

# On the Limits of Rational Choice Theory

by Geoffrey M. Hodgson

Version of 17 November 2011

7900 words

For *World Economics Journal*

University of Hertfordshire Business School, Hatfield, Hertfordshire AL10 9AB, UK  
www.geoffrey-hodgson.info  
g.m.hodgson@herts.ac.uk

## **Key words:**

rationality, utility, preferences, experimental economics, falsifiability, historical specificity

## **ABSTRACT**

The value of rational choice theory for the social sciences has long been contested. It is argued here that, in the debate over its role, it is necessary to distinguish between claims that people maximise manifest payoffs, and claims that people maximise their utility. The former version has been falsified. The latter is unfalsifiable, because utility cannot be observed. In principle, utility maximisation can be adapted to fit any form of behaviour, including the behaviour of non-human organisms. Allegedly 'inconsistent' behaviour is also impossible to establish without qualification. This utility-maximising version of rational choice theory has the character of a universal 'explanation' that can be made to 'fit' any set of events. This is a sign of weakness rather than strength. In its excessive quest for generality, utility-maximising rational choice theory fails to focus on the historically and geographically specific features of socio-economic systems. As long as such theory is confined to ahistorical generalities, then it will remain highly limited in dealing with the real world. We have to move on and consider the real social and psychological determinants of human behaviour.

# On the Limits of Rational Choice Theory

Geoffrey M. Hodgson

Much time has been spent by economists and critics on the elusive concept of rationality.<sup>1</sup> This essay will not end this output, but it will hopefully divert some of the wasted energy into more useful occupations. There are multiple prominent versions of rationality, not one (Sen 1987). One version upholds that rationality is essentially about consistency of behaviour. Another sees rationality as the maximization of explicit (typically pecuniary) rewards. So much empirical and experimental evidence has been marshalled against the second (more restrictive) version that a significant number of economists have now abandoned the idea, and at least six critics of rationality (Friedrich Hayek, Gunnar Myrdal, Herbert Simon, Ronald Coase, Amartya Sen and Daniel Kahneman) have been awarded Nobel prizes in economics. Whether the existence of altruistic behaviour challenges the idea of rationality depends on the definition that is adopted. Definitions of rationality that accommodate altruism end up being unfalsifiable.

This essay is divided into three sections. Section one discusses possible meanings of rationality and distinguishes between payoff maximisation and utility maximisation. Payoff maximisation has been refuted by many experiments. Section two shows that the prominent notion of rationality as utility-maximisation is strictly unfalsifiable and does not offer a viable causal explanation of behaviour. Section three selectively reviews some prominent criticisms and defences of the rationality assumption in the light of the unfalsifiability of utility maximisation. Section four concludes the essay.

## 1. The slippery concept of rationality

One popular notion of rationality is *thoughtful deliberation*. This is not the meaning that is associated with the ‘choice’ concept of Lionel Robbins (1932) or the ‘as if’ methodology of Milton Friedman (1953). But when Herbert Simon (1957) argued that rationality was ‘bounded,’ he sometimes used this term to refer to limited computational and deliberative capacity.

A similar meaning of rationality is *acting for reasons*. Although etymologically accurate, this is alternatively described as the *rationalist concept of action*, and criticized in a different manner (Hindess 1977, Hodgson 1988). The use of different meanings by economists has confused much of the debate surrounding the concept of rationality.

Another prominent notion of rationality is that people try to do the best they can in their circumstances. An important corollary is that agents respond to incentives. There is nothing wrong with this idea. But it fails to tell us how people interpret their situation or identify ‘the best’ goal. It does not acknowledge that different interpretations of situations and hence

---

<sup>1</sup> This essay uses some material from Hodgson (2012).

different goals are often possible. The problem with this ‘doing one’s best’ notion of rationality is that it lacks the necessary explanatory detail concerning agent cognition and goal-formation. It might also be used to buttress self-interested notions of ‘the best’ that are in fact undermined by the evidence. The ‘doing one’s best’ notion of rationality is not entirely vacuous, because it does point to the necessity of problem-solving. But it lacks vital detail. To avoid confusion, ‘doing one’s best’ notion would better be described as *following incentives* or *adaptation to circumstances*. But more specific detail would also be required.

We now consider notions of rationality that dominate modern mainstream economics. I define some additional terms. A payoff is a reward in a game that *has an explicit expected worth* (such as a declared monetary reward) *that is known to the analysts of the game and to all of its players.*<sup>2</sup> By payoff rationality or payoff maximization, I mean the maximization of such explicit payoffs by players, given the information available to them plus their assumption that other players are also payoff maximisers. If one is committed to the axioms of payoff rationality, then logically one is also committed to the idea that rationality involves consistent behaviour. But utility maximization is not necessarily payoff maximization, unless there is a monotonic relation between utilities and payoffs.

Vernon Smith (1982) and others have addressed the problem of the possible absence of a monotonic relation or ‘parallelism’ between overall utility and monetary payoffs. To relate payoffs to utilities, the possibility of additional, subjective utilities that are unrelated to one’s own monetary payoffs, such as utility derived from the satisfaction of others or from taking risks, has to be substantially diminished. The player’s own money payoffs have to ‘dominate’ their decisions. In order to make experiments ‘work’ in this sense, Smith proposes a number of ‘precepts’ of experimental assumption and design constituting an ‘induced value procedure.’ These precepts include nonsatiation, sufficiently large and obvious rewards, restriction of communication between subjects, and so on. But Smith (p. 929) himself is the first to admit that these precepts cannot *guarantee* any monotonic correspondence between observable monetary rewards and preferences, which are ‘not directly observable.’ In fact, we can never know if the precept has been effectively applied. The idea that Smith’s precepts ‘work’ is an article of faith, placed so far under surprisingly little methodological scrutiny.<sup>3</sup>

Despite the headway made by its critics, some economists retain a notion of payoff rationality. They argue that while there are empirical deviations from its norms, if pecuniary rewards are sufficiently large and agents are given long enough to learn the game, then payoff maximization will become established as an approximate behavioural rule (Harrison 1989, Binmore 1994, 1998, 1999, Binmore and Shaked 2010).

