Güth, Werner; Kliemt, Hartmut

Working Paper

What ethics can learn from experimental economics - if anything

Jena Economic Research Papers, No. 2008,062

Provided in Cooperation with:
Max Planck Institute of Economics

Suggested Citation: Güth, Werner; Kliemt, Hartmut (2008) : What ethics can learn from experimental economics - if anything, Jena Economic Research Papers, No. 2008,062, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:
https://hdl.handle.net/10419/31740

JENA ECONOMIC RESEARCH PAPERS

2008 – 062

What Ethics Can Learn From Experimental Economics – If Anything

by

Werner Güth
Hartmut Kliemt

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact m.pasche@wiwi.uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Jena Economic Research Papers 2008-062

Werner Güth* and Hartmut Kliemt**

What Ethics Can Learn From Experimental Economics – If Anything

Abstract

Relying on the specific example of ultimatum bargaining experiments this paper explores the possible role of empirical knowledge of behavioural "norm(ative) facts" within the search for an inter-personal (W)RE – (Wide) Reflective Equilibrium on normative issues. Assuming that pro-social behaviour "reveals" ethical orientations, it is argued that these "norm-facts" can and should be used along with stated preferences in justificatory arguments of normative ethics and economics of the "means to given ends" variety.

JEL Classification: D673, D64, D7, K00, Z13,

Key words: Meta-Ethics, Experimental Economics, Reflective Equilibrium

1. Introduction and overview

Economists are used to look at human interaction as strategic. Since they have been exposed to game theory for a long time, the use of game experiments for testing hypotheses about human behaviour in interactive situations seems rather natural to them. Though the external validity of austere game experiments may often be doubtful there can hardly be any doubt that the experiments cast doubts on the universal validity of the classical Homo oeconomicus model: The game experiments in the laboratory are real in that they provide real, typically monetary, incentives for rational Homo oeconomicus behaviour. The behaviour is real human behaviour and not merely a model of it.

In particular, if individuals forego opportunities – as defined in substantive rather than utility payoffs – this can be directly observed. That they do so may be either due to constraints of cognitive abilities and/or of commitments.

* Max Planck Institute of Economics, Jena, ** Frankfurt School of Finance and Management. We would like to express our gratitude to the CESifo institute for organizing a great conference on the relationship of ethics and economics under the auspices of Vesa Kanniainen and Manfred Holler. We thank the participants of the conference for their discussions and helpful criticisms.

Subsequently we are not so much interested in the effects of cognitive limitations but more in commitments and how the latter contribute to the emergence of what economists classify as "pro-social behaviour".

To that end we shall not only ask what "laboratory experiments measuring social preferences reveal about real moral problems" [1] but also what they can contribute to developing answers to moral problems in terms of prescriptive moral theory. In traditional philosophical terminology we address the "meta-ethical" question whether and if so how the results of laboratory experiments can contribute to justificatory arguments of normative ethics. In doing so, we acknowledge that value and normative judgments cannot be derived from judgments on matters of fact alone. However, within the economic means to given ends framework [2] the *observation* that individuals do subscribe to normative and value judgements or show behaviour that expresses such a subscription is relevant. Put again in more traditional philosophical terminology, the justification of hypothetical imperatives that suggest how given aims, ends, or values should be pursued can be rationally justified within normative economics. Only about the "ultimate" aims, ends, or values, normative economics must remain silent.

Many economists seem unaware that there is an important ethical tradition that does not venture beyond the Robbinsian limits of rational normative argument.[3] In this "sceptical" tradition, normative ethics is restricted to the justification of hypothetical imperatives that point out means to "given", aims, ends, or values of the addressee of the justificatory argument (not the norm so justified!). Even

[1] We are not claiming that the ultimatum game is the only "game in town." But we happen to know this one rather well. It is also not by chance that Levitt and List start their recent useful discussion whose title is echoed here with the ultimatum game; see Levitt and List (2007).
[2] As laid out canonically in Robbins (1935)
[3] The first mature presentation of this kind of ethics emerged from the internal discussion among the British Moralists, see for classical excerpts. Hume (1964 (Reprint of the new edition London 1886)) is to the present day almost an canonical presentation; commented on in Mackie (1980), Hardin (2007), Kliemt (1985) and worked out in some of the strategic details in Binmore (1994),Binmore (1998), Binmore (2005), Sugden (1986), Skyrms (1996), Taylor (1976).

though the ultimate aims, ends, or values, may be "given" in the sense that they are accepted without further justification, their implications are not "given" in another relevant sense but must be construed.

Sceptical economists and ethical theorists both have to cope with the *multiplicity* of given aims, ends, or values that all are to be pursued under the scarcity constraint(s). Pursuing such a vector of given aims, ends, or values under scarcity constraint(s)[4] implies that opportunity costs emerge. The trade-offs between the alternative fulfilments of ends must be determined. To represent such trade-offs economists would typically take resort to indifference curves.[5] To assume that not only the underlying dimensions of value, but also the indifference curves are "given" along with the aims, ends, or values is a rather daring assumption, however. With respect to real behaviour it is in fact grossly inadequate.

Cognitively constraint, boundedly rational decision makers have to go through a complicated process of trial and error to actually construe the relevant trade-offs along any indifference curve. In particular if individuals do endorse certain ethical aims, ends, or values along with their material interests it will become extremely complicated to fix the relevant indifference trade-offs explicitly. In neo-classical economics this search is reduced to the solution of an economic maximization under constraints problem. But in a realistic bounded rationality perspective the search process would be one in which a satisfactory solution fulfilling simultaneously all aspirations – perhaps after adapting some of them "downwards" – is found.[6]

Philosophers tend to acknowledge that boundedly rational ethical actors will necessarily have to enter complicated reflections and an ongoing search process.

[4] Assume for the sake of the argument that all desires are insatiable in the senses that none of the aims can be completely fulfilled under the scarcity constraint.
[5] A point very nicely made in the beginning section of Barry Barry (1965).
[6] See on this Simon (1957), Simon (1985), and from our point of view Güth (2000)

In philosophy this has led to a conception that can be seen as a bounded rationality approach to justification of prescriptive judgments. As so many important developments in practical philosophy of the second half of the 20-th century such a trial and error process has been made popular – though not necessarily been invented – by John Rawls.

