

## THE REASON OF RULES AND THE RULE OF REASON\*

HARTMUT KLIEMT  
Universidad de Francfort

The economic theory of property rights, public choice theory, and constitutional economics are all combining an interest in social or institutional rules with the behavioral assumption that within the rules individuals will act as rational economic men. This dual approach of the new institutional economics<sup>1</sup> has yielded fascinating new insights into the working of social institutions. In particular we now better understand why rational individuals who know that they will choose according to the exigencies of particular situations should want to live under general rules constraining their expedient choices.

As far as constitutional economics is concerned Ge-

\* A former version of this paper was presented to the first Friedrich August v. Hayek Symposium on Knowledge, Information, and Competition that was organized by professor S. Pejovich in collaboration with Dr. A. Bosch and Dr. R. Veit of the Walter Eucken Institut in Freiburg. I am grateful for the invitation to this conference and the criticisms I received there by numerous participants. I would also like to express my gratitude to M. Baurmann for his thought provoking and helpful comments.—To economize on time as the scarcest of all resources of the readers I have relied more than otherwise on footnotes to the paper.

<sup>1</sup> Cf. for instance *Zeitschrift für die gesamte Staatswissenschaft* 140 (1984), for a collection of papers under the heading of "new institutional economics"; for a particularly helpful general statement of some of the more philosophical issues involved, cf. Albert 1977.

offrey Brennan's and James Buchanan's book on "the reason of rules"<sup>2</sup> sums up a long discussion and adds stimulating new insights to it. Still, to demonstrate the reason of rules does not amount to the same thing as a demonstration of how rules are actually brought into existence and how they are sustained.<sup>3</sup>

The new institutional economics can pursue at least three different basic research strategies in dealing with the existence of rules and, thus, with the so-called problem of social order.<sup>4</sup> It can take the rules as exogenously supplied, it can use two different models for explaining the existence of rules and the behavior within the rules, and it may finally try to apply the same model of individual behavior to explain both, what is going on within the rules and how the rules themselves arise from individual choices.

The first, traditional approach simply sidesteps the crucial issue. As far as the choice among the remaining two research strategies is concerned coherence strongly recommends that we "examine the political process in the same general terms as we examine markets".<sup>5</sup> We

<sup>2</sup> Modern classics of the field of course are v. Hayek's (1960), *The Constitution of Liberty*, and (1973-79), *Law, Legislation and Liberty* or, Buchanan's (1975a), *The Limits of Liberty*.

<sup>3</sup> One might express the same distinction referring to the normative problem of why we should want that rules exist on the one hand and the descriptive question of why and how rules do exist on the other hand.

<sup>4</sup> This problem has been emphasized in particular by Talcott Parsons who took up Durkheim's older criticisms of individualistic or, as he uses to call them "utilitarian" explanations of social order; cf. Parsons 1968. At other places I have tried to refute constructively with game theoretic tools the thesis that the existence of a social order cannot be explained within the confines of an individualistic approach—cf. Kliemt 1986a and Kliemt and Schauenberg 1984. In the present paper I stick to individualism but criticize a certain form of consequentialism that is usually, though without logical necessity, associated with it.

<sup>5</sup> Brennan and Buchanan 1985, 15.

should try to use the same model of man throughout and should therefore adopt the third research strategy.

Noting that the model of rational economic man has been applied with great success to the explanation of what is going on within existing sets of rules and also to the normative purpose of institutional design rational social scientists may be tempted to go one step further and to use this model for explaining the existence of the rules too. Contrary to this I will argue subsequently that "rational economic man" cannot serve as a behavioral model uniting *both* the analysis of the rules and what is going on within the rules.<sup>6</sup> In a second step I will indicate a way of solving the problem. Although, regrettably, I am not yet in a position to offer a formal new behavioral model I will refer to some formal models that have been proposed in the relevant literature and outline how they might be adapted to my own views. Finally I will sketch how the argument relates to Herbert Hart's theory of positive law and how it applies to some central aspects of Friedrich August v. Hayek's theory of the evolution of positive law.

## I. *Rules, Commitment, and Reason*

### I.1. Rules and Commitment

Thomas Hobbes convincingly demonstrated that without a state, ordered interaction amongst the many would be impossible. He showed that a state or government is necessary to sustain the rules of ordered large

<sup>6</sup> Some of the results will closely resemble those reached by Ron Heiner, *cf.* 1983, about the origins of predictable behavior and those presented by Nelson and Winter, *cf.* 1982, in their argument in favor of a new evolutionary economics. The results of the present paper will be reached from a completely different starting point, though. They rely on a general theory about the preconditions of a viable legal order as has been developed by Herbert Hart, in 1961. Therefore, hopefully, these approaches may lend support to each other.

numbers' interaction. But, he did not succeed to show how at all a large government organization could be possible under his own premises.

The rules of social order are artificial ones.<sup>7</sup> They neither fall down from heaven nor are they wired into the natural make up of our minds. They have to be produced as social (side) effects of individual actions. In this production process free riding cannot only occur as far as the primary obedience to the rules is concerned but also on the secondary level of enforcing the rules. Therefore some old questions immediately arise: Why should anybody care to enforce rules, why not take a free ride in the enforcement process? The enforcement of rules would be a public good not only to the collectivity of the ruled but also to the collectivity of the rulers. If rulers collectively should derive a rent from ruling the ruled, ruling would be a public good to *them*. Why, then, should anybody of the rulers participate in the enforcement process?<sup>8</sup>

Further, if the rulers could get committed to the rule of enforcing the rules without enforcement why should not the ruled themselves be able to become commit-

<sup>7</sup> Of course, I am alluding here to the distinction between the natural and the artificial (laws of nature *vs.* conventions or institutions) which was so prominent among the British Moralists; *cf.* for a general survey Raphael 1969, Schneider 1967, and Mackie 1980—for a fine discussion of the roots of this distinction in classical antiquity Heinmann 1945 (1987).

<sup>8</sup> I have left out of account the more elementary problems of a direct conflict of interest about the content of the rules. I implicitly assume that the existence of rules would form a Pareto-improvement—according to the weak Pareto-principle—to the community of all individuals. Then the interests of all coincide as far as the existence of the rules is concerned though participation in the production of rules is subject to free-riding. Of course the new institutional economics is well aware of such problems like “shirking”, “opportunism”, etc. But the concept is rarely applied to the production of the rules themselves. Holler 1986 points out some more subtle problems that might arise here.

ted to the rules without enforcement? If no further arguments are offered the evolution and existence of a social order without government agencies would be as (un)plausible and as (im)possible as a social and legal order which is organized and enforced by government agencies.

To cast the problem into slightly different terms: If a large state or government organization involving many members is viable would not for the same reasons anarchy amongst the many be viable too? If we should think that anarchy would not be viable would not the very existence of states as forms of ordered large scale interaction show that this tenet must be mistaken?

Finally, rational insight that from the point of view of ordered social interaction it would be better for everybody if "we" had other preferences keeping us away from the temptation to take a free ride as such is of no help. This is true regardless of the fact that we may be well aware that under given preferences we might end up with Pareto-inferior states and thus develop meta-preferences to change our preferences such as to avoid social dilemmas. Again, if we were free to commit ourselves to different preferences than we actually have or if we could change at will our preferences according to our rationally developed meta-preferences then completely different forms of human interaction than those which as a matter of fact have evolved would have been viable.<sup>9</sup> There could and—I presume—there would be large civilized societies which would be ordered truly anarchical.

<sup>9</sup> In a stimulating paper Heggemann, Raub, and Voss—*cf.* 1986—try to overcome this difficulty with pretty strong informational assumptions. To my mind the paper should be read as demonstrating that it would be of great value should we have access to appropriate commitment facilities.

If unlimited perfect self-commitment would be feasible then not only anarchism but also the constructivist approach to society would be a sound one. Except for exogenous shocks there would be no clearance for a gradual non-teleological social evolution. Rational individuals after becoming aware of improvements would immediately bind themselves to restrictions of individual strategy spaces leading to the realization of the improvements.<sup>10</sup> Society could be purposefully construed out of rational contracts presupposing no non-contractual social rules (forming the institution of contract itself). Reason could completely rule society. All changes of rules would be contractual and directed toward anticipated gains.

As we all know our world is not like this. Almost routinely people tend to blame "limited rationality" or "unreasonable human behavior" for this fact. But, contrary to that it is rather the other way round. Human reason at least in a sense is too powerful.<sup>11</sup>

<sup>10</sup> At least if a Thompson-Faith mechanism of truly perfect information would be involved; *cf.* 1981.

