The Rationality Postulate
in Economics:
Its Ambiguity, its Deficiency
and its Evolutionary Alternative

Viktor J. Vanberg
03/3
The Rationality Postulate in Economics: Its Ambiguity, its Deficiency and its Evolutionary Alternative

Viktor J. Vanberg
03/3
The Rationality Postulate in Economics: Its Ambiguity, its Deficiency and its Evolutionary Alternative

by

Viktor J. Vanberg

Albert-Ludwigs-Universität Freiburg
Wirtschafts- und Verhaltenswissenschaftliche Fakultät
Abteilung Wirtschaftspolitik

und

Walter Eucken Institut

1. Introduction
2. Rationality Principle vs. Rationality Hypotheses
4. The ‘Economic Approach’ Between Rationality Principle and Rationality Hypotheses
5. An Evolutionary Alternative to the Rational Choice Paradigm
6. J.H. Holland’s Theory of Rule-Based Adaptive Agents
7. Rationality and Markets
8. Conclusion

Abstract: Though the rationality postulate is generally considered the paradigmatic core of economics, there is little agreement about its specific content and methodological status. This paper seeks to clarify some of the ambiguity surrounding the postulate by drawing a distinction between the non-refutable, purely heuristic rationality principle on the one side and refutable rationality hypotheses on the other. An alternative, evolutionary outlook at purposeful, problem-solving human behavior is outlined that captures much of what makes the rationality postulate attractive to economists while avoiding the kinds of problems that have made it the subject of enduring criticism.

Key words: Rational choice theory; rationality principle; rule-following; program-based behavior.

* This paper has largely been written during my visits to the School of Economics at the University of Queensland and the Research School of Social Sciences at the Australian National University, August-December of 2002. I would like to thank both institutions, in particular John Foster and Geoffrey Brennan, for their support and hospitality as well as for the stimulating intellectual environment they provided.
1. Introduction

The rationality postulate is generally considered the paradigmatic core of economics and it has, in recent decades, gained increasing prominence in other social sciences as well, in particular in political science and in sociology. It has also been, and continues to be, the principal target of long-standing and enduring criticism of the “economic approach” from heterodox quarters within the economics profession as well as from other social and behavioral sciences. Given its central paradigmatic role as well as the long lasting and extensive debate on its validity, it is somewhat surprising that, among its advocates and its critics alike, considerable ambiguity persists both as to the specific content and the methodological status of the rationality postulate (Laville 2000: 407; Lindenberg 2001: 635f.). Many of the postulate’s advocates treat it as an empirically testable explanatory theory, while others portray it as a non-refutable axiomatic doctrine or a “metaphysical statement” (Boland 1981: 1034), and still others regard it as a normative perspective that “tells us how, as rational agents, we ought to choose” (Sugden 1991: 752). The same ambiguity is to be found among those who criticise or reject the rational choice model. While some critics accuse it of being empirically empty, others blame it for being in blatant conflict with reality, and others still consider it normatively inadequate.

1 R. Sugden (1991: 751): “In mainstream economics, explanations are regarded as ‘economic’ to the extent that they explain the relevant phenomena in terms of the rational choices of individual economic agents.” – Hogarth and Reder (1986: 2): “To provide an economic explanation of observed behavior is to show that the actions of the relevant decision makers conform to the rational choice paradigm and that the behavior of multi-person aggregates is the result of individual choices made according to that paradigm.” – D.K. Foley (1998: 23): “The cornerstone of received economic theory is the idea that human agents behave rationally. Rationality is supposed to underlie the predictability of human behavior, and thus to establish it as a candidate for systematic scientific investigation.”
3 See M. Hechter and S. Kanazawa 1997, D. Heckathorn 1997 and the journal Rationality and Society, founded by J.S. Coleman, the most prominent proponent of rational choice sociology.
4 For an informative, detailed discussion of the main criticisms mounted against the rationality postulate in economics see F. Laville 2000.
The rational choice approach as it is commonly understood includes a number of ingredients that can, at least conceptually, be distinguished from each other. In particular three components seem to be generally associated with rational choice theory, namely methodological individualism, the self-interest assumption, and the rationality postulate itself, i.e. the assumption that individuals pursue what they consider their interest in a rational manner. I want to stress at the outset that my concern in this paper is only and exclusively with the last of the three noted components of rational choice approaches. I do not intend to take issue at all with methodological individualism or the self-interest assumption but shall focus only on the role that the rationality postulate itself can play in economics and, more generally, in the social sciences. And it is the ambiguity of the rationality postulate in this specific sense that is of concern here.

How one judges the usefulness and validity of alternative interpretations of the rationality postulate obviously depends on what the “theory” entailing it is intended to accomplish. The persistent disagreement within economics on the particular content of this postulate is certainly due, in part, to the lack of agreement among members of the profession on what “economic theory” is meant to achieve. Taking a Popperian stance in this matter I shall focus in this paper exclusively on the role the rationality postulate can play in an economics that views itself as an empirical, explanatory social science. If economics is to explain real world social phenomena it must employ empirically refutable explanatory hypotheses as its nomological basis, and if the rationality assumption is to be the fundamental behavioral law that constitutes the nomological basis of economics it must qualify as an empirically refutable conjecture.

Within the confines noted the purpose of the present paper is twofold. First, I shall seek to clarify some of the ambiguity surrounding the rationality postulate, and to identify the reason for some of its shortcomings as an explanatory conjecture. And, second, I shall outline an evolutionary alternative to rational choice theory, namely the paradigm of adaptive rule- or program-based behavior, a paradigm that captures much of what appears to make the rationality postulate so attractive to economists, without being burdened with the kinds of problems that have made the rationality postulate so controversial.
2. Rationality Principle vs. Rationality Hypotheses

Rational choice theorists generally share the notion that human behavior is self-interested in the sense that humans seek to advance what they consider to be in their own interest, given the constraints that they face. Yet, one finds significantly less agreement within the rational choice camp when it comes to the question of how, specifically, we are to model the behavior of the human agents that populate the social world. On closer inspection it becomes obvious that the seemingly simple formula, that humans are to be modelled as agents who rationally choose what they consider to serve their interest best, is interpreted by advocates and critics in a variety of, sometimes quite different, ways. A main reason for the ambiguity surrounding the use of the rationality postulate can, in my view, be located in the failure, on part of its advocates as well as of its critics, to clearly distinguish between the non-refutable, purely heuristic rationality principle on the one side and refutable rationality hypotheses on the other.

Before discussing the difference between them, a few remarks on the general notion of rationality in human action are in order. Inherent to rational choice theory as well as to everyday concepts of rational behavior is the notion that human action is purposeful, forward-looking, governed by its expected consequences or anticipated effects. For a person to act rationally means to select, in a given choice situation, from among available alternatives the action that she expects or predicts to lead to the most preferred consequences. Such predictions or expectations of consequences quite obviously have to be based on beliefs about the world, i.e. on conjectures or theories about what outcomes will actually follow from actions that are contemplated as potential alternatives. That expectation or anticipation is involved means nothing other than that behavior cannot be a function only of external constraints and preferences over outcomes. K. J. Arrow (1996: xiii) points to the relevant issue here when he notes: “Choice is over sets of actions, but preference orderings are over consequences.” A choice reflects not only the actor’s subjective preferences over

---

6 K. Binmore (1998: 357): “A rational person uses his reason to choose an action from his feasible set that maximizes his expected utility relative to his beliefs.”

7 The distinction I want to draw attention to is similar to, but not identical with, distinctions drawn by other authors, such as e.g. K.D. Opp's (1999: 172) contrast between a "narrow" and a "wide" version of rational choice theory, or J. Ferejohn's (1991: 282) distinction between a "thin" and a "thick" version.
outcomes but also his “knowledge of the relation between actions and consequences” (ibid.). In other words, all purposeful action is preference- as well as theory-guided.

By *rationality principle* I mean the notion that human action is rational, given the purposes (or preferences) and beliefs (or theories) of the actor at the time of action. This notion is about the *subjective consistency* of human action or, more precisely, about its *local* consistency, “local” in the sense that reference is only made to the actor’s purposes or preferences and beliefs at the moment of choice. In essence, the rationality principle says no more about human action than that it is *purposeful* and that it can be “understood” as meaningful in light of the preferences, goals or purposes and of the beliefs or theories that inform the actor’s choice. Whatever the nature of an actor’s preferences or beliefs may be, in terms of the rationality principle his actions count as rational as long as they are consistent with what his preferences and beliefs happen to be at the moment of choice.

A rational choice theory that were to include only the rationality principle as its 'explanatory conjecture' would clearly be empirically empty, since the rationality principle, in and by itself, does not rule out any conceivable event within its explanatory domain, i.e. purposeful human action. It is difficult for us even to imagine a potentially falsifying case. This is not to say that the rationality principle has no useful role to play. There are essentially two ways in which it can be interpreted. Either as a *definition*, specifying what we mean when we speak of "purposeful human action." Or as a *heuristic principle* that tells us how we should seek to explain human action.

In the context of Ludwig von Mises' (1949) praxeological approach to economics, for instance, the rationality principle is clearly used as a mere definitional statement. As Mises notes: "Human action is purposeful behavior. Or we may say: Action is ... aiming at ends and goals" (ibid.: 11). To say that human action is

---

8 R. Boyd and P.J. Richerson (1993: 133): “Everyone must acquire beliefs about the world before they can optimize.” – Most often in definitions of the concept of rational choice the belief- or theory-component is not explicitly separated from the evaluative preference-component. The theory component is, for instance, “hidden” in the phrase “most effective means” when Hargreaves Heap et al. (1992: vii) define: “A rational choice is one which selects the most effective means to satisfy the agent’s preferences.”

9 The terms “purpose” and “preference” are used here in parallel even though they are, of course, not synonyms. They both refer, though, to the evaluative component in human choices. “Preference” can be said to be the economists’ technical term for the evaluative dimension of the term “purpose.” What all these terms have in common is that they are about the criteria by which agents evaluate the (expected) outcomes of their actions.
"rational" is, according to Mises, the same as saying that it is subjectively meaningful, purposeful or goal-directed. He therefore concludes: "Human action is necessarily always rational. The term 'rational action' is therefore pleonastic and must be rejected as such" (ibid.: 18). As a definitional statement the rationality principle specifies what we mean when we speak of “purposeful action.” Behavioral responses that would not qualify as “rational” in the sense of the rationality principle, would simply fall outside that category. They would have to be classified as belonging to a different category of behavioral responses, such as unintended body movements, like somebody stumbling accidentally, or purely mechanical reflexes, such as the famous knee-jerk reflex.

