Approximate Truth in Economic Modelling

by Geoffrey Brennan*, Werner Güth** and Hartmut Kliemt

November 30, 2004

Abstract (94<100 words)

Economic intuitions concerning rational behaviour in interactive social situations are shaped by idealized models which are regarded as “approximately true”. But ideal models cannot be meaningfully deemed approximately true unless asymptotically convergent processes imply them as limit cases. We illustrate by various examples – infinitely patient customers on durable monopoly markets, homogeneity of commodities, super-games etc. – how this *necessary* methodological requirement is almost routinely neglected. On this basis we draw some conclusions concerning the continuity between abstract and less abstract models on the one and the world modelled by them on the other hand.

*Economics RSSS, Australian National University, Canberra, **Max Planck Institute for Research into Economic Systems. Strategic Interaction Group. Kahlaische Strasse 10, D-07745 Jena, Germany, *** Philosophy-Department, University of Duisburg-Essen

1. Introduction and overview

The social sciences, according to a variety of commentators (of whom Max Weber is perhaps the most influential), necessarily involve the imposition of a certain view *of* the world *on* the world. The attempt to develop an account of social phenomena that has broad explanatory scope and power necessarily involves some strategic abstraction from reality. Social theorising requires a big canvas approach, in which matters of detail must necessarily be set aside in the interests of the larger picture.

Sometimes, the abstractions in question are developed in an explicitly speculative, "as if" mode. That is, the approach involves an exploration of the implications of deliberately 'unrealistic' assumptions to see just how illuminating the exploration can be. Examples of this approach within economics might include:

- (Harsanyi and Selten 1988) investigation of the logical implications of "rationality and common knowledge";
- (Arrow and Hahn 1971) objective of understanding what is "conceivable" in principle and what not;
- the desire to spell out the *a priori* logic of human action, as in (Mises 1949/1966);
- the analysis of the emergence of order in general evolutionary settings (see (Alchian 1950)) as well as in neighbourhood structures (see (Eigen and Winkler 1975; Schelling 1978));
- the exploration of what might be the outcome of a unanimous social contract in society at large (see (Buchanan and Tullock 1962); and (Brennan and Buchanan 1985)).

On this approach, the relation between model and reality should be conceived less in terms of, say, a painting by Leonardo, and more as an attempt to reveal essential traits of "reality" in the manner, say, of the elder Picasso. Illuminating abstraction rather than true representation is the name of this "economic modelling game".

To play this game makes perfect sense in some contexts. But to formulate illuminating abstractions is not the only game in the economists' town.[1] Sometimes the object is just to make the minimal trade-off between realism and other valued properties of a proper 'scientific' account – such as simplicity, and breadth of explanatory power.

Consider for example the assumption, central to the 'economic approach', of agent rationality. When confronted with the many deviations from rational behaviour, both in real life and in economic laboratory experiments, most adherents of strict neo-classical rational choice modelling will not dispute the evidence. They will concede that the rationality assumption is strictly speaking false. Nevertheless, neo-classical economists will give up neither rational choice modelling nor the focus on rational expectations equilibria in predicting outcomes. They believe that it is possible to rescue the rationality assumptions by claiming, either that the assumptions, though false, are useful instruments of "prediction" or that they are "approximately true".

[1] For instructive examples from other fields that illustrate how far distortions might go (see Fleck, L. (1935/1980), while for criticisms of "Platonism in economic modelling" (see Albert, H. (1998) and as a relevant selection more recently (Mäki, U. (2002).

In defence of this view, many economists point out that physics, for instance, uses extreme idealizations like frictionless motion even though there is no absolutely frictionless motion. Physics does this because it is useful to formulate ideal models under this assumption. In view of the success of physics, the economist asks why his own discipline should not be entitled to rely on similar abstractions. On the face of it the argument seems plausible. However, the analogy between physics and economics applies only within limits. In physics it is in general believed that approximation is dealing with models that represent the true causal mechanisms leading to approximately true predictions and explanations. But most neo-classical economists agree that behaviour can at best be described *as if* being rationally motivated, while admitting that the true springs of action are significantly different. They tend to lump together the "approximate truth of predictions" reached on the basis of speculative "as if" models, with the approximate truth of predictions derived from models that claim to be well-specified as basically correct representations of underlying causal factors.

If models of perfect rationality – except for some mistakes that humans inevitably make – would represent the true psychological factors explaining behaviour, this would be rather different from a situation in which approximately true predictions are derived on the basis of regularities that are acknowledged to be false. We regard the old fashioned Friedman type instrumentalism (as expressed in (Friedman 1953)) underlying such views to be methodologically inadequate; but will not further comment on Friedman's views here. Suffice it to note that such claims are distinct from claims to "approximate truth". In the latter case, it is assumed that rational choice as modelled is in fact basically operative: the neo-classical rationality assumptions are taken to describe the true forces that cause behaviour. Any discrepancies between model and reality are due to distorting factors but not to a deliberate – though allegedly constructively useful – mis-specification of the model.

We think that most economists would want to hedge their bets here – and place at least some of their reliance on the "approximate truth" defence. But as a general matter the usefulness of approximations in model specification depends on an underlying continuity in the mapping from model to prediction. That is, for approximation to be unproblematic, small changes in assumptions must involve correspondingly small changes in predictions. In particular, predictions for the idealisation must be equal to the limit of predictions as assumptions approach the ideal-type case. In fact, however, very slight changes in assumptions seem to generate very significant changes in predictions. There is an excessive sensitivity that comes into play as the ideal type of complete rationality is approached. That being so, the notion of "approximate truth" in relation to the rationality assumption deserves a fuller treatment. Such is the aim of the present paper.