Already in his original "outline of a decision procedure for normative ethics"[7] Rawls acknowledged the relevance of "normative facts" or of the judgements individuals do as a matter of fact accept from a moral point of view. Later Rawls embedded his theory of justice into the justificatory framework of the search for what still later became the search for a (wide) reflective equilibrium.[8] It is impossible and – hopefully – unnecessary to discuss the details of Rawls' own procedural proposals for searching a reflective equilibrium of all particular and general normative judgements here.[9] Suffices it to note that we intend to go beyond Rawls' approach in two regards. First, we do not restrict ourselves to the search from an impartial spectator's point of view. In our particularist rather than universalist framework the person we have in mind is in search of an equilibrium in pursuit of *all* her given aims, ends, or values. Second, also contrary to traditional universalist as well as Rawlsian universalist ethical theory the search for an equilibrium is not restricted to judgemental normative facts – "stated preferences" – but includes behavioural normative facts – "revealed preferences".[10]

We start with a brief rehearsal of existing proposals to frame ethical deliberation in close parallel to scientific deliberation (2.).[11] Next we introduce a class of games which seem to be particularly interesting with respect to justice related

[7] Rawls Rawls (1951)
[8] See Rawls 1970, 1974, Daniels (1979)
[9] A crisp account can be found in Hahn (1998)
[10] On the related concepts of revealed and stated preferences, see Louvierre et al. (2000). In our context it will not be preferences but rather rule following behaviour that is at stake. We acknowledge, however, the problem of identifying the decision rules "driving" overt behaviour, see lucidly on this, Manski (2002).
[11] see Rawls (1974) Daniels (1996) Hahn (2000) The approach in substance but without the term is used in Goodman (1978).

behaviour (3.) and sketch some central justice related normative facts as emerging especially from ultimatum experiments (4.). We then discuss to what extent the search for intra- as well as inter-personal reflective equilibrium can make sense despite the striking heterogeneity of factual justice related behaviour (5.). We conclude with some remarks on less idealized "bounded justice"[12] as naturally embedded in a bounded rationality framework (6.).

2. The wide reflective equilibrium approach to ethics

2.1. The search for an equilibrium in scientific methodology

In former times norms of good scientific practice were developed in a top down approach. Typically they evolved more or less on the basis of a priori arguments out of some epistemologically motivated philosophical conception. Such philosophical conceptions still do play some role. Yet nowadays due respect for established scientific practice – for what the sciences in fact do or have done – serves as the main springboard for normative considerations.[13] This leads to an a-posteriori or experience based process of developing norms of good scientific practice out of a stylized account of scientific practice itself.

The process of finding "best practice standards" is to some extent circular. It starts with a specific practice that prevails in the realm of science. To serve as authoritative evidence the practice must, first, be classified as "successful" according to some very broad evaluative standard. Then, second, a stylized account of the practice is given, or as philosophers tend to say, it is "rationally" reconstructed. Third, certain aspects are identified as likely causes of success. These are, fourth, presented in an idealized or stylized form to serve as (or at least as a basis of) normative standards of "good" science.

[12] Borrowing Volker Schmidt's apt term, Schmidt (1994).
[13] see for the background of this, of course, Fleck (1935/1980), Kuhn (1962), Lakatos (1978); see in the same spirit but closer to experimental economics and to the experiments discussed below, Binmore and Shaked (2007).

To put the same thing slightly otherwise, the established scientific practice gains a special normative status simply because it is an accepted established practice that is deemed successful by the "practitioners".[14] Though successful practices determine what good practice is, an established practice can be corrected in a piece-meal way by the very generalizations and norms that are developed out of observations of that practice.[15] A quest for substantial coherence is the driving force of this "rationalizing" process which can go back and forth between the general and the particular until coherence is reached.[16]

Though it may temporarily come to a halt, reflection can nevertheless always start all over again.[17] The search for reflective equilibrium will stop only temporarily once a "sufficient" level of coherence – meeting some aspiration level concerning the required coherence – is met.[18] In this as in other aspects the reflective equilibrium approach is merely an idealized form of daily trial and error practices of justifying judgement on issues of scientific practice.[19] What is good enough for the paradigm rational practice of science should be good enough for other human endeavours as well. So let us turn to the analogous justificatory method that Rawls proposes for the purposes of justifying normative judgements.

---

[14] see on this Kliemt (2004) And also other contributions in the same issue of CPE.

[15] this is, of course, close to a critical rationalist account, too, see in particular Albert (1978) But it is more coherentist than the critical rationalist approach.

[16] In a strategic context there are clear relations to the concept of theory absorption as in Morgenstern (1972), Morgenstern and Schwödiauer (1976), Dacey (1976) on the one hand and to the dynamics of rational deliberation as in Skyrms (1990) on the other. In a non-strategic context implied consent models may be relevantly related to Lehrer and Wagner (1981). We are, however, interested in kinds of deliberation that are close to actual boundedly rational processes of deliberation as originally in Rawls (1951). These do not rely on the extreme idealizations of the aforementioned in other respects quite inspiring models; for considerations somewhere inbetween, see Güth and Kliemt)

[17] The more extreme puzzles of ethical theory are not very telling with respect to workable ethics. Like the proverbial hard cases that make bad law they may make for good training of ethical theorists in a university setting, but not for good moral theory.

[18] This account of the "decision procedure for normative ethics" is in the spirit of Simon (1985), Simon (1957). In view of cognitive dissonance theories, see Festinger (1957), one might try to measure degrees of such dissonance as emerge from incoherence and then fix a threshold that must be met before we can assume that any remaining dissonance would be insignificant. However, since such considerations are beyond the scope of our present analysis let us simply assume that there is a satisfactory level of coherence which, for the time being, leads to an end of further search processes.

[19] Since we do not regard its application as constitutive for "truth" the WRE metaphor is fully coherent with a realistic conception of science and scientific truth, see the critical rationalist treatise Albert (1985).

2.2. The quest for substantive normative coherence

In search for a coherent normative system[20] of general and specific judgments the approach must start with some basic "normative facts". Traditionally these facts have been taken to be basic "normative *judgements*". For instance in the most simple case of his search for a personal reflective equilibrium on matters of justice Rawls relied basically on introspective evidence. He wondered what he himself – and, as he implicitly speculated, other competent addressees of his argument[21] – would find intuitively appealing.[22]

As Richard Hare objected early on, this seems a bit too much of circularity: "Rawls' POP [people in the original position] come to the decisions that they come to simply because they are replicas of Rawls himself … It is not surprising, therefore, that they reach conclusions that he can accept" (Hare (1973), p. 249). And, with Frohlich and Oppenheimer, we may add that "(t)he traditional philosophical methodology for dealing with justice has called for introspection and argument about these issues. We believe that this narrowly introspective approach has limited progress in the field of ethics because it has not allowed philosophers to introduce the diversity and fine details to obtain the balance sought. For that a broader strategy is needed." (Frohlich and Oppenheimer (1992), 2-3)

