HUMAN AFFAIRS, *8*, 1998, 2, 101–111

A VIEW OF RATIONALITY AT THE END OF THE MODERN AGE[1]

Jozef Kelemen
Department of Applied Informatics, University of Economics, 852 35 Bratislava, Slovakia and Institute of Computer Science, Silesian University, 746 01 Opava, Czechia

Rationality has accompanied people since they have set out on their journey through history. Every age perceives the route in a different way and attaches a different importance to it. The following lines are an attempt at introducing one of the possible views of rationality in the modern age, the age we have lived in since the advent of rationalism. It will concern specific views of rationality dictated by the struggles for formally correct accounts of what we observe in human actions and in organized communities, economic units in particular, and the struggles for technical reconstruction of rational systems we meet in the area of advanced technologies, for example in artificial intelligence and robotics. I will ultimately try to find connections between these modern approaches to rationality and the possible understanding of rationality in the period, the character of which is discussed within the framework of postmodernism and which will come only after the modern age has been internally exhausted.

1 Introduction – rationalism and rationality

The dominant position of rationality in Western civilization and philosophical tradition dates from the time when René Descartes explicitly placed it in its key position within his conception of philosophy. His conception of philosophy – rationalism – is based on respect for and the confidence in scientific knowledge, which he defines with his own means as an activity having rationality in its essence. Conceptions and criteria of the proper functioning of society were built on rationalistic foundations and the education of experts also found its orientation on a rationalistic basis. Our primary effort is to lead these experts to effective solutions and procedures; the society puts confidence in these experts in the decision making on the problems important to both society and individuals.

[1] This is an enlarged English written version of the author's talk delivered at the Department of Philosophy of the Prague University of Economics, Prague, October 16, 1997.

In our civilization *to act rationally* means the same as to act correctly. This is in the core of the philosophy of rationalism. *To think rationally* means to approach decisions how to act in a correct way. *Logic* tries to reveal and postulate the correct way of thinking. In principle, it is such a procedure which does not leave out of account the demand not to put into contradiction what *we want* and what is *possible*. This is the business of the old modern Cartesian *science*.

I will try to deal with rationality in the most universal sense possible but also in the efficiently technicizible sense. Systems with possibility of gaining information from their environment and of influencing this environment in certain ways will be discussed. For the purpose of conciseness, these systems will be called *agents*; the reader can understand *agents* as the self, and robot or an economic unit, for instance a corporation. I will try to show the development of approaches to rationality from the initially formalistically rigorous efforts to analyse this phenomenon through the pursuit of the aim to synthesize (technically (re)construct) artificial rational agents up to one of the possible understandings of rationality in the coming postmodern age.

In spite of a number of possibilities how to understand the *rationality* of an agent, I think that there is a certain core of the meaning of this term which can be revealed within each of the acceptable definitions. The rational behaviour of an agent always presupposes adequacy of its functions in the environment to the aims pursued.

The aims can again be widely perceived: the aims are not only predicates which 'should become' true as a consequence of the performance of a certain sequence of actions in the environment (e.g. for a certain robotic, through grasping object c not only the predicate 'GRASPED' (c) becomes true but also predicates, the continual maintenance of the validity of which is the aim of the agent (e.g. to select macroeconomic steps in a country k, so that they would be continual – at every moment t – the true predicate STATE-BUDGET-BALANCED (k, t)). The two types of aims mentioned are often combined into a demand to reach a particular state of matters and to maintain it after reaching it.

Both types of aim assume *dynamic*, changing environments. They are changing at least as a result of the activities of the agents operating there. This is, however, a rather great idealization of the real environments changing very often unpredictably enough for the agent. It happens frequently that the aim to maintain continuously a certain state of matters does not evoke convincingly enough the idea of rationality of the agent that achieves it. In a dynamically changing environment an impression can arise that the agent actually does nothing autonomously, it does not change the environment, in which it operates, in any way, although implicitly it adapts very actively its behaviour to changes taking place in the environment. From another perspective accepted here, it is irrelevant to the agent whether it acts so as to change the state of matters or because the state of matters has changed. It is important whether it has *at least two opportunities to act*, whether at least one of the opportunities is, from the point of view of the agent's aim, *more advantageous* than the others and whether the *rational* agent will *perform the (most) advantageous task.*

