, after Hume, developed into domain specific theories rather than focusing on human nature and on universal behavioral assumptions about human actors.⁵ Economics is no exception to this. From its beginnings at least to the middle of the 20-th century in particular neo-classical economics saw itself as a domain specific theory confined to market behavior, to exchange and commercial relationships as narrowly defined. Only in the second half of the 20-th century neo-classical economics ventured into new territory beyond the realm of the "commercial". The theory of games developed, public choice theory originated as a theory of non-market decision making (the aforementioned journal "Public Choice" started out as "papers on non-market decision making"), the law and economics movement and the new institutional (and constitutional) economics took off (with contributions reaching from (Buchanan, James M. and Gordon Tullock 1962), over (Milgrom, Paul and John Roberts 1993) to (Young, H. Peyton 1998)). Since then the so-called "new world of economics" (see (McKenzie, Richard B. and Gordon Tullock 1979)) laid out 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, Gary 1976)).

⁵ Incidentally, Hume's treatise was translated into German language by a psychologist who indeed saw Hume as one of the "giants" (in the sense of Merton) on whose shoulders he tried to stand as a psychologist.

In all contexts from family to politics the same *rational, expecting, maximizing, men* (REMM, see (Meckling, W. 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. Thereby social science seems to reach a stage of universality and mathematical precision comparable to that of, say, Newtonian mechanics.

If it were in fact true that the behavioral assumptions underlying “rational economic man” would form a universally applicable model of human behavior leading to valid explanations in social realms as diverse as competition of parties at the polls and competition of firms for nearby 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 its unity of language and of description is not matched by unifying behavioral laws applicable to several realms.

True enough, humans command the faculty to act opportunistically on the basis of a model of the action situation. Therefore the over-socialized model of man according to which human behavior is to be explained *exclusively* in terms of internalized values, norms and rules is certainly mistaken. In that regard the economists and more generally rational choice theorists are right. But this is not to say that they have a more valid theory to substitute the theories that they criticize. The economic approach that models behavior as universally opportunistic case-by-case maximization of utility is at least as far off the mark as the classical sociological one. It either has no empirical content or in its concrete specifications is as fragmented as other social theories.

If the behavioral model of the economic approach to human behavior is reduced to the maximization of pecuniary income (classical profit rather than utility maximization) the model has empirical content. But in this narrow traditional interpretation it is obviously neither universally applicable to all realms of human behavior (in particular not to those in which pecuniary motives play no

role) nor is it the sole valid explanation of behavior in the presence of pecuniary incentives. Motives other than pecuniary ones do play a role throughout – even when pecuniary motives are present. To rescue its model of behavior from the most obvious counter arguments the neo-classical economic analysis must either be confined to very narrow domains like special competitive markets in which it applies “in the limit” or it must be rendered void of most of its empirical content by auxiliary assumptions that are often ad hoc.

A richer motivational structure as such would still be compatible with the aim of formulating general theories if domain specific circumstances were merely initial conditions for quite general motivational and behavioral laws; e. g. “in contexts of type x in which the ‘profit motive’ is perceived as a legitimate concern it will be dominant while in other contexts of type y the ‘fairness’ motive dominates”. In that case the richer structure would link different realms containing types of situations x, y as characterized in general terms. But instead of this desirable linking of realms we rather find bits and pieces of theories of motivation that might explain behavior in specific realms without really transcending the limits of those realms or linking them at least to each other.

In particular to include new components in the utility functions of individuals as seems fit according to the exigencies of diverse situations taken separately is not a convincing strategy.⁶ Only if a serious endeavor to formulate a more general or even ‘universal’ such function is made eventually the pragmatic efforts of developing a richer theory need not be a bad thing. But the issue of empirical content and the search for true behavioral laws must be kept firmly in view.

⁶ If we incorporate motives other than the pecuniary ones into the utility functions of individuals then neo-classical economics becomes fragmented beneath the surface of its unified decision theoretic language. The diverse specific motives that, pending on situational complexities, sometimes do and sometimes do not trigger behavioral responses must be taken into account; e.g. sometimes the profit motive is operative and dominant and sometimes not.

Economists for all their claims that they are searching for an empirical science of behavior do not seem to keep focused on the pursuit of behavioral laws. In particular the thesis that maximization under constraints can be an answer to the fundamental theoretical challenge of developing general theories is as strange as it is common among economists.⁷ Things that quite naturally could and should be expressed otherwise are put into the Procrustes bed of “economic maximization under constraints”. Even though very clearly there is no such maximization process in the human psyche things are made to look “as if” actors maximize under constraints across the board.

However, to say that individuals act “as if” they maximize is not good enough. First, it is, to put it mildly, very unlikely that even the weaker thesis of “as if” maximization would with some 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). For it is obviously true that conscious and intentional maximization of a given objective function is not at root of human choice

⁷ Of course, one might remark here that Gary Becker’s approach is different in this regard. Is he not a critic of “ad hocery”? Indeed he insists that in the last resort human motives have to be treated as invariant between situations. In the last resort there is one human nature that is inter-individually invariant. To explain differences in overt behavior economists have to rely on observable external constraints. The internal technology of producing the ultimate satisfactions of basic needs differs but not the needs and ultimate motives as such.

Without going into the details of the Beckerian methodology let us acknowledge that it tries to cope with some of the more obvious criticisms of the neo-classical approach to human behavior in a quite ingenious way. However, it still insists on the “maximization subject to constraints” paradigm. It rescues mathematical elegance and unity of mathematical method at the price of substantial unifying theories. In particular, Beckerian differences in human capital, in the technology of creating ultimate satisfactions seem quite arbitrary ways to “explain” differences in choice making behavior.

making and thus cannot be the true causal law. Third, human behavior is sometimes so-well adapted to circumstances that an omniscient maximizer could not do better. Yet, again, the well-adapted behavior is definitely not the result of conscious maximization. It is in all likelihood the outgrowth of other processes that somehow led 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. Optimality as well as stability rather require 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 as for example (Heiner, R. 1983)).

Wherever results corresponding to the predictions of full rationality should be observed – and we are very reluctant to concede that they are ever observed – an explanation in behavioral laws of boundedly rational behavior 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 a phenomenon under general behavioral laws? If optimality in the sense of “as if” rationality is widespread how can we explain this extraordinary and astonishing fact? What are the general behavioral laws that can explain behavioral 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 perhaps be done let us start with a more specific brief look at the traditional theories of perfect rationality that were involved in making stronger claims to generality and then gradually bring in a more realistic perspective on choice making.

3. (Commonly known) Perfect Rationality and its difficulties

Let $I := \{1, 2, \dots, n\}$, $n \geq 1$, be the set of players and let Γ denote the class of all games G in a domain with

$$G = (S_1, S_2, \dots, S_n; u_1, u_2, \dots, u_n);$$

where for all $i \in I$ the

$S_i \neq \emptyset$ are finite strategy sets,

$S := \prod_{i=1}^n S_i$ is the set of all strategy profiles $s = (s_1, s_2, \dots, s_n)$

u_i are mappings $u_i: S \rightarrow \mathbf{R}$, $u_i(s) \in \mathbf{R}$, which represent the individuals' preferences by a conventional cardinal utility measure.

If $n=1$ optimality of choice behavior requires that the single actor, 1, must choose $s^* \in S$ such that $u = u_1$ is maximized over S ; i.e.

$$s^* \in \operatorname{argmax} \{ u(s) : s \in S \}.$$

In cases with $n \geq 2$ it is not clear what optimal behavior in the sense of maximizing u_i requires of each actor from $i \in I = \{1, 2, \dots, n\}$. If the behavior of all others is known to i and fixed then, i needs to maximize against $s_{-i} := (s_1, s_2, s_{i-1}, s_{i+1}, \dots, s_n)$ such that an optimal result is reached. The information condition (see on how "truly perfect information" in hierarchies can guarantee Pareto efficiency (Thompson, Earl and Roger Faith 1981)) that the actions of all others are fixed and known can conceivably be fulfilled only for one individual. In all other cases the question of how to deal with situations in which the actions of other actors are not known (either not fixed yet or unknown for other reasons) emerges at least for some actors. This raises the question which criteria of optimal behavior could be applied by those individuals.

Though there is no commonly accepted answer to this question most economists would tend to require as a minimum that choices be in equilibrium (but even this is contested, see for instance (Sugden, Robert 1991)). A strategy vector $s^* \in S$ is in equilibrium (see (Cournot, Auguste 1838), (Nash, John F. 1951)) iff

$$\forall i \in I: s^*_i \in \operatorname{argmax} \{ u_i(s_i, s^*_{-i}) : s_i \in S_i \}.$$

As stated not all individuals can be informed about all choices as *fixed* or *made* by all other individuals. In that sense ignorance as to some of the choices of others necessarily (in a logical sense of necessity) applies when the other $n-1$ of the choosers make their choices. For instance, in a two person sequential game

one actor must move first and thus without knowing the choice of the other actor. He can have views on which actions the other actor will take but he cannot know in advance the other choice as made. In simultaneous move games, to which we confine attention for the time being, all n choice-makers must fix their choices without knowing the choices of others. So in making these choices none can maximize against the *known* choices as made by others.