<sup>11</sup> If we entertain some strong notion of reason which goes beyond that of the individually rational pursuit of subjectively given ends then things might look otherwise. If, on the other hand, we stick to a concept of reason in practical affairs which is compatible with the economic point of view proper we should look at rational behavior as individually rational choice among (future) risky prospects according to the subjective values of the chooser. But, then, we will have to blame reason itself for our limited commitment capabilities. Following such theoreticians of ethics like Hume (1948), Mackie (1977), and Hoerster (1982, 1983b) I would claim that subjectivism is the proper view for practical philosophy too; but, to go into the details of an argument supporting subjectivism and skepticism in matters practical would lead astray at this point of analysis; *cf.* on the relationships between economics and ethics from such a point of view Kliemt 1987.

## I.2. Commitment and Reason<sup>12</sup>

Without the institutional umbrella shielding human interaction rational actors could agree on some rules. They could even “promise” to obey the rules. But, without preexisting institutions of commitment the mere words of a “promise” would not alter their pay offs nor narrow down their choice sets. Neither “threats” nor “promises” to that purpose would be credible without commitment facilities. For rational actors after the “promise” all alternatives of the original problem of individually rational choice would still be present as alternatives and would still offer basically the same pay offs as before.

In the traditional economic model it is assumed that actors endowed with reason will have full control over their own future decisions<sup>13</sup> and that they—according to a fundamental rationality precept which might be called the principle of intervention—distinguish between what they can and what they cannot causally influence by their own choices.<sup>14</sup> This has two implications. First, in a present decision situation the decision maker can causally influence incentives of future decision situations<sup>15</sup> but he cannot *now* control the future decision itself. Secondly, if full control over alternatives is presupposed for every decision situation, taken sep-

<sup>12</sup> This part of the paper is strongly influenced by Brennan and Buchanan 1985, chap. 5, by Baurmann 1983, and of course, though in a more general way, by Schelling 1956.

<sup>13</sup> It is worthwhile to note here that the expectations of the choosing actor do not contain probability distributions over the actor's own actions.

<sup>14</sup> Of course, there are some links to the “independence of irrelevant alternatives” assumptions here.

<sup>15</sup> This is of particular relevance if individuals create specific long term investments; cf. e.g. Alchian 1984. They get committed by modifying future pay offs or values.

arately, then strictly speaking we have a sequence or intertemporal team of decision makers in the same person. Therefore, the multiple-self models which are quite *en vogue* today seem to be implied already in the traditional model of expedient choice.<sup>16</sup>

These implications of the model of expedient choice become particularly clear in game theory. It is assumed that the players perceive each other as rational actors. In the standard approach of game theory a player who tries to anticipate what another player will do does not take resort to psychology or neurophysiology. Though, in the last resort, the behavior of the other player may be and presumably is governed by neurophysiological laws in classical game theory the other player is not modeled as a biological machine which is controlled by the laws of nature.<sup>17</sup> Instead of this we have a chooser who has control over his choices.<sup>18</sup>

To be sure, I am not discussing the ancient problem of determinism and free will here. From the point of view of an omniscient external observer the mind of a human being may be as determined and as predictable as a computer program. But, the model of individual behavior that is used in rational choice explanations is not designed to predict human behavior from an ex-

<sup>16</sup> Cf. for a recent anthology Elster 1986.

<sup>17</sup> Philosophers like P.F. Strawson, cf. 1974 esp. 9, have pointed out the distinction of the perspective of the participant and the external on-looker quite frequently. As such distinctions form a part of the philosophy of Austrian economics one might say that the origins of game theory are quite Austrian too.

<sup>18</sup> Characteristically the game theoretic assumption of complete information does not amount to the same thing as a predictive external theory of individual behavior. Players only know the kinds of incentives that will operate on themselves and others in each decision situation. They do not know the decision itself. How individuals will decide is analysed from the internal point of view of each rational actor faced with the incentives of the situation.

ternal point of view.<sup>19</sup> Instead of this these choices are treated as intentional or teleological expedient choices which are discussed from the internal point of view of the chooser<sup>20</sup>—regardless of whether the decision maker is located in our own or someone else's body. From this point of view decisions always could be otherwise.

At least the chooser, before choosing cannot know how he will choose.<sup>21</sup> Therefore the decision maker located in a body faces a problem of team coordination. The members of the team are present and future rational choosers who are linked by common interests but are deciding on their own.<sup>22</sup>

<sup>19</sup> We should not be distracted here by the concept of utility that we usually employ. A Ramsey-v. Neumann-Morgenstern utility function as such does not explain anything. Contrary to classical utility which had an emotional quality and formed a motive of action modern utility is merely a convenient short hand for describing the results of motivation. Only a very thoroughgoing orthodox behaviorist could avoid to make such a distinction between overt behavior and internal motivation.

<sup>20</sup> This procedure provoked a lot of argument about intentional explanations and the notion of a specific social science having access to methods of prediction which would go beyond the methods of natural science. Contrary to those who assume that social science has access to specific methods of explanation I would insist that for an explanation or a prediction properly so called we will always need a law of nature. This would have to state from an external point of view that individuals will act according to the incentives of a situation as they are observed by themselves. This would relate the internal and the external point of view. Relying on human motives is a substitute for forms of explanations that would be superior if we had only the knowledge; on the latter cf. Rosenberg 1980.

<sup>21</sup> Baurmann 1983 gives a careful account of the informational assumptions which would be needed in a complete discussion of this complex issue. For instance it seems to be necessary that an external observer would not pass on all information to an observed individual if the actions of the observed person are to be "predicted" in the proper sense of that term.

<sup>22</sup> That the individual I have in mind is not made out of straw but is indeed the common rational chooser can be illustrated quite easily by the fact that the whole discussion about credible threats and credible commitments in game theory is based on the implicit assumption of full control over decisions in subgames of a larger game. If, so to say,

This perspective on sequential rational choice might seem quite strange or even perverse to some economists. Therefore, as a pedestrian economist, I should like to draw attention to the words of authorities. Brennan and Buchanan sum up their discussion of "time, temptation, and the constrained future"<sup>23</sup> with the remark that as rational actors "...we cannot, in the present, make choices in future time. Nonetheless, the choices that we make now must embody the recognition that we will also face choices at some later date".<sup>24</sup> There will be a new chooser at every future point in time. "The chooser at  $t_0$  will know that the person alive at  $t_1$  must exhibit a will and a personality, a set of preferences, that are 'all his own'. The new person, emergent only in  $t_1$ , may find it in his power to destroy or modify seriously any plans that may be carefully reflected in the forward-looking choices made at  $t_0$ ."<sup>25</sup>

It might be helpful here to look at a simple decision tree of a sequential decision and to compare the strategic alternatives open to a rational human actor as modeled by economic theory with a computer program that is designed to simulate a rational strategy. The computer program may be written as complex and as flexible as one should like. Nevertheless, it *can* be as inflexible as we like it to be. The program, so to say, can pick branches from a decision tree. Contrary to that

a present self could commit a future self in advance to some retaliatory action which at the time of execution would be contrary to the incentives of the moment then the threat would be credible. The fact that game theorists assume that it is not credible to use such threats demonstrates that they assume that the decision maker in a strategic interaction has full control in each situation of choice taken separately and therefore cannot strictly speaking program future decisions.

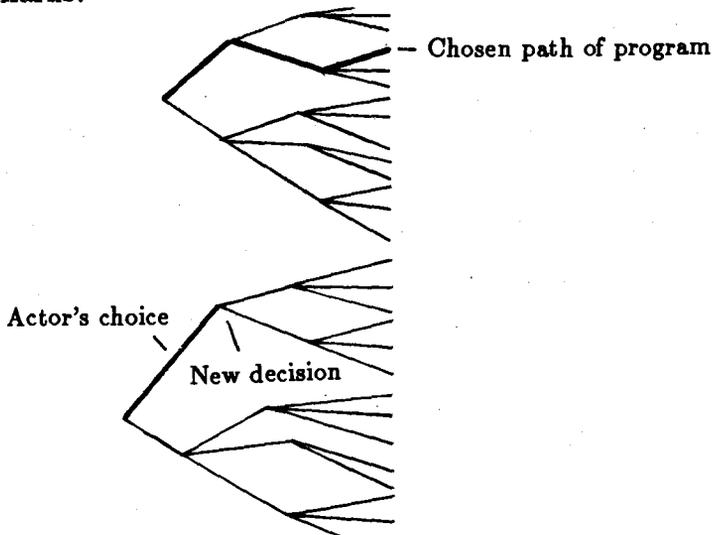
<sup>23</sup> This is the title of chapter five in Brennan and Buchanan 1985.

<sup>24</sup> Brennan and Buchanan 1985, 68.

<sup>25</sup> Brennan and Buchanan 1985, 71.

the inhabitants of the new world of (institutional) economics can only pick the next node of the tree together with all following branches. They can of course predict the incentives which will operate on them in alternative future decision situations. Nevertheless, as rational beings they cannot—strictly speaking—*predecide* their own subsequent decisions.