Man is considered to be a prototype of the rational agent. An effort to analyse the agent's activities was present at the birth of every effort to understand rationality as a specifically human activity. The logic of human thought and argumentation by speech was the background of Aristotle's drafting of *Organon*. *Logic* and *experiment* gradually became the basic means of *legitimization* of scientific knowledge in modern science understood by Lyotard (1979) as the fact that a certain statement has to fulfil a set of conditions in order to be accepted as scientific. According to Lyotard legitimization is the process, through which the "law-maker" judging scientific discourse takes on justification to prescribe the mentioned conditions, required for a particular statement to be part of this discourse and to be taken into account by the scientific community. The foundations of this science were laid by R. Descartes. Human knowledge, its background, specifics and quality were the issues of his consideration when writing his *Discourse on Method*. He thus laid foundations of the view within the framework of which, for example, *good* or *beauty* started to be perceived as something subjective and truth became the only means through which people give an account of the external world. J.-F. Lyotard, as far as I remember, called this shift *depersonalization of the universe.*

The background of human rationality can also be revealed in much later efforts to explain human behaviour, for example in (cognitive) psychology but also in semiotics. U. Eco alerts us to the prerequisite for the rational element in every reader of a literary work, writing that in each argument, in which qualitative names or a certain description occurs, it is assumed that the reader or the listener takes the existence of the subject about which something is stated, for granted (Eco, 1994). In these researches the presupposition that a particular person thinks and acts rationally has a large explanatory and predicative strength. Last but not least, I will also mention a prerequisite of the rationality of the behaviour of whole human communities, which is accepted, for example, in economics. It is this context within which very precisely formulated formal theoretical frameworks were created for research on rationality towards the end of the first half of our century.

2 Traditional frameworks – analysis as formalization

Contemporary science studies rationality from diverse positions in various meanings. I will try to draft a sort of classification of the problems.

First, if speaking about rationality, we have to decide whether we are interested in the behaviour of a *single agent* or whether we focus on the rationality of the behaviour of a *society of agents*.

If our interest is directed towards the issues of rationality of a society of agents, we should realize whether the elements of this society pursue contradictory aims or whether we are interested in the cognition of societies whose components act in the interest of a certain common welfare, that is they pursue common aims. In the former case, research belongs primarily to the domain of *game theory* and in the latter case it is probably ethics that deals with the issue of rationality.

If an individual or a society is interested in rationality, our interest is directed to whether decisions are made under the conditions of *certainty* or *uncertainty*. In the former case – rather idealized – the issue belongs to the area of *utility theory* and in the latter to *decision theory*.

The basic principles of the game theory and the decision theory are based on the assumptions that:

– a set of the states of the environment, in which the agent can find itself and is able to recognize them, is, from the perspective of the agent's decision-making, describable in a sufficiently precise way,

– the agent chooses in the situation, in which it has just found itself, a task, which it performs, from among the finite number of potential possibilities how to act,

– the agent has a certain idea of the utility of the performance of a certain action in a certain state of the environment,

– the agent knows the utilities following from the accomplishment of the states of the environment on the assumption that the agent performed a task and, consequently, entered this status of the environment – it is acquainted with the so-called utility function, and

– the agent has an idea of the expected utility of the performance of a certain task following from its knowledge of the utility of this status and of the probability of the occurrence of this status, assuming the performance of the particular task – it is acquainted with the so-called function of the expected utility.

The aim of the rational agent in decision theory is, under such pre-conditions, optimization (maximization) of its function of the expected utility.

Game theory deals with the situations when the rational agent – the player – has to make an optimum decision on the assumption that another agent also makes a rational decision as follows: The sequence of tasks performed by players is controlled by the strategy of each of them. The strategy – actually means the rules according to which players select tasks – moves – and which they later perform. Let us assume that player I plays according to his strategy I. Let us assume that keeping to this strategy is for him, within the framework of the game with a playmate, the highest utility. In such a case strategy I is the best response of player I to the strategies of other players (there can be more of them but eventually a lot) – his optimal strategy.

The situations interesting in the game theory are those in which the strategy of each player is optimal with respect to the strategies of other players. In such games a sort of equilibrium is achieved – the so-called Nash equilibrium examined by J. Nash, who was awarded the Nobel Prize for his contribution to the field of game theory in 1994. Nash found that every finite game has at least one such equilibrium (in fact, however, real games usually have many more of them).