More specifically we aim to illustrate by appeal to several examples that some of the most influential economic models show that oversensitivity to small variations that makes the claim to approximate truth very precarious. Relying on ideas originally developed by Reinhard Selten (see (Selten 1965), (Selten 1975)) we intend to take the adherents of strict rational choice modelling at their word and to scrutinize the idea of "approximation" in the context of rational choice modelling. We start in Section 2 with some simple models of market processes. After demonstrating the discontinuity between finite and infinite degrees of customer patience in durable monopoly situations (2.1), we focus on the case of the relationship between imperfect and perfect homogeneity of products (2.2). In section 3, we turn to examples drawn from game theory. We start with very simple stage games (3.1.) and use so-called super-game analyses and the famous (or, depending on the point of view, notorious) Folk theorem whose view of infinite horizon games cannot meaningfully be seen as approximations of games with long, yet finite horizons (3.2.). We finally demonstrate – again by way of a specific example – that speaking of "approximate truth" in complex models with multiple causal mechanisms may be almost meaningless since, depending on the sequence of taking limits, completely different structures may be selected (3.3). We take the general thrust of these results to be rather discomforting for the "approximate truth" defence of rationality assumptions. In section 4, we draw some tentative, more general conclusions on issues of idealization, approximation and (mis-)specification.

2. Examples of extreme market idealizations

Idealized models of choice making can be conceived as being embedded in a generic class of models. "Around" each model there is a neighbourhood of structurally similar models that are "close by" in a rather precise sense. Mathematical concepts can characterize what it means to approach the limiting model. If in considering a sequence of models nothing qualitatively new emerges when we take the limit, we have one precise justification for claiming that the ideal model *can* be approximately true of the broader class. In metaphorical terms, the broader class forms a theoretical net with which we can hope to catch some real fish or a glimpse of a reality that is necessarily only approximated by the ideal.

2.1. Asking for some impatience

Consider a monopoly (see (Coase 1972) that persists for more than one period. The value $v \in (0,1)$ that the customer attributes to the indivisible and unique commodity is uniformly distributed. The monopolist is facing constant marginal costs which, without loss of generality, can be set to 0. Choosing appropriate units of measurement we can represent demand on the market by $X(p) = \int_p^1 dv = 1 - p$. Neglecting fixed costs as strategically irrelevant (i.e. as "sunk"), profit is given by $\pi(p) = pX(p) = p(1-p)$. As implicitly assumed in conventional analyses the monopolist first sets the price and the buyer cannot respond strategically to the price charged. This obviously implies the standard profit maximizing monopoly price $p = 1/2$ and the monopoly profit $\pi = 1/4$.

Assume now that the preceding situation is repeated but that a customer buys at most one unit. There are several decisions to be made and interdependence of decisions rather than separability of the decisions might apply. In each period the monopolist must first announce his price. Then the buyer, who has not bought yet, has to decide on whether or not he will buy in that period at that price. However, when considering this, the buyer – like the seller – should be aware at any period t –

other than the final one – that there will be later periods. If he buys only once why should he not hope for a lower price later on? After all, a strategy of price differentiation (think of the cheaper paperback of an expensive earlier hardcover book) would suggest that the monopolist would try to satisfy buyer types with $v \geq p_1 = 1/2$ in period 1 and then go on to satisfy the other types in period 2 at a lower price. But what if the buyer types who otherwise would buy in period 1 would understand this? Would they not then wait until period 2 so as to buy at the lower price?

If we confine ourselves to exactly two sales periods the market decision process is as follows:

- First, the seller, not knowing v, states her price p_1 for period 1, which the customer can either accept or not. If the customer accepts, his gain is $v - p_1$ and that of the seller p_1 (assuming that her costs are 0). After a deal the game ends for that customer. Otherwise the game continues in period 2 with that customer.

- Second, in period 2 the seller, still not knowing v, knows that the customer did not buy in period 1. She states her sales price p_2 for period 2 which the customer can accept or not. If so, the customer receives $\delta(v - p_2)$ and the seller ρp_2 with δ and ρ being the customer's, resp. the seller's discount factor satisfying $0 < \delta, \rho \leq 1$. If not, both earn 0.

Notice that the seller will not increase her price (if she would do, she could only sell in period 1 at the monopoly price of ½ but then she could gain by setting p_2 lower than the monopoly price). For $p_1 \geq p_2$ the customer with value $v \geq p_2$ prefers to buy in period 1 rather than in period 2 if

$$v - p_1 \geq \delta(v - p_2) \text{ or } v(1-\delta) \geq p_1 - \delta p_2.$$

In the generic case of $\delta < 1$ this determines the upper range of v leading to trade in period 1

$$v \geq v^* = \frac{p_1 - \delta p_2}{1 - \delta}$$

and the lower range of v leading to trade in period 2

$$v^* > v \geq p_2$$

For $v < p_2$ there is no trade.

Knowing v^* one can easily derive equilibrium play. First solving the generic $\delta < 1$-cases and then determining the limit of these solutions for $\delta \to 1$ (see (Güth and Ritzberger 1998) leads to definite results. If one studies, however, the limit case $\delta = 1$

directly the critical value v^* separating the two neighbouring value ranges leading to trade in period 1, resp. 2, is not defined at all. With time-indifference, i.e. infinite patience, any price difference will induce all customers with $v \geq \min\{p_1, p_2\}$ to switch instantaneously to buying at the lower price while for $p_1 = p_2$ it is completely undefined who of those with value $v \geq p_1 = p_2$ buys when.

In the preceding model of a two period durable monopoly, the required continuity between the slightly distorted and the fully idealized model is lacking. It is in fact completely unclear what the conclusion from the extreme idealization should be. Unless an asymptotically convergent process towards the limit in the slightly distorted models "close" by is considered, we cannot even say within the logic of those models what the conclusions might be. Assuming perfect patience $\delta = 1$ of customers is not an innocuous "simplifying assumption" that we might make preliminarily to reach some first "rough conclusions". The assumption is far from harmless. We cannot infer anything from it about the likely behaviour of customers who are imperfectly patient. Only a discussion of the class of models with imperfect patience can help us here. Moreover, assuming that as a matter of fact some deviation between ideal and real always applies, only the common structure of the broader class of imperfect patience models can be representative of the structure of the real world. An even greater challenge is posed by the still more extreme case $\delta = 1$ and infinitely many successive sales periods as postulated by the "Coase-Conjecture" (see (Coase 1972). But we will discuss the problem of multiple limits later. Let us first consider another more simple example to get a better feeling for what is at stake.