As part of their "broader strategy" Frohlich and Oppenheimer send the impartial spectator to the laboratory.[23] This takes normative facts of real world practices more seriously than Rawls' arm chair empiricism. However, it is still rather close to the original Rawlsian ways. In particular the participants of Frohlich's

[20] Ideally a normative system would have an elaborate logical structure but here a much looser use of the term is intended; see for a strict analysis the seminal Alchurron and Bulygin (1971)
[21] See on the requirements of competence Hoerster (1977)
[22] There is not only arm chair economics but also arm chair philosophy and not only "in the great library above".
[23] We come back to the relationship to politics; for the time being, see Brennan and Lomasky (1985)

and Oppenheimer's experiments are exposed to "impartiality situations" in which they operate behind some veil of uncertainty. Individuals who do intend to deliberate from a moral point of view will accept this experimentally imposed veil as expressing their impartial intentions. It induces them to take into account all social positions in their joint deliberations in the laboratory[24]. The veil of uncertainty about their own later positions is real. So it is not outrageously optimistic to expect participants to agree unanimously[25] in a calculus of consent[26] manner on a constitutional decision.[27]

The experimental set up of Frohlich and Oppenheimer enables them to test the acceptability of moral principles by factual acceptance under idealized conditions (i.e. by means other than introspection). The individuals are forming their opinions in communicative situations of joint deliberation.[28] In the situations specific strategic aspects play a role because agreement must be found under a Buchanan type unanimity principle[29] and therefore every participant is endowed with veto-power.[30]

This is a possible way of framing decision making on ethical principles. It can generate useful information concerning those given aims, ends, or values that represents the moral point of view of an individual. However, the analysis is a partial one in two closely related senses. Firstly, it is biased towards the moral point of view and, secondly, it contains only some of the given aims, ends, or values. Contrary to that human actors are always making their choices in

[24] Because in the experiment the uncertainty is real they have to do so as a matter of fact whereas in the Rawlsian thought experiment of the original position ignorance is entirely fictitious.
[25] A condition which as a matter of fact is not met in social reality where individuals outside small groups operate under conditions of individual insignificance; see on the hidden collectivism of the unanimity principle also Kliemt (1994)
[26] See, of course, Buchanan and Tullock (1962)
[27] As envisioned in Brennan and Buchanan (1985)
[28] Alluding to the fashion of our day, one might also refer to it as "deliberative democracy in the lab" operating under special knowledge and agreement conditions; for a collection, see Elster (1998).
[29] e.g. Frohlich and Oppenheimer (1992), p. 28, p. 40
[30] It is too often overlooked that there are two completely distinct forms of unanimity: on the one hand the agreement of any number which leaves all who do not join the agreement without a say (club with endogenously fixed membership) and on the other hand the agreement of all in which each has a veto (democratic community with exogenously fixed membership).

situations in which both the strategic and the non-strategic, the partial and the impartial interact with each other. Only if we take this into account can we adequately understand the workings of morality. Therefore, we should consider such situations in which the moral and the non-moral points of view are inseparably intertwined (at least when it comes to action).[31] Accordingly, we suggest to confront theories of justice with real justice related behaviour in situations were impartiality along with partiality is operative.

Propositions that describe and theories that explain morally motivated *behaviour* are of the greatest importance for developing an unbiased normative argument. Empirical observations concerning the behavioural trade-offs can be brought into play via ultimatum games. They have distinct advantages for our purposes: First, in ultimatum game interactions justice and equity concerns express themselves quite directly. Second, in the class of games to which ultimatum bargaining games belong we can be pretty sure that in a wide sense "moral motivations" do play a role and compete with "non-moral motivations".[32]

3. Ultimatums, retributive emotions and morals

In the simple games of proposal and response we consider, two actors, a proposer and a responder, can split a fixed sum or pie p.[33] Proposer X assigns shares $x,y \geq 0$ of the pie such that $x+y=p$. The share of the proposer X will be x, while y will be the share of the responder Y. After learning what the proposal (x, y) is responder Y can accept or reject the proposal.[34] If she accepts, then the rewards are assigned as proposed to the two participants, i.e. X receives x and Y

[31] Though there is a lot of behavioural evidence used even Konow is trying to separate judgment and justice concerns per se from other motives; see Konow (2003).
[32] Note that we use the terms "moral" and "non-moral" without passing any judgment on moral rightness. The same applies in our view to "ethical" and "unethical" which should also be disentangled from "ethically" right as opposed to "ethically wrong". If somebody wants to make a claim about right and wrong she or he should explicitly say so.
[33] A broader general description is presented in Manski (2002), 883.
[34] There are also yes/no experiments in which the responder is not informed about (x, y), see Gehrig et al. (2007).

receives y. Should the responder *reject* the proposal then the rewards will be (αx, βy) with $1 \geq \alpha, \beta \geq 0$.

It is instructive to look briefly at the four extreme parameter combinations (α, β), α, $\beta \in \{0, 1\}$. If α=β=1 then independently of its acceptance or rejection by the responder the reward allocation will be (x, y). The response of the responder is completely inconsequential for the material or substantive payoffs of both participants. In short, if α=β=1, the proposer is in a dictatorial position. If α=0 and β=1 then the responder can reject a proposal without forgoing any material payoff to herself. Her acts are substantially (as measured in material payoffs) inconsequential for herself while maximally consequential for her co-player. If, however, α=1 and β=0 expressing resentment will have no direct monetary impact on the proposer. The proposer can do whatever he chooses with impunity. The rejection by the responder is inconsequential for the proposer while – relative to the proposal y – it is maximally consequential for the responder to say no. Finally, with α=0 and β=0 the responder can express her resentment but only at the full cost of entirely forgoing y. Her rejection of the proposal is maximally consequential for both proposer and responder since it will transform the proposed outcome (x, y) into the realized one of (0, 0).

Overview over the parameter constellations that give rise to different types of justice related interactions in simple proposal response games

I. First mover X (proposer), second mover Y (responder)
pie, p,
proposal by X, (x, y) with x+y=p, addressed at Y

If Y accepts → (x, y) as payout
If Y vetoes → (αx,βy) as payout

II. Games that emerge from extreme parameter cosntellations

I. (α=1,β=1) → Dictator game
II. (α=1,β=0) → Impunity game
III. (α=0,β=1) → Bribe game
IV. (α=0,β=0) → Ultimatum game

V. α,β ∈(0, 1) → intermediate cases of games

In the conventional experimental setting the participants are carefully instructed and have to answer control questions to make sure that any of the preceding rules of the game as happen to apply in a specific setting are rendered common knowledge. Under this proviso the first case corresponds to a dictator game. The last case, α=0 and β=0, corresponds to a situation in which the proposal (x, y) amounts to an ultimatum which can be rejected by the responder only with the consequence that the pie is altogether forgone.