In game theory, Nash's name is also connected with the division of the games into cooperative and non-cooperative. In cooperative games players strictly observe mutual agreements. In non-cooperative games players are not able to achieve consensus.

The strategy of the player in a non-cooperative game is stable when Nash's equilibrium is involved in the particular game. It is evident since in such a case the strategy of every player is optimal with respect to the strategies of other players. In the case of the cooperative game the strategy of the player can also be stable when Nash's equilibrium is not reached in this game. This happens when players conclude an agreement on playing a certain strategy combination.

3 Traditional artificial intelligence – the top-down approach to synthesis

Although at the beginning only implicitly, a new field has directed its attention towards the issues of rationality since the middle of this century. It was artificial intelligence, developing at first within cybernetics and the emerging computer science. Artificial intelligence involves the automation of the procedures leading to the solution of as wide a range of problems as possible, the problem being understood as a discrepancy between the state of matters as they actually are and the state in which we want them to be. Primarily it was supposed that to be intelligent means to have reliable and generally valid criteria for a successful procedure of problem solving. However, from the philosophical point of view, this is in principle a demand for the rationality of agents. The engineering aim of artificial intelligence was and still is to construct a rational agent. Problems associated with a suitable definition of probably the most general understanding of rationality (intelligence) of such an agent have accompanied artificial intelligence from its very beginning. The first attempt at a definition was advanced by A.M. Turing (1950).

Both in economics and in artificial intelligence there was a relatively great amount of justified criticism of the definition and functioning of rationality as a capability to find optimal solutions. The critiques of the two Noble Prize winners, F.A. von Hayek (1990) and H.A. Simon (1986), can serve as examples. Moreover, the existing mathematical formalizations affect the environment in which the rational agent has to decide in its, in a way denaturated, narrowed shape. They work with an abstraction of reality and decisions, to which such models lead, being optimally again in the abstract environment at the most and not in that which was the model for the mathematical abstractions. Such results are, however, rarely as useful in practice as in models. The real rational agent – whether a man or an economic organization – can also be rational in a different sense. It is not necessarily the optimality of one's decisions that is at issue but rather their acceptability. It concerns the maintenance of one's relationship to the environment in a certain reasonable corridor of admissible fluctuations.

Apart from the orthodox account of the concept of rationality another interpretation is possible: "rational" can mean according to this view more "reasonable" than "methodological". An effort to reach optimality of the decisions can be in many real situations replaced by an effort to achieve their practical acceptability. This side of rationality was reflected in economics by H.A. Simon through introducing a concept of the so-called bounded rationality and subjecting it to systematic investigation (see e.g. Simon 1960, 1986).

The conception of bounded rationality in artificial intelligence was successfully applied by Simon and his co-workers in the field of automatic problem solving. It concerned in principle the idea of such processes whose aim was not to gain the attributes of theoretically justifiable resultats, aggregation or effectiveness in search of the solutions to problems of a particular type. It was the acceptability of the found solutions to the majority of problems occurring in practice that was required. Such processes could be identified and successfully reconstructed into the form in which they could be computer-processed. Heuristic programming was thus created as the field that played an important role in artificial intelligence, particularly in the sixties. The results of these researches were embodied, for example, in the system of computer programs named General Problem Solver (GPS) which was presented in a monograph (Ernst, Newell, 1969).

The conception of the authors of the GPS system represented in artificial intelligence a shift from the efforts to build rationality on the basis of the theoretical knowledge of decision-making as such towards the efforts to formulate precisely and eventually reproduce in computers the procedures of the rational behaviour of man. These tendencies are clearly identifiable in the monographs (Newell, Simon, 1972 and Newell, 1990), the contents of which are also relevant to the issues of cognitive psychology.

Another specific feature of traditional artificial intelligence in its approach to the construction of the rational agent was the hypothesis that what is processed by rational methods one way or another and on the basis of which decisions are made and problems are solved, is knowledge. Artificial intelligence was developing in the seventies as a discipline studying conditions, boundaries and, within them, the technical possibilities of the ability of codifying the agent's experiences (Stefik, 1995). Man served as a model with his knowledge and ways of applying it in problem-solving or formal logical systems, it analysed them and looked for pragmatically applicable proposals of the rational systems that can be created artificially. It was the unique rationality of man capable of solving problems, that is able to adapt his environment to his visions of what the environment should be like through his activities that served as a functional model of such systems in both cases.