By contrast, others follow ‘behavioural economists’ and argue that the evidence is sufficient to undermine payoff rationality. Behavioural economics has now spread to the mainstream, and is evident in some of the most prestigious journals in the discipline.<sup>4</sup> Payoff rationality and self-interest were regarded as articles of faith among mainstream economists from the 1950s to the 1990s, and to question them was enough to lose one’s credentials as an

---

<sup>2</sup> Note that the definition of payoff includes those formulated in probabilistic terms. For simplicity I shall ignore games where possible payoffs are known to some players but not others. Including this possibility would not change the principal conclusions below. In fact, it would make the concept of rationality more difficult to define straightforwardly, as in some types of game theory (Sugden 1991).

<sup>3</sup> For a critical discussion of Smith’s precept of parallelism see Siakantaris (2000).

<sup>4</sup> See Earl (2010) for a critical comparison of mainstream and Simonian behavioural economics.

economist. Subsequently, in the face of massive, accumulating, evidence of agents who do not maximize pecuniary rewards, economics has changed.

It is impossible to review all the evidence here.<sup>5</sup> There are ‘framing effects’ when rankings of options change when equivalent choices are presented in different terms. Although payoff rationality means that bygones should be ignored, people often take them into account. Individuals are presented with two gambles – one with a certainty of winning a modest sum of money and the other with a low probability of winning a large sum of money. Even when the expected value of the risky option is greater, people often prefer the certain reward (Slovic and Lichtenstein 1983). In the face of such evidence, strict payoff rationality has been abandoned by many.<sup>6</sup>

Although the axioms of payoff rationality imply consistency of behaviour, the reverse is not true. Without logical contradiction, one can abandon payoff rationality and still uphold that behaviour is consistent, and even utility-maximizing. Herbert Gintis (2007, 2009) is an exponent of this position.

Gintis is a co-author of a fascinating set of cross-cultural studies that show that players rarely reach a Nash payoff solution in ultimatum games (Henrich *et al.* 2001, 2004). One of two players in an ultimatum game is asked to divide an amount of money between herself and the other player. If the second player rejects the division, then both players get nothing; but if he accepts, then they each receive their allocated amounts. If the second player is a payoff maximiser, then he will accept the lowest possible positive allocation when it is offered: payoff maximisers always prefer something to nothing. This is a subgame perfect Nash equilibrium; it gives the best expected payoff outcome for both players, each assuming that the other player is also a payoff maximiser. Consequently, no player has anything to gain by unilaterally changing strategy on their own. But experiments often do not lead to this Nash payoff outcome: players do not always maximize payoffs in this way. Instead, their behaviour is consistent with taking additional, intangible, non-pecuniary factors into account, such as honour, custom and fairness, even when they cannot bargain with one another and the game is not repeated. The cross-cultural studies of Joseph Henrich and his colleagues also showed that the actual pattern of play can vary significantly from one cultural setting to another.

While abandoning payoff rationality, Gintis defends a broader concept of rationality, defined as consistency of behaviour. Given behavioural consistency (or transitivity) – along with the other standard assumptions such as independence and continuity (von Neumann and Morgenstern 1944, Fishburn 1970) – it is possible to construct a standard ordinal utility function where behaviour is consistent with expected utility maximization. Gintis (2007, 2009) considers much of the experimental evidence and points out that the absence of payoff maximization does not mean that these players are behaving inconsistently or failing to

---

<sup>5</sup> See the extensive works of Nobel Laureate Daniel Kahneman and his colleagues (Kahneman 1994, 2003a, 2003b, Kahneman *et al.* 1982, 1986a, 1986b). Bowles and Gintis (2011) provide an excellent overview of the evidence.

<sup>6</sup> But controversy has not ended. For example, different results – sometimes closer to payoff maximization – are obtained from some field experiments (Gneezy and List 2006). The contrasting outcomes of laboratory and field experiments show that the social and institutional context matters, and individuals cannot be taken in isolation (List 2006).

maximize utility.<sup>7</sup> For Gintis (2006, p. 17) this behavioural consistency is rooted in genetically rooted instincts and drives that have evolved over the millennia and dispose us to respond in specific ways to specific cues. Consequently, Gintis (2006, p. 7) argues that ‘utility maximization should be a central tool in analysing human behaviour, even if humans are not self-regarding.’<sup>8</sup>

## 2. Fitting everything and explaining nothing

Gintis does not acknowledge the following key difficulty. When the young Paul Samuelson (1937, p. 156) discussed utility maximization, he understood that ‘all types of observable behaviour might conceivably result from such an assumption.’ Because utility is unobservable, all kinds of behaviour can be ‘explained’ in terms of the idea, without fear of refutation. As Sidney Winter (1964, pp. 309, 315) and Lawrence Boland (1981) have also remarked, no evidence can possibly refute the theory that agents are maximizing some hidden or unknown variable (such as utility). Amartya Sen (1977, p. 325) has similarly pointed to the circularity of explaining behaviour ‘in terms of preferences, which are in turn defined only by behaviour.’ Sen (1987, p. 73) notes elsewhere that the description of choices in terms of utility ‘does not give any independent evidence on what the person is aiming to do or trying to achieve.’

Defending ‘self-interest, rightly understood’ against its critics, Teppo Felin and Nicolai Foss (2009, p. 622) say it is consistent with ‘cooperation, organization, community-building, trust, or for that matter, any other individual, relational, or organizational virtue.’ Rather than selling the assumption, this rather gives the game away. An assumption that is consistent with everything describes little and delimits nothing.<sup>9</sup>

If experiments show that some consumers appear to prefer a monetary reward that is less than the expected outcome, or appear to have intransitive preference orderings, or defy the independence axiom, then we can always get round these problems, and make the evidence consistent with utility maximization, by introducing other explanatory variables.<sup>10</sup>

---

<sup>7</sup> There is experimental evidence that people are not very good at logical problems, especially when they are posed in abstract terms or involve probabilities. Gintis (2007, pp. 11-12) brushes this evidence aside with the questionable conclusion that ‘most individuals do not appear to have difficulty making and understanding logical arguments everyday life.’ The evidence of Wason (1983), Cosmides (1989), and Cosmides and Tooby (1994a, 1994b) suggests otherwise. Even from Gintis’s standpoint, this argument seems somewhat superfluous, because if an individual made the same logical errors over and over again, then he or she might be behaving consistently, and might be ‘rational’ by Gintis’s criterion.