The following two trees may serve as an elementary illustration of the foregoing and some of the subsequent remarks:



For further illustration it may be helpful to compare a rational choice perspective of the iterated prisoners' dilemma with a genuinely evolutionary one as for instance in Axelrod's approach.<sup>26</sup> In Axelrod's discussion strategies are represented by programs. Selection and comparative advantage will operate on the programs and not on separate decisions in repeated interaction.

<sup>26</sup> Cf. Axelrod 1984.

There is an irreversible decision *before* entering the sequence of choice situations forming a supergame.

To assume that such a decision can be made before the actual play of the game is incompatible with the game theoretic model of a rational chooser who in principle can make a new choice in every normal game of a supergame. As long as we stick to game theory proper the notion of a supergame strategy is not referring to a program *guiding* choices but rather a convenient shorthand *describing* a sequence of separate choices. Though the sequence of expedient choices may exhibit some regularity there is no clearance for genuine rule following behavior or a kind of self-programming that supersedes situational expedient choices. The analogy between a supergame strategy and a computer program or a human disposition (trait of character) is therefore at least partly misleading.

In Axelrod's as well as in other models of evolutionary game theory this is often concealed by the fact that the retaliatory strategies involved use a notion of punishment that is based on a strategy that is dominant in every normal game ( $D_t$  in the prisoner's dilemma). In punishing the rational chooser chooses an alternative that is from the point of view of expediency tempting in any normal game anyway. But, in the real world punishment typically may involve additional costs and thus will be built on strategies that will be (weakly) dominated themselves.<sup>27</sup> The rational "temptation" in the normal game is to avoid the costs of punishment<sup>28</sup> and thus the rational chooser cannot avoid the prob-

<sup>27</sup> Cf. for a very stimulating discussion Witt 1986, the relationships to Gauthier 1986 should be clear too.

<sup>28</sup> This, of course, is well known from the discussion about deterrence. Relationships to Buchanan's Samaritan's Dilemma are obvious too; cf. 1975.

lem of credible threats.<sup>29</sup> As a fully rational being he cannot—as a program could—commit himself in advance to actions that hurt his own rational interests at the time of their execution.

The economist may still argue that economics is not the same thing as game theory. The economic perspective on rational choice has always been aware of the distinction between short and long term interests. The rational individual will resist the temptations of the moment if this is necessary to reach long term ends of superior value. The rational chooser can take into account the future. There may be a discount rate—due either to uncertainty or to the fact that present and future selves are not completely identical. But, then, future consequences will simply have a discounted bearing on present choices. The economist therefore might insist that he does not see why commitment is such a problem<sup>30</sup> and he might add that even the “cooperative solutions” of uncooperative games like the prisoner’s dilemma are brought about by the very “reason” that guides individuals in overcoming short term temptations for the sake of future gains.

Two things should be observed here. First, the problem of credible threats in game theory typically arise if after deterrence has failed it may not be wise to execute the threat. Bygones are bygones and carrying out the threat may bear no future consequences. Second, according to the already mentioned “principle of inter-

<sup>29</sup> There are many links to other discussions here that I cannot pursue any further at the present moment; *cf.* for a fine presentation of the general game theoretic background Friedman 1986, 68 ff.

<sup>30</sup> Professor Karl Brunner has done this in a forceful criticism and helpful oral comment on an earlier version of the paper. Though in the last resort I would not agree with his criticism completely it made me understand my own position better and to add—what I hope are—some improvements to the original argument.

vention" rational choosers will in general confine attention to causal consequences of their actions. Therefore, quite independently of the specifics of game theory, the execution of inconsequential acts poses severe problems to any consequentialistic theory of rational action.

The economic conception of rationality clearly is consequentialistic. In particular the argument that long run consequences of individual actions should and can be taken into account by the rational chooser is consequentialistic. The rational chooser as modeled by economic theory can take into account future discounted consequences of present decisions. But, according to the same model he cannot give up full control over the alternatives he is confronted with *in* a decision. If his views about his long term interests change he can still instantaneously change his choices even if they were made to serve what he originally or formerly regarded as his long term interests.

Without giving up the economic model itself the essential assumptions of full control and the principle of intervention cannot be suspended. Therefore, the rational chooser of the economic model cannot, should he come to the conclusion that *this* would serve his interests, decide to forgo full control of expedient choices or to decide irrespective of causal consequences. If the economic model of individually rational choice is to retain any content then it cannot allow for the possibility that rational choosers renounce at will the basic faculties of choosing rationally. In any case it would be clear that after such a general choice to waive these faculties specific situational choices cannot anymore be explained by the model of *expedient choice in these situations*.<sup>31</sup>

<sup>31</sup> Sen 1987, gives many arguments against the consequentialistic model of choice. On the explanatory level his analysis certainly has point though one may doubt whether he can defend his far reach-

It would be much more appropriate then to explain individual actions simply as rule following behavior.

This being said we may return now to the so-called problem of social order, to the question of how at all rules are brought into existence, and to the problem of why anybody might be expected to enforce rules without an external incentive to do so in every instance. The rational actor of our rational choice models cannot adopt by some magic traits of character or behavioral programs keeping him away from certain choices<sup>32</sup> which otherwise would be open to him. The resourceful, evaluating, maximizing man is so resourceful that he will take each situation separately and always do his best.<sup>33</sup> Using this idealized behavioral model we could hardly explain how a system of rules can exhibit that amount of stability that as a matter of fact can be observed in social reality. Stability presupposes commitments of actors which are incompatible with the assumption of expedient choice as a universal explanation of behavior that holds true in any instance of choice.

ing conclusions for the methodology of normative economics on these grounds.

<sup>32</sup> It is exactly here that Axelrod's elegant discussion of the evolution of cooperation is misleading if applied to cultural evolution. If human beings could deliberately and credibly adopt behavioral traits which would be binding as perfectly as a computer program, then the character of social interaction would be fundamentally changed. Above all any conditionally retaliatory strategy—added to the p.d. strategy spaces—involving extra costs to the retaliator could be made credible in a sequential game. This would change the whole game.

In a fascinating forthcoming book on *the passions within reason* Robert Frank, *cf.* 1988, analyzes in detail the possibility that natural or so to say wired in commitments to non-rational ways of choice that violate the assumptions of full control and the intervention principle may exhibit a kind of rational "function" on a higher level of analysis. But, even then it would be very doubtful, as Frank is well aware, to look at these phenomena as the result of choices.

<sup>33</sup> *Cf.* on the notions of REMM and RREEMM respectively Mecking 1976 and Lindenberg 1985a.

The *dual* approach of the new institutional economics should be prepared to find an answer to the question of how the enforcement of rules and the rules themselves can emerge as equilibrium outcomes of a system of interacting individuals.<sup>34</sup> We have to ask to what kind of a game the creation of rules is the equilibrium outcome. It is precisely here that the rule of reason as assumed by rational choice models runs into difficulties in explaining the existence of rules. In particular the concept of a gradual evolution of rules becomes partly incompatible with the concept of expedient choice. Coherence of both notions can be reached only if some assumptions about individually rational behavior which are typically made by rational choice models like those of game theory and economic theory are relaxed. After all what has been said the crucial issue may be formulated in a simple question: Is there anything in human nature that amounts—or is akin—to the same thing as the commitment facilities of a computer program?

## II. *The Past and the Future*

If we want to indicate in a handy formula the direction of necessary modifications of the model of rational choice then we should follow Dennis E. Mueller's presidential address to the Public Choice Society in Baltimore in 1986. Summing up his argument he simply observed "that man has not only a future but also a past".<sup>35</sup>

<sup>34</sup> I am alluding here to the fundamental methodological principle of constitutional economics formulated by Brennan and Buchanan in 1985, 15. Whatever we intend to bring about by rules: not our intentions but the equilibrium emerging within the rules from individually rational adaptive behavior will determine the final outcome.

<sup>35</sup> Mueller 1986a, 19.

Professor Mueller, of course, could have learnt from "sound economics" that for a rational decision the past as such never matters. According to the principle of intervention it is irrelevant from the point of view of reason. It is bygone and that is all to it. Sunk costs are sunk. Therefore it is quite interesting to see that an eminent economist is arguing to the contrary—at least in a sense.

He is referring to behavioral learning theory in this context and suggests to take a closer look at learning processes especially educational ones. These form the past experiences of the individual decision maker and are not easily forgotten. If for reasons of human nature this "acquired behavioral technology"<sup>36</sup> cannot be given up at will by human actors then, in a sense, commitment will come about by the laws of nature.

Nobody who has acquired certain skills and a certain amount of knowledge can decide not to have them anymore. We are endowed by nature with the capacity of acquiring traits of character which cannot be freely suspended afterwards. There will be time lags before these traits will be extinguished again. To train an individual—either human or non-human—will take some time whether the aim of training is to acquire skills or the reversal.<sup>37</sup> According to the laws of behavioral psychology the same should hold true in general for almost all human dispositions.