The results of such researches have found wide application particularly in the area of the development of the so-called knowledge-based and expert systems. The creation of these urgently needed computer systems[2] means in principle a rigorous recapitulation of the codified conceptual side of a particular section of human intellectual activities – the so-called expert domain and the transformation of the result of the recapitulation into the structures offered by the particular development environment in a computer. The applications of traditional artificial intelligence in robotics are far from achieving success comparable to that of knowledge-based systems. The so-called novel artificial intelligence is a reaction to the failure of the traditional cognitive robotics.

4 Novel artificial intelligence – the bottom-up approach to synthesis

Let us now have a look at rationality as an ability of an agent to maintain the truth of a predicate about the agent's relationship to its environment in spite of dynamic changes in this environment. I have already mentioned that in such a case it sometimes escapes our notice that the agent not only acts in such an environment but that it even acts rationally. The dynamic change of the environment can in very simple cases be caused, for example, by the change of the agent's position in this environment. This happens for instance with a mechanical toy – ladybug – which, thanks to its construction, does not fall off the table surface. It might seem that the surface of the table is in this case a stable environment. However, it actually changes because the ladybug on it is embodied. It is actually the body of ladybug and the surface of the table that create the environment in which the ladybug develops its activities. The change of the position of the ladybug on the surface of the table should therefore be understood as a dynamic change of the environment.

I used the example of the mechanical ladybug in various places and for the purpose of the illustration of various principles (see e.g. Kelemen, 1994a, 1994b). Judging from the behaviour of such a ladybug, nothing can prevent our expecting the representation of the knowledge, inference module, strategy of problem-solving used by GPS, etc. in the background of the behaviour of ladybug; my experience of teaching tells me that it is methodologically quite suitable to use such a simple example for explaining the principles of traditional artificial intelligence. A glance at the ladybug, however, makes it clear that it does not use these principles for generating its behaviour. It is obvious because mechanical toys like the ladybug are well known to the majority of us from childhood and its environment is trivially simple for us. I have given (Kelemen, 1996) a theoretical analysis of the general architectural principle on the basis of which the ladybug was constructed and the behaviour produced by it can be considered to be a specific level of rational behaviour, which I called there low-level rationality.

---

[2] The urgent need for knowledge-based and expert systems has been discussed before in Human Affairs (Kelemen, 1993).

The mechanical ladybug is a good example of the sort of system, which began to be studied in the mid-eighties by some experts in the field of artificial intelligence and robotics. These mostly pragmatically oriented engineers looked with a sort of denial at the efforts of traditional artificial intelligence to create rational systems according to the human model. At the beginning of the 1990s, R.A. Brooks (1991, 1993[3]) came up with a programme, which cast doubt on the need for representation and thinking as prerequisites for at least some – practically already interesting – rationality of systems. This doubt came from – apart from other causes – awareness of the practical uselessness of machines, which would be rational in these terms – at the highest level of rationality which we meet in nature. Their interest returned to simple equipment (such as the above-mentioned mechanical ladybug) or to natural systems (insects, even bacteria) which "survive" very effectively in changing environments. This at first led to an assumption that representationalism and the anchorage of artificially created intelligent agents in symbols represent an inadequate approach to the construction of traditional agents. However, some positive connections between the two currents emerged a little later; they are described in more detail, for example, in the works of Kushmerick (1996), Vera, Simon (1993) or Kelemen (1996c).

In these new conceptions rationality became an attribute of effective adaptation to the environment in which the agent "survives"; in other words, the agent maintains its relationship to the environment to preserve the validity of a certain predicate in spite of the changes taking place in the environment and which, without its interventions, would lead to circumstances under which this predicate would become invalid. (For instance, the mechanical ladybug maintains the validity of the predicate which could briefly be formally expressed as (mechanical ladybug) IS-ON-THE-TABLE.