Even in very idealized circumstances in which the game is not only known to all players but common knowledge among them – i. e. they know that all know that all know ... that all know the game – players cannot maximize in any simple sense that would correspond to the case $n=1$. As a consequence of this, though in equilibrium each player maximizes against the choices of each other individual, it is not clear how even fully rational individuals would get to the equilibrium by their actual choices.

To put it slightly otherwise: What kind of play could we recommend to players who intend to act as rational utility maximizers in such situations? This question has no easy answer if any. For example, consider a standard battle of the sexes game:

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

Figure 1

An omniscient outside observer might conceivably be able to predict the choices of all individuals and to feed each of them the correct information on what the others 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 to feed that anticipation back to player 1. This will lead to valid predictions only if player 1 will not alter his plans upon receiving the information about player 2 ...

If we assume that players respond rationally to the known choices of others according to the maximization paradigm of a single actor – the case $n=1$ – then s_1^1 must already be the optimal response to the optimal response of player 2 and so on indefinitely. Thus the omniscient outside observer should predict that player 2 chooses s_2^1 contingent on the information that the choice of player 1 will be s_1^1 . Only in that case the dictates of isolated actor rationality under full information about all other pay off relevant parameters and predictions of interaction decisions can be consistent.⁸

⁸ 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 in fact very much at 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; see on this Albert, Hans. 1964. "Probleme der Theoriebildung," in *Theorie und Realität*. Hans Albert ed. Tübingen: Mohr.

Dacey, Raymond. 1976. "Theory Absorption and the Testability of Economic Theory." *Zeitschrift für Nationalökonomie*, 36:3-4, pp. 247-67.

Dacey, Raymond. 1981. "Some Implications of 'Theory Absorption' for Economic Theory and the Economics of Information," in *Philosophy in Economics*. Joseph C. Pitt ed. Dordrecht: D. Reidel, pp. 111-36.

Güth, Werner and Hartmut Kliemt. 2000. "From full to bounded rationality. The limits of unlimited rationality." Center for Interdisciplinary Research (ZiF): Bielefeld.

Morgenstern, Oskar. 1972. "Descriptive, Predictive and Normative Theory." *Kyklos*, 25, pp. 699-714.

Now, an omniscient outside observer bringing about the “right” choices of all actors by signals (see on related issues (Aumann, Robert 1990; Aumann, Robert and Adam Brandenburger 1995), (Brandenburger, Adam and Eddie Dekel 1987)) is even more of a fiction than the assumption of common knowledge of the game and of full rationality of the individual players. Human beings interact without any external omniscient source of information. Theories of how to play must substitute the information process of the outside observer.

More specifically, besides common knowledge of the game individuals must have access to a theory to co-ordinate their behavior. If individual players are fully rational in the sense of individual maximization of utility then it must be possible that the theory is commonly known to them. Otherwise individuals might speculate that other individuals could follow other theories. Moreover, under common knowledge of the game and the theory of rational play no individual must have an incentive to deviate from the precepts of the commonly known theory if all others follow the theory. And finally the theory must give definite advice to each and every player such that an equilibrium will be selected (see for the most refined such approach (Harsanyi, John C. and Reinhard Selten 1988)).

These are the minimum ideal conditions that must be fulfilled if an equilibrium is not only to exist but also to be predicted as the outcome of rational play. The corresponding equilibrium play can be derived as the outcome of the deliberate rational choices of players in one-off interactions only if we have a theory of what the players think, what they think what others think ..., and what their cognitive processes are. To assume that they use the same theory commonly known to them is a way of approximately capturing what should be on the players' minds.

To illustrate, let us pursue the battle of the sexes game somewhat further. For instance, 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, too, and know that they are, when a theory of rational play that suggests that the players play (s_1^1, s_2^1) is laid out to both players. Then they are brought into separate rooms to actually play the game without further communication. In this situation it is safe to assume that the matrix of figure 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 the simplest cases 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 the 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 even in strategic contexts involving multiple equilibria. 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 absolutely unlikely that a theory of equilibrium selection per se could provide the true behavioral explanations. To assume that it would be necessary to claim that the theory of equilibrium selection would have to be present in the mental processes of the players. Interpreted literally, as is necessary if we aim at true explanations, this is absurd.

But even less demanding rationality assumptions may be quite absurd. In particular the seemingly innocuous premise of the utility representation of preferences, that players can and do know their own preferences as well as those of others, is not harmless. Even in the simplest extreme case in which the number of decision makers is $n=1$ it is rather problematic to treat preferences as

“given”. Assume that the stakes are very high. Say a pedestrian decision maker after winning in a state lottery plans to invest into a portfolio the sum of \$ 1.000.000.

To invest a million is a nice prospect for the individual but does she know her preference ranking among all portfolios? How can she – or we – assign a utility to such yet unknown preferences? Even if the investor would focus merely on a subset of some few “prominent” portfolios she would have to “construe” her preferences first. She would start to compare alternatives along several dimensions of evaluation and structure that process by specific rules of thumb like “a third in bricks, a third in bonds, a third in stocks”. If she is clever she might also say that she wants to have independent risks and therefore invest internationally etc. But at the same time she may – and will in all likelihood – fall prey to familiarity and other biases (for a survey of experiments where complexity renders optimality extremely unlikely, see Anderhub and Güth #####).

More generally speaking we all act frequently the “fast and frugal way” (see (Gigerenzer, Gerhard 1996;1997;Gigerenzer, Gerd, Jean Czerlinski, and Laura Martignon 1998)). We are facing constraints inducing boundedly rational decision-making throughout. These constraints are even more marked in strategic interaction than in single actor contexts. But as the example of the investor shows 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 species homo sapiens as opposed to the economic world of homo oeconomicus.

Most economists are still sticking to the model of utility maximization even though most of them are aware of the preceding line of argument. That they do so, and engage the task of providing “neo-classical repairs” is in itself a remarkable fact. This fact should not be dismissed lightly. So let us turn to some of the more recent efforts to rescue the traditional maximization paradigm.

4. Some neo-classical repairs and their difficulties

4.1. Homo oeconomicus as exception

If ever there might be a universal behavioral model then it will in all likelihood frame human behavior as both drawn by the future and driven by the past (see programmatically, if still to some extent in the spirit of neo-classical repairs (Güth, Werner and Hartmut Kliemt 1998) and on other methodological aspects of evolutionary economics (Nelson, R. R. and S. G. Winter 1982) and (Witt, Ulrich 1987)). Expectations of the effects of individual acts that are taken strictly separately as well as rules comprising a class of actions inform our choice making. Human behavior is, in short, boundedly rational. It is rational in the shadow of the future and bounded by the shadow of the past. Homo sapiens is not a straightforward maximizer but torn between the extremes of strategic calculation and blind rule following. But he can and does seize certain opportunities.

To indicate “exceptional” instances in which we would expect forward-looking rational choice to play a role imagine decision making in firms. If such decision-making would exclusively rely on past experience and be a mere trial and error process the business world would be a different place. True enough, management may be chiefly “management by exception” (see Simon and March #####). Routines as for instance in planning the amount of goods to be kept in stock typically develop in a process of trial and error through time. Routines may indeed be selected and thereby form the basis for evolutionary and certainly some learning theoretic modeling in economics. But it is also clear that the exceptional management efforts do matter. Forward looking strategic choice making does clearly play a tremendous role in such processes like introducing new products, entering a new market, in reacting to exogenous shocks in a market and, very importantly, in reforming the routines of a company (constitutional economic choice making in the sense of (Buchanan, James M. 1999)) ranging from outsourcing, over vertical integration decisions, to acquiring, founding or selling other firms.

The preceding view would render homo oeconomicus and rational choice analyses exceptional. It would give up any claim to universal applicability. The theory would be rescued for special class of decisions. A somewhat more general way to do the same emerges if we rely on the conventional distinction between low and high costs.

4.2. Low and high costs in a maximization framework

In view of the human ability to seize opportunities the economists' special focus on opportunistic, in particular monetary motivation may have some merit. At least in studies of experimental economics it is often possible to exert some control over the opportunistic inclinations as triggered by the motive to earn money by strategic behavior. The rules of the art of economic experimentation therefore suggest that strong monetary incentives be linked with behavior in experiments. It is required that subjects are presented with incentives such that it can be hoped that individual proclivities to show certain forms of behavior increase monotonically with the size of the expected monetary reward from showing such behavior. It is also suggested that the monetary reward must be salient in the sense that it is directly and prominently linked to the choice making in the experiment. Finally dominance of the monetary as opposed to other motives should be aimed at (see (Friedman, James and Shyam Sunder 1994)).

