

Autonomous Agents, AI and Chaos Theory¹

George Kiss

Human Cognition Research Laboratory
The Open University
email: gr_kiss@vax.acs.open.ac.uk

HCRL Technical Report No. 71
September, 1990

¹ In: J.-A. Meyer and S. Wilson (Eds) From Animals to Animats, Proceedings of the First International Conference on Simulation of Adaptive Behavior. Cambridge, Massachusetts: MIT Press, 1991

Abstract

Agent theory in AI and related disciplines deals with the structure and behaviour of autonomous, intelligent systems, capable of adaptive action to pursue their interests. In this paper it is proposed that a natural reinterpretation of agent-theoretic intentional concepts like knowing, wanting, liking, etc., can be found in process dynamics. This reinterpretation of agent theory serves two purposes. On the one hand we gain a well established mathematical theory which can be used as the formal mathematical interpretation (semantics) of the abstract agent theory. On the other hand, since process dynamics is a theory that can also be applied to physical systems of various kinds, we gain an implementation route for the construction of artificial agents as bundles of processes in machines. The paper is intended as a basis for dialogue with workers in dynamics, AI, ethology and cognitive science.

1 Introduction

Agent theory is a branch of artificial intelligence (Kiss, 1988). Its domain is the theory, design and implementation of artificial systems, similar to animals or people, that are capable of autonomous, rational actions through which to pursue their interests and goals. Aspects of this theory cover, among other things, how actions are related to knowledge, how plans for actions to reach goals can be formed, how goals are formed, what the role of intentions for action is, how the state of the world is perceived, and many others.

The abstract formulation of agent theory can be stated in many different languages, both informal and formal. Much current work in this field has made use of formal logical languages (Georgeff and Lansky, 1986). Although these specialised logics are convenient and expressive, often it is difficult to formalise their semantics, or the semantics that have been offered have undesirable properties. An example of this is the possible-world semantics of epistemic logics which unfortunately makes agents omniscient.

The implementation of theories expressed in such formal languages has additional problems. When agent implementation is done by direct mechanisation of the logic, for example as a theorem-prover, the resulting systems turn out to be inefficient. This is a natural consequence of the expressiveness of the language. On the other hand, the languages are sometimes not expressive enough to deal with some concepts that seem needed to describe agents. An example is the expression of quantitative magnitudes for describing strength of belief in an agent.

Refinement of these logics and their formal semantics, and their efficient implementation, is of course an ongoing enterprise. This paper is intended as an informal preliminary to such work, offering some intuitions about the interpretation of agent theory through the general theory of process dynamics.

Such an interpretation can also provide a strategy for implementation. The situation is analogous to the relationship between the abstract Boolean algebra of classes, the propositional calculus, and hardware logic circuits. The abstract algebra is defined in terms of classes and operations on them; intersection, union, complementation, etc. One interpretation of the Boolean algebra is propositional logic, where the variables range over propositions and the operations are truth-functional manipulations, etc. The possibility of implementation arises from the fact that another interpretation of Boolean algebra can be found in the operation of physical electrical circuits. Because of this, the operation of the circuits can thus be described by propositional logic, or stated conversely, the circuits are an implementation of the logic.

Let us represent this by the following schema:

Propositional logic -> Abstract Boolean algebra ->
Electrical circuits

This suggests that we should look for an abstract mathematical theory such that both agent theory and some suitable physical systems can be interpretations of it. The abstract theory can then be used as an intermediary between agent logic and the physical systems that can be used as efficient implementations. This paper explores the possibility of using abstract dynamics as such an abstract theory and physical dynamic systems as implementations of agents, as shown by the schema:

Agent theory -> Abstract Dynamics -> Physical dynamic systems

A theoretical foundation has already been laid by Rosenschein (1985, 1986, 1989) for the epistemic (informational) analysis of agents regarded as processes. We propose that Rosenschein's framework is to be extended in two senses. First, the process-based interpretation of agents is to be extended from the epistemic to praxiologic (action-related) and axiologic (value-related) concepts. Second, processes are to be analysed in terms of their state space dynamics in addition to the correlative relationships between states of a process and an environment.