2.2. From product heterogeneity to homogeneity

Otherwise homogeneous commodities that are traded in different periods can be treated as different commodities distinguished by time. For the durable monopoly case, it has already been indicated how by increasing patience, the limit case of perfect homogeneity of goods is approximated since buying in period 2 becomes a perfect substitute for buying in period 1. In conventional markets for products that are asymptotically homogeneous along dimensions other than time roughly the same argument applies.

For the symmetric heterogeneous market with linear demand functions[2]

$x_i = 1 - p_i + \gamma\left(\overline{p}_i - p_i\right)$ with $\gamma > 0$

for the sellers $i = 1,...,n$ with individual sales prices p_i and other sellers' price average

$\overline{p}_i = \sum_{j \neq i} p_j / (n-1)$ for $i = 1,...,n$

---

[2] Linearity may, of course, put non-negativity of x_i at risk. This can be avoided by appropriately restricting the price range.

homogeneity can be approximated by $\gamma \to \infty$. For sufficiently large γ a small price discrepancy $\overline{p}_i \neq p_i$ will transfer all the demand to seller i (if $p_i < \overline{p}_i$) or will let seller i lose all his demand (if $p_i > \overline{p}_i$).

If all n sellers face the same constant unit cost c with $1 > c \geq 0$ profits are given by

$$\pi_i = (p_i - c)x_i.$$

The market equilibrium prices are easily derived as

$$p^*_i = \frac{1+(1+\gamma)c}{2+\gamma} \text{ for } i = 1,...,n.$$

Since

$$\lim_{\gamma \to \infty} p^*_i = c \text{ for } i = 1,...,n$$

we have derived the well-known "price=marginal cost" textbook result of micro-economic theory for homogeneous goods markets.

It is helpful to compare the preceding analysis with conventional textbook treatments of quantity competition on homogeneous markets. Conventionally it is argued that prices cannot differ and that total demand X is a function of a uniform sales price p, i.e.[3]

$$X = X(p) = n(1-p).$$

Sellers $i = 1,...,n$ allegedly choose quantities $x_i \geq 0$ rather than sales prices. This might conceivably be plausible under certain institutional designs similar to a stock exchange. But quantity fixing is clearly not the prevailing method on ordinary markets. We would rather expect offers with stated prices.[4] But the typical textbook approach takes quantity competition in case of homogeneity for granted. Assuming again that all n sellers have the same constant unit cost c with $0 \leq c < 1$ profits are given by

$$\pi_i = (p - c)x_i \text{ for } i = 1,...,n \ (\geq 1).$$

[3] For equal prices $(p_i = p$ for $i = 1,...,n)$ one has $p_i = \overline{p_i}$ for all $i = 1,...,n$ so that total demand X becomes $X = nx_i = n(1-p)$ according to the individual demand equations above.

[4] Kreps, D. M. and J. Scheinkman (1983) have provided a rather plausible justification for assuming quantity competition. $x_i \geq 0$ is seller i's sales capacity. After deciding on their capacities x_i, the sellers i must choose sales prices knowing that they will have to produce within the constraints of their production capacity. This can also be assumed for heterogeneous markets with $0 < \gamma < \infty$ (see Güth, W. (1995) who views capacities not as rigorous sales restrictions but rather as thresholds beyond which delivery is decisively more costly).

Now market clearing implies

$$X(p) = n(1-p) = \sum_j x_j$$

and thus the inverse demand function

$$p = 1 - \sum_j x_j / n$$

such that

$$\pi_i = \left(1 - c - \sum_j x_j / n\right) x_i \text{ for } i = 1, ..., n.$$

Inserting the equilibrium sales

$$x^*_i = \frac{n(1-c)}{n+1} \text{ for } i = 1, ..., n$$

into the inverse demand function yields

$$p^* = \frac{1+nc}{n+1}$$

which, contrary to our former analysis, shows that $p^* > c$ (due to $c < 1$).

Considering homogeneous markets as limit cases is helpful. Finding solutions for the heterogeneous market games and then determining the limit solution for vanishing heterogeneity of the products yields results other than those of the typical textbook approach. And the results, so derived, are the only ones that should be regarded as convincing, if we intend to think of the ideal of full homogeneity as an approximately true limiting case of heterogeneity.

3. Limits of rational play in games

One might think that arguments like the preceding merely show the superiority of a game theoretic, strategic approach over more conventional ones. We think that there is something to be said in favour of a thorough going game-theoretic foundation of market theory. But this advantage notwithstanding, game theory itself is not impervious to basic problems of continuity between ideal and real. On the contrary, fully rational behaviour and full rationality in thinking about behaviour in game theory must be embedded in a “neighbourhood” of “perturbed” structures as well. Otherwise, we may be led astray as badly in game theory as in the conventional analyses discussed so far.

Since game theory is in itself a family of theories, we restrict ourselves to non-co-operative game theory as the most precise type of rational choice modelling. If we intend to relate solutions of a stylized game model to the real world we should be reasonably sure that solution behaviour would not qualitatively depend on the most minute details of modelling. Therefore solutions can be convincing only if we can find a generic class of "slightly" distorted (perturbed) games whose solutions converge to that of the ideal as a limiting case. If by *a priori* formal reasoning, qualitatively different results can be shown to emerge under small variations of assumptions we should not speak of approximate truth of the ideal model at all. In that case *a priori* reasoning shows that continuity between real and ideal must be lacking or at least be extremely precarious (and this argument applies independently of the *a posteriori* issue of mis-specification of the model as a whole). Seen in this light, the requirement that the game solution can be reached by an asymptotically convergent process is a *necessary* (not a sufficient) condition for regarding an idealized model (or its implications) as approximately true.

3.1. Approaching rationality

Imagine that you are playing the trivial one player game of Figure 1. The obvious backward induction solution is (L_1, R_2, L_3) implying the solution play of L_1.