Further, as far as only low costs are involved it will not be rational to act as a rational decision maker in every instance of choice. The decision maker will have

<sup>36</sup> Cf. for this view of internalized moral norms as part of the "technology of behavior" Vanberg 1986a.

<sup>37</sup> Nelson and Winter 1982, esp. chap. 4, are quite right in pointing to skills in their proposal of a new evolutionary approach to economics.

to follow some rules of thumb, some standards of correct or even decent behavior that will be routinely applied.

This itself at least in part may be viewed as the result of an economizing process and scholars like Herbert A. Simon indeed would insist that it has to be. On the other hand the process of acquiring traits of character is not completely subject to individually rational decisions of acquisition. To become an honest man is not the same thing as buying some means of technical equipment. It is learnt behavior which gradually evolves within a process of social interaction without any explicit investment decision.<sup>38</sup> And, what is even more important in the present context, it cannot easily be renounced because the actor "cannot really shift his behavioral gears"<sup>39</sup> if expediency should suggest it. The behavioral technology of individuals shows a certain amount of rigidity across different situations and different points of time. There are costs of change that a "purely rational" chooser does not face. We have to take into account fixed costs of behavior or, so to say, behavioral inertia.<sup>40</sup>

The latter view very well fits into the picture not only of a behavioral theory of human behavior as proposed by Dennis Mueller but is also congenial to concepts of social evolution as that of Friedrich August v. Hayek. For, evolution requires some kind of rigidity or iner-

<sup>38</sup> The relationships to Buchanan's notions of "natural and artifactual man" should be obvious; *cf.* Buchanan 1979, 93 ff.

<sup>39</sup> I owe this fine formulation to a commentator whom I will leave anonymous here because I do not want him to come up from the morass of his paper with mud on his head too.

<sup>40</sup> If we add to behavioral inertia the notion of interaction specific behavioral investments there seem to be some obvious relationships to the theories of Williamson. But, at least as far as I can see as a layman I am going in a different direction here.

tia of the objects on which selection operates.<sup>41</sup> Thus, learning theory gives us some clue of how in the last resort cultural evolution and the fundamental rules of social interaction as its most important products may be rooted in human nature and a human capability to get committed to certain courses of action.<sup>42</sup>

Nevertheless, the economist should prefer an approach that is closer to main stream economics than learning theory. For, first, economists are right in stressing the human capability of deviating from past experience. Man is not "run" by a program and cannot be programmed completely.<sup>43</sup> A model starting from learning theory could take account of this fact but it is more in line with common experience to look at it from the point of view of expedient choice. Second, if we enter the field of learning theory the "zoo" of diverging idiosyncratic motives is lurking around the corner. Contrary to that the utility function representation of the individual makes it possible to study the effects of expedient choice on social interaction regardless of the motives giving rise to a utility function.<sup>44</sup> Observing that this feature of the economic approach is very convenient for the purposes of a *general* theoretical analysis of interac-

<sup>41</sup> This point is prominent in the book of Nelson and Winter 1982 too. They draw attention to the notion of a "program" in the approach of March and Simon, *cf.* p. 79, and to "routines", *cf.* 96 ff, etc.

<sup>42</sup> Evolutionary changes may involve shifting commitments (involvements?) too.

<sup>43</sup> Nevertheless, I would not argue that human behavior and human decision making could not be simulated by a computer program. But, such a program will be quite different in complexity from the present ones. In a sense it will be as difficult to predict the decisions of such a program as those of a human actor.

<sup>44</sup> One could also argue here that from the point of view of a cautious research strategy the short hand of representing rational actors simply by their utility functions which has proved so fruitful within the *general* analysis of social interaction should be defended in institutional analysis as far as possible.

tion we may wonder whether it is possible too to study the effects of the human capability to become committed independently of the motives that originally give rise to commitments. I will now turn to some ideas about such an approach.

### III. *Limited Natural Commitment Power*

As all human capabilities the natural proclivity to become committed to certain courses of action is limited in scope. We are not completely programmed by our past commitments neither individually nor socially. One of the traits of human behavior is that in principle human individuals can deviate in particular situations from their acquired behavioral technology which is routinely applied otherwise.<sup>45</sup> They *can* follow the precepts of reason and, thus, the exigencies of situations. On the other hand, "(m)an is as much a rule-following animal as a purpose-seeking one".<sup>46</sup> Obviously we should try to combine in one behavioral model two human faculties simultaneously: that of acting according to some rules and that of choosing according to the exigencies of the moment. Combining these two dispositions in a meaningful way requires that we specify the circumstances under which the one rather than the other form of behavior may be expected to occur.

As a first step to accomplish this task we may assume that *individuals will show different kinds of behavior in low and in normal or high cost situations*. According to this simple empirical hypothesis there is a kind of

<sup>45</sup> If authors like Merton look at deviant behavior as innovative then this fits in quite well here.

<sup>46</sup> v. Hayek, 1973, 11. v. Hayek himself points to the philosopher R.S. Peters and his well known book about *the concept of motivation* here.— Presumably some economists will say that he became distracted by a philosopher.

discontinuity in human behavior. Individuals switch between different modes of conduct according to the costs involved.<sup>47</sup>

The hypothesis amounts to roughly the same thing as to looking at rational self-management as management by exception. From this point of view the old controversy between so-called social and so-called economic man as basic concepts of social theory can be elegantly resolved.<sup>48</sup> Both men are combined in one person.

If the opportunity costs of a mistaken routine decision or the opportunity costs of following individual convictions about right and wrong are low we have to expect that a behavioral technology acquired in the past will be routinely applied.<sup>49</sup> If opportunity costs go beyond certain thresholds individuals endowed with reason will follow precepts of expediency. The future will enter the picture again. Instead of being motivated by accepted rules individuals will be motivated by reason and act purposefully according to the exigencies of the situation then.

<sup>47</sup> Rereading Heiner 1983, I became aware that he uses the notion of switching too in describing a related phenomenon.

<sup>48</sup> Both explanatory models suffer from severe deficiencies. As far as the sociological model is concerned I will not have to say much about this point. It has been convincingly criticized quite often. Following Karl-Dieter Opp 1986 I should like to draw attention to only one crucial issue namely that of conflicting norms. How, if not by reference to expected costs could the sociological model explain why one of the conflicting norms is observed whereas the other one is violated? The opportunity costs of obedience to the one consist in violating the other norm. Any acceptable model of human behavior must take account of this fact. Insofar as the sociological model of behavior does not do this it is not acceptable without modification. It must be amended by some elements modeling expedient choices. But, as is argued in the present paper the model of economic man must be modified too.

<sup>49</sup> Behavior will be determined in part by the acceptance of rules and not only by tastes that have been acquired in the past. The view also does not amount to the same thing as to adding new terms to a utility function as for instance Mueller 1986a, p. 7 (eq.(1)) proposes.

The distinction between low and normal or high costs is of crucial importance in understanding the game of life correctly. Even if it is quite commonsensical and almost trivial the results of consequently applying the model of switching are illuminating. The model will at once solve a variety of problems that otherwise might puzzle social philosophy as well as constitutional economics. And, of course, this explanatory power offers at least some empirical evidence supporting the general hypothesis of switching.

For instance, people do vote. This kind of behavior cannot plausibly be explained as the result of expedient choices among future prospects. There is some evidence that people vote because they feel obliged to do so.<sup>50</sup> The feeling is strong enough to overcome the adverse inclination to avoid the costs of voting.

Judges act according to the rules of the legal order without any strong external incentive to do so. In particular they will act according to what they regard as valid law even if this runs against their personal views. This often will occur without selective incentives because judges are deliberately put into a disinterested position by special institutional rules. To put people in a disinterested position or at least one in which opportunity costs of alternative decisions are low would be a strange institutional tactic if all human behavior would be the result of individual maximization.

People do retaliate. They are willing to incur some though not any cost just to punish or reward some other person to whom they perhaps will not have future relationships<sup>51</sup>—i.e. they play (at least) weakly dominated strategies. Quite frequently people praise and blame

<sup>50</sup> Cf. Mueller 1986b on this; see also for a general survey of some related issues Kliemt 1986b.

<sup>51</sup> Cf. on this again the stimulating paper of Witt 1986.

other people without aiming at future gains—at least not to themselves. Engaging in the activities of praising and blaming<sup>52</sup> will frequently involve costs. At the same time the individual actor has no reason to expect that the investment will pay off sufficiently in the future to make it worthwhile. The effects of praising and blaming are public whereas the costs must be borne privately.