2 Concepts in Agent theory

The main idea an agent theory attempts to capture is purposiveness: that agents execute actions in order to reach goals. Refinements of the theory are concerned with optimality issues. Rational agents are often described as executing actions which maximally satisfy their goals.

To provide more structure, agents are also often described in mentalistic, intentional terms by attributing to them (propositional) attitudes. Some of the main examples of such attitudes, perhaps a minimal set of them, are wanting, knowing, liking (related to preferring), and intending.

The common-sense interpretation of these concepts is briefly as follows. Knowledge characterizes how the world is, from the agent's point of view. Likes characterise how the agent likes the world to be. Wants characterise the agent's commitment to reach a goal. Intentions characterise the commitment of an agent to an action.

3 Concepts of Abstract Dynamics

The main concepts of dynamics deal with the structure of state spaces. Abstractly, the theory can be formulated in terms of functional iteration. The functions which define dynamical systems are also called mappings or maps. The main concern of the abstract theory is with the asymptotic behaviour of iterative mappings. The iteration of a function is a discrete process. If the process is continuous, the description is often given in the form of differential equations to describe the behaviour of the solution over time.

In a geometric interpretation, the iterative process maps points into points. The points correspond to the states of the process. The process is then said to go through a trajectory or orbit of points. The main concern of dynamics is to understand the nature of all trajectories of a system and to classify them as moving to a fixed point, being periodic, asymptotically periodic, etc. We shall now turn to an informal summary of some of these concepts. For more detail, see, for example, Abraham and Shaw (1981), Devaney (1986), Thompson and Stewart (1986), or Cvitanovic (1984). Cvitanovic also contains an extensive bibliography. The field is developing very rapidly under the designation of chaos theory, which is a specialised branch of dynamics.

The *state space* of a system is generally a topological surface (manifold) on which the possible states of the system are located. This can be just three-dimensional space, or some curved surface, for example, like a doughnut (torus).

It is normally assumed that there is a *vector field* acting at all points of the state space. This vector field determines the *dynamics* of the system by constraining the *trajectories* to certain directions at each point of the state space. When typical or many trajectories of the system have been drawn, we get a *phase portrait* of the system.

Closed trajectories produce *cyclic behaviour*. Trajectories can otherwise take many shapes, like spirals, straight lines or any kind of curve.

The focus of interest is in the *asymptotic behaviour* of trajectories. *Limit sets* of state spaces are sets of points

towards which the trajectories move asymptotically. Limit sets may be solitary points, or cycles, or more complicated distributions of points. Limit sets which are solitary points, are called *fixed points*.

Fixed points of functions are points x for which $f(x)=x$. That is, the fixed points are mapped into themselves by the function. Fixed points are important in dynamics, because they correspond to equilibrium (steady) states of systems. Once a system has somehow got to a state which is a fixed point, it will not move from that state under the iteration of the function f .

It is of interest to ask how a system may get to a fixed point. The simplest case is that the system may start from an initial state that is a fixed point, and there will be no further change. More interestingly, trajectories starting at other states may lead to a fixed point after a number of transitions. In such cases we say that the fixed point *attracts* the trajectory. The set of states from which trajectories lead to an attractive fixed point are called the *basin of attraction* of the fixed point. It turns out that a fixed point is attractive if the slope (derivative) of the function f is less than 1 at the fixed point. The magnitude of the slope characterizes the strength of the attractor: the greater the strength, the faster the trajectory approaches the fixed point. Two different kinds of behaviour in the neighbourhood of a fixed point are illustrated in Figure 1.