The aim of the exercise of designing experiments according to such rules of the art is to have some control over the preferences of actors. Experimentalists intend to induce the classical behavior of rational economic man who is interested exclusively in monetary gain. But regardless of their efforts they almost never succeed in this completely. Always motives other than direct monetary gain influence human choice making. This seems to apply even in some high stakes experiments in which the argument from low opportunity costs that would not trigger case-by-case maximization does not apply. It is intuitively plausible that the size of opportunity costs will influence behavior (see on some high stakes experiments (Harrison, Glenn W. 1989), (Slonim, Robert and Alvin

E Roth 1998)). The higher the pecuniary rewards the stronger the tendency towards opportunistic behavior should be.⁹

In view of the preceding economists might want to endorse the standard view of ethical theory that human behavior is different if stakes are low than if they are high and may legitimately be expected to be so (see on this see (Brennan, H. Geoffrey and James M. Buchanan 1984; Brennan, H. Geoffrey and Loren Lomasky 1984), (Heyd, David 1982), (Selten, Reinhard 1986), (Urmson, J. O. 1958)). In view of this a general theory of cost asymmetries could be formed (see for a brief sketch (Kliemt, Hartmut 1986)). According to that theory behavior in low will differ from behavior in high cost situations. The economists could come in and stipulate that their theory applies chiefly to high cost situations. Actors will start to maximize only when it really matters or if that effort really pays. Economists should again be warned, however, that with these modifications rational choice modeling has been given up as a universal method. Even if individuals start to maximize deliberately if it really matters this does not imply that only these situations matter. Quite to the contrary, much of social structure in fact does depend on the systematic exploitation of cost asymmetries and the human proclivity to show norm-guided behavior in low cost situations. Therefore a large and essential realm of human behavior – typically norm guided behavior in low cost situations – must be explained according to

⁹ The remaining ambiguity of the motivational structure could be controlled to a certain extent by systematic variations of the size of monetary rewards either with the same individuals on different rounds of experimentation (within subjects design) or with different subjects who engage the same task with different monetary incentives (between subjects-design). For instance, if we would double the monetary rewards, then obviously x units in the first will be $2x$ in the next treatment. The “elasticity” of behavior in response to such variations would give us some clue as to the degree to which the monetary motive does in fact dominate. Regrettably the latter procedure is seldom applied.

principles other than opportunistically rational choice.¹⁰ Moreover, even in high cost situations human individuals at least sometimes do massively deviate from their direct pecuniary interests (see for a more conventional account for instance (Frank, R. 1988)).

Still the neo-classical theorist will not give in here. He will discount maximization of a monetary objective as old fashioned and insist on utility maximization. Even though the pitfalls of that kind of move have been pointed out again and again the argument is still around and we therefore need to deal with it briefly if very reluctantly before we can move on.

4.3. Utility maximization and all that

If we can describe behavior *as if* it were the outcome of a conscious maximization effort of a more complicated utility maximand this is interesting but raises more questions than it answers. It is an explanandum not an explanation since the fact that predictions derived from the maximization assumption cohere with the data is no explanation. Behavior needs to be explained in motivational categories other than maximization.

According to the modern notion of utility an alternative a is not preferred over an alternative b because a has higher utility than b. Quite to the contrary we assign a higher utility to a than to b just to *represent* the higher ranking of a. This higher ranking depends on values and motives other than utility. The utility function is not among the reasons for evaluation or action it is the outcome of such reasoning and evaluation.

Utility is only “representative utility”. Therefore maximization must be representative as well, namely of a process that is driven by motives other than

¹⁰ Just imagine as a most typical example a judge who can create high costs for others without incurring such.

maximizing the utility function. In view of both it is an astonishing fact if behavior does indeed coincide with the maximization of a utility function. Only if monetary rewards are dominating evaluations and reasoning the mystery vanishes to some extent. In that case attitudes to risk may still pose interesting questions in the ordering of alternatives but the focus on monetary rewards that dominate in the framing of the decision situation renders it plausible that the individual indeed is behaving in a maximizing manner.

However, if besides the selfish income-motive other, rather complicated motives like for instance numerical measures of inequality come into play it is less plausible that these measures are also “minimized or maximized”. Who in the world would be consciously maximizing or minimizing such a possibly complicated measure? More importantly, could it ever be plausible that anybody would be consciously maximizing a weighted sum of say monetary income and some such measure of distribution? To say here that individuals do not maximize such measures per se but only the utility derived from them is a “*petitio principii*” since utility is only representing preferences and thus the outcome of a ranking process that – by assumption – in itself cannot be based on the maximization of a weighted sum.

Going back to the behavioral axioms is no solution either but merely postpones the problem. Accepting that preferences can be represented by a utility index whenever certain axioms are fulfilled we must again wonder how it may be that those axioms should ever be fulfilled by choice behavior that in itself is not motivated by maximizing a numerical function. Economists cannot have it both ways on the one hand insist that utility is merely representative of preferences and avoid the question of how the mysterious fact of behavior that looks as if it were driven by utility maximization can be explained.

4.4. Hyper-rational behavior of the simple minded

Peyton Young in his basic text on institutional evolution very nicely presents some essentials of the preceding to motivate why he is pursuing what may be called the “evolutionary alternative”:

“In neoclassical economic theory – especially game theory – agents are assumed to be hyper-rational. They know the utility functions of other agents (or the probability that other agents have these utility functions), they are fully aware of the process they are embedded in, they make optimum long-run plans based on the assumption that everyone else makes optimum long-run plans, and so forth. This is a rather extravagant and implausible model of human behavior, especially in the complex, dynamic environments that economic agents typically face. Moreover it represents a peculiar aberration from traditional ways of thinking in economics. One of the central messages of the pure theory of exchange, for example, is the ability of prices and markets to coordinate economic activity *without* assuming that agents are anything more than naïve optimizers acting on limited information.” ((Young, H. Peyton 1998), p. 5).

Though we agree with Young in what he says we also insist that “the ability of prices and markets to coordinate economic activity *without* assuming that agents are anything more than naïve optimizers acting on limited information” cannot be used to defend rational choice modeling. It is not true that the rational choice approach has explanatory power since it “predicts” equilibrium results of markets. The standard arguments of economists who invoke “trial and error” as experienced in several games and the like are often interesting (see as an example (Roth, A.E. and J. Erev 1995)). Likewise the classical argument as in particular presented by Armen Alchian (see (Alchian, Armen A. 1950)) that even individuals without any ability to make forward looking choices (“zero intelligence traders” in the sense of (Gode, Dhamanjay K. and Shyam Sunder 1993)) would reach equilibrium results on markets with appropriate competitive conditions does not support the behavioral model underlying neo-classical economics. In the present context of defending strict rational choice modeling such arguments are entirely irrelevant. They basically reject the rational choice

approach by introducing substitutes that explain why results may look “as if” brought about by rational choice. Rather than presenting the explanation of observed equilibrium behavior in terms of rational choice they “reduce” it to an explanation in terms other than rational choice namely adaptive ones.

The question of how and through which motives preferences that can be represented as if amounting to the maximization of an objective function could indeed emerge has no easy answer unless a simple-minded dominance of the monetary income motive can be assumed. Since the experiments of economists as well as everyday experience show that such simple-mindedness is exceptional the neo-classical framework runs into serious difficulties. Let us look more closely at some before we turn to some tentative ways to deal with them more adequately than in the “neo-classical repair shop”.

5. Two Paradigms of Experimental Economics

Even though much of experimental economics is still pursued within the traditional neo-classical framework its most basic results are not well in line with the behavioral model of homo oeconomicus. Experimentalists with a background in economics did not, however, strive to drive home this point. They were in general very reluctant to abandon the maximization framework completely and rather tried to incorporate experimental results into the established corpus of neo-classical economic theories by incorporating them into the utility functions representing individuals.

A particularly instructive recent example from that “neo-classical repair shop” is Gary Bolton’s and Axel Ockenfels’ theory of equity, reciprocity and competition, ERC ((Bolton, Gary and Axel Ockenfels 2000)). This theory tries to account for a wide class of experimental game theoretic results within a modified neo-classical „maximization of utility“ approach. To accomplish this the basic utility functions of individuals are calibrated on the results of dictator and ultimatum games and then, with these functions in hand, experimental outcomes of other game experiments are „explained“.

ERC represents neo-classical ways of dealing with experimental results in exemplary manner. However, since we do not accept the background thesis that neo-classical economics was basically an explanatory success we are not committed to piecemeal repairs of the neo-classical approach to explaining human behavior at all. Quite to the contrary we shall subsequently argue that reward allocation, dictator and ultimatum games suggest that more radical moves towards theories of bounded rationality are necessary if real progress is to be made eventually.¹¹

¹¹ Rubinstein, Ariel. 1998. *Modeling Bounded Rationality*. Cambridge. MA: Cambridge University Press. claims to model bounded rationality but he is clearly still very close to the more traditional economic approach.