If it is not meant as a purely definitional convention, the rationality principle may, alternatively, be viewed as a heuristic device, i.e. as a principle that tells us how we should go about explaining purposeful human action. It tells us that we ought to explain such action in terms of the actor's purposes and beliefs, and that we should do so under the presumption that purposeful actions are consistent in terms of the actor’s purposes and beliefs (at the moment of choice). It suggests, in other words, that we should ignore the possibility of “irrational” or “inconsistent” actions, concentrating our attention instead on identifying the purposes and beliefs that "make sense" of the actions that are to be explained.

Whether it is meant as a definitional statement or as a heuristic device, in neither case does the rationality principle qualify as an empirically contentful, 

---

10 In reference to non-adapted, i.e. instrumentally inadequate action Mises (1949: 20) notes: "An action unsuited to the end sought falls short of expectation. It is contrary to purpose, but it is rational, i.e., the outcome of a reasonable - although faulty - deliberation and an attempt - although an ineffectual attempt - to attain a definite goal."

11 Mises (1949: 11): "Conscious or purposeful behavior is in sharp contrast to unconscious behavior, i.e., the reflexes and the involuntary responses of the body's cells and nerves to stimuli." - "The opposite of action is not irrational behavior, but a reactive response to stimuli on the part of the bodily organs and instincts which cannot be controlled by the volition of the person concerned" (ibid.: 20).

12 In K. R. Popper's (1994: 181) view the adoption of the rationality principle "reduces considerably the arbitrariness of our models, an arbitrariness which becomes capricious indeed if we try to do without this principle." – In the context of his comments on the methodology of the social sciences and, in particular, on the explanatory approach that he calls “situational logic or situational analysis” Popper refers to the “rationality principle” as the central “animating law” (ibid.: 169) contained in the social theorists’ situational models. Popper’s failure in this context to distinguish between the non-refutable rationality principle and empirically contentful rationality hypotheses seems to be the source of the inconsistencies in his concept of “situational logic” that have been pointed out by critics such as N. Koertge (1979), B.J. Caldwell (1991: 13ff.) and Vanberg (1975: 109ff.). – For a detailed discussion of this issue see Vanberg (2002: 18ff.).
falsifiable conjecture. It cannot be true or false, it can only be more or less useful for its intended purpose. Only when the rationality principle is supplemented by additional assumptions about actors’ purposes and beliefs can we speak of rationality hypotheses in the sense of empirically refutable claims. And only then can the rationality postulate serve as the nomological basis of an empirical, explanatory social science.

The rationality principle may easily be mistaken for an empirically contentful conjecture because its specific ‘applications’ – in the sense of statements such as: “Person A did X because she sought to achieve purpose Y and held the belief Z” – make assertions that can, indeed, be factually true or false. To deny that the rationality principle as such is empirically refutable is to argue that no refutable claim is made by the assertion that actors do what they consider to be their best choice, given their preferences and beliefs. The conjecture that a person A did in fact exhibit action X because of her purpose Y and her belief Z may, of course, be a refutable statement in the sense that the particular assertions made about her purposes and beliefs may be false. Yet, that such statements are nothing other than singular statements about time- and place-specific events. As such they do, of course, not qualify as general hypotheses or nomological statements. And the fact that such specific ‘applications’ are, at least in principle, refutable does not mean that the rationality principle itself is necessarily refutable. The fact that a heuristic principle instructs us to look at variables about which refutable statements can be made does not make the principle

---

13 In order to qualify as an empirically contentful conjecture a statement must exclude as factually impossible conceivable events, i.e. events that the human mind can imagine. One can imagine, and in fairy tales this happens, that humans can fly if they move their arms like birds move their wings. The theory of gravity is empirically contentful, and therefore has explanatory power, because it rules this out as factually impossible. One can also imagine that governments act as omniscient and benevolent maximisers of the common weal, and in traditional welfare economics this seems to have been a not uncommon implicit assumption. Because the conjecture that political agents are boundedly rational, and self-interested agents rules this out as factually impossible it qualifies as an empirically contentful hypothesis.

14 When it is argued that “the maximization hypothesis is neither directly … nor indirectly testable” (B. Caldwell 1983) one has to take into account that there is, of course, a difference between a hypothesis being stated in a way that makes it in principle unfalsifiable (because it does not exclude any conceivable state of the world) and a hypothesis making claims that are extremely difficult or even impossible to refute, given the present state of our knowledge and our presently available research methods. Since persons’ purposes and preferences as well as their beliefs are not directly observable but can, at best, be inferred from observable facts that are, on theoretical grounds, considered to be reliable indicators of purposes, preferences or beliefs, empirical testing of conjectures of the noted type may be difficult. But this is only of secondary importance in the present context. The principal issue is whether assertions are made that are at least in principle refutable, independent of the difficulties that actual attempts at refuting them may pose.
itself a testable conjecture. In order for that to be the case, the principle itself would have to rule out events that could potentially happen. The rationality principle does not rule out any conceivable combination of preferences and beliefs, requiring only that they are logically consistent with the chosen action. The only actions it rules out, namely those inconsistent with the actors' preferences and beliefs, seem impossible for us to even imagine as potential realities.

In order to qualify as an empirically contentful, refutable rationality hypothesis the rationality postulate must go beyond the rationality principle in the sense of including substantive claims about the kinds of preferences and beliefs that inform human action, thereby imposing certain general constraints on what in our explanations of actions we are allowed to assume about the actor’s purposes and beliefs, constraints that go beyond requiring mere local consistency. It is only to the extent that the rationality postulate includes such conjectures that some general discipline is imposed on the range of permissible assumptions about preferences and beliefs that can be employed in rational choice explanations.

There are, in particular, two kinds of such substantive conjectures about preferences and beliefs that have been employed, in different specific versions and various combinations, in rational choice approaches. These are, on the one hand, claims about the content of persons’ preferences, purposes or goals (i.e. their “utility functions”) and claims concerning the overall, not just “local” (i.e. at the time of action), consistency of their preferences and beliefs. And these are, on the other hand, claims concerning the factual adequacy of the “beliefs about the world” that people hold.

According to the first claim human preferences and beliefs are not only locally consistent with actions at the time of actions, but a person’s entire system of preferences and beliefs is consistent in a more inclusive sense, across domains and over time. Rationality hypotheses of this kind state that human action is rational, given the actor’s entire system of preferences (or purposes) and beliefs. The conjecture is that what an actor does is consistent with all the various purposes he pursues and all the beliefs he holds, not just the purposes and beliefs “active” at the moment of choice. By contrast to the notion of “local” consistency implied in the rationality principle, rationality hypotheses of this kind can be said to be about “global” consistency of purposes and beliefs. An action, for instance, that is carried out as an “emotional response” to a situational stimulus (e.g. an insult, a sensual temptation, a
threat) is rational in the sense of being “locally consistent” with the agents preferences and beliefs at the moment of action, but it may well not be “globally consistent” in the sense that it is inconsistent with the agent’s more inclusive system of preferences and beliefs or, in simple terms, with his well considered judgement.

According to the second claim human action is rational, given the factual nature and state of the world. Implied in this claim is that purposes or preferences and beliefs are not merely internally consistent but that humans act on “objectively” adequate [preferences and] beliefs about the world, i.e. [preferences and] beliefs that allow them successfully to operate in the world as it is. Rationality hypotheses of this kind are about the actual adaptedness (Zweckmaessigkeit) of beliefs to the objective environment in which the actor operates.

By contrast to the rationality principle, rationality hypotheses are empirically refutable statements about facts. They can be true or false. Such hypotheses may come in many varieties, and it is not always immediately obvious, but may have to be judged in terms of the respective context, whether a particular statement represents, in fact, a refutable rationality hypothesis or is, in essence, not more than a more elaborate restatement of the rationality principle. This is true not least for the maximization hypothesis of neoclassical economics. In its most extreme interpretation, claiming perfect global consistency of preferences and beliefs as well as their perfect adaptedness to the agent’s problem environment, it is a very demanding conjecture, and one that is in such apparent conflict with reality that it is hardly considered by anybody to be descriptive of actual human behavior. Yet, the maximization hypothesis can also be, and often is, interpreted in ways that make it compatible with all conceivable behavior – thus turning it into the non-refutable rationality principle.

---

15 A person’s ability to “successfully” operate in the world as it is, in the sense of achieving her long term enlightened interests, depends, of course, not only on the adequacy of her beliefs about the factual nature of the world but also on the prudence of the immediate goals she pursues or on her immediate preferences. This issue of the adaptedness of a person’s preferences to the nature of the environment in which she lives is not separately discussed here because it does not raise any new problems that could not be discussed in terms of the adaptedness of beliefs, relative to a person’s ultimate or fundamental preferences.

16 D.K. Foley (1998: 66): “Social theory based on the rational choice paradigm seeks to explain agents’ representations of their social world as more or less accurate reflections of its true nature.”

17 A strong version of the rationality hypothesis is implied when J. Elster (1986: 16) enumerates as requirements that “a rational-choice explanation of an action” would “ideally” satisfy: “The action is the best way for the agent to satisfy his desire, given his belief; the belief is the best he could form, given the evidence … . Both the belief and the desire must be free of internal contradictions.”

18 See the extensive debate on “anomalies” observed in behavioral experiments (D. McFadden 1999; see also Laville 2000: 399f.). I shall return to this issue below.
All that is needed for such transformation is to leave the content of the “objective function” or “utility function” unspecified, to be filled by any substantive assumption that the analyst may find convenient in applying the maximization “hypothesis.”


Whether particular versions of rational choice theory do, in fact, entail refutable rationality hypotheses and, thus, go beyond the purely heuristic rationality principle, is by no means always immediately obvious. A case in point is G.S. Becker’s economic approach to human behavior, an approach for which he claims that it “is applicable to all human behavior, be it behavior involving money prices or imputed shadow prices, repeated or infrequent decisions, emotional or mechanical ends, rich or poor persons, men or women, adults or children, brilliant or stupid persons, patients or therapists, businessmen or politicians, teachers or students” (Becker 1976: 8). As I shall seek to show, despite its appearance as a particularly rigorous modern interpretation of the “economic model of man” Becker’s rationality concept ultimately represents no more but a refined version of the non-refutable rationality principle.

In order to avoid "weaknesses in the received theory of choice" (Becker 1976: 133) that result from its "reliance on 'changes in tastes' in interpreting observed behavior" (ibid.), Becker proposes a new approach to the theory of consumer behavior that puts "greater emphasis on income and price effects" (ibid.: 144), reinterpreting consumer behavior as productive activity that uses ordinary market goods as inputs for the production of ultimate goods or pleasures (Becker 1996: 5), that are the only direct objects of utility.

The essential implication of Becker's restatement is that the observable demand for ordinary (market) goods - and, in fact, the "demand" for any kind of behavior - is interpreted as derived demand, derived from intra- and inter-personally constant preferences for ultimate goods and intra- and inter-personally variable capacities for producing these goods. The critical link between ultimate preferences and derived demand is provided by the consumer's (household's) production function for ultimate goods:

---

19 D.K. Foley (1998: 23): “The hypothesis of rationality puts no observational restrictions on an agent’s actions. We can always rationalize behavior by positing an appropriate objective function.”
In this equation, \( x_i \) denotes inputs in the form of ordinary goods, \( t_i \) denotes the consumer's own time input, and \( E \) refers to so-called "environmental variables" including, in particular, the production technology that consumers employ in producing ultimate pleasures.