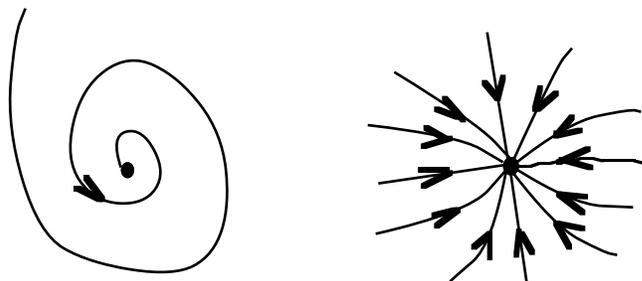


Figure 1. Attractive fixed points

A periodic point is a generalisation of the concept of the fixed point to the case when a trajectory cyclically visits a point after every n iterations of the function f .

If the iteration is run backwards, trajectories would appear to diverge from an attractive fixed point. In this situation the fixed point is called a *repellor*. Such fixed points correspond to unstable equilibria in physical systems. Slight disturbance from the equilibrium starts the system on a trajectory leading away from the equilibrium state. Conversely, attractive fixed points correspond to stable equilibria.

An interesting situation is shown in Fig. 2. There is a limit point which is attractive for points on the left and repelling for points on the right. The two heavy lines show trajectories which are not in the basins of attraction

and repulsion, and are deflected by the presence of the limit point without going through it. This illustrates how the global attractors and repellers influence the local shape of the trajectories of the system even when the trajectories do not touch them.

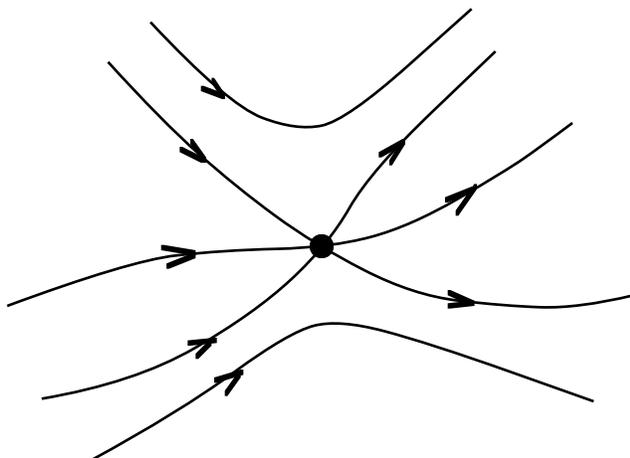


Figure 2. Deflected trajectories

4 Dynamics and Information

Dynamics and information theory are connected. The connection with the concept of information can be found when we consider the information needed to *specify* or *measure* the state of a dynamic system at a point in time. For example, when we want to start up a dynamic system from some initial state, we need to specify that starting state to some degree of precision. Conversely, if we try to find out what the state of the system is at some later instant of time, we can only measure its state to some limited degree of accuracy, determined by the resolution of our measuring instruments or sense organs. We can then speak of the amount of information needed to make the specification, or the amount of information gained from the measurement.

The convergence and divergence of trajectories near limit sets are associated with information gain and loss. Divergence of trajectories is associated with greater uncertainty about the actual state of the system and therefore with information loss, as can be seen from Figure 3. Consider the trajectories that go through the smaller circular area on the left. In order to specify or locate a particular trajectory, we have to specify the position of a point within this area. The accuracy of specification or measurement defines a coordinate grid over the area. The amount of information involved is that needed to select a cell in this grid. The larger circular area on the right shows the effect of divergence in the phase space. If the accuracy of specification or measurement remains constant, a larger amount of information is needed to specify the location of a trajectory, and a larger amount of information is yielded when the state is measured, because there are a larger number of cells in the grid.