<sup>8</sup> Gintis distinguishes between ‘self-regarding’ and ‘self-interested’ preferences. With ‘self-regarding’ preferences one takes account of one’s own situation only. Hence for Gintis a charitable act is not ‘self-regarding’ but it may be ‘self-interested’ because of increased utility gained by the ‘warm glow’ or satisfaction of giving. By contrast, it is argued here that notions of utility-maximization or ‘self-interested’ behaviour are in principle unfalsifiable, and of little use.

<sup>9</sup> Significantly, Felin and Foss (2009, p. 622) continue: ‘the type of ‘enlightened’ self-interest we have in mind should be completely decoupled from ethics.’ In contrast to their defence of the self-interest assumption, their critique of social constructivist and performativity arguments in their article is much more robust.

<sup>10</sup> Hausman (1992, ch. 13) documents several attempts to explain the apparent anomalies that have been revealed by the experimenters, notably by pointing to other possible sources of utility. But in some of these cases the independence axiom is abandoned in attempts to rescue the idea of utility maximization.

For example, preference reversals can be regarded as consistent with expected utility theory. Assume that a subject is faced with a choice between \$10 with certainty, and \$1,000 with a probability of 2 per cent. Experiments with real subjects indicate that in such situations the \$10 option is sometimes chosen, despite the fact that the expected value of the second option is higher at \$20 (Slovic and Lichtenstein 1983). But preference reversals also fail to falsify expected utility theory, once we accept that (expected) utility is not necessarily measured in terms of the monetary payoffs in the experiment. If we assume an added disutility associated with involvement in a risky and low probability choice, then the theory that people are maximizing their utility is not overturned by these experiments. A risk-averse actor may not maximize expected monetary value but still be maximizing expected utility. By appropriate functional manipulation, the choice of \$10 can be made perfectly consistent with the maximization of expected utility, rather than the maximization of the expected monetary value of the payoff.

Gintis and others might respond that inconsistent behaviour would refute utility maximization. The problem here is one of identifying inconsistent behaviour in empirical terms. Note that the utility maximand is unobservable. For example, if an experiment shows that option *A* with an expected value of \$4 is preferred to option *B* with an expected value of \$5 then we can simply assume that there are additional attributes of option *A* (for example, we may enjoy losing or gain pleasure from seeing others win) that are consistent with the view that it yields higher overall expected utility for the subject.

On repeated visits to the same restaurant, we may prefer steak to fish one day, and fish to steak on another. Is this behaviour inconsistent? Maybe. Maybe not. We may discover that the steak is not as good as expected. Or we may have seen an alarming television report about mad cow disease that causes us to switch to fish. The two choice occasions were different in terms of circumstances and knowledge. Hence they do not necessarily imply inconsistency.

The empirical detection of preference intransitivity is also problematic. An experiment may seem to reveal preference intransitivity, by showing that while *X* is preferred to *Y*, and *Y* is preferred to *Z*, *Z* is preferred to *X*. But this result can be explained away by showing that the three pairwise comparisons did not take place under identical conditions, or were separated in time or space. Extraneous factors may account for the apparent intransitivity. All we have to do is indicate in some way that the two *Z*s in the above comparisons are not quite identical. The two *Z*s could be slightly different in timing, substance, or their informational or other contexts. We then get the result: *X* is preferred to *Y*, *Y* is preferred to *Z*<sub>1</sub>, and *Z*<sub>2</sub> is preferred to *X*. In these circumstances, transitivity is no longer violated. The defender of utility maximization may conflate *Z*<sub>1</sub> with *Z*<sub>2</sub>, whereas they were in fact different.

It may be objected that if preferences are assumed stable, then evidence on revealed preference could reveal inconsistent preferences. But this would not be the case if utility depended on other factors in the environment. Consider the utility function  $U = f(X, E)$ , where *X* is a vector of consumption inputs and *E* is a vector of environmental or contextual conditions. Assume the function *U* is perfectly stable. But *E* can never be strictly held constant. Some part of the environment, however remotely or slightly, will inevitably alter. Hence, in practice, intransitivity (or intertemporal inconsistency) in the rankings of the elements of vector *X* alone would not reveal preference inconsistency because some elements in the vector *E* would also have changed, even by the tiniest amount. Strictly, the environment is never constant. Consequently, because we cannot strictly and identically replicate the *E*

conditions, intransitivity or inconsistency of  $X$  choices can never falsify the assumption of fixed preferences.<sup>11</sup>

Given that we can never in principle demonstrate that some unobserved variable (like utility) is not being maximized, then the theory is invulnerable to any empirical attack. No amount of evidence can establish non-existence. Hence the standard core of expected utility theory is *unfalsifiable*.<sup>12</sup>

The utility-maximization assumption is unfalsifiable, but it is not a tautology in the logical sense because it is *conceivably false*.<sup>13</sup> Logical tautologies – such as a triangle has three sides – are true by definition. By contrast, it might be the case that individuals are not maximizing anything. But we can never establish this on the basis of empirical evidence.

This does not necessarily mean that the utility maximization framework is useless or wrong. We do not have to uphold falsifiability as the mark of science – a criterion attributed to Karl Popper, who in fact adopted a more nuanced position (Ackerman 1976). Neither tautological nor non-falsifiable statements are necessarily meaningless or unscientific.<sup>14</sup>

A key problem with utility maximization is that it is so general that it can explain anything; consequently its explanatory power in specific instances is dramatically diminished. Its explanatory success is an illusion. Close inspection of its proclaimed achievements reveal that the results always depend on additional assumptions. For example, Gary Becker (1976a, 1991, 1996) contends that standard rationality assumptions generate a number of testable predictions concerning human behaviour. But all of Becker's 'predictions' depend on assumptions *additional* to his core axioms of utility maximization. Indeed, because it is difficult to conceive of evidence that falsifies these axioms, such models must depend on auxiliary assumptions to generate specific results. As Mark Blaug (1992, p. 232) puts it: 'The rationality hypothesis by itself is rather weak. To make it yield interesting implications, we need to add auxiliary assumptions.'