5 Emergence of rationality – the role of embodiment and societies

The possibility of investigating the process of mutual subsumption and thus the process of the formation at least of low-level rational agents from irrational components follows from the fact that rational agents in the form they are perceived by novel artificial intelligence, are created by properly mutually subsumed components which are not necessarily individually rational. The possibility of the formation of low-level rational agents from irrational ones was, as I have already mentioned, theoretically described by me (Kelemen, 1996a). Later (in Kelemen, 1996b) I drafted a connection between the rationality understood in this way and the so-

[3] The work (Brooks, 1995) is a book version of the plenary lecture delivered by R. Brooks at the 12th International Joint Conference on Artificial Intelligence in 1991 in Sydney.

called substantive rationality characterized by H.A. Simon (e.g. in Simon, 1976 or Simon, 1986) as a rationality based on the immediate reaction, in contrast to the so-called procedural rationality based on consideration preceding an action. Some connections between novel artificial intelligence and these conceptions of rationality, in principle following from economics, have been analysed, for example, in the works of Liday, one of my students at the Bratislava University of Economics (Liday 1996, Liday 1997).

Let us return to our example of the mechanical ladybug and examine the possibility of its formation from irrational components of certain types. Let us assume that there are two types of components on the surface of the table, which create the environment of the low-level rational mechanical ladybug. One moves directly on the table surface and the second rotates on the table surface. If any of them approaches critically the table edge, it falls off.

Let us assume that if two different types of the components mentioned approach critically on the surface, they join to form one machine, which creates the system similar to mechanical ladybug and the system so created will be able to identify the table edge and to avoid the fall. In this way rational components may be created on the table surface from irrational ones. This process has, however, certain interesting properties on the assumption that agents are "embodied".

If, during the motion on the table surface two components of the same type meet or if a machine of the type of mechanical ladybug meets any component, let both systems stop working after their meeting, and stay in the place where they stopped working. If they are "embodied", they will create a barrier after meeting on the table surface. Let us realize that the ladybug formed by joining two components, is not able identify or avoid this barrier sensorially. In this way the number of fatal obstructions can increase on the table surface and their growth can also cause the disappearance of rational agents, which might have originally been formed in the environment. The growth of the number of rational agents or components can further complicate this process and already at this very simple level the behaviour of agents in a dynamically changing environment can be made very unpredictable. Nowadays there is a lot of experimentation with such situations carried out in computer-simulated environments; I am best informed about the experiments I have just mentioned, which are conducted by my student at the Silesian University in Opava; some results have been presented (Štýbnar, 1998).

Under the conditions I mentioned above, an agent or society of agents should probably be more rational if it is able to survive as long as possible in a dynamically changing environment.

There is now a relatively great optimism concerning the agents situated in sufficiently complex environments. Here is the opinion of the pioneer of these efforts, R. A. Brooks: "My feeling is that thought and consciousness are epiphenomena of the process of being in the world. As the complexity of the world increases, and the complexity of processing to deal with that world rises, we will see the same evidence of thought and consciousness in our systems as we see in the people other than ourselves now. Thought and consciousness will not need to be programmed. They will emerge." (Brooks, 1995, p. 70).

6 Conclusion – rationality at the end of the modern age

In the preceding sections I tried to outline some dominant ways which contemporary science has found and used for the analysis of rationality (primarily economic, that is human organizational units) and those found for the synthesis of rational agents (although in this case not only human rationality was a model).

The rational agents subjected to analysis were silently assumed to have as the aims of their activities changing of the environment in which they are, so that the changes would lead to the accomplishment of a particular aim of their activities, that is to the fulfilment of a predicate which had not originally been fulfilled but was fulfilled as a consequence of their activities.

It is noteworthy that the given approach to rational human action was also adopted by some philosophers. Rationality meant, using the formulation of R. Rorty (R. Rorty, 1989), to have criteria of a successful procedure worked out in advance. To find such criteria or to formulate the applicable methods for their discovery is the role of modern science. In the cited essay, Rorty constructs the idea of the science of the postmodern age, which should replace the idea of science as a means of changing the existing state of matters, that is, of modern science. He therefore proposes another interpretation of rationality, where rational means to be reasonable rather than methodical. As for the character of society, in which such a rationality will function, Rorty outlines a vision of a community which will have higher aspirations than the maintenance of itself and the preservation of civilization. In such a community rationality will be identical with the efforts to sustain it and not with objectivity. In his opinion, no stronger foundations than mutual loyalty will be needed.