5.1. Reward allocation games

A simple class of experiments originally introduced by social psychologists as “reward allocation games” (see for instance, (Shapiro, E.G. 1975) or (Mikula, Gerold 1973)) illustrates that straightforward maximization of monetary rewards can hardly be typical human behavior.¹² As performed by social psychologists the experiments typically adopt the following 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 contributed a fraction $c \in (0, 1)$ 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 $x, y \geq 0, x + y = p$,
- the two individuals do 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 five of the Nicomachean Ethics – and allocate proportional to the contribution c (see for a standard philosophical account (Frankena, William K. 1966)); i.e. X chooses $(x = cp, y = (1 - c)p)$. If the work effort is rather trivial then, since p more or less

¹² A variant of such games became more widely known as “dictator games” among economists.

drops down as manna from heaven, we may also find the allocation $(\frac{P}{2}, \frac{P}{2})$ as the most frequently chosen alternative.¹³

If the monetary payoffs present in the situation are salient and dominant such that experimental subjects, c.p., prefer higher to lower payoffs homo oeconomicus behavior should be expected at least approximately. The situation is sufficiently simple to expect that. Nevertheless this kind of behavior is in fact rarely observed. Therefore the rational choice approach seems clearly falsified.

Still, all theories with empirical content – among them the most useful ones – are running into trouble somewhere. They all are falsified in some sense of that term. Therefore falsification per se does not mean too much in practical terms. In particular it does not imply that the falsified theory is to be given up immediately. In the case of hitherto successful theories the wise practical strategy suggests to modify those theories such as to account for the falsifying evidence rather than to give up the successful theory completely.¹⁴ Along these lines the neo-classical economist will typically argue that neo-classical economics was in fact successful. Therefore, he will say, economic theory should not be given up lightly in view of such evidence as resulting from reward allocation experiments.

¹³ Letting manna drop 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 and Königstein in ##### ... Königstein, Manfred. 2000. *Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production*. Berlin Heidelberg New York: Springer.).

¹⁴ Meaningful falsificationism as implying the norm that our theories should be *formulated* such that they are most easily falsifiable – that counter evidence is as readily at hand as possible if there is such – is completely different from the non-sensical advice to give up theories instantaneously once counter evidence is found.

Even though we have our second thoughts about the alleged explanatory success of standard decision theoretic economic theory we readily admit that trying to modify the rational choice approach is a *prima facie* plausible research strategy – if only for the reason that the rational choice approach is as a matter of fact prevalent. More specifically, at least those who for the preceding practical reasons are leaning towards the rational choice approach anyway may seek for ways to rescue it. To do so 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 into the broader picture of human motivation).

To start with the latter, the danger of “ad hocery” is looming large here. However, it should also be seen that the inclusion of additional motives need not be that. It is quite respectable to start with an austere model of human motivation like the classical homo oeconomicus model and gradually add motivational factors such as to develop a fuller account of human motivation. If the modifications intend 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. The criterion for avoiding ad hoc repairs is that suggested modifications are aiming beyond the problem that they are supposed to fix at a class of situations.¹⁵

For the sake of illustration let us look here merely at the minimum such modification by one other factor with an application to one other type of social interaction slightly more complicated than the reward allocation game. In doing so we consider as additional factor that of so-called “inequality aversion” (see (Fehr, Ernst and Klaus Schmidt 1999)) and as additional interaction situation the “ultimatum game” (see on this conceptually (Güth, Werner 1976) and experimentally (Güth, Werner, Rolf Schmittberger, and Bernd Schwarze 1982)).

¹⁵ In case of the reward allocation game it would be purely ad hoc if the modification would amount to no more than that the theory holds except in case of reward allocation games.

– Let us first sketch the game and then discuss how inequality aversion may “explain” observed results.

5.2. 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 , $x, y \geq 0$ 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.¹⁶

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 $0 \leq y \leq p/3$

¹⁶ 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 Hoffmann ##### [Bolton, 1995 #1102]).

- Proposers X in general offer $y > p/3$, mostly an equal split $y = p/2$ of the pie which nearly all responders accept.

These observations¹⁷ 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 responders or, if not so, the proposers must be intrinsically motivated not to allocate rewards (too) unevenly.¹⁸

One way of dealing with the observations in a rational choice frame work relies on the aforementioned motive of "inequality aversion". Forming utility functions $u_x(x, |x-y|)$, $u_y(y, |x-y|)$ increasing in the first and decreasing in the second argument¹⁹ 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

¹⁷ The recent results of a newspaper experiment with more than 1000 participants can be studied in Güth, Werner, Carsten Schmidt, and Matthias 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: Jena.*

Güth, Werner, Carsten Schmidt, and Matthias Sutter. 2003. "Fairness in the Mail and Opportunism in the Internet - A Newspaper Experiment on Ultimatum Bargaining." *German Economic Review*, forthcoming. while older evidence is discussed in Roth, Alvin E. 1995. "Bargaining Experiments," in *The Handbook of Experimental Economics*. John H. Kagel and Alvin E Roth eds. Princeton: Princeton University Press, pp. 253-348.

¹⁸ In a fuller account the comparison between dictator and ultimatum games on the one hand and reward allocation games and ultimatum games with a preceding joint effort as in the reward allocation case would be appropriate.

¹⁹ Such effects need not be continuous, though. Neo-classical repairs should also allow for lumpiness, see for instance Khalil #####.

motivations are included in a systematic way specifying their relative weight. Finally the qualitative results derived are quite well in line with observational data.

In the preceding regards the argument from inequality aversion seems to be impervious to the most 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 older explanations of, the sleepiness of opium smokers by the “dormitory power” of opium (where dormitory power does not mean more than that it makes you sleepy to smoke the stuff).

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 in the role of a second mover who re-acts 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 all 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.²⁰ We have some evidence in day-to-day experience that such emotions do play a role in motivating human behavior. But they do not explain the observations in full. Retributive emotions directly apply only to the motivation of the responder in an ultimatum game. The anticipation of that motive by the first moving player would explain behavior in the ultimatum game. In the reward allocation game retributive emotions towards somebody who has contributed to a joint product could conceivably be operative as well and explain some tendency towards rewarding the other individual in the allocation. However, retributive emotions would not explain the observations in dictator games as typically used in experimental economics.²¹

May be that there is a deeper theory of motivation that implies both forms of behavior conditional on whether the action is or is not a response to some action of somebody else. It is one of the great merits of experimental game theory that it raises such questions in a systematic manner inducing a dynamic research agenda. As far as this is concerned it should be noted that inequality aversion is clearly not conditional in the relevant sense. Since human beings act differently contingent on preceding histories that lead to results rather than focusing on results per se (they are *not* consequentialists *throughout*, see also (Güth, Werner, Hartmut Kliemt, and Axel Ockenfels 2002)) we should have some second thoughts and in fact some reservations with respect to theories that focus on consequences (results) and explain them as directly intended.

²⁰ For instance philosophers like John Mackie believe that retribute emotions are fundamental to ethics in general, see Mackie, John L. 1985. *Morality and the Retributive Emotions..*

²¹ The desire not to disappoint others could be useful in that regard (Gnea., Pearce and Stacchetti, ..., Rabin, Matthew. 1993. "Incorporating Fairness Into Game Theory and Economics." *American Economic Review*, 83, pp. 1281-302., ..., Falk and Fischbacher ...Dufwenberg and Kirchsteiger#####) since it would explain first mover behavior in the dictator game as well.

Again, we do not deny that sometimes humans aim at consequences. Quite to the contrary we insist that this be taken into account. Human beings command the faculty to make choices strategically; i.e. in view of the expected causal consequences of those choices. They can seize opportunities if they recognize them as evaluated according to multiple criteria of evaluation. If there is independent evidence that individuals are motivated that way we should frame our explanations and predictions accordingly. But independent evidence that behavior is motivated in this way is necessary.

All the preceding concerns behavior that is “rational” to some extent but in effect merely “boundedly rational”. What can be observed in the field also shows up in the laboratory. Even though much of experimental economics is still seeing itself within the neo-classical framework the experiments of experimental economics can as well be interpreted as case studies in boundedly rational behavior. The results of economic experiments allow for 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).²² So let us turn to an account of behavior in dictator and ultimatum games in categories of bounded rationality.

6. 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 do hardly play any role in approaches based on perfect rationality. Individuals show up only as represented by the stenographic device of utility

²² At the risk of beating this to death let us re-iterate the demonstration that behavior is “as if” perfectly rational in the sense of utility maximization is not good enough for rescuing homo oeconomicus.

functions. Vice versa, rejecting the axioms that guarantee the existence of the utility function we eliminate the individual as perceived by neo-classical economic theory from the picture entirely. No wonder that a kind of “horror vacui” (the fear of the void or the vacuum) takes hold of many economists if utility maximization is taken away.

Obviously we, too, need to put something in place of the rejected “explanations”. Models of bounded rationality are clear candidates. However, not too much has been accomplished yet in this field. 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, Herbert A. 1957), (Simon, Herbert A. 1985)). 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 be the best in terms of profit” 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 not costless and not perfect).