The name "environmental variable" would clearly seem to suggest, that "E" refers only to inter-subjectively observable aspects of the "environment in which the production takes place" (Becker 1976: 135). In fact, however, the E-variable and its main component, "production technology," are interpreted by Becker in a way that does very little to restrict the range of admissible explanatory arguments. If, as Becker explains (ibid.: 145; Stigler and Becker 1977: 82, 84), "production technology" includes such things as personal skills and human capital as well as people's knowledge - "whether real or fancied" (Stigler and Becker 1977: 84, emphasis added) - it seems that "rationality" in Becker's account entails, ultimately, no stronger claim than the rationality principle, namely (local) consistency between people's actions and their preferences and beliefs. The presumption of invariable ultimate preferences does not impose effective limits on what we may assume about people's derived demands if what people "correctly or incorrectly" (ibid.) believe about the world provides the critical link between the two, allowing all kinds of subjective constraints to determine behavior in addition to objective constraints.20

In a further developed version of his approach to "utility-maximizing forward-looking behavior" (Becker 1996: 140) Becker has put special emphasis on the "human capital perspective" (ibid.), drawing attention to the impact of people's current choices on their future "human capital," i.e. their future capacity for producing ultimate goods and their future (derived) preferences for ordinary goods.21 The purpose of this "extension of the utility-maximizing approach (is) to include endogenous preferences" (ibid.: 4). It is meant to retain "the assumption that individuals behave so as to maximise utility while extending the definition of individual preferences to include

---

20 Of course, objective constraints trump subjective constraints. The laws of gravity, for instance, will overrule a person's beliefs that he can fly. Differences in beliefs can have relevant effects on behavior only within the limits set by objective constraints.

21 Becker (1996: 7): "Current behavior may raise future personal capital ... (F)orward-looking persons recognize that their present choices and experiences affect personal capital in the future, and that future capital directly affects future utilities."
personal habits and addictions, peer pressure, parental influences on the tastes of children, advertising, love and sympathy, and other neglected behavior" (ibid.).

While the new extension may enrich Becker's conceptual and analytical apparatus, it does not alter the fact that his "economic or rational choice approach to behavior" (ibid.) rests, as far as its conjectural foundations are concerned, on no more than the purely heuristic rationality principle. This means, whatever empirical content specific "applications" of this approach may be said to have, such content cannot be derived from the irrefutable, but empirically empty, notion "that individuals maximize welfare as they conceive it" (ibid.: 139, emphasis in the original). It must derive from other sources.

In his Nobel Lecture of 1992, "The Economic Way of Looking at Life," Becker noted that in his work he had "intentionally chosen certain topics - such as addiction - to probe the boundaries of rational choice theory" (ibid.: 155). Yet, to show that even addictive behavior is covered by the rationality principle, i.e. that it is rational given the person's preferences and beliefs at the moment of choice, does not seem much of a "probing," in particular if addictions are acknowledged to decrease "the capacity to anticipate future consequences" (ibid.: 11), or to "reduce the attention to future consequences" (ibid.: 120). Ultimately, Becker seems to claim no more than consistency between an individual's actions and his preferences and beliefs when he insists that addictions, habits or other forms of seemingly "irrational" conduct are "rational" in the sense that they are forward-looking, responsive to "changes in prices and wealth" (ibid.: 123), and do "not imply a reluctance to 'calculate'" (ibid.).

Though he insists that his approach maintains the assumption of forward-looking, maximising and consistent choices (ibid.: 22f.), Becker notes that his "type of rationality ... is quite different ... than that found in standard models" (ibid.: 23). And, indeed, there is not much left of the rigid constraints that the "hard" version of rational choice theory seems to impose on our explanatory efforts if one concedes, as Becker (ibid.: 9) does, that rationality "does not imply perfect foresight, or even accurate calculation of the probabilities of future events," but "implies only that

---

22 On the relation between the "extended" and the previous version of his "economic approach" Becker (1996: 5) notes: "In a more fundamental approach, utility does not depend directly on goods and consumer capital stocks, but only on household-produced 'commodities,' such as health, social standing and reputation, and pleasures of the senses. The production of these commodities in turn depends on goods, consumer capital, abilities, and other variables."

23 Becker (1996: 11): "I believe that even extreme forms of addictive behavior, such as heavy smoking or drinking, involve forward-looking, consistent utility-maximization."
individuals try as best they can to anticipate the future consequences of their present behavior. Becker softens his concept of rationality even further when the acknowledges the relevance of habitual behavior. Noting that "cooperation can be sustained more easily ... when individual behavior is habitual" (ibid.: 17), Becker explains: "For if individuals are habitual, and if they were cooperative in the past, they might continue to be cooperative even if they could gain an advantage from uncooperative behavior" (ibid.).

With regard to such extensions one may well doubt whether Becker's version of the assumption of "utility-maximising behavior" can be more than an empty formula, accommodating all conceivable kinds of behavior, if it allows for the fact that, in their efforts to "try as best as they can," individuals "may have imperfect memories, ... may discount the future 'excessively,' ... (and) may make erroneous calculations" (ibid.: 22), as well as for the influence of "habits, childhood and other experiences, and culture" (ibid.: 23). In fact, when Becker describes his own rational choice approach as "a method of analysis" (ibid.: 139) it seems that he himself regards it as a heuristic device rather than an empirical theory that consists of refutable conjectures about human behavior. To be sure, if it is meant as a heuristic device, Becker may justly claim "that no approach of comparable generality has yet been developed" (ibid.: 155). But such "generality" is simply due to the fact that the core principle, "the assumption of individual rationality" (ibid.: 156), is specified in a way that makes it an empirically empty – even if heuristically useful - principle. Such "generality" should, however, not be confused with the explanatory power of an empirically refutable theory that makes widely applicable claims about the nature of the world. The relevant test of the latter is not the ease with which evidence can be provided that is compatible with the theory. In fact, every conceivable behavior is compatible with the empirically empty rationality principle. The relevant test is in the

24 Becker (1996: 9): "I believe the main reason habitual behavior permeates most aspects of life is that habits have an advantage in the biological evolution of human traits." – Such acknowledgement of the relevance of habitual behavior, and its evolutionary origins, would seem to lend support to the heuristic of "program-based behavior" that I discuss below (section 5) as an alternative to the rational choice heuristic.

25 Becker (1996: 6) leaves significant room for interpretation when he notes about the methodological status of his approach: "Our assumption that extended preferences are stable was intended not as a philosophical or methodological 'law,' but as a productive way to analyze and explain behavior."

26 Becker (1996: 156) speaks of "the analytical power provided by the assumption of individual rationality."
ease with which one can imagine potentially observable evidence that would *contradict* or refute the theory.

4. The ‘Economic Approach’ Between Rationality Principle and Rationality Hypotheses

In their statements on what the rationality postulate entails economists seem to oscillate between the non-refutable rationality principle and the amply refuted maximisation hypothesis, depending on whether they seek to defend the rationality postulate against critique or whether they wish to emphasise its analytical rigor.

In response to criticism of the “economic” or “rational choice” approach to human behavior economists often seem to feel called upon to defend the rationality principle (Laville 2000: 417). Yet, the rationality principle is obviously not in need of any defence. It is not disputed and, in fact, cannot be disputed. For any conceivable action, including the most bizarre, there are purposes and beliefs (also very bizarre) conceivable that make it consistent. Often economists, when they defend the rational choice model, argue for what one may call “soft” rationality hypotheses, such as the claim that people normally do not act in outright inconsistent ways, or that they are not oblivious to incentives. Such “soft” rationality hypotheses are, to be sure, disputable, but many (if not most) of them are in fact hardly under any dispute within the behavioral sciences. Hardly any serious student of human behavior outside of economics would be prepared to take issue with the claim that humans “take advantage of obvious opportunities.”

While in their efforts to demonstrate the plausibility of their rational choice approach economists tend to appeal to the unquestionable rationality principle or to uncontroversial soft versions of the rationality postulate, when their ambition is to stress the rigor and analytical power of their formal models they typically take recourse to “strong” versions of the rationality hypothesis, such as the maximisation hypothesis of orthodox neoclassical economics. It is these strong versions of the

---

27 R.E. Lucas (1977: 15): “Even psychotic behavior can be (and today, is) understood as ‘rational’ given a sufficiently abnormal view of relevant possibilities.”

28 As Simon (1986: 26) notes: “I emphasize this point of agreement at the outset – that people have reasons for what they do – because it appears that economics sometimes feels called on to defend the thesis that human beings are rational. Psychology does not quarrel at all with this thesis. If there are differences in viewpoint, they must lie in conceptions of what constitutes rationality, not in the fact of rationality itself.”

29 P. Krugman (1998: 120): “Economic actors are intelligent in the sense that they take advantage of obvious opportunities.”
rationality hypothesis that are the principal subject of controversy, such as, in particular, the assumption of perfect (global) consistency and perfect adaptedness in human action or, as it is usually called, the assumption of perfect rationality. This assumption is in such apparent conflict with empirical reality that, even among its proponents, it is rarely ever defended on the grounds that it adequately depicts how real human beings behave. In fact, as R.J. Aumann (1997: 2) has noted, “economists have long expressed dissatisfaction with the complex models of strict rationality that are so pervasive in economic theory,” as witnessed by Jean Tirole who, in his presidential address to the European Economic Association in 2001, frankly declared that “economists have of course long been aware of the crudeness” (Tirole 2002: 634) of the rational maximization hypothesis and its “deliberately simple-minded description of human preferences and behavior” (ibid.). The name that for a long time has been the symbol for the critique of the empirical inadequacy of the rationality hypothesis in economics is, of course, that of Herbert A. Simon who, over decades, has persistently argued that, if it is to gain explanatory power as an empirical science, economics must replace its model of “Olympian rationality” by a theory of “imperfect” or “bounded” rationality, that provides a more adequate account of how real people go about solving the choice-problems they face.

There are essentially three ways in which economists have tended to respond to the issue of the empirical inadequacy of the maximization hypothesis. One common response has been to argue that, even though it may not be descriptive of human behavior in all its dimensions, the maximization hypotheses is a good approximation of human behavior in the particular kind of social environments that economists focus their attention on, namely the competitive environments of markets. I shall postpone a

---

30 In answer to the question of why “economic analyses come in for so much criticism both from within and outside the profession” Krugman (1998: 120) notes, that it is because of “extreme formulations” such as the assumption “that intelligent behavior goes all the way to strict maximization of some objective function.”