It seems that rationality in the variety of forms which we have been able to give to it in different periods will also accompany us in the postmodern age. Within the new context outlined by Rorty, the understanding of rationality within the intentions of novel artificial intelligence can become not only a technically interesting and useful tool of the further development of technology on the basis of modern science, but it can also become a certain forerunner of technicization of rationality as (maybe) postmodern science will understand and approach it.

By generalizing what we have discussed so far, it would probably be worth trying to characterize individual periods of our civilization through the understanding and functions of rationality which dominated. Such a look might uncover something of the mechanisms of the functioning of individual periods of human civilization and some causes of the changes in these mechanisms. The understanding of rationality could then be stratified and rationality could gain a different position rather than being a characteristic of only one period of the development of our civilization – roughly the period of the European modern age.

REFERENCES

BROOKS, R.A.: Intelligence without representation. *Artificial Intelligence 47* (1991) 139-159

BROOKS, R.A.: Intelligence without reason. In: *The Artificial Life Route to Artificial Intelligence.* (L. Steels, R.A. Brooks, eds.), Lawrence Earlbaum, Hillsdale, NJ., 1995

ECO, U.: *Six Walks in the Fictional Woods.* Harvard University Press, Cambridge, Mass., 1994

ERNST, G.W., NEWELL, A.: *GPS – A Case Study in Generality and Problem Solving.* Academic Press, New York, 1969

HAYEK von, F.A.: *The Fatal Conceit – The Errors of Socialism.* Routledge, London, 1990

KELEMEN, J.: Expert systems as a necessity. *Human Affairs 3* (1993) 3-10

KELEMEN, J.: *Strojovia a agenty* (Living Machines and Nonliving Agents). Archa, Bratislava, 1994a

KELEMEN, J.: On rationality of behavior-based agents. In: *Cybernetics and Systems '94.* (R. Trappl, Ed.), World Scientific, Singapore, 1994b

KELEMEN, J.: A note on achieving low-level rationality from pure reactivity. *Journal of Experimental & Theoretical Artificial Intelligence 8* (1996a) 121-127

KELEMEN, J.: How reactive agents do right things. *Acta oeconomica pragensia 5* (1996b) No. 1, 105-120

KELEMEN, J.: Towards robots with minds. *Human Affairs 6* (1996c) 97-110

KUSHMERICK, N.: Cognitivism and situated action – two views on intelligent agency. *Computers and Artificial Intelligence 15* (1996) 393-417

LIDAY, M.: Towards rational decision support agencies. *Acta oeconomica pragensia 5* (1996) No. 1, 39-57

LIDAY, M.: Expertise management as a multi-agent activity. In: *Proc. 7th Intern Conf. Artificial Intelligence and Information-Control Systems of Robots '97.* (I. Plander, Ed.), World Scientific, Singapore, 1997

LYOTARD, J.-F.: *La condition postmoderne.* Editions de Minui, Paris, 1979

NEWELL, A.: *Unified Theories of Cognition.* Harvard University Press, Cambridge, Mass., 1990

NEWELL, A., Simon, H.A.: *Human Problem Solving.* Prentice-Hall, Englewood Cliffs, NJ., 1972

RORTY, R.: Science as solidarity. In: *Dismantling Truth – Reality in a Postmodern World.* (H. Lawson, L. Appignanensi, eds.), Weidenfeld and Nicholson, London, 1989

SIMON, H.A.: *The New Science of Management Decision.* Harlper & Row, New York, 1960

SIMON, H.A.: From substantive to procedural rationality. In: *Method and Appraisal in Economics.* (S.J. Latsis, Ed.), Cambridge University Press, Cambridge, 1976

SIMON, H.A.: *The Science of the Artificial (2nd Edition).* The MIT Press, Cambridge, Mass., 1986

STEFIK, M.: *Introduction to Knowlegde Systems.* Morgan Kaufmann, San Francisco, Cal., 1995

ŠTÝBNAR, L.: Achieving low-level rationality from pure reactivity – some empirical observations. In: *Proc. 14th European Meeting on Cybernetics and Systems,* Vienna, Austria, April 1998 (to appear)

TURING, A.M.: Computing machinery and intelligence. *Mind 59* (1950) 433-460

VERA, A.H., SIMON, H.A.: Situated action – a symbolic interpretation. *Cognitive Science 17* (1993) 7-48