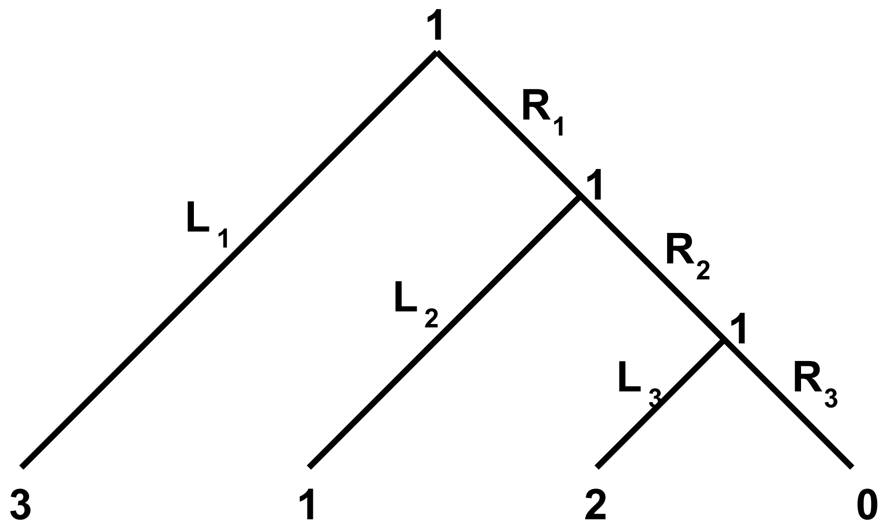

Figure 1

But assume that you are supposed to choose between L_2 and R_2. Backward induction suggests to choose R_2 in the perfect confidence of L_3 later on. But does not the fact that you have to choose between L_2 and R_2 at all render this reasoning precarious? After all you would not have to choose between L_2 and R_2 if you would be rational at all stages of the game. It seems that reaching the second decision node at all would prove to you that you are not rational. Yet if you have to conclude to be non-rational why should you trust in the rationality of your future choice making and draw any inference about the subsequent game on the basis of rational choice theory?

Allowing for perturbances (in the sense of positive minimum choice probabilities, see (Selten 1975) removes the necessity to rationally conclude that one is not rational, see also (Samuelson 2004). When analyzing the game tree one need not anymore anticipate that the conclusion of irrationality would have to be drawn if some node would be reached. If I know that I unintentionally can choose R_1 although I wanted to play L_1 I can conclude from the need to choose between L_2 and R_2 that I have chosen R_1 unintentionally if I allow for such a possibility at all. Attributing the need to decide to pure chance I should still go for R_2 and rely on my expectation that L_3 is very likely, provided that I assume that my behaviour is approximately rational, i.e. that the likelihood of perturbances to occur is small. If ε with $0<\varepsilon<1/2$ is the probability to choose unintentionally R_3 the expected payoff of choosing R_2 with $2(1-\varepsilon)$ still exceeds the expected yield of the choice of L_2.

It is obvious that the same kind of argument would apply in games with more than one personal player like the two person game of figure 2 below.

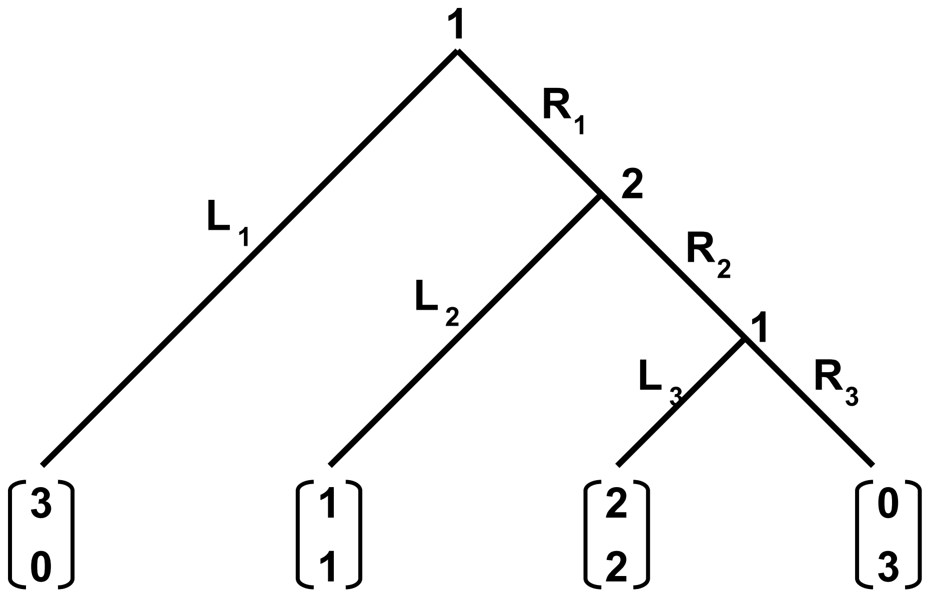

Figure 2

Player 2 has to decide between L_2 and R_2. If only approximately rational behaviour is assumed and it is assumed that rational players would endorse that assumption as well then player 2 need not infer from the fact that he is to move that player 1 is not approximately rational. Player 2 can attribute the fact that he must decide at all to 1's unintentional choice of R_1. As long as players attribute to each other merely a kind of rationality that implies a positive minimum choice probability for deviations from rationality no dramatic inferences about the rationality of another player must be drawn from observed deviations.

The preceding illustrates that backward induction solutions are at least not logically incoherent for a perturbed game. Using asymptotic convergence $(\varepsilon \to 0)$ the backward solution emerges also for the limit, i.e. the unperturbed game. In that sense the argument from backward induction is not only consistent but could conceivably be approximately true. That it in all likelihood is as a matter of fact not true (who could reason backward for instance when playing chess?) is due to specification problems rather than any *a priori* aspects of the models on which backward induction logic is based.

Still, if we let the numbers of stages of a game increase and if it is unreasonable that later stages are reached at all it becomes increasingly implausible that they are reached by mistake. In contexts in which the number of stages of a stage game itself goes to infinity the probability that later rounds are reached by mistake would converge to zero. Nevertheless, as long as we stick to the issue of logical possibility, then for any finite number of repetitions there is a positive probability in any perturbed game that a later stage of the game will be reached unintentionally.