31 K.J. Arrow (1986: 213): “The main implication of this extensive examination of the use of the rationality concept in economic analysis is the extremely severe strain on information-gathering and computing abilities. Behavior of this kind is incompatible with the limits of human beings.” – R.R. Nelson and S.G. Winter (2002: 29): “The specific kind of rationality that economists usually build into their theories typically implies, or at least connotes, careful deliberation and attempted foresight. Real actors, however, simply do not have the vast computational and cognitive powers that are imputed to them by optimization-based theories.”

32 In numerous contributions H. A. Simon has pointed to “the discrepancy between the perfect human rationality that is assumed in classical and neoclassical economic theory and the reality of human life” (Simon 1992: 3), arguing that neither people’s “knowledge nor their power of calculation allow them to achieve the high level of optimal adaptation of means to end that is posited in economics” (ibid.).
discussion of this argument until later in this paper (section 7) when the context for its adequate evaluation is better prepared.

Another not uncommon response has been to insist on the analytical value of the perfect rationality assumption despite its apparent lack of realism, simply by denying the adequacy of a falsificationist methodology for economic theory. According to this interpretation, and this is the interpretation that seems to underlie de facto the whole of formal general equilibrium analysis, it is the very purpose of economic theory to work out the implications that follow if one assumes that economic agents are perfectly rational. On this interpretation economic theory is not seeking to directly explain real world phenomena, it is, instead, about explicating the functional properties of a world populated by “completely rational individuals,” i.e. individuals who have “the ability to foresee everything that might happen and to evaluate and optimally choose among available courses of action” (Kreps 1990: 745). The (internal) validity of such a theory is, obviously, totally independent on whether or not real human agents come even close to being “completely rational.” It is an axiomatic exercise that informs us about how the world would function if it were populated by perfectly rational individuals. Yet, such a theory does not per se tell us anything about the “real world,” i.e. the world populated by people as they de facto are.

The third strategy economists tend to employ in response to the empirical inadequacy of the maximization hypothesis is to modify their behavioral model with the aim of making it more realistic without giving up its basic logic. This is typically done by “enriching” the utility function (i.e. the assumptions about agents’ goals) in ways that accommodate observations incompatible with more standard interpretations of the maximization hypothesis. The problem with such attempts to rescue the rationality postulate is that they tend to rely on “ad hoc assumptions, for the reason that they are designed in order to account for the latest empirical violations” (Laville 2000: 404). Such unprincipled ad hoc “enrichment” of the rationality hypothesis in effect renders the hypothesis itself non-refutable, emptying it of any empirical content and transforming it into the unobjectionable rationality principle.

---

33 D. M. Kreps (1990: 746): “Most economic theory concerns completely rational individuals. ... Completely rational individuals ... populate the economies of general equilibrium.”
That a strategy of an unprincipled, ad hoc adjustment of the maximization hypothesis to explanatory needs is not without problems has, of course, not escaped the attention of economists. One field in which economists have been most sensitive to this issue is game theory, especially where it has been extended into experimental research. Authors like R. Selten (1998) and W. Gueth (1995) have in recent years expressed a concern that a principled alternative to, rather than ad hoc modifications of the rationality postulate is needed if one is to provide a satisfactory explanatory account of well documented experimental findings.

Selten draws a distinction between “normative game theory” and “descriptive game theory” (Selten 1998: 22f.), defining the former as an enterprise that aims at working out the implications of the assumption “of ideal full rationality in interactive decision situations” (ibid.: 28). According to Selten, normative game theory is a purely “philosophical,” not an empirical enterprise: “Empirical arguments are really irrelevant in this discussion” (ibid.: 23). By contrast, descriptive game theory aims at explaining “observed behavior” (ibid.). It deals with empirical problems “and only empirical arguments count, nothing else” (ibid.). The need to draw the distinction between the two types of game theory arises, according to Selten, because there is experimental evidence from human game players which refutes the assumption of ideal full rationality and which requires us “to develop a descriptive game theory about human players” (ibid.). His own search for such a theory has led Selten in the direction of a theory of “bounded rationality,” a “learning-direction theory” (ibid.: 24) that is “based on a principle of ex-post rationality” (ibid.), concerned “with the analysis of the past” (ibid.), a theory which is not merely a modified version of the model of “ideal full rationality” but exhibits a “different structure” (ibid.: 25, 31).

W. Gueth’s long lasting research with ultimatum bargaining experiments has led him to insist, like Selten, on “the need to supplement normative game theory by a

---

34 Tirole (2002: 636) recognizes these problems when he notes: “Clearly, we, as a profession, should not impulsively add a new element into the utility function every time we cannot readily explain a behavior or an apparent concern.”

35 J.C. Harsanyi (1977: 16) also notes that the assumption of perfectly rational behavior as used in game theory “is a normative (prescriptive) theory rather than a positive (descriptive) theory.”

36 R. Selten (1998: 11): “Behavior in such situations is hence not really an empirical question, you just deduce it from the principle of rationality. This is the position I call naïve rationalism. It has prevailed in economics for a long time and it is still a very strongly held position in economics – but it is weakening.”

37 Selten (1998: 28): “The results of our experiment suggest that the structure of the bounded strategic rationality of experienced human players is quite different from that of ideal strategic rationality.”
behavioral theory of game playing” (Gueth 1995: 338), an insight, he argues, that should not be surprising to anyone who recognizes game theory for what it is, a “theory of perfectly rational individual decision making which pays no attention at all to the cognitive limitations of human decision makers and which therefore can hardly be expected to explain human decision making” (ibid.). As Gueth notes, a common response of economists to the kind of “anomalies” that ultimatum bargaining experiments or other sources of empirical evidence present to them is to include “additional arguments of utilities” (ibid.: 342), such as a “preference for fairness,” to get a better fit. While recognizing the contribution that may be made by “such attempts to explain experimental phenomena” (ibid.), Gueth doubts that this is the route that leads to a systematic improvement of our theoretical account of actual human behavior. As he puts it: “Very often this type of research resembles, however, a neoclassical repair shop in the sense that one first observes behavior for a certain environment and then defines a suitable optimization or game model which can account for what has been observed” (ibid.). As Gueth suggests, what is needed for empirical explanatory purposes is to replace the perfect rationality model of normative game theory by a behavioral theory of actual human decision-making. Again, in his own words: “Of course, the assumption of perfect individual rationality is simply wrong and can be justified at best as an ‘as if’-explanation. Although we do not deny the need for a normative theory like game theory and, more generally, neoclassical theory, we prefer the natural psychological categories of human decision making over their artificial analogues resulting when they are represented in the typical neoclassical framework of utility maximization based on subjective beliefs” (ibid.).

What authors like Selten and Gueth in essence are looking or calling for is a principled alternative to the received rational choice paradigm, an alternative that accommodates the empirical evidence of real world human behavior not by ad hoc adjustments of the contents of a utility function that rational agents supposedly

---

38 In ultimatum game experiments a fixed amount of money is to be divided between two players. One player is assigned the right to propose a division between himself and the second player. If the second player accepts the amount is divided as proposed, if he refuses to accept the proposal both players receive nothing. In experiments proposers typically offer larger amounts to the second player than the standard rationality postulate would lead one to expect, a fact that, as Gueth (1995: 329f.) notes, “inspired a lively and still ongoing debate about the predictive role of game theory and, more specifically, about how fairness considerations influence behavior.”

39 In a more recent paper Gueth (2000) advances a general theoretical outlook at boundedly rational decision making that is, in many regards, similar to and compatible with the “evolutionary alternative” outlined below.
maximize, but by a systematic theoretical account. Another author who has made a similar plea is Dennis C. Mueller who, in his 1986 presidential address to the Public Choice Society (Mueller 1986), suggested to his colleagues to replace their overly simplistic model of “rational egoism” by a behavioral model of “adaptive egoism” that has a more solid foundation in psychological theory. The key difference between the kind of behavioral approach that he had in mind and “the usual application of rational actor models” Mueller sees in the fact that the behavioral approach “forces the investigator to examine the past histories of the people whose behavior he wishes to explain, and not just focus upon the entries in the different cells of the game’s payoff matrix. Human behavior is viewed as being adaptive and only approximates the purely forward-looking behavior depicted in rational choice models” (Mueller 2002: 667). As Mueller notes, following the route that he suggests would mean to “take a long step away from the pure forms of rational choice analysis” (ibid.: 668).

5. An Evolutionary Alternative to the Rational Choice Paradigm

Despite the longstanding and enduring criticism of the rationality postulate and notwithstanding their own skepticism about the empirical adequacy of the maximization hypothesis, economists have been extremely reluctant to even consider the possibility of parting with their accustomed outlook at human behavior and adopting a “more psychological” approach. There seem to be two principal reasons for their unshakable loyalty to the rational choice paradigm. One reason surely is the intuitive plausibility of the notion of rational choice. It obviously captures what, introspection and everyday experience tell us, are essential features of human behavior: Its intentional, purposeful, goal-seeking or forward-looking nature and its instrumental adaptedness to the problem-environment in which actors operate. Introspection tells us that our own behavior is purposefully aimed at solving problems we encounter and we make sense of other persons' behavior by employing a folk

---

40 In his own post-1986 work, though, Mueller himself does not seem to have heeded very much his own advice, and he seems much closer to the traditional doctrine when he argues: “The rational choice approach does not require that the rational choice scholar refrain from using knowledge from other disciplines that might help her to specify the goals of individuals. … Once the rational choice analyst has specified the arguments of the objective function and the relevant constraints, she can maximize this function” (Mueller 2002: 659).

41 M. Blaug (1992: 230): “So strong and persuasive has been the hold of the rationality postulate on modern economics that some have seriously denied that it is possible to construct any economic theory not based on utility maximization.”
version of rational choice theory,\textsuperscript{42} interpreting what others do as a - from their perspective - reasonable response to whatever choice problems they face. Successful interaction with others would scarcely possible without it. The other reason, connected with the first, is that the rational choice paradigm provides a unifying general theoretical outlook that applies to all domains of human behavior, and that economists fear that to allow the complexities of actual human psychology to play an explanatory role would require them to give up on the ambition to explain human action in terms of one unifying theory.\textsuperscript{43} There is, economists typically suspect, no really attractive alternative in sight.\textsuperscript{44} As J. Tirole (2002: 641) has put it: “Rationalists are not ignorant of the shortcomings of the homo economicus paradigm. Rather, and oversimplifying, they are concerned that the behavioral approach will prove too impulsive for its own sake and may forget what economics is all about, namely parsimony and normative analysis.”