The notion of utility maximization is so capacious that it goes beyond the parameters of human decision. Experimental work with rats and other animals (Kagel *et al.* 1981, 1995) has 'revealed' that animals have downward-sloping demand curves, supposedly just like humans. Becker (1991, p. 307) proposes that: 'Economic analysis is a powerful tool not only in understanding human behaviour but also in understanding the behavior of other species.' Similarly, Gordon Tullock (1994) has claimed that organisms – from bacteria to bears – can

---

<sup>11</sup> There is a standard argument that intransitive preferences would involve a 'money pump' – an agent with intransitive preferences would accept a series of trade offers that leaves her worse off to the benefit of the other trader. But given the sequential separation of each choice in time, strict intransitivity may never apply. This becomes evident when we leave the timeless world of neoclassical economics to the real world with historical time.

<sup>12</sup> This argument is redolent of the so-called Duhem-Quine thesis, which claims that it is generally impossible to falsify any single hypothesis because we always have to adopt additional hypotheses in the analysis of any set of observations (Harding 1976). Consequently, we can never be sure that the main hypothesis is being targeted and tested on its own, and that other auxiliary hypotheses are not complicating the picture.

<sup>13</sup> Several important authors, from Simon (1986, p. S222) to Field (2001, p. 6), mistakenly confuse tautological with non-falsifiable propositions.

<sup>14</sup> Indeed, it is widely accepted in the philosophy of science – including by Popper – that some unfalsifiable propositions are necessary for science itself. These include the principle of determinacy (every event has a cause) and the assumption of the uniformity of nature. Without these prior assumptions, science is impossible.

be treated as if they have the same general type of utility function that is attributed to humans in the microeconomics textbooks. Utility maximization is applied to humans in all forms of society since the origin of our species, and to a large portion of the animal kingdom as well. Seemingly, we now have ‘evidence’ of the ‘rationality’ of everything in evolution from the amoeba onwards. As a consequence such assumptions are telling us very little about what is specific to human nature and human society.

Arguably, human societies are partly differentiated from other animals in terms of developed institutions and cultures. The authors cited in the preceding paragraph thus demonstrate that these distinctive elements are effectively separated from the utility-maximizing picture of ‘rational economic man.’ Consequently ‘rational economic man’ bears no mark of any specifically human culture or institution. The causal mechanisms through which culture and institutions mould and constrain human agents remain unexplored in their paradigm. Human psychology is likewise neglected. Essentially, there is no adequate and substantial theory of *human* agency at the core of the standard theory. It tells us nothing of significance that is specific to the human psyche, human interaction, human nature, or human society. With respect to specifically human characteristics it is causally vacuous. Its very weakness, when applied to the human domain, stems from its excessive scope.

The non-falsifiability of the concept of rationality-as-behavioural-consistency-or-utility-maximization sustains an *epistemic* critique. It does not clinch the matter. One has also to consider the *theoretical* limitations of this stance. Here rationality-as-behavioural-consistency-or-utility-maximisation falls down for at least two reasons. First it neglects the problem of *explaining the causes* of behaviour. Second it fudges the question of the individual *development* of capacities and dispositions.

In a prominent defence of rationality, Richard Posner makes his neglect of psychological or other causes of behaviour explicit. Posner (1980, p. 5) sees the ‘rationality of ‘economic man’’ as ‘a matter of consequences, not states of mind.’ In discussing ‘economically rational’ human agents he declines ‘any statement about their conscious state ... behavior to an economist is a matter of consequences rather than intentions’ (p. 53 n.). Here the problem of explaining behaviour, by reference to psychology or other matters, is openly abandoned.

For related reasons, claims that there is an evolutionary basis for utility maximization (Robson 2001, Gintis 2006, p. 17) do not pass muster. It is insufficient to show that the behavioural outcomes of evolution are consistent with some utility function. Ultimately this claim is trivially true, because one can always find a function that fits. One has to show that utility maximization is useful causal account of behavioural motivation. This is problematic, for reasons elaborated below.

Indeed, it is rather odd to claim simultaneously that evolution has produced individuals that maximize utility and are also capable of altruism, as a consequence of inclusive fitness or whatever. Altruism is typically defined as costly for the individual concerned but beneficial for others. This sits uneasily with a utilitarian framework, and consequently the definition of altruism is out-of-equilibrium and has to be constantly clarified (e.g. Bowles and Gintis 2011). Utilitarians working in an evolutionary framework might awkwardly depict altruism as simultaneously involving a fitness cost and a utility gain for the agent, to preserve the near-vacuous dogma that all agents are utility maximisers.

Other defenders of rational economic man – notably Becker (1996) – treat the individual ‘as if’ she is born with a sophisticated but fixed meta-preference function. The process of human development is then regarded as a matter of gradually acquiring information about underlying ‘true’ tastes. Sure enough, some such meta-preference function can always be

stretched and twisted to fit the data. But as an account of the developmental process it is untenable. The fixed meta-preference function has no place in our current understanding of the neural system. Although many dispositions are inherited biologically, our further development from birth depends on the formation of many neural structures and connections, which are contingent on our environment and our past development. Although there is dispute concerning the details, psychologists and neuroscientists agree that there is considerable flexibility and plasticity in the developing brain (Penn and Shatz 1997, Marcus 2004, Sarnecki 2007). This neural flexibility and plasticity goes against the idea of an entirely inherited and fixed meta-preference function.

Placed within some versions of modern game theory, the ‘as if’ argument is stretched beyond the limits of credulity. It is not simply assumed that agents act ‘as if’ they are rational, but also that they act ‘as if’ they consider the rationality of others, and ‘as if’ others respond rationally with such common knowledge (and so on ...), somehow without necessarily making any assumptions about their deliberative behaviour. Retaining the ‘as if’ argument in this context requires us to treat individuals as capable of emulating incredible super-calculators with unbounded cognitive capacities, without any consideration of how they would manage to do this.

Past economists have tolerated the ‘as if’ neglect of real phenomena, but it no longer satisfies scholars in this new age of exploration for evolutionary understandings of origin and development. We are interested specifically in the human mind and human social organization. We obtain little insight in this respect from overly-capacious and unfalsifiable principles that apply to any organism or behavioural entity.