Again, one might say, even if it is logically possible that later stages be reached in a long series of mistakes the fact that they are reached would cast strong doubts on the specification of the model as a whole. But if that is so it should be anticipated when forming the game model. Rational players would themselves take into account the possibility that their model of the game as a whole, as well as of the rationality of players in particular, may not be approximately true. They would deem it possible that the model is mis-specified and more fundamental deviations from its assumptions occur.

Though the latter can be taken into account by a specification of the model with appropriate sets of player types etc. we need to address the question whether forming a sequential game by repeating an identical base game is a meaningful idealization at all. After all, the history of the play of a game matters and might influence the specification of later sub-games. Identical repetition of a base game assuming that a payoff-identical game is played regardless of any previous history is in itself an extreme idealization. For this to be the case we must assume that game theoretic payoffs that represent preferences of actors “all things considered” will not be changed over the course of time.[5] But even if we would grant the idealization of an identical repetition of a base game, a still more fundamental problem would be encountered. For it is impossible that a game formed by finite numbers of repetitions of a base game becomes approximately like a game created by repeating the base game indefinitely. We think that it is instructive to see in more detail why this is impossible and why therefore approximation fails in the case of so-called Folk-Theorems.

3.2. Limits of Folk-theorems

[5] The repetition of an identical game is trivially possible as long as we only look at the game form in objective payoffs. But we are dealing with games and thus with subjective payoffs here (see Weibull, J. W. (2004).

We all have only finite lives to live. Since we generally do not know when we will die we have possibly long and uncertain time horizons. It is assumed that the basic aspects of the length of the time horizons and the uncertainty of the end point may be captured by simply assuming an infinite time span. But the conclusions drawn from infinite time horizon models can approximately apply to real world conditions only if the distinction between finite and infinite can be shown to be irrelevant for the issues at hand. If not so we would have an a priori reason to discard the infinite repetition model and the conclusions derived from analyzing the model as irrelevant for purposes of providing (approximately) true explanations. Continuity between the model and real structures would be missing for a priori reasons.

To give a specific example, if the prisoners' dilemma game of figure 3 as so-called base game is repeated indefinitely the corresponding super-game emerges. In that super-game even constant play of (C_1, C_2) – where overt behaviour is completely co-operative on each round of play – is a sub-game perfect equilibrium outcome provided that a present gain of 1 payoff unit "today" is outweighed by a constant loss of 1 payoff unit starting from "tomorrow". Under this assumption grim strategies (prescribing for i=1, 2 the choice of C_i whenever (C_1, C_2) has been played before without exception and D_i otherwise) support the outcome of constant (C_1, C_2)-behaviour as a sub-game perfect equilibrium.

	C_2	D_2
C_1	3, 3	1, 4
D_1	4, 1	2, 2

Figure 3

Infinity is clearly the (upper) extreme regarding increasing, yet finite numbers of repetitions. Therefore, if the case of repeating the base game indefinitely is meant to form the limit of the sequence of cases with finite but ever increasing numbers of repetitions then nothing qualitatively new should emerge when the "limit" from finite to infinite is eventually "transgressed". However, in the generic case of finitely many repetitions solution behaviour results that is qualitatively different from the case of repeating the base game indefinitely. With any finite number of repetitions the only sub-game perfect equilibrium outcome is constant play of (D_1, D_2) from beginning to end. The asymptotically convergent solution of the super-game (with infinitely many repetitions of the base game) therefore amounts to constant play of (D_1, D_2).

The "explosion" of the set of sub-game perfect equilibrium outcomes when going from finite to infinite horizons is restricted to the limit case alone. It is ruled out if we require that for the infinite horizon games only solutions are acceptable which can be approached by solutions of finite horizon games (see, for instance, (Selten 1965) and (Güth, Leininger et al. 1991)). Comparable conclusions would follow for all solution concepts that generalize sub-game perfectness. And practically all plausible solution concepts of classical non-co-operative game theory – e.g. perfectness (Selten 1975) or sequential rationality (Kreps and Wilson 1982) – generalize sub-game perfectness. Therefore the result is fairly general.

Since game theory is about what is going on in the minds of players one might still try to rescue the argument of the Folk theorem in terms of uncertain expectations of the players. It is, so the argument runs, common knowledge that all players – except perhaps for the gods or God (for the corresponding game theory see (Brams 1983)) – will die. But it is not known when that will be. This uncertainty amounts to a probability that the game will end in each period and the complementary probability that it will go on. If the likelihood that the game will go on is constant this would indeed transform the analysis of the super-game into one that is isomorphic to that of a truly infinite super-game with constant discounting. However, it is implausible that the probability to go on is assumed to be constant among individuals who commonly know that their life span is finite. For the probability of another round of play to vanish it is not necessary that all players have the same deterministic or stochastic expectation about the end point of the game. The situation is to be classified as a case of finite repetitions if a finite upper bound for the number of repetitions is commonly known. Accepting that beyond the finite upper bound probabilities are zero the standard justification for assuming infinity regardless of the fact that all men are and know to be mortal falls to the ground.[6]

Rational behaviour in finitely repeated games with a large, unknown number of repetitions is not approximately the same as in games repeated indefinitely. No other justifications for the assumption of indefinite repetitions in a finite world seem in sight. Therefore models relying on infinite sequences of repeated base games cannot give rise to conclusions that can reasonably be held to be approximately true in the real world.

3.3. Ambiguity of "approximate truth"

Information is often imperfect. Therefore conventional 'perfectness' conditions must be approachable in the case of information as well. To see what is at stake here consider the following simple trust game of figure 4

[6] Even Ken Binmore's otherwise quite ingenious epistemic logic reformulation of "reasoning about knowledge" will not get us out of the problem eventually, but we cannot here dwell on this any further, see Binmore, K. (1992).