To refer 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.²³ We think that again – as can be

²³ 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 perhaps

learnt from the difficulties of rational choice approaches – perceptions and the framing of situations (see, of course, (Kahneman, Daniel and Amos Tversky 1984)) do play a crucial role in triggering certain responses and cognitive-cum-motivational processes.

6.1. The framing of reward allocation and dictator games

Rules (of thumb) that guide individuals in their boundedly rational choice making specify conditions under which the rules apply. Whether behavior x or x' is adequate depends on how the action situation is perceived. If we put people in a position in which “lying” is the “right” thing to do in a game show because, say, the task is to tell an untrue story that is seemingly true then most of them will lie even though they resent lying otherwise. If we put somebody in a contest of winning as great a share of a pie p then she will act accordingly. If she perceives a situation in which selfish behavior is normatively characterized as adequate then she might show that behavior because it is the norm. She acts selfish because she feels entitled to or because she thinks that this is the expected behavior not because she is naturally completely selfish. Likewise if a certain situation is characterized as an occasion to show fairness then the actor may perceive it in a completely different frame and again act accordingly because she thinks that this appropriate not because she is naturally completely fair.

More formally:

Let G be a class of situations which are to be classified as action situations. For the sake of simplicity and in order to keep things parallel with the discussion of preferences based on self- and other-regarding motives – like monetary gain and inequality aversion in the distribution of such gains – let us assume that

also about the process in which they are formed or modified through time (starting from some initial set up) in some exemplary contexts.

situations may be perceived merely in two ways by actors from a set of actors $I:=\{1, 2, \dots, n\}$:

$\emptyset \neq G_i^n \subset G$; non-ethical (opportunistic) behavior appropriate;

$\emptyset \neq G_i^e \subset G$; ethical (non-opportunistic) behavior appropriate;

$\emptyset \neq G_i^u := G - \{G_i^n \cup G_i^e\}$; no standard of behavior directly applies.

By implication $G_i^n \cup G_i^e \neq G$. Let us assume that $G_i^n \cap G_i^e = \emptyset$.²⁴ For any situation $g \in G$ we have $g \in G_i^n$, $g \in G_i^e$, or $g \in G_i^u$. In the first case rules or norms of appropriate pursuit of self-interest will apply as dominant considerations, in the second case ethical rules, like rules of fairness, retribution, beneficence etc. will dominate or at least guide the deliberation process, while in the third case additional deliberations that in turn take into account different considerations must be performed.²⁵ To illustrate this process briefly let us look more specifically at the three active decision making roles in the dictator and the ultimatum game:

(D) dictator X in a reward allocation game with contribution $c \in (0, 1)$;

(P) proposer in an ultimatum game;

(R) responder in an ultimatum game.

6.1.1. Some boundedly rational deliberations in role (D)

²⁴ The possibility that two standards, either coherent or conflicting, of which one is ethical and one non-ethical apply is excluded here at least initially.

²⁵ In the hierarchical way humans seem to reach their decisions the classification of situations according to some stereotype seems to be lexicographically pre-ordered.

If an individual finds herself in role (D) in a dictator or reward allocation game g she must classify the situation as to whether it is of type G_i^n , G_i^e or G_i^u . If the situation is framed as a reward allocation game r with “earned entitlements” that amount to c it seems hardly ever the case that individuals classify their task as $r \in G_i^n$. From $r \notin G_i^n$ it does not necessarily follow that $r \in G_i^e$ applies. But in the reward allocation experiments of psychologists almost all subjects $i \in I$ were choosing allocations $(p/2, p/2)$ or $(cp, (1-c)p)$ and thus according to an “ethical” rule of “equal” or “proportional” shares (see again (Shapiro, E.G. 1975) and (Mikula, Gerold 1973)).

The entitlement to the reward seems to trigger a classification according to $r \in G_i^e$.²⁶ In the dictator games d without entitlement in which experimental economists let the pie drop like manna from heaven individuals were less sure about the framing of the situation (see (Hoffman, Elizabeth, and Matthew L. Spitzer 1985)). Since manna quite rarely falls on human individuals they do not have much prior experience with such situations and are quite naturally uneasy about how to proceed. We may – with some support from the data – speculate that in case of economic dictator games individuals were less sure how to frame the task than in psychological reward allocation games. There were individuals who obviously were going for $d \in G_i^e$ while others indeed chose allocations according to $d \in G_i^n$. Yet there were also individuals who would classify according to $d \in G_i^u$. They were uncertain which rules and standards would apply to the problem at hand.

²⁶ An analogous reason may perhaps form part of the explanation for the prevalence of producer as opposed to consumer interests. Within the standard neo-classical framework we could merely point out factors like information and organization costs here leaving open many questions.

It seems that many individuals when undecided are inclined to go along with whatever they perceive as dominant in the group. Also some who are undecided themselves might nevertheless think that most others would not be uncertain and perceive the situation either as $d \in G_i^e$ or $d \in G_i^n$. However, it is much more plausible that in some kind of “false consensus” uncertain individuals would take their own uncertainty as indicating that others feel likewise (see on false consensus effects Engelmann and Strobel #####). Unable to proceed either according to $d \in G_i^e$ or $d \in G_i^n$ such individuals might try to go for some kind of “convex-combination” of rules. This all being said, it should be noted that, more often than not $d \in G_i^e$ seemed to emerge and the ethical component or frame of decision making did play a role in the deliberations of real decision-makers (see [Bolton, 1995 #1102], Güth and Huck #####). Even though less dominant than before allocations $(p/2, p/2)$ formed modal behavior.

The preceding is supported by some experimental evidence. But we readily concede that there is no powerful theory of boundedly rational behavior here.²⁷ Somewhat more general elements show up in the stress that is laid on classification and framing (aspects of human behavior well known and well studied in cognitive psychology) as preceding choice making. It should be noted well, too, that human choice making is modeled as “rule-bound” rather than proceeding case by case in an opportunity taking manner. Moreover, being rule-bound does not imply that all individuals are bound by the same rules. There may be basically heterogeneity at two more extreme ends of the behavioral spectrum and even those who compromise may have gotten there due to their failure to find an appropriate rule and not because of their desire to balance ethical and non-ethical values in an opportunistic manner.

²⁷ To argue in common sense categories may be deemed appropriate in view of the fact that we are, after all, dealing with cognitive processes to which individuals bring nothing but their common sense.

6.1.2. Some boundedly rational deliberations in the responder role (R)

In ultimatum games u the participants must classify the situation also as to whether $u \in G_i^u$, $u \in G_i^e$ or $u \in G_i^n$ applies. Now active decision makers X and Y are playing in two roles, the proposer and responder roles respectively.

Starting with the responder it should be noted first that – at least in those cases in which the experiment does not employ the strategy method – the responder will know not only u but also that X offered y in u to her. Let us refer to the information that an offer of $y=p/2$ was made in an ultimatum game u by $u(p/2)$. Then $u(p/2)$ should be sufficient to trigger a positive response from the responder. For $u(p/2) \in G_i^n$ this is obviously what opportunism (or, perhaps, rather non-ethical rules of behavior) would suggest. For $u(p/2) \in G_i^e$ the ethical reasoning about fair divisions will lead to the acceptance of the offer $y=p/2$. In cases with $0 < y \neq p/2$ the moral response of rejection might be forthcoming. Interestingly enough in the aforementioned newspaper experiment (see (Güth, Werner, Carsten Schmidt, and Matthias Sutter 2002; Güth, Werner, Carsten Schmidt, and Matthias Sutter 2003)) one could observe that not only in cases $y < p/2$ rejection occurred but also in cases $y > p/2$ if less frequently. Along with a quite distinct tendency to accept very low offers to an extent that goes beyond most other ultimatum game experiments this is remarkable (about a third of the participants accepted in their strategies offers as low as a 10% share of the pie).

Speculating about the reasons for the remarkable results of the newspaper experiment we should like to point out the following: In the experiment the strategy method was employed; i.e. individuals had to specify for each offer in a discrete space of possible offers whether they would accept or reject the offer. This induced obviously a tendency of participants to distance themselves from direct emotions (assisted by complete anonymity of participants, wide social and local distance, wide gap in time, purely hypothetical proposals).

Looking for rules in a sober-minded unemotional way may have furthered both the tendency to go for the “right solution” and therefore in the ethical framework to insist on $y=p/2$ and a tendency to reason opportunistically rational and thus to accept all offers. Since individuals can do both, act opportunistically and rule bound, such heterogeneity as observed is to be expected. It should also be noticed that the framing and perception of situations is such a delicate matter influenced by minor details of the perceived situation and the perceiving subject that heterogeneity is to be even more expected (see Güth, Huck and Müller ##### for a study of the influence of minor deviations from fairness).

6.1.3. Some boundedly rational deliberations in role (P)

An individual X in the proposer role (P) in an ultimatum game is confronted not only with the task of classifying himself the game u as $u \in G_X^u$, $u \in G_X^e$ or $u \in G_X^n$ but also the task of anticipating how the responder (R) will classify the situation. Will Y perceive the situation as $u \in G_Y^u$, $u \in G_Y^e$ or $u \in G_Y^n$?