On the basis of experimental evidence, some neuroeconomists (Platt and Glimcher 1999, Glimcher *et al.* 2005) make the strong claim that the utility function exists as a physiological reality inside the brain. This claim is scrutinized by Jack Vromen (2010), who argues that at best the neurological evidence exhibits consistency with the predictions of expected utility theory. There is no evidence of actual computation of utility. Given the argument here that any observed outcomes can be made consistent with some utility function, the consistency claim is hardly powerful or surprising. But existence claims are unsupported. After an extensive review of the evidence, Colin F. Camerer, George Loewenstein and Drazen Prelec (2005, pp. 54-55) are also sceptical of the claim that neuroscience supports a standard model of rational choice. The evidence that carries some weight relates to simple decisions only, not the ‘abstract, complex, long-term tradeoffs which are the traditional province of economic theory.’

Overall, the long debate over whether behaviour is ‘rational’ has generated more heat than light. Sometimes the antagonists have misunderstood one another, particularly by confusing the falsifiable notion of payoff maximization with the unfalsifiable propositions of utility maximization or behavioural consistency. Since 1990, many leading members of the economics profession have abandoned payoff maximization. Yet the credo of rationality is preserved in the empty mantra of utility. In these terms it tells us very little. Faced with this explanatory agenda, ‘rationality’ in the broader sense of utility maximization is but a word of little consequence. By contrast, payoff rationality is more meaningful. But it turns out to be wrong.

The important task is to understand the nature and evolutionary origins of our human dispositions. Both genetic inheritance and cultural transmission are relevant to this quest. To understand the motion of the planets or the nature of matter is to comprehend the structures and forces that lie behind events, not to imagine spirits or gods that create every eventuality.

To understand human nature and society is to appreciate human dispositions and interactions, not to fit all observations of behaviour to imagined mathematical functions of ever-expandable correlative capacity. Rationality in the broader sense serves an ex-post rationalization – rather than a materially-grounded causal explanation. A utility function may serve a limited purpose as a formalized preference ordering. Such formal constructions have some benefits. They can be useful shortcuts for modelling or explanatory purposes. But they do not enhance our understanding of human motivation. Utility theory is an elegant way of summarizing what we don't know about human psychology.

Q: Why did the chicken cross the road? A: To maximize its utility. Some economists may be satisfied with this answer. But it tells us nothing about chickens, roads, specific motives, developmental histories, or detailed causal mechanisms. We should also be dissatisfied with summarizing all the complexities of human motivation in terms of a relatively simple preference function. As Sen (1977, pp. 335-6) has famously argued:

A person is given a preference ordering [that] is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behaviour. Can one preference ordering do all these things? A person thus described may be 'rational' in the limited sense of revealing no inconsistencies in his choice behaviour, but if he has no use for these distinctions between different concepts, he must be a bit of a fool.

In sum, a problem with the standard rationality assumption is not that it lacks empirical correlation, but that they could cover every conceivable decision situation and every possible causal mechanism underlying choice. Insofar as there may be common features of every decision situation then it may be possible to extract universal and meaningful propositions. Nevertheless, some important and specific features or causal mechanisms may be excluded by concentrating solely on the common features of every decision situation. In fact, the degree of universality involved is so great that it goes beyond the parameters of mere human decision.

For the neoclassical economist, the fact that utility theory can 'explain' a wide variety of types of economic behaviour is regarded as a strong vindication of this general approach. I take a different view. First, the sheer generality of a theory tells us nothing of its explanatory power. We could conceive different general theories, such as that we all are programmed by aliens from outer space, or that we are all pawns of God. These would be quite general in their scope and could be applied in principle to any behavioural manifestation. But we would rightly be sceptical of their explanatory value. A theory does not explain anything unless it points to an underlying causal mechanism. In the case of individual behaviour, explanations must thus relate to the known mechanisms of the human psyche and human interaction and draw upon psychology, anthropology, sociology and other disciplines. This is precisely what the neoclassical advocates of utility theory refuse to do. They take the utility functions as given and give the job of grounding them theoretically to somebody else. By this refusal they indicate that utility theory itself cannot provide a real explanation.

Arguably, human societies are partly differentiated from other animals in terms of developed institutions and cultures. If utility maximising behaviour not confined to humanity, then these differentiating elements are effectively absent from the universal picture. Whether true or false, this picture can tell us little of importance about historically specific human cultures or institutions. That is the unintended achievement of the exponents of ubiquitous rationality and economic imperialism. The causal mechanisms through which culture and institutions mould and constrain human agents remain unexplored in this paradigm. Essentially, there is no adequate and substantial theory of human agency at the core of the

standard theory. It tells us nothing of significance that is specifically about the human psyche or about human interaction. Outside the realm of the universal, no particular causal mechanism is identified by the theory. With respect to specifically human characteristics and specific human societies, it is causally vacuous. Its very weakness stems from its excessive universality. Indeed, to attain the status of universality it has to be evacuated of much of its real content.

### 3. Weak criticism and false approval

But many critics of mainstream economics have taken a different line of attack. For example, the Post Keynesian economist Alfred Eichner (1983, p. 211), complained that the core assumptions of mainstream economic theory ‘have yet to be empirically validated’ and that they have ‘no empirical counterpart in the observable world’. However, the problem with these assumptions is *not* primarily their lack of empirical corroboration. It is that they are vessels into which *any* empirical content can be filled. The problem with the theory is *not* that it lacks empirical validation but that *any* conceivable fact about behaviour, from church attendance to suicide, can be fitted into the theory.<sup>15</sup>

Just as the critics of neoclassical theory wrongly claim that its basic postulates have been falsified, its exponents misleadingly claim that they have been rigorously confirmed. Jack Hirshleifer (1985, p. 59) went so far as to write: ‘Ultimately we must be ready to abandon the rationality paradigm to the extent that it fails to fit the evidence about human behavior.’ But this apparent concession to empirical confirmation in fact conceals a methodological misunderstanding. Hirshleifer did not have to worry, because no conceivable evidence can ‘fail to fit’ some version of the theory. Both Hirshleifer and the critics of the rationality paradigm share the flawed supposition: that evidence can in principle refute the theory. Both supporters and critics of neoclassical theory have perpetuated the myth that it is susceptible to decisive empirical testing.