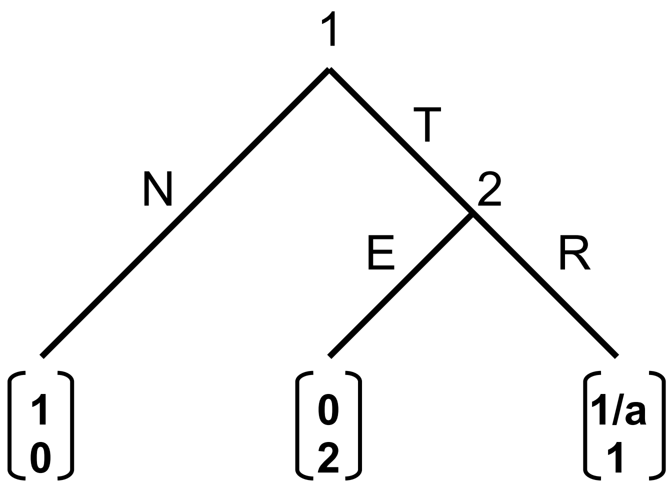

Figure 4: (0 < a < ½)

If the game is perceived by both players as depicted in figure 4 solution behaviour is obviously (N, E). However individuals might suffer from some imperfections of information that are due to incompleteness of knowledge of the game. As we know incomplete information can be captured in terms of imperfect information. Fictitious moves of mother nature can represent what players do not know (see (Harsanyi 1967-8)). The probabilities of those moves are assumed to be (in the so-called consistent case[7]) unique and commonly known.

In the case at hand, substituting Figure 4 by Figure 5 introduces a fictitious (therefore the dotted branches) chance move. This move leads into a sub-game identical with that of figure 4 with probability $1-p$. With complementary probability $p \in (0,1)$ player 2 perceives the game differently, though. Unlike player 2 in figure 4 the p-type of player 2 prefers R over E. The fictitious chance move captures player 1's incomplete information about player 2's (preference type) – presuming that player 2 himself is aware of this. Clearly, the (1-p)-type of player 2 chooses E whereas the p-type would prefer R. The obvious implication is that whenever p>a, player 1 would prefer T over N.

[7] The inconsistent case, in general, relies on individual probability assignments by all players which are commonly known, i.e. players may agree to disagree (Aumann, R. J. (1976).

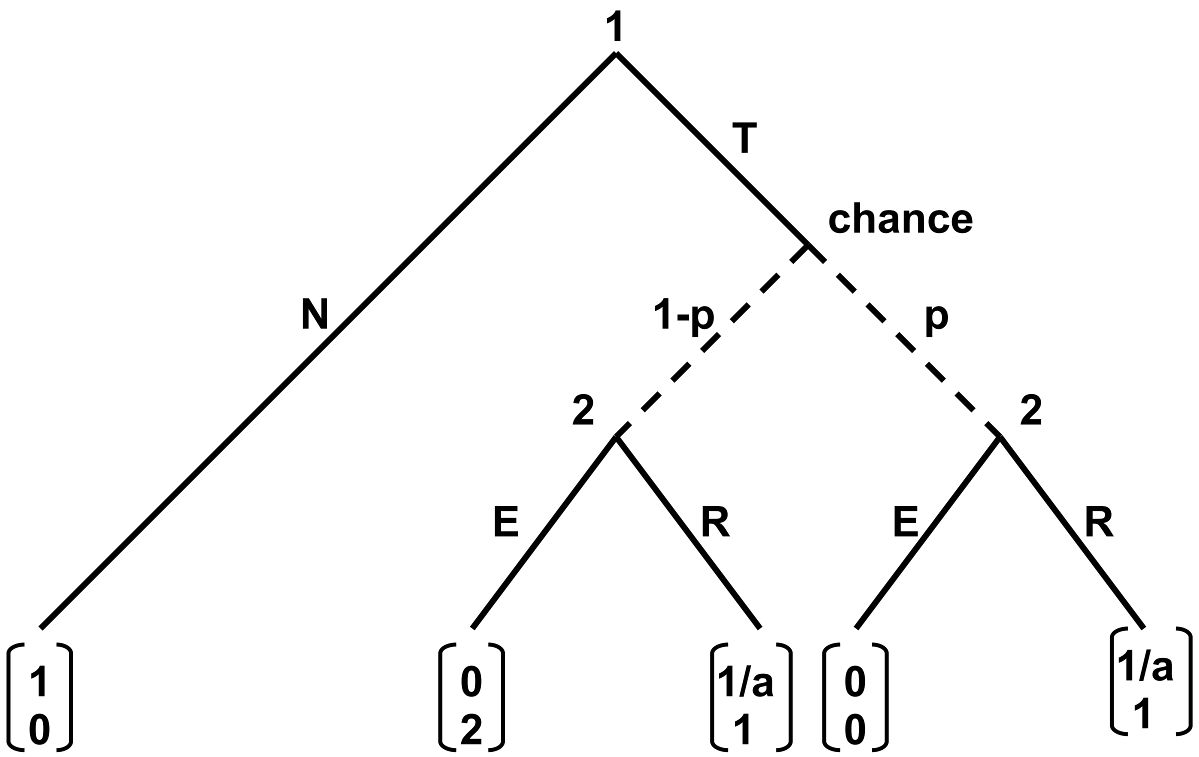

Figure 5

If perfect information is to be viewed as a limiting case of imperfect information it must be possible to approximate the case of perfect information as represented in figure 4 without changing the qualitative conclusions drawn from the model. And, letting p vanish, $p \to 0$, the asymptotically convergent solution of the game in Figure 5 is indeed (N, E). This, however, changes when we assume that the game of figure 5 is repeated. To start with the most simple case of finite repetitions let us assume that the game is repeated once over. We assume that payoffs are simply the sums of payoffs in 1st and 2nd round play and that after an E-choice by player 2, player 1 uses N (assuming that player 1 perceives the probability of an unintentional E-choice as so small as not to treat it as a mistake but as indicating that player 2 has thereby revealed[8] his (1 – p)-type). Let p_+ denote player 1's posterior probability for confronting the p-type of player 2 when player 1 actually has to choose between N_2 and T_2 in the game of Figure 6. We know that 1 will choose T_2 only if $p_+ \geq a$. But how is p_+ depending on p?