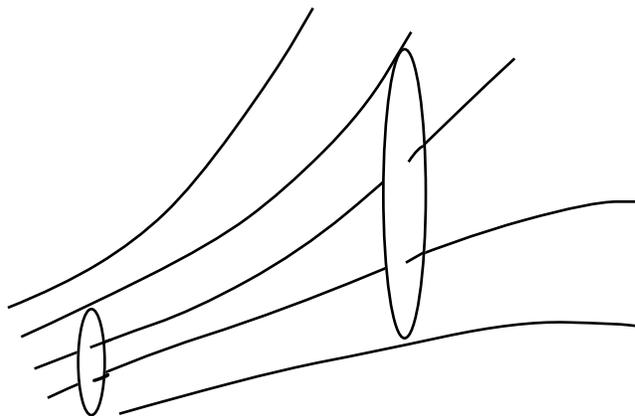


Figure 3. Information loss in divergent flow

Since diverging flows involve a progressive loss of information about the state of the system they lead to sensitive dependence on the initial conditions (the so-called "butterfly effect": the flapping of a butterfly's wings can cause a cyclone later). Although the system is *deterministic*, it becomes *unpredictable* in the long run due to this information loss.

As far as practical applications are concerned, this phenomenon can be used for *amplification and control* purposes in information processing: a small change in initial conditions can produce a large change in later behaviour. The converse effect, gain of information in convergent flows, produces insensitive dependence on the initial conditions and can therefore be used for *classification or recognition* processes in information processing. The convergence of trajectories to a fixed point serves as the recognition of the starting points in the basin of attraction as belonging to the same class.

5 Agent Attributes and Dynamics

We now turn to the central proposal of this paper, which is to interpret agent-theoretic concepts in terms of abstract dynamics.

Rosenschein (1985) has already interpreted agents as processes and offered an operationalisation of epistemic attitudes in terms of states of a process. The epistemic interpretation of states hinges on there being a correlation between states of the environment (the world) and states of the agent. When such a correlation exists, the agent is said to know a fact that can be stated as a description of the corresponding world state.

This interpretation of knowledge as states of a process does not yet offer any insight into why the correlations exist and what their nature is, nor is there any principle that would describe the dynamics of the process that is said to be the agent.

Complex agents are architecturally compositional both structurally and behaviourally. The complex agent structure is produced by assembling simpler component elements. Complex agent behaviour is produced through the interactions between the simpler component behaviours. A vital point is that agents are *nonlinear* systems. The importance of this is in the fact that *only in nonlinear systems* will *qualitatively* new phenomena arise through the interactions between components. In linear systems these interactions will be merely additive accumulations of size (scaling), without new features. Nonlinearity leads to the appearance of new features in the *global* behaviour of the system which were not present in the behaviour of the *local* components and are not just the additive accumulation of the component behaviours.

Concurrency, parallelism and distributed systems become important because structural and behavioural complexity only emerge as the result of iterative accumulation of local components. The local components can be identical and simple in the computations they carry out. Complex structures appear as the result of the many interactions. For real-time operation the execution of the numerous *simple identical steps* can be, and needs to be, parallel. Long behavioural trajectories can then be computed in essentially a single step in the limit of parallelism, or increasingly serially between a single-processor and the maximally parallel system. Proportional speed-up is thus obtainable by adding more processors. This contrasts with the rather more limited gains that can be obtained in the case of carrying out complex structured computations on multiprocessor systems.

The dynamics of the agent process are of importance in relation to all three aspects of agent theory: epistemological, praxiological and axiological.

The epistemic issues are concerned with the way states of the environment process determine states of the agent process and hence produce the correlations referred to above. The trajectory of a part of the environment process (abstractly in a phase-space, concretely in physical space) enters into the region within the boundary of an agent system and then continues as part of the agent process. Since this part of the agent process is causally determined by the environment process by means of the state-to-state transitions, the required epistemic correlations will be produced. In automata-theoretic terms these are inputs. In agent-theoretic terms these are sensations (at the boundary) or perceptions (further inside the agent) or cognitions (deep inside the agent). Looking at this another way, the agent process is simply a part of a global world process and its states are parts of the world states. An agent is just a local phenomenon within the global world process.