My purpose in the remainder of this paper is to show that, counter to the perception prevailing among economists, there exists indeed an alternative theoretical outlook at human behavior that is able to capture the intentionality and adaptedness of human behavior no less than the rational choice paradigm, and that provides a no less unifying and general theoretical framework, without being subject to the kinds of objections that the rational choice paradigm has notoriously invited. The general thrust of this alternative, evolutionary outlook is well captured in the paradigm of program-based behavior suggested by biologist E. Mayr.\textsuperscript{45} The central argument of

\textsuperscript{42} J. Ferejohn and D. Satz (1996: 79) speak of "folk intentionalism" and argue "that successful intentional scientific accounts must 'track' folk intentionalism." As they note: "Social-science explanations must, we claim, be compatible with intentional descriptions of human agents" (ibid.: 74).

\textsuperscript{43} This concern is voiced by J. Tirole (2002: 652) when with regard to possible behavioral alternatives to the received rationality hypothesis he notes that such behavioral models “should start from a single view of the individual’s preferences, cognitive machinery and basic problem-solving strategies.” – D. Mueller refers to this concern when with regard to the behavioral approach that he suggest as an alternative to rational choice modeling he notes: “Thus, anyone who would follow the route that I have suggested would take a long step away from the pure forms of rational choice analysis … . But this approach would not force one to abandon the search for a universal theory of human action” (Mueller 2002: 668).

\textsuperscript{44} A. Sen (1987: 72): “The inadequacy of the traditional assumption of rational behavior standardly used in economic theory has become hard to deny. It will not be an easy task to find replacements for the standard assumptions of rational behavior … . (T)here is little hope of finding an alternative assumption structure that will be as simple and usable as the traditional assumptions of self-interest maximization, or of consistency of choice.” - One of the standard objections against H.A. Simon’s concept of “bounded rationality” has been that “there is no unified theory of bounded rationality” (R.J. Aumann 1997: 3; see also e.g. M. Blaug 1992: 159; F. Hahn 1996: 298; D.K. Foley 1998: 68; F. Laville 2000: 413;).

\textsuperscript{45} E. Mayr (1988: 30ff.; 1992). For a more detailed discussion of Mayr’s approach than will be presented here see Vanberg (2002: 15ff.).
Mayr’s approach is that intentional, goal- or purpose-seeking behavior can be viewed as program-based behavior in the sense of being guided by instructions, encoded in the agent, for what to do (or not to do) in certain types of situation. The programs incorporate knowledge of the contingencies of the agent’s problem environment, allowing the agent to anticipate consequences of potential actions. And the fact that the action-guiding programs are adapted to the relevant contingencies of an agent’s environment is seen as a result of processes of evolution and learning in which, through trial and error-elimination, more adequate programs replace less adequate ones.

A principal shortcoming of rational choice theory is that, rather than providing an explanatory account of the intentionality and adaptedness in human action, it simply treats them as definitional attributes of “rationality,” postulating rationality essentially as an unexamined human capacity. By contrast, a principal advantage of Mayr’s paradigm of program-based behavior is that it shows how, from an evolutionary perspective a naturalistic account can be provided for the very capacity that rational choice theory simply definitionally ascribes to “rational actors,” namely the capacity to anticipate consequences and, thereby, to be guided in one’s problem-solving behavior by expected consequences.

The essential difference between the two paradigms, or the two general outlooks at human behavior, lies in their respective accounts of the adaptedness of human action. While in rational choice accounts the adaptedness that we find in purposeful human action is simply attributed to an unexamined capacity “rationality,” the paradigm of program-based behavior attributes it to the adaptedness of the programs or rules of action that guide human behavior. More specifically, what distinguishes the two paradigms is a shift of explanatory focus from the adaptedness of single actions to the adaptedness of programs or rules for action. The rational choice paradigm is focused on the explanation of single actions. Its ambition is to explain the adaptedness of particular actions as a direct product of “rationality.” By contrast, the paradigm of program-based behavior seeks to explain the adaptedness of single human actions indirectly, as a consequence of – and contingent on – the

46 P.S. Albin (1998: xxvii): “Economics is the study of rational choice and the consequences of choice under restrictions imposed by productive technique, a legacy of institutions, and finite resources. By and large the technology of rationality has remained unexamined.” – G.M Hodgson (2002: 269): “It is (in economics, V.V.) still taken for granted that the existence of human intentionality is sufficient to explain human action, without probing the causes behind intentions themselves.”
programs that guide humans in their situational choices. And it explains, in turn, the
adaptedness of programs as a product of evolution and learning. It allows for the
possibility that single actions may not be “adapted” in the sense the rational choice
paradigm would postulate even though the programs that guide them may well be
adapted to the general features of the problem-environment in which the agent
normally operates.\footnote{The distinction that Aumann (1997: 7) draws between “Rule Rationality” and “Act Rationality”
points in the same direction as the distinction that I want to emphasize here. – J. Tirole does not seem
to realize that he is shifting from a paradigm focused on “act rationality” to one focused on “rule
rationality” when in discussing the issue of how to adjust the economist’s “maximization hypothesis”
in the face of conflicting behavioral observations he notes: “I consider it much more productive … to
try to retrieve information about the rules of thumb, mental representations, analogies and
categorization strategies that we employ. These heuristics are probably quite efficient ‘on average’
even though they may be very inappropriate in specific decision-making environments” (Tirole 2002:
640). That adopting the notion of rule-following behavior is a significant step away from a rational
choice perspective does not seem to be realized by Tirole when, a few pages later, he describes his own
“position on the future of psychology and economics” by saying: “My own and evolving preference is
to attempt to enrich homo economicus by embodying a small number of new ingredients” (ibid.: 642).
} In other words, it may well allow one to provide a systematic
account of observed behaviors that, from a rational choice perspective, one would
classify as unexplainable anomalies.\footnote{As D. Campbell (1986: 174) points out, “the ‘wisdom’ of the products (of evolutionary processes,
V.V.) is wisdom about past worlds” and evolved behavior may, therefore, be maladapted “in an
environment substantially different from the environment that shaped it.” In reference to the kind of
“anomalies” studied by A. Tversky and D. Kahneman (see e.g. Tversky and Kahneman 1986)
Campbell (ibid.) notes that they are “probably best explained as a by-product of generally adaptive …
inference habits, innate and acquired.”}

Significant about Mayr’s approach is that it directs our attention to the issue of
how the adaptedness of the programs themselves can be explained, and of how
programs are “applied” to guide action in particular choice-situations. In other words,
it draws attention to the questions of how the encoding and the decoding of programs
can be understood, where “encoding” means the processes that have shaped and
continue to shape the programs on which agents act and “decoding” means the
mechanisms by which programs as general instructions are applied to ever new
specific choice situations. It is by incorporating specific theories about the processes
of encoding and decoding that the general notion of program-based behavior can be
given specific explanatory content.

Mayr’s paradigm of program-based behavior provides, as I suppose, no less
than the rational choice paradigm a straightforward account of the intentionality and
adaptedness of human behavior, and it provides no less a unifying theoretical
framework, even if it cannot offer the deceptive simplicity that economists like to
praise as the trademark of rational choice theory. Yet, alternative theories should
surely not be judged in terms of their simplicity alone, independent of a their capacity to account for the complexities of the subject area they are supposed to illuminate. What the paradigm of program-based behavior offers instead is an analytical framework that is compatible with, and supported by, numerous other general theoretical approaches in the behavioral and cognitive sciences. It is a paradigm that invites theoretical integration by contrast to the self-imposed theoretical isolation that has shielded the economists’ rational choice model far too long from research developments in other behavioral sciences. In particular, it allows one to integrate, and to relate to each other, a number of established theories that deal in more detail with the various aspects of the encoding and decoding of action-guiding programs (see Vanberg 2002).

In the next section I shall take a look at J.H. Holland’s theory of adaptive agents as an example of an explanatory approach that is quite obviously compatible with, and lends support to, the paradigm of program-based behavior.

6. J.H. Holland’s Theory of Rule-Based Adaptive Agents

“How does an organism use its experience to modify its behavior in beneficial ways (i.e., how does it learn to ‘adapt under sensory guidance’)?” (Holland 1992a: 1)? How can a cognitive system use experience to increase its “problem-solving capabilities” (ibid.). These are the questions that J.H. Holland’s theory of adaptive agents is meant to answer. It looks at “adaptation as a fundamental process, appearing in a variety of guises but subject to unified study” (ibid.: 2), and its ambition is to provide “a unified theory of adaptation” (ibid.: 3).

At the heart of Holland’s theory, as well as of his efforts to design computer simulations that model the relevant processes, is the concept of rule-based systems or classifier systems (Holland 1988: 121), i.e. of information processing systems that accumulate knowledge about their environments by way of refining the categories by which they classify observations and the rules by which they model the regularities

49 H.A. Simon (2000) has commented on this issue: “Perhaps the simplicity we should look for, in place of unattainable classical rationality, will come as we study … how human beings actually adapt to the very severe limitations on their computational powers” (ibid.: 255). “It may be objected that such a theory will be much more complex … than the simple theories of the past. … Such complexity does not imply that … economics should ignore the facts and go back to phantasmoagoria that lack substance. … Even when we attain a tolerable level of realism, economics will not have the (relative) invariance of some other sciences. … Even if this is true, the empirical study of decision making will allow us to treat economics as a science: a historical science, like geology, evolutionary biology, and even astronomy. We should not find this an unattractive prospect” (ibid.: 252).
that characterize their environments. Adaptive agents are looked at as rule-based or classifier systems, i.e. as systems that classify problem-situations they encounter and respond to types of situations based on a repertoire of rules that could be verbally stated as “if …, then …” instructions. Accordingly, an adaptive agent can be “described in terms of rules” (Holland 1995: 10) or, more precisely, in terms of the “cluster of rules … that generates its behavior” (ibid.: 49). In terms similar to E. Mayr’s notion of program-based behavior, Holland notes that “rules so defined act as much as instructions in a computer, the cluster serving as a program that determines the agent’s behavior” (ibid.).

Holland regards his theory explicitly as an “evolutionary approach to learning” (Holland 1996: 282) that is in direct contrast to a theory that, like economics, is “built around agents of perfect rationality – agents that perfectly foresee the consequences of their actions, including the reactions of other agents” (Holland 1995: 85). To take an evolutionary outlook does not mean at all, as Holland (1996: 282) stresses, to deny that “economic agents anticipate,” it means to insist, however, that the capacity to anticipate, a capacity that rational choice accounts simply postulate, is explained in a backward looking manner, as based on past experience.

Adaptive agents owe their “ability to anticipate” (Holland 1992: 20) to the rules on which they operate, “rules that anticipate the consequences of certain responses” (ibid.) and that can be “viewed as hypotheses that are undergoing testing and confirmation” (Holland 1995: 53). Rules “impose regularity on a complex world” (ibid.: 37). They simplify the world and allow agents to cope with the inexhaustible complexity of everything and the impossibility of considering all things by focusing

---

50 Holland et al. (1986: 14): “Our most basic epistemic building block is a condition-action rule, which has the form 'IF such-and-such, THEN so-and-so,' where the IF part is the condition and the THEN part is the action.”