It would be most interesting for our enterprise to include all extreme cases as well as taking samples from the full range of intermediate cases $\alpha, \beta \in (0,1)$ of such games.[35] This would allow for finer discriminations between possible normative convictions of individuals who show different forms of behaviour. We will confine ourselves to the case of "the ultimatum (bargaining) game"[36] as an exemplary and straightforward case. But it should not be neglected that looking at the full class of games might be necessary for identifying the normative principles actually guiding behaviour and how the norms are traded off against other considerations.[37]

The emphasis on the ultimatum game makes also systematic sense because the role of retributive dispositions and emotions that shows up so clearly in that game has traditionally been identified as crucial for the proper workings of moral (and, for that matter, legal) institutions in general.[38] On the one hand, retributive dispositions may assist individuals in overcoming some kind of myopia that may otherwise impede their pursuit of long run interests. This happens if (e.g. in a context in which the folk theorem logic applies) individuals

[35] See for instance Suleiman (1996) who explored intermediate cases $\alpha = \beta \in (0,1)$.

[36] Studies concerning such experiments in different societies can be found in Henrich et al. (2004). Overviews are given in Güth and Tietz (1990), Roth (1995) and more recently Camerer (2003)

[37] See again the fine paper Manski (2002).

[38] See Mackie (1982), Westermarck (1906).

have good long term reasons to show certain kinds of retributive behaviour but suffer from some kind of weakness of mind or will. "Short-sightedness" prevents them from actually executing the sub-game perfect acts required according to the rational "master plan" and emotions kick in to overcome it.[39]

On the other hand, retributive dispositions may induce non-sub-game perfect behaviour that violates requirements of forward-looking opportunistically rational choice.[40] Such non-equilibrium behaviour[41] of certain individuals can support norm compliance in others. When one person sanctions the misbehaviour of others because person acts from an internal point of view according to some rule this will support the workings of moral and legal institutions.[42]

That some boundedness or restriction of opportunistic rationality must be present in ultimatum game experiments is clear from observed rejections of substantial positive offers. No future causal consequences of retributive acts can explain that in terms of substantive payoffs. Even though such in all likelihood "norm-bounded" behaviour violates what may be called the "efficiency axiom" it is not necessarily "moral" in the full sense of that term. However, it is at least akin to moral behaviour and therefore forms a natural starting point for an empirically based ethical study of the phenomenology of moral behaviour.

---

[39] Arguments of a closely related type are found in Frank (1987) Frank (1988) The general relationship to weakness of the will problems as discussed in economics and philosophy is obvious. See on the economic side for instance Strotz (1955), Thaler and Shefrin (1981) Schelling (1984) More to the psychological and to the philosophical psychology side, see Ainslee (2002), Spitzley (1992), Spitzley (2005)

[40] For interesting if somewhat sweeping recent claims concerning the punishment case see, Fehr and Gächter (2002).

[41] The behaviour is out of equilibrium only as far as the game in substantive or material payoffs is concerned.

[42] To understand the full impact of the Hartian analysis it may be helpful to consult the very insightful and clear account in, Barry (1981) See also the classical statement of the alleged Hobbesian order problem which was already the central concern of the British moralists in Parsons (1968) The original treatment of rule following from an internal point of view in the context of legal institutions is, of course, Hart (1961).

4. Some normative facts in ultimatum bargaining experiments

4.1. A first account of responder behaviour

In "standard" ultimatum bargaining experiments (i.e., if $\alpha=0$ and $\beta=0$) many if not most responders (consistently more than 50%) reject offers, y, in the range $p/3>y$ and even more distinctively so if offers fall below 20% of the fixed pie p. As questionnaires show this applies to responders who are fully aware that the interaction is anonymous. They seem to understand, too, that with practical certainty they never will interact again with the same proposer.[43] If so, any rationalization of the observed behaviour in terms of objective external incentives or extrinsic motivation is ruled out. Participants must be bound by some kind of intrinsic motivation to behave in the way they do regardless of the fact that their behaviour will not have any external causal consequences that will affect *themselves*.

The experimental control over material payoffs does not include the subjective framing of the situation by participants. Since the "pie", p, in the original experiments was simply given to the participants without further ado the frame of reference might have been that of splitting up a gift. In such a framework, under conditions of anonymity that excluded value dimensions like desert, merit, need etc. a claim to substantively equal shares seems to be natural. Though the roles of the two actors were quite different the roles were assigned randomly and participants knew that this was so. In view of the fact that offers deemed to be too low were frequently rejected we may perhaps conclude that the role-asymmetry was not perceived as an overwhelming concern in the experimental setting. Though it could have justified interpersonal differences in rewards, responders did not perceive it that way, at least not across the board.

[43] There has always been a certain amount of scepticism concerning this premise since individuals might have deeper gut feelings adapted to repeat interaction. For a small group context see Huck and Oechssler (1999).

The framing of the interaction situation as well as its embeddedness in a larger context matters. For instance, in situations in which the roles had been auctioned off among participants the willingness to accept low offers considerably increased. Likewise, if the pie p was "earned" in a joint effort by the participants rather than "dropping down from the sky" the relative contributions to the effort of earning it induced an increased proclivity to accept asymmetries in assignment.[44] The conclusion from this seems to be that not merely final distributions (end states) lead to justice concerns but history or how the results were brought about are crucial as well. This is what common sense tells us anyway but clearly it is re-assuring that experimental results do not contradict elementary common sense.

4.2. Responder strategies

Experiments based on the so-called strategy vector method offer additional insights. In such experiments participants were presented with two lists, one for the proposer and one for the responder role. In the first complete list of all possible offers they had to select the one that they would make in the proposer role. In the second list they had to decide for each possible offer whether they would accept it or not in the responder role.

For instance, in newspaper ultimatum experiments participants were asked to submit full strategy vectors for both roles[45]. They were informed that merely a few randomly chosen participants would be paid out in real monetary terms. The participants knew that those selected were to be paid according to the play (respectively the result) emerging from pairing the strategies of different individuals. Within the readership of a large nationwide newspaper the probability $1>q>0$ to be paid had to be expected to be low. Though a rather large

[44] Akin to classical beliefs about justice as described in ethical theory, see Frankena (1966) or for that matter the Aristotelian views directly as well as the social psychology literature on which we will not even dare to touch here.