As a result, the mainstream theory is not wrong because it is empirically inaccurate. It is not unrealistic in the sense that it fails to fit the data. Any data can be fitted into it. Hence no data can refute the theory. It cannot be displaced simply by an appeal to the evidence. The experimental evidence of preference reversals and other choice ‘anomalies’ may lead us to search for a different and better theory, but it does not in principle refute the old version based on utility and rational choice.<sup>16</sup>

Critics such as Eichner (1983) based their criticism on an untenable and empiricist view of science that denies that some non-falsifiable and ‘metaphysical’ assumptions are *essential to any science*. In fact, all sciences depend upon some propositions that are untestable. No theory can be composed entirely of empirically validated elements. Prior concepts are required to make sense of any fact. These prior concepts cannot all be ‘tested’ empirically. In any case, any ‘test’ itself relies on prior concepts or categories. As a result, all sciences must unavoidably make extensive use of some untestable and metaphysical assumptions.

For this reason, the empiricist criticism of mainstream economics is untenable. In practice, their denial of the essential role of non-falsifiable assumptions in any theory would disable

---

<sup>15</sup> See Azzi and Ehrenberg (1975) and Hammermesh and Soss (1974).

<sup>16</sup> I am not arguing that evidence is unimportant. Although evidence cannot falsify the theory, the accumulated evidence may provide a context in which the theory is more readily questioned. See Loomes (1998, pp. 485-6).

any of their own attempts at theoretical construction. Given that it is practically impossible to test all assumptions, any theoretical construction would reveal hidden, 'ad hoc' assumptions, privileged to lie beyond empirical test. For reasons outlined above, every theory must involve some untestable assumptions. Hence any theory built on the claim of complete testability would be highly vulnerable to critique by its own canon.

But this does not mean that 'anything goes' and that all criticisms are disabled. There are powerful theoretical criticisms of the rationality assumption. Essentially, the theory lacks adequate theoretical concepts to discriminate, understand and properly explain key phenomena. A problem with the standard assumptions of rationality and expected utility maximisation is their lack of specific theoretical and conceptual content, pertaining to specific causal mechanisms involved in the human psyche and in the structures of specific real world economic institutions.

To repeat: the empirical evidence is valuable and important, but it cannot be used to show that the theory is false. In recent years, there have been attempts to apply models of rational, utility maximising behaviour to a wide variety of phenomena, even beyond the sphere of commerce and markets. Models of utility-maximising behaviour have been applied to politics, marriage, religion, suicide, and much else. Such attempts have been widely resisted. Many tried to defend their academic discipline from the 'economic imperialism' of rational choice models. However, the widespread failure to recognise the non-falsifiability of 'rational' maximising behaviour has weakened many such counter-arguments. They appealed to evidence: it was mistakenly argued that rational choice models did not fit the facts. On the contrary, models of utility-maximising behaviour can always be adjusted to fit the facts. The attempt to resist the incursions of rational choice theory by claiming otherwise was bound to fail. In this instance, appeals to evidence cannot win.

In development economics, for example, there was a debate in the 1970s over whether peasants were or were not 'rational'. Critics of this idea appealed to 'evidence' of 'non-rational' behaviour, without realising that no evidence can strictly falsify the theory. With opponents weakened by their own theoretical position and methodological misunderstandings, the rational choice theorists seemed to win the argument (Popkin, 1979). Similarly weak defences were evident in sociology and political science, as they too were invaded by rational choice theorists. Again and again an attempt was made to resist the incursions of utility and rational choice, on the grounds that its assumptions are not 'realistic'. Such attempted defences against the invasion of rational choice theory are methodologically flawed and ultimately doomed.

The moral here is that mistaken claims concerning the testability of rational choice theory led its opponents to attack it with weak arguments. It would have been much more fruitful if both sides had admitted that the theory was falsifiable and then debated its explanatory value in specific circumstances. Instead, these controversies were entirely confined to claims and counter claims concerning empirical validation. At that primitive level the issue is simple: the assumptions of utility theory cannot be falsified.

#### **4. Conclusion**

But ironically, it has become fashionable for mainstream economists to question some of these core assumptions. Perhaps because mainstream economists have lost the capacity to police their own disciplinary boundaries, in search of a new separate identity they have begun to question their own *raison d'être*. As Kyriakos Kontopoulos (1993, p. 90) has pointed out:

‘Ironically, economists become less economic at a time when sociologists seem to become enamored with rational choice theory.’ Accordingly, some economists are now deconstructing rational economic man. As economist Robert Sugden (1991, p. 783) put it:

There was a time, not long ago, when the foundations of rational-choice theory appeared firm, and when the job of the economic theorist seemed to be one of drawing out the often complex implications of a fairly simple and uncontroversial system of axioms. But it is increasingly becoming clear that these foundations are less secure than we thought, and that they need to be examined and perhaps rebuilt.

One reason for this change of heart is the rise of game theory. In certain types of game the very definition of rationality becomes problematic. Nevertheless, the response of mainstream economists to these problems has largely been to become immersed in the technicalities, rather than to give the economic agents at the core of the theory of human behaviour some real institutional and cultural flesh and blood. Some still cling tenaciously to the principles of rationality, in a manner that is reminiscent of Ptolemaic astronomers, fitting the evidence of the apparent circular movements of the stars into complicated models (Koestler, 1959). Others are not inclined simply to ‘save appearances’; they express their misgivings but seem unaware where to look to find an alternative paradigm.

For some, the move to game theory has led to the questioning of core assumptions. For others it has reinforced the idea that economics itself is a formal game, with little connection to reality. If a theory makes no claim outside a single domain, then there is no aim to use the theory to explain other real world phenomena. The interest in the theory is typically in its mathematical content, rather than its usefulness to help understand reality. Accordingly, there is a move away from former attempts to build a universal theory (which turned out to be unfalsifiable), to the building of exemplifying theories that are designed with limited empirical scrutiny in mind.