The second part of the proposer’s task could be fulfilled in different ways. The proposer could conceivably try to represent the behavior of the second moving responder “as if” it were perfectly rational. As a first mover he would employ the short hand of the utility function to describe her behavior in the second mover role. Assuming that he is aware of the theories of inequality aversion he may speculate that his partner is subject to that aversion and may ascribe to her a utility function giving weight to both selfish opportunism and unselfish fairness. However, it does not seem very likely that a human decision maker who himself is clearly boundedly rational and perceives himself as being so would in fact rely on the theoretical construct of a utility representation to capture his anticipations of the behavior of another individual. Being aware of the boundedly rational character of his own decision-making he will speculate that the second mover’s decision-making is of the same kind.

If X tries to anticipate the reasoning of Y he must decide on whether $u \in G_Y^u$, $u \in G_Y^e$ or $u \in G_Y^n$ applies. If he knows a bit about experimental results he will also know that at least outside of newspaper experiments it is quite unlikely that $u \in G_Y^n$ prevails. His responder will not accept whatever he will offer her. She will in all likelihood classify u after a very low $y \ll p/2$ as $u(y \ll p/2) \in G_Y^e$. In all likelihood the ethical rule will suggest rejection very strongly and distinctly. In such a case a proposer who himself classifies u according to $u \in G_X^n$ has good reason to propose what seems appropriate according to a situational perception $u \in G_X^e$ speculating introspectively that the latter emulates G_Y^e .

But, of course, it is in general exceedingly unlikely that subjects in experiments would know the results of previous studies of the ultimatum game when playing the game themselves. In all likelihood proposers will form their own view of the boundedly rational theorizing of the responder according to some rather primitive theory of how responders reason and respond. In fact individuals in the proposer role may skip reasoning about the reasoning of responders altogether by simply classifying the situation as $u \in G_X^e$ and use some rule of thumb specifying what is the appropriate offer in a situation as u . The most likely case here is that individuals will endorse normative theories that would suggest to offer $y = p/2$ in such cases. At least modal behavior would coincide with this view of the matter.

But what about those who come to the conclusion that an offer of $y < p/2$ is appropriate? Let us initially assume for the sake of specificity that X offers $y = p/3$. It seems likely that such a proposer X is looking at the situation as $u \in G_X^n$ or, possibly, $u \in G_X^u$ and would speculate that his co-player is subject to both motives the ethical one to reject unfair offers supported by retributive emotions – otherwise the proposer should offer $y = 0$ – and the non-ethical temptation to take what is on offer. A clever X will consider other possibilities than $y = p/3$. He will, perhaps, apply some non-Bayesian reasoning to determine a

range of offers $[y^*, p/2]$ which will in all likelihood be accepted and a range of offers $[0, \underline{y}]$ that he expects to be rejected. Clearly $p/2 \geq y^* \geq \underline{y} \geq 0$ must apply but $y^* > \underline{y}$ may also be true. The way boundedly rational individuals typically seek for sufficient (satisficing) alternatives makes it most likely that X will neglect offers both from $[0, \underline{y}]$ and from $[\underline{y}, y^*]$ and focus on y^* if his perception of the situation is $u \in G_X^n$.

6.1.4. Being of two minds

In our discussion of the deliberation process in roles (D), (R), (P) we made the assumption $G_i^n \cap G_i^e = \emptyset$. From an intuitive point of view this assumption may seem quite strong. It seems that many of our decision problems arise from intra-personal conflict. We may be torn between factors as for instance the temptation to break a rule or norm and the requirements imposed by that standard of behavior. This is a very common experience and we all must find ways to deal with such conflicts somehow. A rational way of doing so clearly emerges if we start to weigh opportunity costs at the margin. To adjust our behavior such that conflicting considerations contribute equally at the margin to our overall well-being is appealing from a rational choice point of view. But are such dictates of rationality in any way realistic models of mental processes?

Even what may look “as if” it were the outcome of multi-factorial marginal adjustment and global utility maximization is at root in all likelihood a different animal. It is rule bound or rule guided choice-making rather than consequentialist utility maximization. To some extent we follow one rule and then to some extent another one. We switch categorically rather than adjust at the margin. Putting such behavior through the neo-classical repair shop to have utility maximization written all over it does not make too much sense. Maybe that some modified utility function “explains” the data but in the end it is behavior that we want to explain. To accomplish this we better go for the mental processes that bring about the results observed as data.

Since these processes seem a far cry from marginal adjustment we should explore alternatives more systematically. Let us start with the stepwise adjustment of aspiration levels which fits into a “satisficing” approach.

6.2. The formation of aspiration levels

With respect to the manner in which aspirations become operative considerable evidence seems to suggest that human subjects often proceed in a lexicographic or quasi-lexicographic way. If they have aspiration levels along several dimensions they do not adjust at the margin but satisfy first one of the levels, then the next etc. In view of multiple aspirations individuals are moving on a kind of (multi-dimensional) “grid” (Selten #####). If their aspirations are not met even after an extended effort modification of the levels will be pending. This modification will often follow also some lexicographic pattern.²⁸

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 re-acting individual or responder will not aspire to get more than $p/2$. From the results of many experiments we know that offers of $p/2$ will practically always be accepted and practically never be transcended by the actor (the strategy method in the newspaper experiment provided some data for that eventuality, too, though). 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 interest is the aspiration level y^* such that they will not accept any y with $y < y^* < p/2$.

²⁸ Lexicographic modes of thinking are discussed in the case studies in Ahlert, Marlies and Hartmut Kliemt eds. 2001. *Making Choices in Organ Allocation*. Stuttgart: Lucius und Lucius.

It is quite impossible to say something about the absolute height of individual aspiration levels in general. There is also heterogeneity in any population of individuals in that regard. 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 come forward;

$y^* \leq y < p/2$ acceptance will come forward if grudgingly;

$y < y^* < p/2$ rejection response 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 be reasoning along the following lines (see Güth #####) which we present in dialogic “query and answer” mode to give just one example of a conceivable cognitive process 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 while $x = p/2$ will also satisfy my own feelings of self-esteem.²⁹

²⁹ see Brennan and Pettit 2001##### for an inroad in the territory of an economics of (self-)esteem ...

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 iff $x/y > 2$ ".

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 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 appear not 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 on 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.³⁰ 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 is rather underdeveloped. We readily admit that. At the same time we think that the aim of experimentation in economics should be developing such a

³⁰ The relation to the distinction between low and high cost situations as discussed above is obvious.

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 try to 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 in the beginning. Complications immediately emerge if we try to generalize. For instance, ultimatum games are sometimes also discussed under the heading of “ultimatum bargaining games”. Posing an ultimatum is just the limiting case of bargaining. If we would allow 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} of 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 level adaptation will involve steps of lowering the initially incompatible demands such that individuals step down to the next lower level. 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, Reinhard and Hans-Juergen Weber 1972)).

6.3. 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 here the turf of cognitive psychology. Acknowledging this we still think that economists just cannot keep out of the “psychological territory”. Being well aware of our own deficiencies in these regards as well as of the immature state of the field of cognitive sciences (interesting references to the field are (Clark, Andy ;Hutchins, Edwin)) we shall confine ourselves to a

speculative discussion of the basic examples of reward allocation and ultimatum games.

6.3.1. 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 to 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. The introspection-biases created in reporting itself. The validity and reliability problems of content analyses are not only well known from other branches of social and psychological theory as for instance media research but cannot be neglected in thinking aloud procedures as well. Still, there are at least some remedies for the deficiencies. For instance 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 is not without merit. If for instance the dictator in a reward allocation game would say things like “one has to be fair”, “why not take what I can”, “Y should not go without anything but I deserve more”.

Such 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. For instance, in the case under consideration here, the

speculative arguments that we proposed in 6.1.1. for the reasoning process of the individual in the dictator role of a reward allocation game may be critically assessed. Did we capture typical lines of argument? Are the kinds of argument consistently repeated in think aloud studies by a certain percentage or perhaps even a majority of participants of such studies etc.?

6.3.2. 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 would be asked by an artificial agent (typically a computer program) 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. However, this is a way of information collection that is not without its own risks. Whether she classifies the situation as $r \in G_X^e$ or not may be affected by the question. The same would apply if we would ask individuals specifically about asymmetries in situational control or power before they decide. Again the research could affect the situation or rather its perception by the participants of an experiment.

More generally, the kind of questions asked and the sequence in which they are asked during rather than after an experiment are non-neutral with respect to outcomes. But to use artificial agents – basically computer programs possibly endowed with some kind of artificial intelligence – during an experiment has its advantages, too. It is in all likelihood not neutral with respect to results but it also can create valuable information in a controlled way. After all the artificial agent other than a human agent 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 fully understood nor has it in any depth been explored up to now. In particular if it is

used in re-runs of former experiments it might create some useful insights. Since we come back to that suggestion in the next section in which personal actors are teamed with another personal actor rather than an agent let us directly move on.