[45] i.e. for the responder role it would be said for all (x, y) proposals whether a yes or no would be the response; for instance according to monotonic response strategies all proposals larger than some y* would be accepted while all lesser or equal amounts would be rejected.

pie p was allocated according to the strategies submitted[46], the real payment consequences became concealed behind a veil of uncertainty. To the extent that opportunity cost of fixing strategies one way or other were perceived as low, strategy fixing may itself have been in part an expressive rather than a strategic act.[47]

Regardless of the possible influence of expressive strategy fixing it seems safe to conclude that equity and justice concerns are in fact expressed in responder roles. Such normative orientations presumably operate with increasing relative strength the lower their opportunity costs are. But behaviour that is not in line with maximization of substantive payoffs is shown also when opportunity costs are quite high. The violation of the basic economic assumption of opportunism motivated by substantive payoffs is obvious. Less obvious and more interesting is it to find out why retributive behaviour is shown (expressed) in some instances while not in others. Why do actors actually yield to a retributive impulse sometimes and sometimes not?

4.3. Additional aspects of responder behaviour

One reason for behavioural differences might be that (for x, y>0) "punishment efficiency", x/y, matters to actors.[48] The value of x/y=(p-y)/y would be increasing with decreasing y. However, in cases in which even "too" high offers have been rejected punishment efficiency can hardly be the motive. Another possible argument might be that actors want to express in some way or other their resentment against violations of equity per se. Deviations can go beyond the limits of the tolerable in all directions. What is tolerable is fixed by a kind of aspiration level located in a neighbourhood around the equitable solution of the problem. If the results are deemed intolerable the retributive emotion is aroused.

[46] See Güth et al. (2003), Güth et al. (2007)

[47] See on this in particular Brennan and Lomasky (1985) Brennan and Lomasky (1989) Kliemt (1986). The incentives point in the correct direction, though, as required in Carson and Groves (2007).

[48] For a given proposal (x, y), x,y>0, the influence of punishment efficiency might be tested across games $(\alpha,\beta)\in(0,1)\times(0,1)$ by letting $\alpha\to 0$ or $\beta\to 0$ and considering $\alpha x/\beta y$.

It must somehow find a way to express itself. And it is expressed at the cost y>0. or so the speculative argument may run.

This reading is supported by experiments with α=1 and β=0 in which regardless of the absence of punishment options responders nevertheless chose to reject too low offerings to themselves. It seems also to corroborate such an interpretation that offering an additional cheap talk option, in which individuals in the responder role could voice their complaint to a proposer by whom they felt unfairly treated, reduced rejection rates considerably.[49] Expressive needs do matter and not merely in the "talk is cheap" sense.

4.4. Proposer behaviour

Ultimatum game experiments are interesting not only because they teach us something about retributive behaviour of responders but also because they tell us something about the expectations of proposers. In experiments with α<1 and in particular with α=0, responses will have substantive monetary consequences for proposers. Therefore they should have a substantive incentive to form some view of what the responder, Y, might do in response to a proposal (x, y).

Should the proposer X form a model of the situation in which the responder acted as a fully rational choice maker motivated by monetary payoffs only, he should offer merely the smallest monetary unit y>0. However, in classical ultimatum game experiments, most individuals who are assigned the proposer role X tend to offer more than the minimum amount. Many decide on an allocation of (p/2, p/2). It seems that they want to be assured that the responder would accept the offering. But we cannot be sure of the presence of that strategic extrinsic motive since it may as well be that they are intrinsically motivated to make an offer of (p/2, p/2) regardless of the expected response.

[49] See Xiao and Houser (2005), Güth and Levati (2007).

Some kind of inequity aversion may in fact be operative here in the proposer as well as in the responder role.[50]

In experiments relying on the strategy vector method, individuals who indicated that they would accept meagre offers in the responder role often were nevertheless willing to make rather "equitable" offers in the proposer role (in fact the most frequent strategy vector). That would not be in line with the hypothesis that inequity aversion applies across the board. Others would make low offers in the proposer role X which they themselves would not accept. If they expected similar response behaviour by others this would not make sense as strategic behaviour even though it looks like it.[51] – Whether a bounded rationality approach can help in identifying meaningful normative decision rules is an open question as well.

4.5. Bounded rationality in proposers

The behaviour of proposers can at least conceivably be in line with the consequentialist forward looking rationality concept of standard economic theory. Though it seems rather clear that the individuals do not form beliefs and expectations along the lines suggested by expected utility theory they may still be acting in view of expected future consequences. And we think they often do act in such a teleological manner.

If proposers go about their decision-making in terms of basic rules of thumb which express expectations about acceptance or rejection by the co-player, the following graph might be used to present in a stylized way what is going on:

[50] See Fehr and Schmidt (1999) and also Bolton and Ockenfels (2000), but also the criticism of the work of Fehr and Schmidt in Binmore and Shaked (2007).
[51] See on an experiment in which beliefs were elicited Güth et al. (2007) .

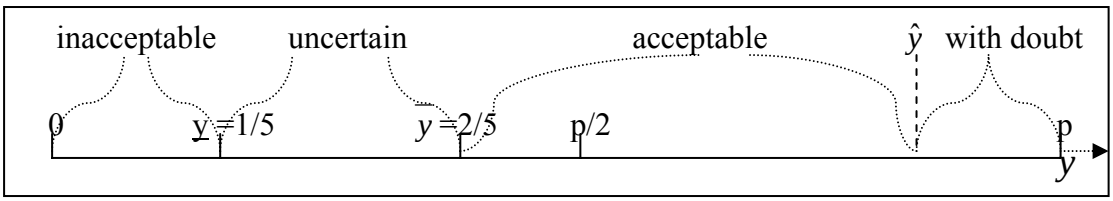

Figure 1

Assuming that $0 < \underline{y} < \bar{y} < \frac{p}{2}$ we can offer the following comments on behaviour and its likely motives in the different intervals:

$y \in \left[\frac{p}{2}, p\right]$, proposal $y < \hat{y}$ is expected to be accepted but since for $y \geq \hat{y}$ offerings may increasingly appear like charitable givings there may be some doubt about responder behaviour in case of offers close to p

$y \in \left[\bar{y}, \frac{p}{2}\right)$, in this realm X expects proposal y to be accepted by Y

$y \in \left[\underline{y}, \bar{y}\right)$, proposal y is such that the proposer just does not know what to expect

$y \in \left[0, \underline{y}\right)$, proposal y is expected to be rejected

When $y = \frac{p}{2}$, proposal y is expected to be accepted with practical certainty. This is in some Non-Bayesian way qualitatively different from all the other assignments. Likewise the inability of a decision maker to say what to expect in the range of $y \in \left[\underline{y}, \bar{y}\right)$ is clearly not in line with common Bayesian assumptions. Genuine uncertainty rather some probabilistic uncertainty prevails.