Rational choice theory has invaded social sciences such as sociology with its devotees declaring triumphantly that it can ‘explain’ nuanced social phenomena such as altruism, honour, trust and duty. The point argued here is that it is ineffective to counter these allegations with denials of their *empirical* validity. This is a weak response because, in principle, any manifest behaviour can be fitted into the rational choice framework. However, once the unfalsifiable nature of rational choice theory is understood, we can meet the triumphant claims with the polite response: ‘So what. We know *in advance* that any behaviour can fit the theory.’ We are then able to move on to the more important question, concerning claims of explanation and their derivation. Explanation, we insist, is much more than empirical correlation. And a perfect empirical fit does not necessarily imply any explanation of the causal processes underlying behaviour. The debate then moves on to the vital question as to how such causal explanations can be obtained.

Economic and social theory can only advance if it takes account of historical, cultural and institutional specificities (Hodgson 2001). The crusade for rational choice theory in modern social science is part of a wider project to develop a universal theory of all social phenomena. The problem with such a theory is that, in its excessive quest for generality, it will fail to focus on the historically and geographically specific features of the socio-economic systems that we wish to study and understand. As long as social theory is confined to generalities it will remain highly limited in dealing with any specific socio-economic system, including the one in which we live. What is required is a theory that is far more sensitive in this regard.

## References

- Ackerman, Robert (1976) *The Philosophy of Karl Popper* (Amherst, MA: University of Massachusetts Press).
- Azzi, Corry and Ehrenberg, Ronald (1975) 'Household Allocation of Time and Church Attendance', *Journal of Political Economy*, **38**(1), pp. 27-56.
- Becker, Gary S. (1976a) *The Economic Approach to Human Behavior* (Chicago: University of Chicago Press).
- Becker, Gary S. (1991) *A Treatise on the Family*, 2nd edn. (Cambridge, MA: Harvard University Press).
- Becker, Gary S. (1996) *Accounting for Tastes* (Cambridge, MA: Harvard University Press).
- Binmore, Kenneth (1994) *Playing Fair: Game Theory and the Social Contract. Volume I* (Cambridge, MA: MIT Press).
- Binmore, Kenneth (1998) *Just Playing: Game Theory and the Social Contract. Volume 2* (Cambridge, MA: MIT Press).
- Binmore, Kenneth (1999) 'Why Experiment in Economics?', *Economic Journal*, 109(2), February, pp. F16-F24.
- Binmore, Kenneth and Shaked, Avner (2010) 'Experimental Economics: Where Next?', *Journal of Economic Behavior and Organization*, 73, pp. 87-100.
- Blaug, Mark (1992) *The Methodology of Economics: Or How Economists Explain*, 2<sup>nd</sup> edn. (Cambridge: Cambridge University Press).
- Boland, Lawrence A. (1981) 'On the Futility of Criticizing the Neoclassical Maximization Hypothesis', *American Economic Review*, **71**(5), December, pp. 1031-36.
- Bowles, Samuel and Gintis, Herbert (2011) *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton, NJ: Princeton University Press).
- Camerer, Colin F., Loewenstein, George and Prelec, Drazen (2005) 'Neuroeconomics: How Neuroscience Can Inform Economics', *Journal of Economic Literature*, **43**(1), March, pp. 9-64.
- Cosmides, Leda and Tooby, John (1994a) 'Beyond Intuition and Instinct Blindness: Towards an Evolutionary Rigorous Cognitive Science', *Cognition*, **50**(1-3), April-June, pp. 41-77.
- Cosmides, Leda and Tooby, John (1994b) 'Better than Rational: Evolutionary Psychology and the Invisible Hand', *American Economic Review (Papers and Proceedings)*, **84**(2), May, pp. 327-32.
- Earl, Peter E. (2010) 'Economics fit for the Queen: A Pessimistic Assessment of its Prospects', *Prometheus*, **28**(3), pp. 209-25.
- Eichner, Alfred S. (ed.) (1983) *Why Economics is Not Yet a Science* (Armonk, NY: Sharpe).
- Felin, Teppo and Foss, Nicolai J. (2009) 'Social Reality, the Boundaries of Self-Fulfilling Prophecy, and Economics', *Organization Science*, **20**(3), May, pp. 654-68.
- Field, Alexander J. (2001) *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity* (Ann Arbor: University of Michigan Press).

- Fishburn, Peter C. (1970) *Utility Theory for Decision Makers* (New York: Wiley).
- Friedman, Milton (1953) 'The Methodology of Positive Economics', in M. Friedman, *Essays in Positive Economics* (Chicago: University of Chicago Press), pp. 3-43.
- Gintis, Herbert (2006) 'Behavioral Ethics Meets Natural Justice', *Politics, Philosophy and Economics*, **5**(1), pp. 5-32.
- Gintis, Herbert (2007) 'A Framework for the Integration of the Behavioral Sciences', *Behavioral and Brain Sciences*, **30**(1), pp. 1-16.
- Gintis, Herbert (2009) *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton NJ: Princeton University Press).
- Glimcher, Paul W., Dorris, Michael C., and Bayer, Hannah M. (2005) 'Physiologic Utility Theory and the Neuroeconomics of Choice', *Games and Economic Behavior*, **52**, pp. 213-56.
- Gneezy, Uri and List, John A. (2006) 'Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments', *Econometrica*, **74**(5), pp. 1365- 1384.
- Hammermesh, Daniel S. and Soss, Neal M. (1974) 'An Economic Theory of Suicide', *Journal of Political Economy*, **82**(1), January-February, pp. 83-98.
- Harding, S. G. (ed.) (1976) *Can Theories be Refuted?: Essays on the Duhem-Quine Thesis* (Dordrecht: Reidel).
- Harrison, Glenn W. (1989) 'Theory and Misbehavior of First-Price Auctions', *American Economic Review*, **79**(4), September, pp. 749-62.
- Hausman, Daniel M. (1992) *The Inexact and Separate Science of Economics* (Cambridge and New York: Cambridge University Press).
- Henrich, Joseph, Boyd, Robert, Bowles, Samuel, Camerer, Colin, Fehr, Ernst, Gintis, Herbert and McElreath, Richard (2001) 'In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies', *American Economic Review (Papers and Proceedings)*, **91**(2), May, pp. 73-84.
- Henrich, Joseph, Boyd, Robert, Bowles, Samuel, Camerer, Colin, Fehr, Ernst, and Gintis, Herbert (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford and New York: Oxford University Press).
- Hindess, Barry (1977) *Philosophy and Methodology in the Social Sciences* (Brighton: Harvester).
- Hirshleifer, Jack (1985) 'The Expanding Domain of Economics', *American Economic Review*, **75**(6), December, pp. 53-68.
- Hodgson, Geoffrey M. (1988) *Economics and Institutions: A Manifesto for a Modern Institutional Economics* (Cambridge and Philadelphia: Polity Press and University of Pennsylvania Press).
- Hodgson, Geoffrey M. (2001) *How Economics Forgot History: The Problem of Historical Specificity in Social Science* (London and New York: Routledge).