51 As far as the relation between his theory of adaptive agents and the economic model of perfect rationality is concerned, Holland (1995: 85) points to the affinity between his own views and those of W.B. Arthur, who in turn points to the similarity between his “behavioral approach to bounded rationality” (Arthur 1991) and a “Holland-type classifier” (ibid.: 355). Noting that “economists have long been uneasy with the assumption of perfect, deductive rationality” (Arthur 1994: 411), Arthur argues that what is at issue is not “whether perfect rationality works, but rather what to put in its place” (ibid.: 406). The theory that he suggests to put in its place he describes as follows:

“Humans … form mental models, or hypotheses, or subjective beliefs. … Each agent will normally keep track of the performance of a private collection of such belief-models. When it comes time to make choices he acts upon his currently most credible … one. … Once actions are taken … agents update the track record of their hypotheses. This is a system in which learning takes place. Agents ‘learn’ which of their hypotheses work, and from time to time they may discard poorly performing hypotheses and generate new ‘ideas’ to put in their place. A belief model is clung to … because it has worked in the past and must cumulate a record of failure before it is worth discarding” (ibid.: 407).
attention on selected aspects of the “torrent of information its environment produces” (ibid.: 44). In an ever-changing environment in which every situation they encounter is in some way novel and unique, adaptive agents handle each situation by combining “rules into clusters that model the environment” (Holland 1996: 283). Rules serve as “building blocks” of internal models that “can be used to generate predictions about the outcomes of potential solution attempts” (Holland et al. 1986: 14).

The internal models on which actions are based may, in Holland’s terminology, be tacit or overt. “A tacit internal model simply prescribes a current action, under an implicit prediction of some desired future state” (ibid.: 33), without any conscious deliberation on part of the agent. A most basic example would be a bacterium that “moves in the direction of a chemical gradient, implicitly predicting that food lies in that direction” (ibid.: 32). By contrast, an “overt internal model is used as a basis for explicit, but internal exploration of alternatives, a process often called look ahead” (ibid.: 33). Overt internal models inform the kind of deliberative choices on which rational choice accounts tend to focus.

The “use of building blocks to generate internal models is,” as Holland (1995: 37) emphasizes, “a pervasive feature of complex adaptive systems.” Since we “cannot have a prepared list of rules for all possible situations” we draw on a “repertoire of everyday building blocks” when we encounter a new situation. We “combine relevant, tested building blocks to model the situation in a way that suggests appropriate actions and consequences” (ibid.). What allows rule-based agents to

---

52 On the issue of what aspects of the world an agent’s simplifying rules or “mental models” draw attention to Holland notes: “The answer is fundamentally pragmatic, and it highlights the link between mental models and problem solving. … The model need only describe aspects of the environment and of the system’s actions that are relevant to the attainment of goal-satisfying states” (Holland et al. 1986: 39).

53 Holland et al. (1986: 15): “In the absence of rules, a system would have to rely on storage, retrieval, and processing of a host of episodes and examples. For efficient operation in a changing environment, a system is much better off if it has the capacity to use rules for quick generation of expectations.”

54 Holland (1998: 26): “This use of models … comes into play in everything from the mundane task of finding an alternative route when roadwork blocks the usual way home, to the generation of sophisticated hypotheses in science.”

55 Holland (1995: 34): “In realistic situations an internal model must be based on limited samples of a perpetually novel environment. Yet the model can only be useful if there is some kind of repetition of the situation modeled. … Indeed, it is evident that we parse a complex scene by searching for elements already tested for reusability by natural selection and learning. … We gain experience through repeated use of the building blocks, even though they may never appear in exactly the same combination.”

56 Holland (1998: 25): “The actual projection of external scenes on the millions of sensory cells in our eye is never twice the same; nevertheless, every scene has some aspects that have appeared before. … It is our ability to discern and use building blocks that makes the perpetual novelty of our world understandable, and even predictable.”
form internal models is their capacity as classifier systems \(^{57}\) to activate many rules simultaneously or concurrently (Holland et al. 1986: 29), a capacity, that Holland calls “parallelism.” \(^{58}\) In the “provision of simultaneously active rules” he sees the clue for understanding “an agent’s ability to handle a perpetually novel world” (Holland 1995: 51). “With simultaneously active rules, the agent can combine tested rules to describe a novel situation” (ibid.).

The principal focus of Holland’s theory and simulation models is on the process by which adaptive agents or systems manage to improve their repertoire of rules and thereby to increase their capability to deal successfully with the kinds of problems they are confronted with. Holland uses the term induction to describe the learning process in which the feedback from (direct and indirect) experience allows an agent to improve its internal models, noting that the “study of induction then, is the study of how knowledge is modified through its use” (Holland et al. 1986: 5).

“Induction” in the sense Holland uses the term does not mean the “derivation” or “construction” of general hypotheses from observations. Instead, it means the process in which experience induces agents to modify or refine the rules or hypotheses on which they act. It is a process in which rules or conjectures come first and are subject to ex-post selection by consequences. \(^{59}\) Holland’s adaptive agents are equipped from the outset with a repertoire of rules, and they adapt to the contingencies of their environment “by changing their rules as experience accumulates” (Holland 1995: 10).

The process in which adaptive agents improve the internal models that guide their problem-solving efforts is explicitly modeled as an evolutionary process of variation and selection by consequences. \(^{60}\) There is always a population of rules

---

57 Holland et al. (1986: 3): “Classifier systems are a kind of rule-based system with general mechanisms for processing rules in parallel, for adaptive generation of new rules, and for testing the effectiveness of existing rules.”

58 Holland (1992a: 174): “Parallelism, the concurrent activity of many rules, is an important aspect of classifier systems. Parallelism makes it possible for the system to combine rules into clusters that model the environment, providing two important advantages. 1. … The system builds a ‘picture’ of the situation from parts, rather than treating it as a monolithic whole. … 2. Experience can be transferred to novel situations. … Building-block rules give the system a capacity for transferring experience to new situations.”

59 In K.R. Popper’s terms, it is a process of “conjectures and refutation,” of “trial and error elimination.” Holland’s “induction” is, therefore, perfectly compatible with K.R. Popper’s views on the “growth of knowledge” (see Vanberg 2002: 22ff.), and is not subject to Popper’s arguments on the “logical problem of induction” (Popper 1989: 52ff.).

60 As Holland (1995: 34) notes on how an internal model is improved upon: “Variants of the model are subject to selection and progressive adaptation.”
present upon which selection can operate, and new rules are continuously generated, mainly through re-combination of component of existing rules. In order for systematic selection to take place that reinforces “successful” rules and works against ones that do not contribute to successful problem-solving, a feedback mechanism must be in place that attributes success in action appropriately to the rules that are “responsible” for the carrying out of the respective actions. This is the problem for which Holland uses the term “credit assignment” and on which he notes: “Credit assignment is not particularly difficult when the system receives payoffs from the environment for a particular action – the system simply strengthens all the rules active at that time (a kind of conditioning). Credit assignment becomes difficult when credit must be assigned to early-acting rules that set the stage for a sequence of actions leading to payoff” (Holland 1996: 285).

It is a significant achievement of Holland’s approach that it specifies a model of how such credit assignment operates (called “bucket brigade algorithm”), a model the general thrust of which can be captured by the metaphor of a market in which not only the final sellers of products are rewarded by the price paid by their customers, but in which the revenue raised in the final product market is transferred back to the producers of these final products, to the producers of inputs for these products, to the producers of inputs for the production of inputs, and so on. Thus, productive activities that “set the stage” for success in the final product market are encouraged, while failure in the final stage translates into inability to reward suppliers of inputs. In similar ways the “bucket brigade algorithm” models the ways in which adaptive agents carry on a “calculus of advantage” at the level of rules of action that assigns credit to – and thus strengthens – rules according to their respective contribution to the agents’ overall success in solving the problems they encounters in their

---

61 Holland (1995: 53): “That is, rules amount to alternative, competing hypotheses. When one hypothesis fails, competing rules are waiting in the wings to be tried.”
62 Holland et al. (1986: 16): “Rules are a natural vehicle for what we take to be the most fundamental learning mechanism: prediction-based evaluation of the knowledge store. … A rule that leads to successful predictions should be strengthened in some way, increasing the likelihood of its use in the future; one that leads to error should be modified or discarded.”
63 Holland (1995: 53): “We want to assign each rule a strength that, over time, comes to reflect the rule’s usefulness to the system. The procedure for modifying strength on the basis of experience is often called credit assignment.”
64 Holland (1995: 56): “This credit-assignment procedure, which I call a bucket brigade algorithm, strengthens rules that belong to chains of action terminating in rewards. The process amounts to a progressive confirmation of hypotheses concerned with stage setting and subgoals.” - For details see e.g. Holland (1992a: 176ff.).
environments (Holland 1996: 285f.). As Holland (ibid.) emphasizes, the “bucket brigade algorithm” makes a task manageable that otherwise would be beyond the capacity of a boundedly rational agent, namely to keep track of the success record of a complex repertoire of rules that are activated in varying combinations as components of internal models of current problem-situations.

The prospects for adaptive agents to improve their problem-solving capacity critically depend on their ability to generate new and superior rules or hypotheses from their original repertoire. There are two ways in which Holland’s theory of adaptive agents allows such generation of new rules to occur, namely the recombination of components of existing rules (crossover) and the random alteration of rules (mutation). Crossover is a process of “generating plausible new rules” (ibid.: 286) by recombining components of rules that have proven to be successful in the past, - similar to the process of genetic recombination in biological reproduction. It is made possible by the fact that rules are decomposable into components that can be recombined in various ways to create new rules. Rules can be described in terms of “IF such and such, THEN so and so”-statements, and such statements may include on their “if”-side as well as on their “then”-side a number of components that qualify as potential building-blocks in the creation of new rules. Crossover is a process in which “experience biases the generation of new rules” (ibid.) because it allows the components of existing rules to be represented in new rules in proportion to the previous success – and thus strength – of the parent-rules. It biases “the rule generation process so that above-average building blocks are favored in the construction of new rules” (ibid.).

The generation of new rules by crossover or genetic algorithms is not a process of “blind variation,” even though success cannot be guaranteed in advance. Each particular recombination is a novel experiment, and only selection by consequences, i.e. ex-post selection based on actual performance, will reveal which among the newly constructed “plausible” rules do indeed work and which do not. By

---

65 Holland (1995: 65f.): “Offspring are not identical to parents, so this is a genuine discovery process. Offspring, in both genetics and rule-based systems, amount to new hypotheses to be tested against the environment. In genetics … crossing over causes the characteristics of the parents to appear in new combinations in the offspring.”