It is true, expected value formation could also explain statistical observations of real proposer behaviour. Assuming that individuals endorse heterogeneous beliefs about responder behaviour roughly the same statistics might emerge.

Nevertheless, there seems to be convincing evidence that the model of boundedly rational decision making is more faithful to the cognitive processes underlying actual human behaviour than the expected value hypothesis.

Additional experiments are necessary to understand the complexities of (in a wide sense) boundedly rational "moral" motivation more fully.[52] The results of such additional empirical research can, of course, not be foreseen. It is a rather safe prediction, though, that heterogeneity between individuals will persist. The same holds good with respect to cultural and situational differences. The generalization from one situation to others will raise additional complicated problems which may require the ability to make prudent judgements in one way or other. We have to take these facts of decision making into account when seeking a reflective equilibrium on justice or equity related matters rather than to insist on some streamlined rational choice model. Explaining away normatively relevant heterogeneity by heterogeneous beliefs while insisting that some basic consensus on aims, ends, or values prevails is not very plausible.

5. Towards behavioural reflective equilibrium?

Since a simple description and explanation of factually observed behaviour would not help in a justificatory enterprise like the search for (W)RE an additional step is needed: The observed behaviour must be put into a "rule" perspective. To that effect we need to postulate norms and rules such that an individual accepting those rules and norms as standards of her own behaviour would plausibly show the observed behaviour. Clearly, norms cannot be tested directly against behavioural facts.[53] Yet it can be tested whether or not the behaviour that should be shown according to the rules and norms imputed to the actors is in fact shown.

[52] In the present case all the reservations laid out in Binmore and Shaked (2007), would kick in. It would perhaps be necessary to design a sequence of experiments to identify which of the rules of boundedly rational choice making are operative.

[53] It is possible to run experiments in which by an observable act of choice participants select a norm which then determines behaviour – without the actors being guided by their understanding of the norm in each case.

The crucial claim is: If actors would follow the rules *from an internal point of view*[54] then they would show overt behaviour of a certain kind. If an actor who allegedly adopts an internal point of view to certain rules does not show the corresponding overt behaviour this falsifies the ascription of the normative theory as an accepted standard of behaviour – at least to some extent. The actor reveals in overt behaviour that she either follows different rules or that the opportunity costs of rule-following are too high.

5.1 Intra-personal incoherence in ultimatum game experiments

Let us start with intra-personal "heterogeneity" which seems to express itself in a kind of "role incoherence". For instance, if the proposer in an experiment employing the strategy vector method is personally inclined to accept meagre offers as a responder and at the same time is willing to offer an equal split of the pie as a proposer, such behaviour, at least at first sight, seems to violate certain principles of role coherence. Should not the morally coherent individual act in ways that would lead to the same result if the individual would adopt both roles in the ultimatum game?

Some ethical theorists as well as pedestrian moralists seem to tend to such views. They would require that an individual should in the proposer role offer what the individual would accept in the responder role and demand in the responder role not more or less than the offer the individual would make in the proposer role. However, already the very first ultimatum game experiment indicated that some participants would not have accepted their own proposal in the responder role. The offers of others would have gone well beyond the threshold demanded by them in the responder role.[55]

A moral philosopher who argues that the consideration of trade-offs between moral and other motives contaminates moral analysis behaves like the rational

[54] In the sense of Hart (1961).
[55] See again Güth et al. (1982).

choice economist who intends to form a theory of rational behaviour without paying due respect to the facts and practices of actual boundedly rational behaviour. Akin to his "brother in guilt"[56] such a moral philosopher may want to render his claims definitional truths by identifying moral behaviour as being motivated by respect for the moral law per se. Such a move may be fine for the Kantian "homo noumenon" but the morals of the "homo phenomenon" cannot abstract away everything but the "moral dimension".[57]

More often than not, there is in fact a trade-off between the requirements of impartiality and partiality. Day to day morals does not require that we grant no weight to motives other than moral ones. It requires that we give moral motives "acceptable" weight. This fits neatly not only with notions of boundedly rational (satisficing) behaviour it coheres well also with commonly accepted requirements to help others if that can be done at low costs to oneself.[58]

Interpersonal comparisons are required. That economists tend to rule them out as non-operational does not imply that real people would not perform such inter-personal comparisons intra-personally all the time.[59] Once we look at it that way it seems unacceptable to eliminate the so-called non-moral motivations from the picture. When seeking a reflective equilibrium on moral matters the trade-off between what we owe to ourselves (as well as to those close to us) and what we owe to others is of the essence of the moral decision problems.

If we include trade-offs then behaviour in the proposer and the responder role that otherwise seems incoherent may become quite coherent. Opportunism

[56] See Sugden (2004).
[57] The somewhat strange Kantian composition of Latin and Greek is used in a rather elaborate way in Kant (1798/1977).
[58] As opposed to Anglo-Saxon law not merely a moral but a legal obligation under German law. Still particularly instructive on this Frellesen (1980).
[59] Assuming that there are as many personal welfare functions for the society as there are individuals each of them might represent the intra-personal comparisons of inter-personal utility trade-offs.

applies differently in the roles of the proposer and the responder and therefore different trade-offs may seem justified.

5.2 Inter-personal (in)coherence in ultimatum game experiments

The existence of consent – or, for that matter, homogeneity – is not supported by observations of pro-social behaviour in the laboratory. There seems to be one exception, though. In the original class of experiments of the ultimatum game type an equal split by the proposer is seen clearly as unobjectionable by practically all in the responder role. So, perhaps here we can identify some minimal moral consensus on equality? Yet, it should be noted that the original situation is framed such that the pie comes as a kind of gift. If the pie had to be earned in a preceding round of interaction then a proportionality norm would have kicked in.[60] Likewise had there been some individual with special needs an equal split might have been rejected.

The moral philosopher may want to draw attention here to an Aristotelian version of proportional assignments of which the equal splits observed form a special case. If there was no preceding round of interaction in which the pie was rendered available both actors were equal in their (then zero) contributions. By imposing anonymity the individuals were made artificially equal in all other regards. Therefore proportionality would suggest an equal split of the pie as a special case of a proportionality norm, or so the argument might run.

In the more general case the dimensionality of the problem would have to be fixed. Is merit, i.e. the effort in contributing to the common work of generating the pie or is rather proportionality of, say, need the crucial factor? What about a host of other value dimensions and subsequent explanations that could apply and might be elicited by appropriate experiments?

[60] See Hoffman et al. (1994)

There is a great heterogeneity of justice related behaviour. In view of this observed heterogeneity of justice related behaviour we should at least prima facie suspect that any alleged hidden consensus of inter-subjectively and in this stronger sense generally accepted normative principles is lacking.