[8] Such beliefs satisfy the requirement of asymptotic convergence since (positive) trembles ε_1 resp. ε_2 for mistakenly choosing R, resp. E by the 1-p, resp. p-type of player 2 imply the posterior probability $\frac{(1-\varepsilon_2)\,p}{\varepsilon_1(1-p)+(1-\varepsilon_2)\,p}$ of confronting the p-type which converges to 1 for $\varepsilon_1, \varepsilon_2 \to 0.$

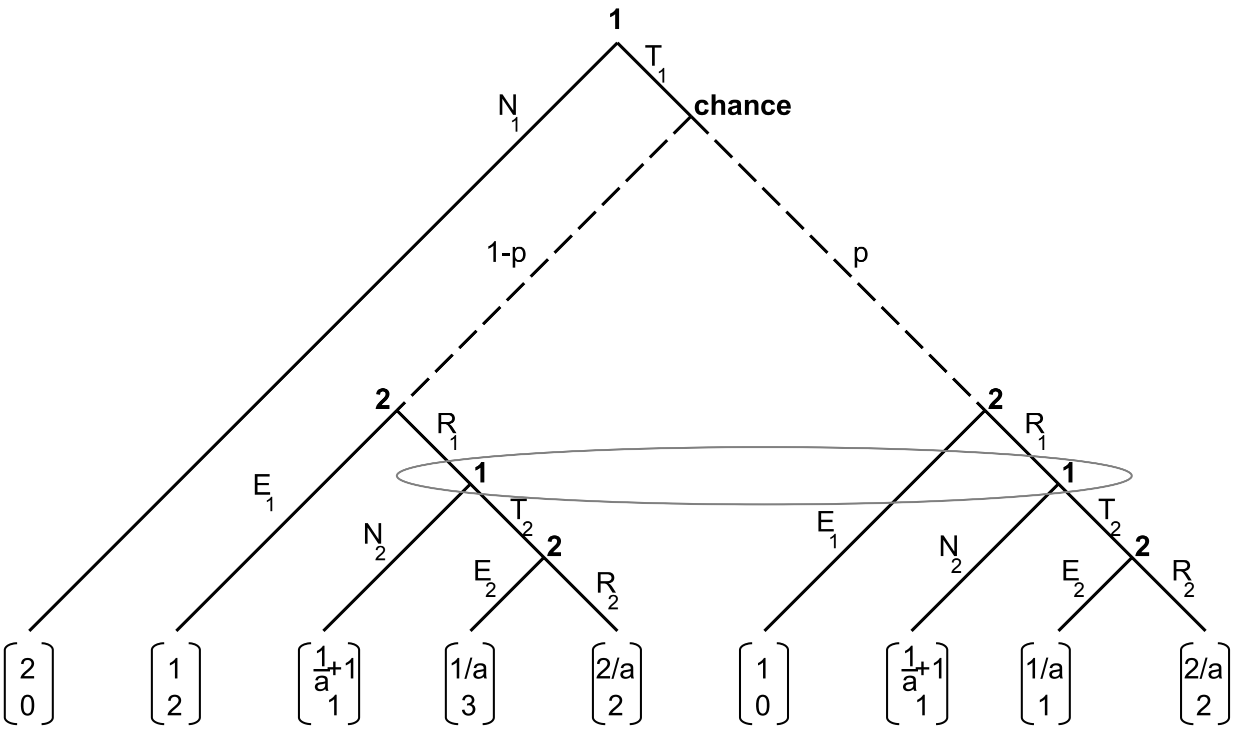

Figure 6

The p-type of player 2 chooses R_1, resp. R_2. Let us now assume that x is the (1-p)-type's probability for selecting R_1 and that player 1 has chosen T_1 with probability $y_1(>0)$. Since the probability of the left decision node in player 1's information set is $y_1(1-p)x$ whereas – assuming for the time being that the p-type chooses with certainty according to type – the right one is reached with probability $y_1 p$. We get

$$p_+ = \frac{y_1 p}{y_1(1-p)x + y_1 p} = \frac{p}{(1-p)x + p}$$

for the posterior probability p_+. The condition $p_+ \geq a$ therefore requires

$$p \geq \frac{ax}{1-a+ax}.$$

Thus only for x = 1 the old threshold $p \geq a$ is preserved. Can we actually induce a solution with x < 1 for which the p-threshold is lower than a? To explore this in more detail let p_+ = a . This implies

$$x^* = \frac{p}{1-p}\frac{1-a}{a}$$

and allows player 1 to freely choose his probability y_2 for selecting T_2. Consider now the (1-p)-type of player 2 when choosing between E_1 yielding the payoff of 2 and R_1 yielding the payoff expectation of $1 - y_2 + y_2 3$.

For indifference one needs $y_2^* = 1/2$. In turn this allows the (1-p)-type of player 2 to rely on the choice probability

$$x^* = \frac{p}{1-p}\frac{1-a}{a}$$

which is required for p_+ = a.
Now the condition $x^* < 1$ is equivalent to p < a. Thus it justifies a final phase of mixing even for 0 < p < a with the p-type of player 2 choosing

$$x^* = \frac{p}{1-p}\frac{a}{a}$$

and player 1 choosing $y_2^* = 1/2$. To guarantee that this really happens one, of course, must additionally satisfy $y^*_1 > 0$. This would require

$$(1-p)\left[1 - x^* + x^*\left(\frac{1}{2}\left(\frac{1}{a}+1\right)+\frac{1}{2a}\right)\right] + p\left[\frac{1}{2}\left(\frac{1}{a}+1\right)+\frac{1}{a}\right] \geq 2.$$

Inserting x^* yields

$$a^2 \geq p\left[\frac{7a^2+9a-2}{2} - \left(a^2+3a-1\right)2p\right]$$

The latter shows that fulfilling the requirement $y^*_1 > 0$ depends on the parameters $a \in \left(0, \frac{1}{2}\right)$ and $p \in (0,1)$.

For a given payoff parameter a the interaction of p and of the number of repetitions leads to the following (see, for instance, (Anderhub, Engelmann et al. 2002):

- For any given $p \in (0,1)$ players will initially play (T, R) until the rather short terminating mixing phase. If we let the number of repetitions approach ∞ the outcome is constant play of (T, R) – including the super-game emerging when the base game is repeated indefinitely.