It would be evidently besides the point to look at all phenomena of public opinion formation from the point of view of free-riding. We trust that scientists in general and constitutional economists in particular will engage in activities of persuading a reluctant public though this goes beyond their own personal interests. To neglect free-riding and related phenomena completely will distort our view of social reality. We will not understand the game of life correctly if we simply discount the counterexamples that typically occur in low cost situations as anomalies.

The examples are taken from both realms, that of “producing” the rules themselves and that of behavior within the rules. The proposed dual model of behavior does not simply amount to the same thing as two models of behavior disguised as only one. For, a definite hypothesis is formulated about the incidence of the two selves within the one self of the rational actor. In this the proposed approach deviates from related models of “discontinuous individual behavior” as for instance have been suggested by Lindenberg and Margolis.<sup>53</sup>

<sup>52</sup> James S. Coleman, *cf.* 1983, has analysed the economics of praise and blame in a very stimulating way but I do not agree with his attempts of adapting the phenomena *completely* to a common rational choice perspective; *cf.* for a related effort also Axelrod 1986.

<sup>53</sup> *Cf.* Lindenberg 1980, 1983b and Margolis 1981, 1982. I should like to state explicitly that these models have some advantages over the one proposed presently. First, they are formalized already. Second, they can take account better of certain kinds of behavior that are

Still, the adherent of the economic approach to human behavior might not be convinced. He may argue that low cost behavior and the kind of self-commitment which is introduced here may be modeled with traditional tools. In a sense I would agree. On the other hand, I would insist that we should be careful about this argument. The process of modeling is not neutral with respect to our view of reality. My argument is not a thesis about models as such but about the processes which are modeled. If these processes are different from what is usually assumed in forward looking theories of rational choice this difference should not be concealed by modeling them as if they were of the same kind as expedient choices. They should be modeled in a way that makes obvious the empirical claims involved.<sup>54</sup>

If, for instance, a rational actor knows that he—as part of his behavioral technology of commitment—is endowed with retributive emotions<sup>55</sup> then this can either be modeled as an alteration of pay offs which are associated with specific branches of a game tree or as a modification of the form of the tree. Assuming a fundamental difference in behavior it should be regarded as more promising to model this fact by modifying the

reserved for saints and heroes in high cost situations. Presumably a merger of the low cost hypothesis and these models may be the most promising avenue of future research. Some of the weaknesses of my argument are also avoided in Coleman 1983 which is of particular interest here if viewed in the light of the seminal philosophical paper of Urmsen 1958.

<sup>54</sup> We could look at certain forms of behavior as non-teleological. We then could regard them as "consumption". Zintl 1986 pursues this line of argument that is quite close to Margolis 1981, 1982 and much in Sen 1982, 1987 too.

<sup>55</sup> Cf. Mackie 1985a on retributive emotions as basic to the working of human moral codes.

tree—e.g. to add branches to the tree and not simply to rely on modifications of pay offs.<sup>56</sup>

In a game in normal form a modification of pay offs of the original matrix would only conceal crucial strategic aspects of the situation. Instead of this the original matrix should be bordered by additional strategies of self-commitment. This would make quite clear that the thesis which is presently defended implies certain tenets about “additional” strategies which are not easily observed in overt behavior.

To be more specific, imagine a repeated prisoners’ dilemma game which is “bordered” by additional retaliatory strategies.<sup>57</sup> We may talk here of a bordered prisoners’ dilemma. It might be objected that it is no “real” prisoners’ dilemma at all. Conceding that it is at least a prisoners’ dilemma with “noise” I would still maintain that most real world processes showing some features of a prisoners’ dilemma contain noise.<sup>58</sup>

<sup>56</sup> On the other hand, retributive emotions are quite specific motives. Introducing them instead of general notions of costs and utility will lead away quite far from conventional economic analysis. At least from the point of view of institutional analysis of strategic interaction it would be better if we would not have to go into such details of individual motives. We need handy formulas to describe the behavior of individual entities in general if we want to study general properties of interaction.

<sup>57</sup> The procedure of including retaliatory strategies works properly only if the interaction comprises a series of at least two prisoners’ dilemma games. Prisoners’ dilemma interactions have to be viewed as coming at least in pairs or there must be some after play interaction. What individuals do in the first game will have causal influence on what will happen in the next game(s) or in the interaction that is expected after the game has been played. Of course, one could argue here that I simply have changed the original p.d. game into something else. But, any iteration of a normal game in a sense transforms the game. Still one might insist that the normal game retains its essential characteristics even if put into a sequence.

<sup>58</sup> To give but one example: participants of an experiment may expect to meet each other after the experiment. From this some external expectations and strategic possibilities will arise.

The simple dilemma is adequate only as an abstraction or a pure case. It will be completely valid only in the limit, i.e. if numbers of participants grow (using for instance Taylor's generalization to  $n$  persons),<sup>59</sup> if the stakes and the pay offs dependent on them rise, if the probability of future interaction with the same partners declines and gathering information about the behavior of others becomes costlier. Within the limits we have to expect certain forms of behavior that deviate from the model of rational expedient choice.<sup>60</sup> They should be taken into account in our models of the prisoners'dilemma and the best way to model empirical facts like these presumably consists in adding appropriate additional strategies.

The argument does not yet contain anything incompatible with the "received view". However, according to the view presently taken "noise" will arise from an internal or intervening process within every actor rather than from external features of the situation. Traditionally neither the game theorist nor the economist were prepared to admit inner commitments as strategic choices. These strategies simply were not there. To add

<sup>59</sup> Cf. Taylor 1976.

<sup>60</sup> This sheds some light on the common view that limited altruism poses most of the social problems with which we are faced in social reality. That "professional good men" raise this claim is of no surprise. For, many of them spend their living on exploiting our tendency to become committed to certain causes. Except for times of emergency, war, etc. when the arousal of emotions will help to overcome obstacles to collective action anyway things will be just the other way round. We will have to take institutional precautions that artificially strengthen our sense of our interest so as to insure ourselves against various forms of altruism "going wild". The severe political evils in the world do not stem from self-interested behavior but from involvements which are altruistic in spirit—though conceptions of the content of altruism diverge interpersonally.

them to the picture makes a big difference to traditional theory even if it can be modeled with traditional tools.<sup>61</sup>

Still, some hard liners might argue that economists never doubted that people can and do act according to rules of thumb, follow routines etc. But they would stipulate that the outcome of this behavior can be predicted by the model of expedient choice. According to these theorists the crucial issue is the predictive value of their theories and not the realism of assumptions. To discuss this theme adequately and in all aspects a whole book would not offer enough space. Within the confines of this paper I can only deal with it very shortly.

First, the problem that is discussed in the present paper arises because the received view does *not* lead to acceptably accurate predictions. If one wants to defend the traditional model here, additional arguments would be necessary. Modeling the human individual as a resourceful forward looking chooser is correct in many situations. But, there are many situations too in which this view leads to inadequate predictions. In any case the defender of the received view cannot have it both ways. On the one hand insist that only predictive power matters and on the other hand, if predictions fail, refer to the alleged realism of his preferred model of man.

Second, noting that the modern notion of utility is motivationally neutral some adherents of the received view argue that explaining behavior as the result of utility maximization is possible without any motive to maximize on the side of individuals. The success of game theoretically inspired models in recent biological research seems to lend strong support to this argument though the economist usually would point to Armen

<sup>61</sup> A large part of the analysis in Robert Frank's forthcoming book (1988) which I regrettably did not know when writing the paper may be interpreted as an effort to add these strategies to the picture.

Alchians seminal article on *Uncertainty, Evolution and Economic Theory*.<sup>62</sup> To be sure, I accept this ingenious argument as far as *profit* maximization within an appropriately structured *preexisting institutional setting* is concerned. Nevertheless, it should be noted that the institutional structure artificially provides an objective selection principle operating on the individual decision units (profits and elimination of the non-profitable instead of “subjective utility”), which substitutes the natural or biological selection. It is the institutional structure which explains the survival of the more profitable and the process of “as if optimization”.

This is an extremely convincing defense of the market as an institutional device for the pursuit of social aims. The best thing about it seems to be that it shows how a market can work efficiently without relying on any assumption about individual motivation.<sup>63</sup> Nevertheless, the explanation of observed behavior is a genuinely evolutionary one. It relies on properties of a system of rules rather than on properties of individual behavior.

Still, in institutional economics we have to ask how the rules come into existence. This, I maintain, cannot be explained by a selection process operating independently of human motives. Further, to refer to choices motivated by considerations of expediency and maximization will not be sufficient. To take account of such choices is of crucial importance. But, additional factors or complementary technologies of human behavior are involved.

The behavioral technology of commitment in low cost

<sup>62</sup> *Cf.* 1950.

<sup>63</sup> There is not only a relationship to Adam Smith’s remark about the benevolence of the butcher but also a strong structural similarity to later arguments as those of Schelling about sorting in human behavior, *cf.* 1978, or those of Eigen and Winkler on the evolution of order in games, *cf.* 1981.

situations will be present all the time but it can be suppressed in certain situations in favor of rational expedient choice or case by case maximization. This not only echoes the remark about self-management as management by exception but also indicates how the argument can be taken one step further.