66 In order to capture the nature of crossover as an experience-guided process of rule discovery Holland uses genetic algorithms in modeling the “rule-generating algorithms for classifier systems” (Holland 1988: 122; 1996: 287). What crossover or genetic algorithms assure is that the components of previously successful rules are more likely to be combined into new rules than the components of previously less successful rules.
contrast, the generation of new rules through random mutation, e.g. because of simple errors that occur in the copying of successful rules, cannot be but “blind” in the sense that it ventures into genuinely new territory. It “is a history-independent operation that does not make use of the system’s knowledge or past history” (Holland 1996: 291). Quite obviously, exploration into “the inexhaustible space of possible hypotheses” (Holland et al. 1986: 9) promises to be more successful if it relies on rule-discovery by experience guided crossover rather than on rule-discovery by random mutation (Holland 1995: 60f.). The principal way in which rule discovery proceeds is by “generalization and specialization of existing rules” (Holland et al. 1986: 86). In essence, generalization makes an existing rule more general by dropping or ignoring components of its “IF”-clause and/or of its “THEN”-clause (ibid.), while specialization restricts the applicability of a rule by augmenting the respective clauses of an “IF …, THEN …”-rule (ibid.: 88).

As Holland points out, looking at economic men as rule-based, adaptive agents would lead one to a conception of the market, and of economic processes more generally, that is significantly different from that suggested by the neoclassical model of perfect rationality (Holland 1996: 293, 295f.). By contrast to the neoclassical focus on average or “representative” economic agents (the representative household or firm) and on equilibrium endpoints, the theory of adaptive agents emphasizes the diversity of individuals and the evolutionary, open-ended nature of processes of interaction among adaptive agents. Markets are, in Holland’s terminology, a special case of complex adaptive systems. i.e. of systems that consist of interacting adaptive agents who follow rules and who adapt to the contingencies of their environment – an environment that is largely made up of other adaptive agents – by changing the rules on which they operate (Holland 1995: 10). And market processes exemplify in a paradigmatic fashion the features that Holland describes as characteristic attributes of complex adaptive systems. It is the diversity of the agents forming the system and the fact that their interactions are governed by adaptable anticipations that make for the complexity of the system.  

---

67 Holland (1988: 122f.): “It is important that recombination rather than mutation is the main driving force of the search. … The fundamental theorem about genetic algorithms shows that new rules are generated primarily by the recombination of parts – the building blocks – of rules that have already been useful to the system.”

68 On the issue of what distinguishes complex adaptive systems (cas) from other kinds of systems Holland (1995: 93) notes: “One of the most obvious of these distinctions is the diversity of the agents that form cas. … The interactions of agents in cas are governed by anticipations engendered by
For complex adaptive systems in general and, in particular, for markets that are composed of intelligent, learning human beings “there is no way to predict the overall behavior by looking at the behavior of an ‘average’ individual” (Holland 1998: 118), because the actions of each individual are conditioned by an environment of other adaptive agents who learn and continuously adapt to changing circumstances. Complex adaptive systems in which, as is true for markets, individual agents continuously adapt to each other are characterized by creativity and perpetual novelty (ibid.: 42), a fact that does not exclude them, though, from systematic analysis. Even though the complexities that result from the co-evolution of the agents’ strategic repertoires makes it impossible to predict the specific outcomes that emerge in such systems, they allow for what Hayek (1967) has called “pattern prediction” and “explanation of the principle.” There are no pre-determinable end-points or equilibria towards which such systems converge. “They continue to evolve, and they steadily exhibit new forms of emergent behavior” (Holland 1992b: 20). And from this insight, Holland argues, we must conclude that it “is the process of becoming, rather than the never-reached end points, that we must study if we are to gain insight” (ibid.).

---

69 As B. Arthur (1994: 408) summarizes the outlook that he shares with J.H. Holland: “Agents differ, and each uses several subjective models … always leading to new hypotheses. … It is also a world that is evolutionary, or more accurately, coevolutionary. … In this world hypotheses … must prove themselves by competing and being adapted within an environment created by other agents’ hypotheses. The set of ideas or hypotheses … coevolves.”

70 Holland (1998: 122): “Under these conditions, the whole is indeed more than the sum of its parts. However, we can reduce the behavior of the whole to the lawful behavior of its parts, if we take the nonlinear interactions into account.”

71 Holland (1998: 43): “The coevolution exhibits the creativity we expect of any evolutionary process. … Though prediction is difficult in these circumstances, it is not a hopeless task. Everything depends on the level of detail we require of the prediction.”

72 Holland (1992a: 184f.): “The system as a whole typically operates far from a global optimum or equilibrium. Standard theories in physics, economics, and elsewhere are of little help because they typically concentrate on ‘end points,’ whereas complex adaptive systems ‘never get there.’”

73 Holland (1992b: 29): “They (complex adaptive systems, V.V.) do not yield to classic, equilibrium-based mathematical approaches, that rely on linearity, attractors, fixed points, and the like. A new kind of mathematical framework is required, one that emphasizes continuing adaptation through recombination of building blocks.”
7. Rationality and Markets

A principal implication of adopting the paradigm of program-based behavior instead of a rational choice perspective in explaining human action is that it forces one to deal with the complexities that arise from the fact that characteristics of the actors, namely their program-repertoires, provide a critical link between the characteristics of the situations they encounter and their behavioral responses. Their program repertoires condition, as intervening variables, how agents will respond to the choice-situations they face. As far as their genetically encoded program repertoires are concerned, we may safely assume that humans are very much alike, due to their common evolutionary history. Learned program repertoires will in part be common among individuals, to the extent that they grew up in similar socio-cultural environments. And in part they will be different from person to person, because of idiosyncrasies in their respective learning histories. To the extent that the program repertoires of persons are different, according to differences in their past cultural environment and personal learning histories, they may behave quite differently in what, from an observer’s perspective, appear to be similar problem-situations. And to the extent that differences in persons’ behavioral repertoires matter for explanations of their current behavior, such explanations may require one to look for clues in persons’ learning histories that can provide plausible foundations for conjectures about the nature of their respective program repertoires. This is why, as D.C. Mueller (2002: 667) has put it, a behavioral approach “forces the investigator to examine the past histories of the people whose behavior he wishes to explain.”

By adopting a theory of perfect rationality economists have afforded themselves the luxury of avoiding the need “to examine the past histories of the people whose behavior they wish to explain.” In a world of perfect rationality there is no need to consider the characteristics of the actor in addition to the characteristics of the situation, because there is only one “objectively optimal response to the situation presented” (Simon 1987: 267), a response that rational actors will choose and that the observing investigator can identify. When above (section 4) I listed three kinds of arguments that economists employ in defense of their use of the rationality postulate, I postponed the discussion of one argument that I shall now look at more closely. According to this argument economists can afford to ignore the particularities of personal histories and can, instead, operate on the presumption that there is only one type of actors, namely the perfectly rational type, because this outlook is reasonably
adequate for the particular kind of social arrangement that economists concentrate their explanatory ambitions on, namely *markets*. The assumption of perfect rationality, a common argument says, is not so much a conjecture about the cognitive and calculative capabilities of human beings per se, but a conjecture about the working properties of markets as social arrangements. Kenneth J. Arrow alludes to this argument when he notes that rationality, rather than being “a property of the individual alone, … gathers not only its force but also its very meaning from the social context in which it is embedded” (Arrow 1987: 69). To the extent that one moves away from the context of competitive markets, Arrow suggests, “the rationality assumptions become strained and possibly even self-contradictory” (ibid.). And he adds, “we need not merely pure but perfect competition before the rationality hypotheses have their full power” (ibid.: 70).

Upon closer examination the argument that the rationality postulate should be understood as a claim about the attributes of markets rather than the attributes of individual human beings per se turns out to be quite ambiguous. At least three critically different interpretations of this claim can be identified. The first and least interesting version of this claim simply says that pure economic theory is about the working properties of *perfect markets*, i.e. markets in which adjustments to changing opportunity costs take place without any friction whatsoever, and that perfect rationality on part of the market participants is just one of the defining characteristics of a perfect market. This version of the noted claim is the least interesting because it turns the whole issue into a purely definitional matter. It defines economics from the outset not as an empirical science that informs us about the world of our experience, but as an analytical enterprise that explores the properties of an imagined, hypothetical world, a theory that tells us what we were to observe if our world were to exhibit the properties that define a perfect market.

From the purely semantic claim that “perfect rationality” is a definitional attribute of “perfect markets” one must distinguish the *factual claim* that real world markets, due to the forces of competition, approximate the properties of perfect

---

74 See also Arrow (1986: 201, 203).
75 Exploring the properties of a “perfect market,” populated by perfectly rational agents, is the research program that L. Walras chose for his “pure economics” (see Vanberg 2001), a research program that has largely been a defining characteristic of the neoclassical tradition in economics. Stanley Jevons, the second principal founder of the neoclassical tradition, defined a “perfect market” as one in which “all traders have perfect knowledge of the conditions of supply and demand and the consequent ratio of exchange” (Jevons 1871: 87).
markets in the sense that they tend to select against non-rational behavior, such that rationality tends to become their dominant attribute, justifying the use of the rationality postulate as a simplifying assumption. The most famous pronouncement of this second version of the notion of rationality as an attribute of markets is the “evolutionary” argument made by A.A. Alchian (1950) and M. Friedman (1953) according to which the rationality hypothesis can rightly be used as an “as-if” assumption, because under the selection-constraints of a competitive market environment “it summarizes appropriately the conditions of survival.” It is essentially a restatement of their argument when R.M. Hogarth and M.W. Reder (1986: 6) note: “The economic paradigm focuses on actions taken in competitive circumstances. The underlying assumption is that through competition the actions of individual agents are subject to feedback that forces them either to become effective or to withdraw from such action. … Economists have little interest in modeling agents who do not behave according to rational principles since they believe that these agents will not survive in the market.”

The “evolutionary” as-if argument in support of the rationality hypothesis is somewhat ambiguous in itself because it can be meant to imply either, or both, of two different claims (Laville 2000: 411, 416). On the one hand it can be meant as an argument about learning effects, namely as the claim that under the conditioning effects of market competition individuals come to act as if they were rational, maximizing agents. On the other hand it can be meant as an argument about

---

76 According to Friedman (1953: 21), under competitive conditions “individuals behave as-if they were seeking rationally to maximize their expected returns … and had full knowledge of the data needed to succeed in this attempt.” – For a discussion of the Alchian-Friedman “as-if” argument see e.g. J. Vromen (199.: 33, 53).

77 Friedman (1953: 22): “Let the apparent immediate determinant of business behavior be anything at all – habitual reaction, random chance or what not. Whenever this determinant happens to lead to behavior consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not the business will tend to lose resources. … [G]iven natural selection, acceptance of the [maximization, V.V.] hypothesis can be based largely on the assumption that it summarizes appropriately the conditions for survival.”