- If for a given finite number of repetitions we let p vanish, the probability p will eventually become smaller than the threshold for inducing initial play of (T, R). We thus will see constant play of N regardless of how often the game is repeated.

To start with $p \to 0$ and then increasing the number of repetitions implies therefore constant play of N. For the reverse order of limit operations things are different. The result for any positive p is constant play of (T, R) when the horizon becomes infinite. There are no unambiguous asymptotically convergent solutions in such cases. In case of such models the claim of approximate truth is prevented by the inherent ambiguity that emerges once we take approximation seriously.

One could argue, though, that perfect information is a less demanding assumption than that a game is repeated indefinitely. Since in case of some "incompleteness" of information already finitely many repetitions induce initial play of (T, R), the result for the infinitely repeated game of Figure 4 should be constant play of (T, R) as implied by first letting the repetition number go to ∞ and then $p \to 0$. But even though the latter argument can be made it seems somewhat weak. In case of idealizations that have qualitatively different results depending on the way we approximate the ideal model we should be highly suspicious about the idealization as a whole. Any claim to approximate truth is vulnerable to the query as to which aspect of the truth is being approximated.

Obviously the preceding troublesome problem could emerge in other cases, since in rational choice modelling our models are typically limiting cases along several dimensions. The classification of a situation as a limit (an ideal) case of a generic one therefore depends on focusing on some dimension and on choosing a generic sub-class. In our initial examples it was always quite obvious which subclass we meant. In discussing Folk-theorems, for instance, we considered the class of all repeated games with the same base game. Thus a specific game in that class is defined by the number of repetitions. We classified the game as generic when the repetition number was larger than one but finite and as ideal when it is infinite (or just one). Similarly, durable monopolies were considered as limit cases if $\delta = 1$, homogenous markets when $\gamma \to \infty$, trust games when $p \to 0$ and super-games of the trust game when both, $p \to 0$ and the repetition number $\to \infty$ applied. More generally speaking, all so-called generic situations are limit cases when seen as members of a more comprising class of situations.

4. Conclusion

Most economists believe that they are not merely interested in speculative social philosophy but in approximately true theoretical representations of the real world. To achieve the kind of generality of explanation economists desire, abstraction and idealization will be necessary. But there must also be, first, a structural correspondence between "model and reality" and, second, a continuity between the real and the model structure. Whether or not there is a correspondence between model and reality is an empirical question to be answered with *a posteriori* methods. This latter specification issue is of course, central to all scientific work – but we have not addressed it here. Our focus has been exclusively on the second issue of continuity between different structures. As the examples we have offered demonstrate, there is a fundamental *a priori* issue of model formulation here: There must be a "continuity" between "ideal and real", and this continuity requirement is violated if even the slightest variations in the values of crucial model variables would imply qualitatively different results. More generally, all claims to approximate truth are precarious if we cannot make sense of the very concept of approximation. An ideal model must be formulated such that its conclusions remain valid in a whole class of models that are "close by". Only then can we be reasonably confident that one model of the class of *models* can at all be “close” to the real world structures represented by the common structure of the class. Unless this necessary condition for making sense of the concept of approximation is fulfilled, there is a strong *a priori* methodological reason to discard any claim to approximate truth. In view of this fact, we economists should become much more cautious when making our rather loose statements about approximation, approximate truth, the necessity of abstractions and so on. And this is not just a methodological nicety. The core commitment to rationality itself seems to be vulnerable to the “closeness” requirement in lots of cases that have been significant in guiding professional intuitions. In this sense, any defence of rationality assumptions on the grounds that they are “approximately true” seems on its face deeply problematic.

References

Albert, H. (1998). Marktsoziologie und Entscheidungslogik.Zur Kritik der reinen Ökonomik. Tübingen, Mohr.

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

Anderhub, V., D. Engelmann, et al. (2002). "An experimental study of the repeated trust game with incomplete information." Journal of Economic Behavior and Organization **48**: 197-216.

Arrow, K. J. and F. Hahn (1971). General Competitive Analysis. New York und San Francisco, Elsevier Science Publishing Company.

Aumann, R. J. (1976). "Agreeing to Disagree." The Annals of Statistics **4**(6): 1236-1239.

Binmore, K. (1992). Fun and Games - A Text on Game Theory. Lexington, D.C. Heath.

Brams, S. J. (1983). Superior Beings - If They exist, how would we know? New York / Berlin / Heidelberg / Tokyo, Springer Verlag.

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

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

Coase, R. (1972). "Durability and Monopoly." Journal of Law and Economics **15**: 143-149.

Eigen, M. and R. Winkler (1975). Das Spiel. Naturgesetze steuern den Zufall. München, Piper.

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

Friedman, M. (1953). The methodology of positive economics. Essays in Positive Economics. Chicago, University of Chicago Press**:** 3-43.

Güth, W. (1995). "A Simple Justification of Quantity Competition and the Cournot-Oligopoly Solution." ifo-Studien, **41/2**: 245-257.

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

Güth, W. and K. Ritzberger (1998). "On Durable Goods Monopolies and the Coase-Conjecture." Review of Economic Design **3**: 215-236.

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

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

Kreps, D. M. and J. Scheinkman (1983). "Quantity precommitment and Bertrand competition yield Cournot outcomes." Bell Journal of Economics **14**: 326-337.

Kreps, D. M. and R. Wilson (1982). "Reputation and Imperfect Information." Journal of Economic Theory **27**: 253-279.

Mäki, U. (2002). Fact and Fiction in Economics: Models, Realism and Social Construction. Cambridge, Cambridge University Press.

Mises, L. v. (1949/1966). Human Action. Chicago, Henry Regnery.

Samuelson, L. (2004). "Modeling Knowledge in Economic Analysis." Journal of Economic Literature **42**(2): 367-403.

Schelling, T. C. (1978). Micromotives and Macrobehavior. New York and London, Norton.

Selten, R. (1965). "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit." Zeitschrift für die gesamte Staatswissenschaft **121**: 301-324 and 667-689.

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

Weibull, J. W. (2004). Testing game theory. Advances in Understanding Strategic Behaviour. S. Huck. Basingstroke, Palgrave Macmillan, 85-104.