It may be stated first that people enter a lot of finitely iterated bordered prisoners' dilemmas in real life. According to the present view individuals enter such situations with fixed behavioral programs like those of Axelrod's experiments.<sup>64</sup> Imperfect discrimination between situations will play a considerable role. In particular there will be a carry over from small group interaction in which most behavior is learnt to broader contexts. Individuals accept some rules of behavior, either rules of thumb or obligatory rules, that may guide their own choices beyond considerations of expediency if this does not prove too costly.

Acceptance of a rule as a standard of behavior is a notion which is independent of the motives inducing acceptance. In this respect, it is not too far away from the notion of utility as representing motives of all kinds. To introduce the notion of acceptance of a rule or norm will also bring institutional economics in general and constitutional economics in particular closer to some of the wisdom of legal theory. Part of this wisdom is that the concept of expedient choice does not suffice to understand the actual working of a legal order.

The classical Hobbesian and Austinian concepts of law have been "economic theories of law" insofar as they tried to reduce the existence of rules to expedient choices of rational individuals.<sup>65</sup> We should take notice

<sup>64</sup> From this point of view the experiments become relevant for discussing human decision making too.

<sup>65</sup> I am following here what I would regard as the logic of the Hobbe-

of the fact here that one of the leading philosophers of law, Herbert Hart, has devoted a large part of his book, *The Concept of Law*<sup>66</sup> to the task of refuting this approach. Hart's arguments show that without relaxing the assumption of expedient choice social theory in general and economics in particular will not be in a position to cope with the existence of a complex set of legal rules. But, without this economics will not be able to deal adequately with Hayek's idea of the evolution of rules either. To these themes I will turn in the remaining part of my paper.

#### IV. *Accepting Legal Rules*

##### IV.1. *Accepting Rules and the Economy of Motives*

The concept of accepting a rule implies that at least sometimes we simply follow the accepted rule without any calculation of effects. We use it as a guide or standard of behavior, as a reason for action which at the very moment when it exerts its influence is not dissolved into calculations of the probable effects of following the standard or rule. Unless that be true sometimes the thesis that we accept a norm, standard, or rule would not be warranted.

To refer to the meaning of a term as has been done before as such does not show that the extension of the term is non-empty. Even if acceptance of a norm according to common usage of the term should exclude considerations of expediency it might well be that no real phenomena correspond to the concept.

sian approach. Otherwise, sticking closer to the text of the *Leviathan*, we would indeed have to take seriously Brian Barry's point about projecting Austin's onto Hobbes' analysis; cf. 1968.

<sup>66</sup> Cf. Hart 1961.

Some amount of self-deception may be involved if an individual allegedly accepts a norm but at closer inspection this might reveal itself as a concealed process of optimization. And, the adherent of the extreme model of economic explanation indeed would insist that a closer scrutiny of any situation apparently structured by norm or rule governed behavior would in the last resort yield an explanation that is based on individual calculations of expected future effects.

As should be clear by now I would not in general object to this perspective on human behavior. Economic approaches to human behavior are much more convincing than many of the orthodox sociological approaches that stipulate that norm orientation is the key issue of social analysis. Nevertheless, it would be grossly misleading to neglect rule or norm oriented behavior completely.

This kind of behavior should be carefully distinguished from selfless behavior. Contrary to what many economists and even distinguished critics of standard economic assumptions like Amartya K. Sen maintain the real issue is *not* "whether there is a plurality of motivations, or whether self-interest *alone* drives human beings."<sup>67</sup> The real issue is whether there is only expedient choice—either selfishly or selflessly motivated—or whether there are other forms of behavior that are not motivated by expected causal consequences of singular actions. On this head I maintain that the extension of the term "acceptance of a rule or norm" is non-empty in social reality.<sup>68</sup> After a rule is accepted the reasons

<sup>67</sup> Sen 1987, 19.

<sup>68</sup> It is not true too that people only behave *as if* acceptance would prevail. Of course sometimes expediency would dictate the same behavior as an accepted rule. The rule may only be the conscious counterpart of an unconscious incentive structure which typically would

that originally gave rise to the acceptance of the rule do not directly influence behavior anymore. At the very moment of action it is the rule itself that motivates behavior.

People sometimes act because they think that a certain action is right. Even if such a motivation of action should be the result of self-deception, even if it should spring from our imagination it would be real. We cannot deny that even a mistaken image may be a motivating factor of real influence on human behavior. As every economist is well aware the consequences of erroneous beliefs are as real as those of correct ones.

Further, even if it could be shown that without exception the acceptance of rules is motivated by higher order considerations about long run individual welfare even then the accepted rule would serve as a kind of intervening motive. As far as rules of thumb are concerned economists routinely admit that they may form a kind of intervening motive. But, even acceptance of rules of thumb and not only of obligatory rules forms an exception to the model of expedient choice.

Still, some economists may suspect that intervening motives are simply fictitious. They might argue that these motives are as mysterious as many other kinds of intervening variables that have been postulated during the evolution of social theories.

Again, I basically agree that there has been a great proliferation of variables and motives in the history of social explanation that did not improve social theories. But, it does not make much sense to economize on explanatory variables if this will conceal rather than make

be stable over time. Then the argument that the behavior of so-called *homo sociologicus* would be shown by rational economic men as limiting case of the model of expedient choice would apply. But, that this may be the case does not amount to the same thing as showing that these conditions will prevail regularly or even always.

apparent the workings of rules in the real economy of human motives. Parsimony being a virtue of theories in general will turn into a vice if it is forced onto reality for reasons of preconceived methodological ideas.<sup>69</sup> This is of particular relevance as far as the basic rules of social order are concerned.

#### IV.2. Hart's Argument

Herbert Hart has argued that the existence of a legal order could not be explained if individuals would rely on case by case calculations of the comparative advantages of alternative actions. As has been mentioned already he is attacking the classical Hobbesian and Austinian view that all norms are orders issued by a higher person or body of persons toward subordinate addressees of the norm. According to this view a norm or rule can be a motive of obedience only insofar as the addressees incur the risk to be visited with an evil should they deviate. Norms of law are only predictive rules indicating the risky prospect of punishment and thus facilitating expedient choices. The concept of obligation can be resolved into the prediction or threat of punishment. If the probability to be punished for deviant behavior converges to zero then the norm ceases to exist.

Of course, within the complex pattern of a legal order the risk of punishment is of great importance. That the supreme authority should be endowed with the power to punish seems to be beyond doubt. Nevertheless this observation is of no value in the present context. For, the topic to be discussed is not whether the fear of punishment sometimes is a decisive motive of individual

<sup>69</sup> Cf. on some related discussions Hirshman 1985.

actions but rather whether it is the only motive which has to be present all the time.<sup>70</sup>

That the latter thesis cannot be true is demonstrated by Hart with his model of "rex". Rex is a superior who is issuing orders and who will punish some individuals in his surrounding if they should fail to punish other individuals who are more distant to him. These more distant individuals in turn may be punished by their own superiors should they fail to punish still other individuals etc.

Even if we grant that such a highly improbable social structure of hierarchically expanding circles could work in principle as long as rex is alive it is quite clear that it would cease to exist at the very moment rex dies. The hierarchy of punishment will break down when the top of it passes away. In turn all norms or rules will cease to exist. For, according to the premises of the model, nobody can commit himself to the observation of rules and thus to the execution of threats beyond considerations of expediency. None of the players involved has a motive to proceed in enforcing the rules. Not incurring the costs of enforcement will be a dominant strategy.

This in turn will exclude the possibility of ordered succession. The successor of rex cannot claim allegiance according to some rule or other. There cannot be such a rule creating obligations independently of rex. All rules depend on rex and the fear of punishment. Rex is the creator of all rules and only rex sustains them.

The model of rex is of course a highly artificial one. But it is the only one that is compatible with a thoroughgoing economic or game theoretic model of the working of a legal order. That the model of rex seems to be almost absurd is a very telling fact. It clearly shows

<sup>70</sup> This remark closely parallels the one of Sen cited before.

that something must be wrong with our basic model of expedient choice if we use it as the *exclusive* basis of social explanations.

If we look at social reality unbiased by theoretical prejudices we will become aware immediately that a simple hierarchy of threats will not explain how a complex legal pattern of rules actually works. Besides standing point of the gun there are other *empirical* facts about the legal order<sup>71</sup> of which acceptance of rules—or, for that matter, principles<sup>72</sup>—is the most important one.