5.3 Behavioural heterogeneity in ultimatum game experiments

If one looks at actual raw data rather than statistical aggregates that often conceal rather than reveal it, heterogeneity is all over the place. On average, behaviour may be of a certain kind and the averages may be similar across time and place. However, for those who, as the contractarians do, emphasize respect for the separateness of persons averages do not matter. Individuals do matter and it is a normatively most relevant fact for assent based ethics that there are distinct types of behaviour because at least prima facie they indicate heterogeneity of "deeper" normative orientations.[61]

We believe that for the moral philosopher in general and the applied ethicist in particular the demonstration of widespread heterogeneity is the most relevant lesson from experiments of the ultimatum bargaining type. There is pro-social behaviour but no homogeneity of the type of that behaviour unless artificially created. If ethicists seek to find agreement outside the Frohlich and Oppenheimer setting and outside their rational consent models they will seek in vein. For, if homogeneity of ethical views in very basic and simple matters of justice and equity does not show itself in homogeneous behaviour even in simple ultimatum bargaining experiments where else? If it is still claimed that seeming disagreement merely conceals a deeper agreement then additional experimental research can perhaps decide the issue and ethics can learn even more from experimental economics.

[61] Heterogeneity of retributive responses shows up pretty strongly in Güth et al. (2001).

6. Conclusions on bounded justice

In developing a theory of justice Rawls should have taken his own empiricism more seriously. Then he might have looked at real actions of real people in justice related situations. However, the project of "a theory of justice" was too ambitious for this. Bringing in empirical facts is easier within a "local justice" approach.[62] In line with such an approach we explored a very specific case: the relationship between efforts to develop a boundedly rational theory of justice and normative facts as appear in the well-known ultimatum game experiments. Thereby the search for reflective equilibrium is anchored not in pseudo-empirical arm chair judgements derived introspectively. Moreover, it is not merely based on what people say they would do but on results about what they do or have in fact done when confronted with justice or equity related real choices in ultimatum bargaining experiments.

We do not deny that moral language articulates justice claims. We also agree that this is in itself a normative fact. However, it is a fact concerning language use and *expressive* habits. As such it is directly predictive for expressive behaviour as in voting only.[63]

If we go beyond the stage where "my say so" is pitted against "your say so" and look at "your do so" as opposed to "my do so" we are addressing complementary issues of behaviour in situations with high opportunity costs. Here we agree that "(s)ince 'actions speak louder than words,' the information conveyed by actions may also be the most credible" (Bikhchandani et al. (1992)). The moral theorist should set out to articulate the normative convictions that might be guiding the justice related actions that are actually observed. As

[62] Elster (1992). Using Volcker Schmidt's apt term, we could speak of a "bounded justice" approach, Schmidt (1994). We use "local" in a broader sense including specification along several dimensions like time, space, context.

[63] See Brennan and Lomasky (1993)

the ultimatum game experiments show, the ethical theorist will find less agreement in the world than the official wisdom would have it.

At least if ethical theorists would use revealed along with stated normative principles in their search for (W)RE there is no hope of inter-personal agreement. Within the broadly speaking sceptical tradition of normative argument – reaching from Hume to Mackie – this result is not too disturbing. For this tradition, all moral argument is ultimately "agent-relative" or "to whom it concerns".[64] If individuals have different concerns they will come up with different views on justice and act differently. However, for the sceptical ethical theorist it makes a difference whether an argument concerns many or few and whether it relates to deeper or more superficial concerns of its addressees. Though arguments from fictitious or conceivable consent may be irrelevant, the factual consent of as many individuals as possible can be most relevant. And, this is an empirical matter in which ethics can learn a lot from experiments along with other empirical research.

Ainslee, G. (2002): Break Down of the Will. Princeton.

Albert, H. (1978): Traktat über rationale Praxis. Tübingen.

Albert, H. (1985): Treatise on Critical Reason. Princeton.

Alchurron, C. and Bulygin, E. (1971): Normative Systems. Berlin et al.

Barry, B. (1965): Political Argument. London.

Barry, N. (1981): An Introduction to Modern Political Theory. London and Basingstoke.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992): A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy, 100(No. 5. (Oct.)), 992-1026.

[64] see Hume (1739/1978), Mackie (1980), Mackie (1977), Harman (1977).

Binmore, K. (1994): Game Theory and Social Contract Volume I - Playing Fair. Cambridge, London.

Binmore, K. (1998): Game Theory and Social Contract Volume II - Just Playing. Cambridge, London.

Binmore, K. (2005): Natural Justice. New York.

Binmore, K. and Shaked, A. (2007): Experimental Economics: Science or What?, London and Bonn, pp. 37.

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

Brennan, G. and Lomasky, L. (1985): The impartial spectator goes to Washington. Economics and Philosophy, 1, 189-211.

Brennan, H. G. and Buchanan, J. M. (1985): The Reason of Rules. Cambridge.

Brennan, H. G. and Lomasky, L. E. (1989): Large Numbers, Small Costs - Politics and Process - New Essays in Democratic Thought. Cambridge.

Brennan, H. G. and Lomasky, L. E. (1993): Democracy and Decision. Cambridge.

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

Camerer, C. (2003): Behavioral Game Theory. Princeton.

Carson, R. T. and Groves, T. (2007): Incentive and informational properites of preference questions. Envirinmental Resource Economics, 37, 181-210.

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

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

Daniels, N. (1996): Justice and Justification. Cambridge.

Elster, J. (1992): Local Justice. How Institutions Allocate Scarce Goods and Necessary Burdens. New York.

Elster, J. (Ed.), 1998. Deliberative Democracy. Cambridge University Press, Cambridge.

Fehr, E. and Gächter, S. (2002): Altruistic Punishment in Humans. Nature, 415(January), 137-140.

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

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

Fleck, L. (1935/1980): Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv, vol. 312. Frankfurt.

Frank, R. (1987): If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience? The American Economic Review, 77/4, 593-604.

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

Frankena, W. K. (1966): Some Beliefs about Justice. Lawrence. Kansas.

Frellesen, P. (1980): Die Zumutbarkeit der Hilfsleistung. Frankfurt/M.

Frohlich, N. and Oppenheimer, J. A. (1992): Choosing Justice. An Experimental Approach to Ethical Theory. Berkeley et. al.

Gehrig, T., Levati, V., Levínský, R., Ockenfels, A., Uske, T. et al. (2007): Buying a pig in a poke: An experimental study of unconditional veto power. Journal of Economic Psychology, 28, 692-703.