78 R.H. Day (1993: 61) summarises, in similar terms, the reasoning behind the belief of the “new classical” school “that when rational thought and competition are given full scope … disequilibrium can be ignored. The implicit assumption on which this hypothesis is based is that opportunities for profit are arbitraged until profits above equilibrium returns are eliminated. It is assumed that rational individuals will do this. If they do not know how, they will learn to do so, and even if they do not learn how, they will act as if they have learned how. In short, according to this school, economic agents in a competitive market economy behave as if they had unbounded rationality.”

79 I am putting the “evolutionary” in quotation marks in order to distinguish the Alchian-Friedman as-if argument from the perspective that I have described above as an evolutionary alternative to the rational choice paradigm.
aggregation effects, namely as the claim that the selection pressure of competition makes markets work as if they were populated by rational, maximizing agents. Both versions of the “evolutionary” argument are intrinsically problematic. They both ignore the very features that characterize markets as complex adaptive systems in which the interaction and co-evolution of adaptive agents are constant generators of novelty. A conditioning of human actors into perfectly rational maximizers might be conceivable in problem-environments that are sufficiently stable over long periods of time, long enough to allow the actors involved to learn everything that is to be learned. Yet, markets - apart, perhaps, from their most elementary incarnations - hardly qualify as problem-environments of such a kind. And the selective weeding out of inefficient agents may, conceivably, in such environments eventually produce a population of optimally adapted agents. But, again, in markets the interaction and co-evolution of intelligent and inventive agents are unlikely to permit competitive selection ever to settle with a population of optimal problem-solvers.

The third version of the notion of rationality as a property of markets seems to me the most reasonable and defensible one. This is the factual and – compared to the “as-if” construct - much more modest claim that markets work as conditioning environments that teach agents – not to be perfectly rational, but – to be more “economically rational” than they would otherwise be. The reason is that competitive markets make agents realize the opportunity costs of their choices more directly and effectively than other social arrangements do, allowing for more effective feedback between choices and consequences. The “rationality” that, according to this view, markets induce is a matter of degree, though. Where, and to the extent that, market competition works as an effective error-eliminating force, the agents involved can be expected to learn to adopt efficient problem-solving strategies. This is, for instance, very likely in regular, everyday economic activities, such as shopping for one’s groceries. It is unlikely that a person continues to buy from the same supplier if a cheaper source is available around the corner (except, of course, other advantages obtained from using the more expensive source outweigh the difference in price). In the case of economic activities where the feedback between choices and outcomes is relatively short-term, experience can quickly work to eliminate ineffective strategies. This is the kind of setting that standard textbook models of equilibrating competitive markets for homogeneous products tend to focus our attention on. By contrast, where the feedback loop from choices to eventual consequences is more indirect and less
transparent, the error-eliminating selective pressure on alternative problem-solving strategies is much less stringent, and a broad range of competing conjectures about potentially promising strategies may persist without convergence towards “optimal” solutions. This is the case, for instance, with genuinely entrepreneurial choices among alternative investment projects with long time horizons. And it is by no means an accident that standard economic theory has very little to say about the entrepreneurial dimension of market processes.

That markets work, in the manner indicated, as social environments that enable agents of limited cognitive abilities and with limited foresight to act more rationally than would otherwise be the case is, of course, the central message of F.A. Hayek’s critique of “the bogey of ‘rational economic man’” (Hayek 1948: 11) and of his theory of the spontaneous order of the market as a system of communication that, through the price mechanism, speedily signals changes in relative scarcities throughout the entire exchange nexus. And it is a message that V.L. Smith endorses when, summarizing the experience of decades of experimental research, he notes that “far from having perfect information, subjects (in experiments, V.V.) know only their own ‘circumstances’” (Smith 1991: 880), and that what matters is the “manner in which institutions serve as social tools that reinforce, even induce, individual rationality” (ibid.: 881). As Smith (ibid.: 894) puts it: “On the basis of cognition alone, without the language of the market and ongoing social interaction with other agents, rational decision is frustratingly illusive.”

The claim that the competitive nature of markets works as an error-eliminating force appears to me not only to be the most defensible version of the notion of rationality as an attribute of markets, it is also a claim that, quite obviously, can be readily accommodated in the context of the paradigm of program-based behavior. If

---

80 Referring to Hayek’s 1945 AER article “The Use of Knowledge in Society” H.A. Simon (1992: 27) applauds Hayek for arguing “that the real importance of the market mechanism is … that it conserves information for all of the economic actors, and allows them to behave rationally with relatively simple computations and on the basis of relatively little information. This is an idea that advocates of bounded rationality can accept with enthusiasm, and it is a pity that mainstream economics didn’t take this proposal of von Hayek with the seriousness it deserves.”

81 As a response to the question of how agents of limited knowledge and cognitive capabilities come to act consistently “in the social context of exchange institutions” (Smith 1991: 894) notes: “It seems that an important part of the answer resided in the properties of exchange institutions and in how privately informed, but globally poorly informed, decision making is mediated by institutions” (ibid.). And in line with Hayek’s account of the rules of the market as an emergent product of cultural evolution Smith (ibid.) conjectures in response to “the question why institutions have the structure and rules we observe”: “I want to suggest that perhaps the structures we observe have survived because of their merit in coaxing Pareto-efficient behavior out of agents who do not know what that means.”
one thinks of the behavioral programs on which people operate as conjectures about what are suitable problem-solving strategies, then an elimination of unsuitable programs and a convergence towards efficient programs can be expected to occur the more speedily, the greater the testability of the behavioral conjectures that are exposed to the forces of market competition or, in other words, the greater the immediacy of feedback between choices and relevant consequences. In this sense, the rational choice paradigm might be looked at not as an alternative to, but as a special case of the paradigm of program-based behavior, namely as the latter’s application to an environment in which the actors involved have learned their lesson and have settled on their optimal program repertoires for dealing with the kinds of problems the environment presents to them, and in which no further change occurs that would challenge their established behavioral programs.

This seems, indeed, to be the kind of role that R.E. Lucas wants to assign to the economic model of man. In discussing the issue of the relationship between psychological and economic views of behavior, Lucas (1986: 217) argues: “In general terms, we view or model an individual as a collection of decision rules (rules that dictate the actions to be taken in given situations) and a set of preferences used to evaluate the outcomes arising from particular situation-action combinations. These decision rules are continuously under review and revision; new decision rules are tried and tested against experience, and rules that produce desired outcomes supplant those that do not. I use the term ‘adaptive’ to refer to this trial-and-error process through which our modes of behavior are determined.” Such statement could easily lead one to expect that Lucas is about to make a plea for replacing the rational choice model by a theory of rule- or program-based behavior. Yet, this is not at all the conclusion that he arrives at. Instead, he argues: “Economics has tended to focus on situations in which the agent can be expected to ‘know’ or to have learned the consequences of different actions so that his observed choices reveal stable features of his underlying preferences. We use economic theory to calculate how certain variations in the situation are predicted to affect behavior, but these calculations obviously do not reflect or usefully model the adaptive process by which subjects have themselves arrived at the decision rules they use. Technically, I think of economics as studying decision rules that are steady states of some adaptive process, decision rules that are found to work over a range of situations and hence are no longer revised appreciably as more experience accumulates” (ibid.: 218). – Economists may, of course, justify
their continued adherence to the rationality postulate by choosing to limit their explanatory ambitions in the way Lucas suggests. Yet, such self-limitation would clearly conflict with the ambition that is behind the most innovative theoretical developments in modern economics, namely to demonstrate the explanatory potential of the economic approach by extending it beyond the study of markets into domains traditionally reserved by other social sciences.

8. Conclusion

Their focus on the study of markets is surely one of the reasons why economists have been relatively satisfied with the rationality hypothesis as their behavioral model despite its recognized shortcomings. And indeed, the world depicted by the standard textbook models of demand and supply analysis is one in which there seems to be little benefit from applying the paradigm of program-based behavior because this world leaves little scope for variation in the program repertoires on which people act. The easy testability of the effectiveness of alternative programs and the immediacy of feedback in the kinds of environment that these models depict make the assumption that the agents involved act on programs that the observing analyst can identify as the “rational” or “objectively appropriate” ones quite plausible. Yet the more we move into dimensions of market activities and into non-market social realms in which the conditions of easy testability and immediacy of feedback are not met, this assumption becomes more and more problematic.82

An economics that does not want to limit itself to the study of “steady states” (Lucas 1986: 218) in which the agents involved have learned all that is to be learned, and have settled on rules optimally adapted to a stable problem-environment, such an economics must recognize the rational choice model for what it is: a special, limiting case of a theory of rule- or program-based behavior. As economics is developing into a general social science, applicable to the non-market as well as to the market realm, and as it aims at explaining political and institutional phenomena no less than market behavior, it can no longer avoid to face the challenges that come with the complexities

82 As A. Clark (1997: 274f.) notes: “In sum, traditional economic theory (invoking the substantive rationality paradigm) succeeds wherever individual choice is strongly constrained by social and institutional scaffolding that has itself evolved subject to selective pressure to maximize rewards. Outside such highly constrained settings, genuine individual thought plays a greater role, and the psychological irrealism of the substantive rationality model takes its toll.”
inherent in the more general paradigm of which the rational choice paradigm is, at best, only a limiting case.

References


00/6 Zoll, Ingrid: Zwischen öffentlicher Meinung und ökonomischer Vernunft: Individuelle Meinungen über Globalisierung und Wettbewerb.
01/3 Stamm, Hansueli: Institutioneller Rahmen des Electronic Commerce: Eine ordnungsökonomische Analyse am Beispiel der digitalen Signatur.
01/6 Vanberg, Viktor J.: Citizens’ Sovereignty and Constitutional Commitments: Original vs. Continuing Agreement.
02/1 Vanberg, Viktor J.: F. A. Hayek und die Freiberger Schule.
02/2 Pelikan, Pavel: Why Economic Policies Need Comprehensive Evolutionary Analysis.
02/3 Märkt, Jörg: Armutsexternalitäten: Verfassungsökonomische Rechtfertigung einer kollek-tiven Grundsicherung.
02/4 Märkt, Jörg: Zur Methodik der Verfassungsökonomik Die Aufgabe eines vertrags-theoretisch argumentierenden Ökonomen.
02/5 Vanberg, Viktor J.: Rationalitätsprinzip und Rationalitätshypothesen: Zum methodo-logischen Status der Theorie rationalen Handelns.
02/6 Schnellenbach, Jan: The Evolution of a Fiscal Constitution When Individuals are Theoretically Uncertain.
02/7 Wohlgemuth, Michael: Schumpeterian Political Economy and Downsian Public Choice: Alternative economic theories of democracy.
02/8 Fischer, Christian: Europäisierung der nationalen Zivilrechte – Renaissance des institutionellen Rechtsdenkens?
03/1 Vanberg, Viktor J.: Die Verfassung der Freiheit: Zum Verhältnis von Liberalismus und Demokratie.
03/2 Goldschmidt, Nils / Berndt, Arnold: Leonhard Miksch (1901–1950) – A Forgotten Member of the Freiburg School.