6.3.3. Teams as unitary decision-makers

Economists are used to treat groups of individuals as so-called corporate actors or unitary decision-makers even though they are well aware that in any literal sense only individuals can decide and act. For instance if a team of three individuals would have to decide collectively on the choice of a program that to be watched by them on their tv-set they might do so by simple majority vote. We would routinely describe that act as a collective choice. But, of course, literally speaking there is no collective choice. No collectivity literally makes the choice of the tv program and for that matter no individual. The individuals can only choose to vote for or against a proposed program. Their options are – in the simplest case – “yes” or “no” but not the choice of the program as such. That choice emerges from the several acts of assent or rejection chosen by the individuals.

Individual decision-making seems to emerge from individual processes of weighing pros and cons too. Individual choice as emergent rather than made seems to be a good way to look at choice making. It suggests that teams of individuals should be “teamed up” in pairs to play simple games (see for instance Henning-Schmidt #####). Recording discussions among the team members by video and audio devices and analyzing that information might reveal some of the reasoning processes. Teams that have to make the decisions jointly must reach some form of consensus. In particular, if the team is brought into a co-operative mood such that internal strategies like “holding out” or “bullying” would not appeal to the team members there is some hope that something can be learnt from the argument in the team for individual deliberations as well.

A design according to which individuals exchange messages over a keyboard in a more formalized manner might work here, too. In the near future speech recognition might reach a state in which spoken messages could automatically be transformed into written ones providing some additional control over the sequence of information exchange and the group dynamics. Controlling for more implicit aspects like the tone of the voices etc. may offer additional insights. Such methods do not enable the researcher to “look into individuals’ heads” but they do bring us somewhat closer to making explicit some interesting processes of deliberation that otherwise remain implicit. On the other hand it may well be that true deliberation processes are not explicit and to force individuals to make them explicit may in itself strongly bias them.

It seems an interesting step to apply the research strategies – that are known among psychologists and other researchers anyway – to some of the research in experimental economics. Repeating some of the experimental games relying on teams rather than persons as players may on the one hand control for biases and on the other hand produce new insights. Since the game experiments have been conducted before by individuals there exists some prior knowledge of individual play or some benchmark already. If the distribution and patterns of results of team play are broadly the same as of individual play one can be fairly sure that team- and explicitness-biases are not too strong. Since there are prior hypotheses on how the results in the experimental games came about as well, we can create new insights by testing these hypotheses in light of the new evidence. Letting teams of players play the same games as personal players did before we may obviously 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 and if team. If personal player behavior roughly coincides it seems fairly clear that the reasoning indicates something about considerations of personal players. Moreover, since many of our decisions and for that matter many of those most important in business life are in fact 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.

As stated already there is nothing new or innovative about the preceding suggestions. Psychologists have been thinking along these lines for a long time. Often they found these kinds of research too clumsy or too unreliable to make them worthwhile. But to use them in the context of prior research of experimental economics may change this assessment to a considerable extent. It seems a quite promising research program to repeat economic experiments performed before with personal players with teams of players who must explicitly reason about their decisions in some formalized format.

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 of p might be transformed into a $2p$ or perhaps $(2+\varepsilon)p$, $1>\varepsilon>0$. An even within team split of payoffs 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.

Recording what individuals say etc. and analyzing it we may hope to gain some insights on what is in fact driving their decisions. It seems quite likely for instance that one can learn something about whether or not such a motive as “inequality aversion” was in fact present. Intentions in the proposer role might become more transparent. In the responder role we may hope to find out more about motives as well. Was an aversion against inequality driving subjects or were retributive emotions inducing individuals to reject offers? More fancy research techniques like connecting individuals to brain-scanners may emerge in the future and may indeed, if the workings of the human brain should be better

understood, lead some additional insights. – Neglecting that let us rather take a broader, final look at the basic perspective and prospects of research on boundedly rational behavior. After going through the rather lengthy preceding argument it seems appropriate to put things into perspective.

7. Conclusion homo sapiens, boundedly rational and expecting man

As remarked before 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 are both in general imposing constraints on them. They are cognitively or normatively bounded in their ability to engage in opportunity taking behavior. Still, sometimes they can in fact overcome boundedness. Consulting experts, engaging in an exceptional effort of situational analysis may be helpful in overcoming cognitive routines. Likewise, those who intend to abandon old habits and the allegiance to some ethical norm of conduct or other may do so as well in some special effort or by simply giving in to a temptation.

We have indicated very tentatively our own views on some possible ways of future research on bounded rationality. 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 in particular look at situations that repeat themselves under several influences including frequent external shocks of minor and sometimes major proportions.³¹ Even though we would reject the typical neo-classical focus of optimization as brought about by selective adaptation in repetitive situations we do not deny that repetition,

³¹ This general experience is expressed in the extended citation from Peyton Young's theory of institutional evolution above.

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.

Experimentally one can respond to this insight by turning to “robust learning” experiments (see Güth 2002#####). In these experiments participants play a variety of structurally related games. Let us again for the sake of specificity look at ultimatum games u and dictator games d (or, for that matter, reward allocation games r). A robust learning experiment would engage participants in the

(1) repeated play of games of the form d with changing participants
after announcing the change

(2) repeated play of games of the form u with changing participants

possibly repeating the sequence of two sequences (1)-(2) over and over again.

Adaptive learning is observable in the repetitions of the same game. If after the announcement of change and switching to the other game drastic changes of behavior occur this will indicate the presence of forward looking rational strategic choices. For instance if after a series of u -games and the announcement of the switch to the series of d -games behavior changes very distinctively then this may suggest that strategic opportunism was influential in determining the behavior in the games of the type u . But even this evidence for opportunistic case-by-case maximization is not too strong. After all, the observed switch could still be a switch between two rules, one deemed appropriate for one class of games and one for the other class. The choice maker then is not maximizing but rather switching between rules typically to meet certain aspirations connected to former experience of applying those rules.

The problem that comes to the fore here again is simply that the case-by-case maximization assumption of the standard perfect rationality model of 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 would cease 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 is nurturing all sorts of illusions is no better in those regards. Worse, it is quite certainly a dead-end of research 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 of individuals. 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, the bounded rationality approach bears more promise for future development than the perfect rationality approach only if mental processes are better understood and modeled. We are very much aware that such remarks as our preceding programmatic proposals are mere gestures. A new synthesis of several strands of research on human decision making is needed if real progress is to be made. Branches of psychology as cognitive psychology are part of or can at least be utilized in the move towards theories of boundedly rational behavior (reaching from (Festinger, L. 1957) to (Gigerenzer, Gerhard 1996;1997)). Cognitive science in the broader sense of that term as inspired by Herbert Simon himself and then pushed on in several directions may bear promise too (see for some non-standard contributions opening new perspectives of research (Clark, Andy 1997;Hutchins, Edwin 1995)). Even though one should be careful not to be carried away by intellectual fashions of the day there is some hope out there. Still, there are no well-established theories of bounded rationality yet. There are case studies, occasional efforts to systematize results and speculations but there is no integrated corpus of such theories.

Obviously our aspiration as boundedly rational theoreticians cannot be the provision of a remedy to such deficiencies of the bounded rationality approach. We, however, dare to suggest that the future lies in designing new experiments and field studies that may shed some light on human cognitive processes and may heuristically inspire the formation of new theories. We think that this should be done along the lines laid out here. For 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 perhaps might be on the right track towards the exemplary understanding of the delicate relationship between norm orientation and opportunism in human behavior.

References

- Ahlert, Marlies and Hartmut Kliemt eds. 2001. *Making Choices in Organ Allocation*. Stuttgart: Lucius und Lucius.
- Albert, Hans. 1964. "Probleme der Theoriebildung," in *Theorie und Realität*. Hans Albert ed. Tübingen: Mohr.
- Albert, Hans. 1967. *Marktsoziologie und Entscheidungslogik*. Neuwied/Berlin.: Luchterhand.
- Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory." *Journal of Political Economy*, Vol. 58, pp. 211 ff.
- Aumann, Robert. 1990. "Nash Equilibria are not Self-Enforcing," in *Economic Decision Making: Games, Econometrics and Optimisation*. J. J. Gabszewicz, J.-F. Richard and L. A. Wolsey eds. Amsterdam: Elsevier Science Publishers.
- Aumann, Robert and Adam Brandenburger. 1995. "Epistemic Conditions for Nash Equilibrium." *Econometrica*, Vol. 63:5, pp. 1161-80.
- Becker, Gary. 1976. *The Economic Approach to Human Behavior*. Chicago und London: University of Chicago Press.
- Bolton, Gary and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, 90, pp. 166-93.
- Brandenburger, Adam and Eddie Dekel. 1987. "Rationalizability and Correlated Equilibria." *Econometrica*, 55:6, pp. 1391-402.
- Brennan, H. Geoffrey and James M. Buchanan. 1984. "Voter Choice: Evaluating Political Alternatives." *American Behavioral Scientist*, 28:No. 2, November/December, pp. 185-201.
- Brennan, H. Geoffrey and Loren Lomasky. 1984. "Inefficient Unanimity." *Journal of Applied Philosophy*, 1,1, pp. 151-63.