Goodman, N. (1978): Fact, Fiction and Forecast. New York.

Güth, W. (2000): Boundedly Rational Decision Emergence - A General Perspective and some Selective Illustrations. Journal of Economic Psychology, 21, 433 – 458.

Güth, W. and Kliemt, H. Bounded Rationality and Theory Absorption. Homo Oeconomicus, 21((3/4)), 521-540.

Güth, W., Kliemt, H. and Ockenfels, A. (2001): Retributive Responses. Journal of Conflict Resolution, 45(4), 453-469.

Güth, W. and Levati, V. (2007): Listen: I am angry! An experiment comparing ways of revealing emotions, Jena Economic Research Paper.

Güth, W., Schmidt, C. and Sutter, M. (2003): Fairness in the Mail and Opportunism in the Internet - A Newspaper Experiment on Ultimatum Bargaining. German Economic Review, 4(2), 243-265.

Güth, W., Schmidt, C. and Sutter, M. (2007): Bargaining outside the lab - A newspaper experiment of a three person ultimatum game, The Economic Journal, pp. 449-469.

Güth, W., Schmittberger, R. and Schwarze, B. (1982): An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior and Organization, 3, 367-388.

Güth, W. and Tietz, R. (1990): Ultimatum bargaining behavior - A survey and comparison of experimental results. Journal of Economic Psychology, 11(3), 417-449.

Hahn, S., 1998, Reflective Equilibrium-Method or Metaphor of Justification? Schriftenreihe der Wittgensteingesellschaft. Hölder-Pichler-Tempsky, Wien, pp. 237-243.

Hahn, S. (2000): Überlegungsgleichgewicht(e). Prüfung einer Rechtfertigungsmetapher. Freiburg i.Br.

Hardin, R. (2007): David Hume: Moral and Political Theorist. Oxford.

Hare, R. (1973): Rawls' Theory of Justice. Philosophical Quaterly, 21(April and July), 144-155, 241-252.

Harman, G. (1977): The Nature of Morality. An Introduction to Ethics. New York.

Hart, H. L. A. (1961): The Concept of Law. Oxford.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. et al. (2004): Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. New York.

Hoerster, N., 1977, John Rawls' Kohärenztheorie der Normenbegründung. In: Otfried Höffe (Ed.), Über John Rawls' Theorie der Gerechtigkeit. Suhrkamp, Frankfurt a. M., pp. 57-76.

Hoffman, E., McCabe, K., Sachat, K. and Smith, V. (1994): Preferences, property rights and anonymity in bargaining games. Games and Economic Behavior, 7, 346–380.

Huck, S. and Oechssler, J. (1999): The indirect evolutionary approach to explaining fair allocations. Games and Economic Behavior, 28, 13-24.

Hume, D. (1739/1978): A Treatise of Human Nature. Oxford.

Hume, D. (1964 (Reprint of the new edition London 1886)): A Treatise of Human Nature and Dialogues Concerning Natural Religion, vol. 1. Darmstadt.

Kant, I. (1798/1977): Die Metaphysik der Sitten, vol. VIII. Frankfurt.

Kliemt, H. (1985): Moralische Institutionen. Empiristische Theorien ihrer Evolution. Freiburg.

Kliemt, H. (1986): The Veil of Insignificance. European Journal of Political Economy, 2/3, 333-344.

Kliemt, H. (1994): The calculus of consent after thirty years. 79, 341-353.

Kliemt, H. (2004): Contractarianism as Liberal Conservatism: Buchanan's Unfinished Philosophical Agenda. Constitutional Political Economy, 15(2), 171-185.

Konow, J. (2003): Which Is the Fairest One of All?

A Positive Analysis of Justice Theories. Journal of Economic Literature, XLI(December), 1188-1239.

Kuhn, T. (1962): The Structure of Scientific Revolutions.

Lakatos, I. (1978): The Methodology of Scientific Research Programmes. Cambridge.

Lehrer, K. and Wagner, C. (1981): Rational Consensus in Science and Society. Dordrecht.

Levitt, S. D. and List, J. A. (2007): What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives, 21(2), 153-174.

Louvierre, J. J., Hensher, D. A. and Swait, J. D. (2000): Stated choice mehtods: analysis and application. Cambridge.

Mackie, J. L. (1977): Ethics. Inventing Right and Wrong. Harmondsworth.

Mackie, J. L. (1980): Hume's Moral Theory. London.

Mackie, J. L. (1982): Morality and the Retributive Emotions. Criminal Justice Ethics, 1982, 3-10.

Manski, C. F. (2002): Identification of decision rules in experiments on

simple games of proposal and response. European Economic Review, 46, 880-891.

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

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

Parsons, T., 1968, Utilitarianism. Sociological Thought. International Encyclopedia of Social Sciences, New York und London.

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

Rawls, J. (1974): The Independence of Moral Theory. Proceedings and Addresses of the American Philosophical Association, 48, 4-22.

Robbins, L. (1935): An Essay on the Nature and Significance of Economic Science. London.

Roth, A. E., 1995, Bargaining Experiments. In: John H. Kagel and Alvin E Roth (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 253-348.

Schelling, T. C. (1984): Choice and Consequence. Cambridge, MA.

Schmidt, V. H. (1994): Bounded Justice. Social Science Information, 33(2), 305-333.

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

Simon, H. A. (1985): Models of Bounded Rationality (1&2). Cambridge, MA.

Skyrms, B. (1990): The Dynamics of Rational Deliberation. Cambridge.

Skyrms, B. (1996): Evolution of the Social Contract. Cambridge.

Spitzley, T. (1992): Handeln wider besseres Wissen. Eine Diskussion klassischer Positionen. Berlin/New York.

Spitzley, T. (Ed.), 2005. Willensschwäche. Mentis, Paderborn.

Strotz, R. H. (1955): Myopia and Inconsistency in Dynamic Utility Maximization. Review of Economic Studies, 23, 165 ff.

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

Sugden, R. (2004): What Public Choice and Philosophy Should *Not* Learn From Each Other. American Journal of Economics and Sociology, 63(1), 207-211.

Suleiman, R. (1996): Expectations and fairness in a modified ultimatum game. Journal of Economics Psychology, 175, 531–554.

Taylor, M. (1976): Anarchy and Cooperation. London u. a.

Thaler, R. H. and Shefrin, H. M. (1981): An Economic Theory of Self-Control. Journal of Political Economy, 89/2, 392 ff.

Westermarck, E. (1906): The Origin and Development of Moral Ideas I and II. London.

Xiao, E. and Houser, D. (2005): Emotion expression in human punishment behavior. Proceedings of the National Academy of Sciences, 102(20), 7398-7401.