The rules that are constitutive for the institutional order must influence our imagination in a way that transgresses their use as means of prediction<sup>73</sup> of future evils or rewards. If constitutional economics turns to such wider questions as demonstrating the reason of rules, as developing criteria for the design of rules etc. it cannot afford to neglect the actual working of social institutions. It must be based on a model of the human individual which is more adequate for institutional analysis than the usual *homo oeconomicus* model of expedient choice.<sup>74</sup>

<sup>71</sup> I am not arguing in favor of natural law here. I subscribe to legal positivism and the separation of law and morals; *cf.* for a clear presentation of different concepts of validity of norms or rules Hoerster 1983a.

<sup>72</sup> Though Ronald Dworkin who pointed out the importance of principles would not agree. I still would claim that this notion is compatible with legal positivism; *cf.* on this Baurmann and Kliemt 1986. Of particular interest in the context of the present discussion is Mackie 1985b.

<sup>73</sup> It should be observed here that David Hume as one of the early "champions" of institutional analysis was very well aware of the effects of the human imagination. As Gerhard Streminger, 1981, has convincingly argued the whole *Treatise* can be interpreted as a critique of the human imagination.

<sup>74</sup> Presumably it is unnecessary to stress here that the assumption that individuals will behave towards rules or within the clearance left

## *V. A Defense of v. Hayek's Concept of Cultural Evolution and its Application to Social Rules*

In a paper on the relationship between v. Hayek's libertarian evolutionism and Buchanan's contractarian constitutionalism that was prepared for the Liberty Fund Conference on "Philosophical and Economic Foundations of Capitalism" in Freiburg in 1981 Viktor Vanberg has argued convincingly that v. Hayek's antithesis of evolutionistic and constructivistic views is asymmetric.<sup>75</sup> On the one hand we have the thesis that constructivism maintains that *all* social rules or institutions are or should be the outcome of intentional design. On the other hand we have the somewhat weak thesis that at least *some* rules evolved as unintended side effects of individual actions which were not directed to an intentional enactment of these rules.

According to the observations made before the Hayekian description of evolutionism may now be considerably strengthened. The weak thesis that only some and not all rules of social order have gradually evolved is compatible with the thesis that these rules necessarily include the most fundamental rules which make stable social order viable at all in that they make people provide the "first" or "fundamental" selective incentives for other individuals without themselves having a selective incentive to do just this.

In the light of what has been said before about law and normal cost decisions this strengthened thesis may be defended quite easily. The model of the rational actor who is making expedient choices is first relaxed to the

by existing rules as rational economic men may be of the greatest importance if we want to design rules; *cf.* on this especially Brennan and Buchanan 1985, chap. 4.

<sup>75</sup> *Cf.* Vanberg 1981, 17 f. and also for a related general discussion Vanberg 1986b.

model of an actor who is showing different behavioral dispositions in low and high cost situations. In low cost situations the actor will be expected to follow rules without a selective incentive to do so.<sup>76</sup> These rules will then guide behavior as accepted standards that are not merely used as means to predict probable consequences of behavior. In his imagination the actor is perceiving what according to the rules has to or should be done. He is committed to the rule and will follow his perceptions as long as he will not be faced with considerable costs that could make him switch to innovative behavior or to case by case optimization.<sup>77</sup>

Taking a closer look at the partition of labor in the enforcement of rules it can be easily observed too that the fundamental institutions of the legal order involve asymmetric costs. Individuals can inflict high costs on others incurring only low costs themselves.<sup>78</sup> This will typically hold true for the legal staff which is crucial for the working of the legal order. Especially the most fundamental rule of a legal order, the so-called rule of recognition,<sup>79</sup> which is used as a standard of identification of valid law is applied and interpreted by the highest courts of a system in low cost situations.

<sup>76</sup> These rules may themselves be rules of thumb that economize on decision costs. But, regardless of this fact such satisficing behavior is changing the situation.

<sup>77</sup> To put it slightly different: The individual is conditionally committed only—the condition being that costs will not rise too much. The actor is both potentially obeying the rule as well as potentially deviating from the precepts dictated by the rule. But because of the condition of costs this is not an empty assumption. There is a quite clear condition for switching.

<sup>78</sup> It has been pointed out frequently that such asymmetries are characteristic for power relationships. We now may observe that if institutions which order social interaction on a larger scale are necessarily based on cost asymmetries it comes as no surprise that power relations cannot be eliminated from social life.

<sup>79</sup> Cf. Hart 1961, 97 ff.

From the point of view of personal interest the opportunity costs of preferring decision *a* to decision *b* are negligible for the judges—at least we hope so. The same will hold true in general if the rule of recognition or some of its more subordinate rules are applied by lower courts and other officials of the system.

Further, analogous remarks will apply too to other forms of social institutions like non-government organizations within an existing legal order or to government institutions of a system which does not rely on courts. If hierarchies of stable rules exist at all then according to Herbert Hart's arguments at some basic stage of the hierarchy the explanation of regular behavior will not be reducible to expedient choices. Some individuals will have to accept some rules for other reasons than future expected costs. They will have to apply them without calculation.

If we concede that the latter hypothesis is correct then it is clear how besides and beyond technical costs like set up costs and long term specialization of resources some amount of social inertia or observation of rules may arise. As long as the partition of labor in the enforcement of rules will secure that enforcers will not be faced with high opportunity costs of enforcement enforcers will in general follow the rules of enforcement without considerations of expediency.

The acceptance of rules as an empirical fact can also explain how cultural and especially social evolution as a process of selection may be construed. On accepted rules as fixed programs for a whole sequence of choices some kind of selection may operate. Rule acceptance leads to behavioral inertia and thus provides the kind of stable pattern (meme) on which genuinely cultural evolution may operate. Without the ability to accept rules and to act accordingly rational individuals would

be victims of their own capability to choose according to the exigencies of the moment and social evolution could not get started nor could the results of the process be stable.

Hayek's forceful remark that "we can either have a free parliament or a free people"<sup>80</sup> fits in very well here. He assumes throughout his writings that *all* actors within a legal order properly so-called are bound by rules which are not subject to arbitrary alterations according to the exigencies of the moment.<sup>81</sup> This could not be realized if man besides being a purpose-seeking animal would not be a rule-following animal too.

The notion of accepting rules is central to both a proper understanding of the process of gradual evolution of rules and the explanation of how systems of stable binding rules can exist at all among resourceful, innovative individuals who are able to adapt purposefully in situations of expedient choice. What I have in mind is a merger between the Social Darwinist and Hayekian idea of the gradual evolution of a system of rules with Hart's explanation of the working of the outcome of this process. Uniting both approaches with each other and the traditional model of rational economic man seems to be necessary if we eventually want to form an adequate view of the system of rules governing social life and rules must limit the rule of reason if we want to live in a free or "great society".

<sup>80</sup> v. Hayek 1979, 102.

<sup>81</sup> On this point I have commented elsewhere using Hart's theory; cf. Kliemt 1978.

## REFERENCES

- Albert, H. 1977: "Individuelles Handeln und soziale Steuerung", in: Lenk, 1977, 177 ff.
- Alchian, A.A. 1950: "Uncertainty, Evolution, and Economic Theory", in: *Journal of Political Economy*, Vol. 58, 211 ff.
- 1984: "Specificity, Specialization, and Coalitions", in: *Zeitschrift für die gesamte Staatswissenschaft*, Vol. 140, 34 ff.
- Axelrod, R. 1984: *The Evolution of Cooperation*. New York.
- 1986: "An Evolutionary Approach to Norms", in: *American Political Science Review*, Vol. 80, No. 4, 1095 ff.
- Barry, B. 1968: "Warrender and his Critics", in: *Philosophy* 48, 117 ff.
- Baurmann, M. 1983: "Zweckrationalität und Strafrecht". Inaugural Dissertation des Fachbereichs Gesellschaftswissenschaften der J.W. Goethe-Universität. Frankfurt/M.
- and H. Kliemt 1986: "Rechtspositivismus auf die leichte Schulter genommen?", in: *Österreichische Zeitschrift für Soziologie*, Vol. 10/2, 121 ff.
- Brennan, G. and J.M. Buchanan 1985: *The Reason of Rules*. Cambridge.
- Buchanan, J.M. 1975a: *The Limits of Liberty*. Chicago.
- 1975b: "The Samaritan's Dilemma", in: *Phelps* 1975, 71 ff.
- 1979: "Natural and Artifactual Man", in: *What Should Economists Do?*, Indianapolis, 93 ff.
- 1986: "Better than Plowing", in: *Banca Nazionale del Lavoro Quarterly Review* (Roma), No. 159, 359 ff.
- Coleman, J.S. 1983: "Free Riders and Zealots"; in: *Sodeur* 1983, 135 ff.
- Diekmann, A. and P. Mitter 1986: *Paradoxical Effects of Social Behavior*. Essays in Honor of Anatol Rapaport. Heidelberg and Wien.
- Eigen, M. and R. Winkler 1981: *Das Spiel*. München.
- Elster, J. (ed.) 1986: *The Multiple-Self*. Cambridge.
- Frank, R. (forthcoming; 1988): *The Passions Within Reason*. Norton Publ.
- Friedman, J.W. 1986: *Game Theory with Applications to Economics*. Oxford.
- Garzón-Valdés, E.: "Die gesetzliche Begrenzung des staatlichen Souveräns", in: *Archiv für Rechts-und Sozialphilosophie* LXVIII/4, 431 ff.