- Hodgson, Geoffrey M. (2012) *From Pleasure Machines to Moral Communities: An Evolutionary Economics without Economic Man* (Chicago: University of Chicago Press), forthcoming.
- Kahneman, Daniel (1994) 'New Challenges to the Rationality Assumption', *Journal of Institutional and Theoretical Economics*, **150**(1), pp. 18-36.
- Kahneman, Daniel (2003a) 'A Psychological Perspective on Economics', *American Economic Review (Papers and Proceedings)*, **93**(2), May, pp. 162-8.
- Kahneman, Daniel (2003b) 'Maps of Bounded Rationality', *American Economic Review*, **93**(5), December, pp. 1449-75.
- Kahneman, Daniel, Knetsch, J. L. and Thaler, Richard H. (1986a) 'Fairness as a Constraint on Profit Seeking: Entitlements in the Market', *American Economic Review*, **76**(4), September, pp. 728-41.
- Kahneman, Daniel, Knetsch, J. L. and Thaler, Richard H. (1986b) 'Fairness and the Assumptions of Economics', *Journal of Business*, **59**, pp. 285-300.
- Kahneman, Daniel, Slovic, Paul and Tversky, Amos (eds) (1982) *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge and New York: Cambridge University Press).
- Koestler, Arthur (1959) *The Sleepwalkers: A History of Man's Changing Vision of the Universe* (London: Hutchinson).
- Kontopoulos, Kyriakos M. (1993) *The Logics of Social Structure* (Cambridge: Cambridge University Press).
- List, John A. (2006) 'Field Experiments: A Bridge between Lab and Naturally Occurring Data', *Advances in Economic Analysis and Policy*. **6**(2), Article 8, <http://www.fieldexperiments.com/uploads/97-full.pdf>
- Loomes, Graham (1998) 'Probabilities vs Money: A Test of Some Fundamental Assumptions About Rational Decision Making', *Economic Journal*, **108**(1), March, pp. 477-89.
- Marcus, Gary (2004). *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexity of the Human Mind* (New York: Basic Books).
- Neumann, John von and Morgenstern, Oskar (1944) *The Theory of Games and Economic Behavior* (Princeton: Princeton University Press).
- Penn, Anna A. and Shatz, Carla J. (1999) 'Brain Waves and Brain Wiring: The Role of Endogenous and Sensory-Driven Neural Activity in Development', *Pediatric Research*, **45**(4), pp. 447-458
- Platt, Michael L. and Glimcher, Paul W. (1999) 'Neural Correlates of Decision Variables in Parietal Cortex', *Nature*, July 15, **400**(6741), pp. 233-8.
- Popkin, Samuel L. (1979) *The Rational Peasant* (Berkeley: University of California Press).
- Posner, Richard A. (1980) 'A Theory of Primitive Society, With Special Reference to Law', *Journal of Law and Economics*, **23**(1), pp. 1-53.
- Robbins, Lionel (1932) *An Essay on the Nature and Significance of Economic Science* (London: Macmillan).

- Robson, Arthur J. (2001) 'Why Would Nature Give Individuals Utility Functions?', *Journal of Political Economy*, **109**(4), August, pp. 900-14.
- Samuelson, Paul A. (1937) 'A Note on the Measurement of Utility', *Review of Economic Studies*, **4**(2), February, pp. 155-61.
- Sarnecki, John (2007) 'Developmental Objections to Evolutionary Modularity', *Biology and Philosophy*, **22**(4), September, pp. 529-46.
- Sen, Amartya K. (1977) 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory', *Philosophy and Public Affairs*, **6**(4), pp. 317-44.
- Sen, Amartya K. (1987) 'Rational behaviour', in Eatwell, John, Milgate, Murray and Newman, Peter (eds) (1987) *The New Palgrave Dictionary of Economics*, (London: Macmillan), vol. 4, pp. 68-76.
- Siakantaris, Nikos (2000) 'Experimental Economics Under the Microscope', *Cambridge Journal of Economics*, **24**(3), May, pp.267-81.
- Simon, Herbert A. (1957) *Models of Man: Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting* (New York: Wiley).
- Simon, Herbert A. (1986) 'Rationality in Psychology and Economics', *Journal of Business*, **59**(4), Part 2, October, pp. S209-S224.
- Slovic, Paul and Lichtenstein, Sarah (1983) 'Preference Reversals: A Broader Perspective', *American Economic Review*, **73**(4), September, pp. 596-605.
- Smith, Vernon L. (1982) 'Microeconomic Systems as an Experimental Science', *American Economic Review*, **72**(5), December, pp. 923-55. Reprinted in Smith, Vernon L. (1992) *Papers in Experimental Economics* (Cambridge: Cambridge University Press).
- Sugden, Robert (1991) 'Rational Choice: A Survey of Contributions from Economics and Philosophy', *Economic Journal*, **101**(4), July, pp. 751-85.
- Tullock, Gordon (1994) *The Economics of Non-Human Societies* (Tuscon, Arizona: Pallas Press).
- Vromen, Jack J. (2010) 'On the Surprising Finding that Expected Utility is Literally Computed in the Brain', *Journal of Economic Methodology*, **17**(1), March, pp. 17-36.
- Wason, Peter C. (1983) 'Realism and Rationality in the Selection Task', in J. St. B. T. Evans (ed.) *Thinking and Reasoning: Psychological Approaches* (London: Routledge and Kegan Paul).
- Winter, Sidney G., Jr (1964) 'Economic "Natural Selection" and the Theory of the Firm', *Yale Economic Essays*, **4**(1), pp. 225-72.