Buchanan, James M. 1999. *The Logical Foundations of Constitutional Liberty*. Indianapolis: Liberty Fund.

Buchanan, James M. and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Clark, Andy. 1997. *Being There. Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

Cournot, Auguste. 1838. *Recherches sur les principes mathématiques de la théorie des richesses*. Paris: L. Hachette.

Dacey, Raymond. 1976. "Theory Absorption and the Testability of Economic Theory." *Zeitschrift für Nationalökonomie*, 36:3-4, pp. 247-67.

Dacey, Raymond. 1981. "Some Implications of 'Theory Absorption' for Economic Theory and the Economics of Information," in *Philosophy in Economics*. Joseph C. Pitt ed. Dordrecht: D. Reidel, pp. 111-36.

Dreyfuss. ????. *The limits of artificial intelligence????*

Fehr, Ernst and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, pp. 817-68.

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

Frank, R. 1988. *The Passions within Reason: Prisoner's Dilemmas and the Strategic Role of the Emotions*. New York: W. W. Norton.

Frankena, William K. 1966. *Some Beliefs about Justice*. Lawrence. Kansas: The University of Kansas.

Friedman, James and Shyam Sunder. 1994. *Experimental Methods. A Primer for Economists*. Cambridge: Cambridge University Press.

Frohn, Joachim, Werner Güth, Hartmut Kliemt, and Reinhard Selten eds. 2001. *Making Choices I & II*. München: Accedo.

Gigerenzer, Gerd, Jean Czerlinski, and Laura Martignon. 1998. "How Good Are Fast and Frugal Heuristics?," in *Decision Research from Bayesian Approaches to Normative Systems: Reflections on the Contribution of Ward Edwards*. James Shanteau, Barbara Mellers and David Schum eds. Dordrecht: Kluwer, pp. 1-23.

Gigerenzer, Gerhard. 1996. "Reasoning the Fast and Frugal Way." *Psychological Review*, 103, pp. 650-69.

Gigerenzer, Gerhard. 1997. "Bounded Rationality: Models of Bounded Rationality." *Swiss Journal of Economics and Statistics*, 133, pp. 201-18.

Gode, Dhamanjay K. and Shyam Sunder. 1993. "Allocative Efficiency of Markets With Zero Intelligence Traders: Markets as a Partial Substitute for Individual Rationality." *Journal of Political Economy*, 101, pp. 119-37.

Güth, Werner. 1976. "Towards a More General Analysis of von Stackelberg-Situations." *Zeitschrift für die gesamte Staatswissenschaft*,:, pp. 592-608.

Güth, Werner and Hartmut Kliemt. 1998. "Towards a Fully Indirect Evolutionary Approach." *Rationality and Society*, 10:3, pp. 377-99.

Güth, Werner and Hartmut Kliemt. 2000. "From full to bounded rationality. The limits of unlimited rationality.". Center for Interdisciplinary Research (ZiF): Bielefeld.

Güth, Werner, Hartmut Kliemt, and Axel Ockenfels. 2002. "Retributive Responses." *Journa of Conflict Resolution*, forthcoming.

Güth, Werner, Carsten Schmidt, and Matthias 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*: Jena.

Güth, Werner, Carsten Schmidt, and Matthias Sutter. 2003. "Fairness in the Mail and Opportunism in the Internet - A Newspaper Experiment on Ultimatum Bargaining." *German Economic Review*, forthcoming.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3, pp. 367-88.

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

Harrison, Glenn W. 1989. "Theory and Misbehavior of First-Price Auctions." *American Economic Review*, 79, pp. 749-62.

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

Heiner, R. 1983. "The Origin of Predictable Behavior." *American Economic Review*, 73/4, pp. 560 ff.

Heyd, David. 1982. *Supererogation. Its Status in Ethical Theory*. Cambridge et al.: Cambridge University Press.

Hoffman, Elizabeth, and Matthew L. Spitzer. 1985. "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice." *Journal of Legal Studies*, 14:2 (June), pp. 259-97.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge: MIT Press.

Kagel, John H. and Alvin E. Roth eds. 1995. *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kahneman, Daniel and Amos Tversky. 1984. "Choices, Values and Frames." *American Psychologist*, 39:April, pp. 341-50.

Kliemt, Hartmut. 1986. "The Veil of Insignificance." *European Journal of Political Economy*, 2/3, pp. 333-44.

Königstein, Manfred. 2000. *Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production*. Berlin Heidelberg New York: Springer.

Luce, D. and H. Raiffa. 1957. *Games and Decisions. Introduction and Critical Survey*. New York et al.: John Wiley.

Mackie, John L. 1985. *Morality and the Retributive Emotions*.

Maital, S. and S. L. Maital. 1984. *Economic Games People Play*. New York: Basic Books.

McKenzie, Richard B. and Gordon Tullock. 1979. *The New World of Economics*. ????????

Meckling, W. 1976. "Values and the Choice of the Model of the Individual in the Social Sciences." *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 112/4, pp. 545 ff.

Mikula, Gerold. 1973. "Gewinnaufteilungsverhalten in Dyaden bei variiertem Leistungsverhältnis." *Zeitschrift für Sozialpsychologie*, 3, pp. 126-33.

Milgrom, Paul and John Roberts. 1993. *Economics, Organization and Management*. Englewood Cliffs: Prentice Hall.

Morgenstern, Oskar. 1972. "Descriptive, Predictive and Normative Theory." *Kyklos*, 25, pp. 699-714.

Morgenstern, Oskar and Gerhard Schwödiauer. 1976. "Competition and Collusion in Bilateral Markets." *Zeitschrift für Nationalökonomie*, 36:3-4, pp. 217-45.

Nash, John F. 1951. "Noncooperative Games." *Annals of Mathematics*, 54, pp. 289-95.

Nelson, R. R. and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Rabin, Matthew. 1993. "Incorporating Fairness Into Game Theory and Economics." *American Economic Review*, 83, pp. 1281-302.

Roth, A.E. and J. Erev. 1995. "Learning in extensive form games: experimental data and simple dynamic models." *Games and Economic Behavior*, 8:1, pp. 164-212.

Roth, Alvin E. 1995. "Bargaining Experiments," in *The Handbook of Experimental Economics*. John H. Kagel and Alvin E Roth eds. Princeton: Princeton University Press, pp. 253-348.

Rubinstein, Ariel. 1998. *Modeling Bounded Rationality*. Cambridge, MA: Cambridge University Press.

Selten, Reinhard. 1986. "Institutional Utilitarianism," in *Guidance, Control, and Evaluation in the Public Sector*. Franz-Xaver Kaufmann ed. Berlin und New York: de Gruyter, pp. 251 ff.

Selten, Reinhard ed. 1991a. *Game Equilibrium Models I - Evolution and Game Dynamics*. Berlin / Heidelberg / New York / London / Paris etc.: Springer Verlag.

Selten, Reinhard ed. 1991b. *Game Equilibrium Models II - Methods, Morals, and Markets*: Springer Verlag.

Selten, Reinhard ed. 1991c. *Game Equilibrium Models III - Strategic Bargaining*. Berlin / Heidelberg / London etc.: Springer-Verlag.

Selten, Reinhard ed. 1991d. *Game Equilibrium Models IV - Social and Political Interaction*. Berlin / Heidelberg / New York etc.: Springer-Verlag.

Shapiro, E.G. 1975. "Effects of future interaction on reward allocation in dyads: equity or equality." *Journal of Personality and Social Psychology*, 32, pp. 873-80.

Simon, Herbert A. 1957. *Models of Man*. New York: John Wiley and Sons.

Simon, Herbert A. 1985. *Models of Bounded Rationality (1&2)*. Cambridge, MA: MIT-Press.

Slonim, Robert and Alvin E Roth. 1998. "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica*, 66, pp. 569-96.

Sugden, Robert. 1991. "Rational Choice: A Survey of Contributions from Economics and Philosophy." *The Economic Journal*, 101:July, pp. 751-85.

Thompson, Earl and Roger Faith. 1981. "A Pure Theory of Strategic Behavior and Social Institutions." *American Economic Review*, 71, pp. 366 ff.

Tietz, Reinhard and Hans-Juergen Weber. 1972. "On the Nature of the Bargaining Process in the KRESKO-Game," in *Beiträge zur Experimentellen Wirtschaftsforschung*. Heinz Saueremann ed. Tübingen: Mohr, pp. 305-34.

Urmson, J. O. 1958. "Saints and Heroes," in *Essays in Moral Philosophy*. I Melden ed. Seattle/London: University of Washington Press, pp. 198 ff.

Witt, Ulrich. 1987. *Individualistische Grundlagen der evolutorischen Ökonomik*. Tübingen: Mohr.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure. An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.