- Gauthier, D.P. 1986: *Morals by Agreement*. Oxford.
- Hart, H.L.A. 1961: *The Concept of Law*. Oxford.
- v. Hayek, F.A. 1960: *The Constitution of Liberty*. Chicago.
- 1973: *Law, Legislation and Liberty*. Vol. 1, *Rules and Order*. London and Henley.
- 1976: *Law, Legislation and Liberty*. Vol. 2, *The Mirage of Social Justice*. London and Henley.
- 1979: *Law, Legislation and Liberty*. Vol. 3, *The Political Order of a Free People*. London and Henley.
- Hegselmann, R., W. Raub and Th. Voss 1986: "Zur Entstehung der Moral aus natürlichen Neigungen. Eine spieltheoretische Spekulation", in: *Analyse und Kritik*, Vol. 8/2, 150 ff.
- Heiner, R. 1983: "The Origin of Predictable Behavior", in: *American Economic Review*, Vol. 83/4, 560 ff.
- Heinimann, F. 1945: *Nomos und Physis*. Darmstadt.
- Hirshman, A.O. 1985: "Against Parsimony", in: *Economics and Philosophy*, Vol. 1/No. 1, 7 ff.
- Hoerster, N. 1982: *Rechtsethik ohne Metaphysik*, Juristenzeitung, 265–272, and 714–716.
- 1983a: "'Wirksamkeit', 'Geltung' und 'Gültigkeit' von Normen. Ein empiristischer Definitionsvorschlag", in: Mayer-Maly, D. und Simons, P.M. (ed.), *Das Naturrechtsdenken heute und morgen*, Berlin, 585 ff.
- 1983b: "Moralbegründung ohne Metaphysik", in: *Erkenntnis*, Vol. 22, 225 ff.
- Holler, M.J. (ed.) 1984: *Coalitions and Collective Action*. Würzburg.
- Holler, M.J. 1986: "Moral Sentiments and Self-Interest Reconsidered", in: Diekmann, A. and Mitter, P. 1986, 223 ff.
- Hume, D. 1948: *Moral and Political Philosophy* (ed. Aiken). New York and London.
- Kliemt, H. 1978: "Can There Be Any Constitutional Limits to Constitutional Powers", in: *Munich Social Science Review*, 4, 99 ff.
- 1986a: *Antagonistische Kooperation*. Freiburg.
- 1986b: "The Veil of Insignificance", in: *European Journal of Political Economy*, 2/3, 333 ff.
- 1987: "Ethik und Ökonomik", in: *Wirtschaftswissenschaftliches Studium*, Vol. 16/3, 113 ff.

- and B. Schauenberg 1984: "Coalitions and Hierarchies", in: Holler, 1984, 9 ff.
- Lenk, H. (ed.): *Handlungstheorien interdisziplinär*, IV. München.
- Lindenberg, S. 1980: "Marginal Utility and Restraints on Gain Maximisation: The Discrimination Model of Rational Repetitive Choice", in: *Journal of Mathematical Sociology*, Vol. 7/2, 289 ff.
- 1983a: "The New Political Economy: Its Potential and Limitations for the Social Sciences in General and for Sociology in Particular", in: Sodeur, 1983, 7 ff.
- 1983b: "Utility and Morality", in: *Kyklos*, Vol. 36/3, 450 ff.
- Mackie, J.L. 1977: *Ethics. Inventing Right and Wrong*. Harmondsworth.
- 1980: *Hume's Moral Theory*. London.
- 1985a: "Morality and the Retributive Emotions", in: *Persons and Values*. Oxford, 206 ff.
- 1985b: "Rules and Reason", in: Stewart, 1982, 31 ff.
- Margolis, H. 1981: "A New Model of Rational Choice", in: *Ethics*, Vol. 91/2, 265 ff.
- 1982: *Selfishness, Altruism, and Rationality. A Theory of Social Choice*. Cambridge.
- Meckling, W. 1976: "Values and the Choice of the Model of the Individual in the Social Sciences", in: *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, Vol. 112/4, 545 ff.
- Melden, I. (ed.) 1958: *Essays in Moral Philosophy*. Seattle and London.
- Mueller, D.C. 1986a: "Rational Egoism Versus Adaptive Egoism as Fundamental Postulate for a Descriptive Theory of Human Behavior", in: *Public Choice*, (51), 3 ff.
- 1986b: *The Voting Paradox*. Unpublished Manuscript.
- Nelson, R.R. and S.G. Winter 1982: *An Evolutionary Theory of Economic Change*. Cambridge and London.
- Opp, K.-D. 1986: "Das Modell des Homo Sociologicus. Eine Explikation und eine Konfrontierung mit dem utilitaristischen Verhaltensmodell", in: *Analyse und Kritik*, Vol. 8/1, 1 ff.
- Parsons, T. 1968: "Utilitarianism. Sociological Thought", in: *International Encyclopedia of the Social Sciences*, New York and London.
- Phelps, E. (ed.) 1975: *Altruism, Morality, and Economic Theory*. New York.

- Raphael, D.-D. (ed.) 1969: *British Moralists*. Oxford.
- Rosenberg, A. 1980: *Sociobiology and the Preemption of Social Science*. Baltimore and London.
- Schelling, T.C. 1956: "An Essay on Bargaining", *American Economic Review*, 46, 281 ff.
- Schelling, T.S. 1978: *Micromotives and Macrobehavior*. New York and London.
- Schneider, L. (ed.) 1967: *The Scottish Moralists on Human Nature and Society*. Chicago and London.
- Sen, A.K. 1982: *Choice, Welfare, and Measurement*. Oxford.
- 1987: *On Ethics and Economics*. Oxford.
- Sodeur, W. (ed.) 1983: *Ökonomische Erklärungen sozialen Verhaltens*. Duisburg.
- Stewart, M.A. (ed.) 1982: *Law, Morality and Rights*. Dordrecht.
- Strawson, P.F. 1974: *Freedom and Resentment*. London.
- Streminger, G. and E. Topitsch 1981: *Hume*. Darmstadt.
- Taylor, M. 1976: *Anarchy and Cooperation*. London et al.
- Thompson, E. and R. Faith 1981: "A Pure Theory of Strategic Behavior and Social Institutions", in: *American Economic Review*, Vol. 71, 366 ff.
- Urmson, J.O., 1958: "Saints and Heroes", in: *Melden*, 1958, 198 ff.
- Vanberg, V. 1981: "Liberaler Evolutionismus oder vertragstheoretischer Konstitutionalismus?" (Walter Eucken Institut: *Vorträge und Aufsätze* 80), Tübingen.
- 1986a: "De Moribus Est Disputandum. Ethics in Economic Explanations." Paper prepared for presentation at the Interlaken Seminar on Analysis and Ideology (unpublished), May 1986.
- 1986b: "Spontaneous Market Order and Social Rules. A Critical Examination of F.A. v. Hayek's Theory of Cultural Evolution", in: *Economics and Philosophy*, Vol. 2, 75 ff.
- Williamson, O.E. 1983: "Credible Commitments: Using Hostages to Support Exchange", in: *American Economic Review*, 83/4, 519 ff.
- Witt, U. 1986: "Evolution and Stability of Cooperation without Enforceable Contracts", in: *Kyklos*, Vol. 39/2, 245 ff.
- Zintl, R. 1986: "Ökonomisches Rationalitätskonzept und normorientiertes Verhalten", in: *Jahrbuch für Neue Politische Ökonomie*, Vol. 5, 227 ff.

## RESUMEN

La explicación de los fenómenos colectivos a partir de decisiones individuales racionales es el objetivo del enfoque económico de la teoría social. Este enfoque parece ser correcto en su núcleo. Sin embargo, su suposición básica, es decir, que el decisor racional posee un control pleno de las alternativas de la respectiva situación de decisión, es inconciliable con un comportamiento guiado por reglas. Esta incoherencia impide que el enfoque económico pueda llegar a una adecuada comprensión del origen y mantenimiento de las instituciones sociales. Si la existencia de reglas y el comportamiento según reglas ha de ser explicado de acuerdo con el mismo modelo de comportamiento, es necesario introducir una modificación básica del "modelo del comportamiento económico". En este trabajo se esboza la manera como puede llevarse a cabo esta modificación dentro del marco de consideraciones filosófico-sociales generales acerca de la razón de las reglas.