Praxiological issues are concerned with the way agent processes eventually (at the boundary of the agent) exert causal influence on the environment process. In automata-theoretic terms, these are outputs. In agent-theoretic terms,

these are actions (at the boundary), volitions, willings or intentions (progressively deeper within the agent).

Axiological issues are concerned with the dynamics of the agent-process trajectory in its phase-space and, in particular, with the directional nature of the process. The teleological nature of agent behaviour is one of the central examples of such issues. In terms of nonlinear system theory, the dynamics can be described in terms of the movement of the system state *towards* stable equilibrium states and *away from* unstable equilibrium states. Teleological agent behaviour is to be identified with movement towards stable equilibria which are in this sense *preferred* states of the system: the agent *likes* to be in these states. Aversive agent behaviour is to be identified with movement away from unstable equilibria which are in this sense *disliked* by the agent. In the terminology of dynamic system theory, these states are *attractors and repellers*. Unstable equilibria arise mainly through competition between attractors and represent boundaries between the basins of attraction of those attractors.

The correspondence between the various main characteristics of *purposive* agents and dynamic process-related concepts can be sketched as follows:

- Agents correspond to atomic or structured processes.
- What the agent knows corresponds to the information contents of a state.
- An agent's actions correspond to a change of state in the state space.
- The likes and dislikes of an agent correspond to the global (high-dimensional) attractors and repellers of the state-space. In complex agents these are elements of a value system.
- A goal of an agent corresponds to a local (low-dimensional) attractor in a basin of attraction.
- A want of an agent corresponds to a *trajectory* which converges to a local attractor in its basin of attraction.
- Hedonic states of pleasure and pain correspond to satisfaction and dissatisfaction of the constraint system constituted by the attractors and repellers of the state-space, i.e. to *distance* from them.

Following Rosenschein's conceptual framework, processes will be regarded as trajectories over time and spatial locations. A trajectory is specified by a function $w : L \times T \rightarrow D$, where T is a set of times, L is a set of locations of some physical system and each location a can take on values from some set D_a . D is the union taken over all sets D_a . An agent will be taken to be a process or a

structured set of processes. The structuring is defined in terms of constraints between the process states.

Interpreting goals as local attractors can be seen as a more special case of trajectories being shaped by the existence of attractors and repellers in the state space. Attractors and repellers determine the direction of movement, i.e. the direction of agent action. It is natural to interpret the pro- and anti-attitudes of agents with this kind of directionality. Movement according to attractors and repellers leads to the notion of the movement *satisfying* the constraints applied by them. We propose to tie this notion of constraint satisfaction to the notion of hedonic satisfaction in complex agents. Pleasure and pain constrain agent action in the same structural sense as attractors and repellers constrain dynamic systems. We assume that due to the physiological structuring of living organisms attractors and repellers are created in their behavioural space. By analogy, it should be possible to create attractors and repellers in non-living computational systems through appropriate construction or programming.

Wants in agent theory express the notion that an agent is committed to carrying out some action or series of actions to reach a goal. This commitment can be naturally interpreted as embarking on a trajectory towards a local attractor in a basin. Weaker forms of desire, like wishing, can be interpreted as a belief that some goal state is attractive, without commitment to actions.

For simplicity we do not distinguish here between belief and knowledge as epistemic states of an agent. Adopting the theory proposed by Rosenschein (1985, 1986), we interpret knowledge as the information content of an agent's state. The information content expresses the relationship between an agent's internal state and the corresponding state of the environment. An agent is said to know a proposition p in a situation in which its internal state is s , if in all possible situations in which the agent is in state s , p is satisfied by the environment. Recall that agents are interpreted as processes, so that agent states correspond to process states. An agent can only tell what state the environment is in by examining parts of its own internal state, i.e. the state of its sensory apparatus. Since there are causal constraints between the state of the environment and the state of the agent, the agent's internal state contains information about the environment. What the agent knows is the discriminatory power of this information in distinguishing between possible states of the environment. The agent's knowledge defines what states of the environment are indistinguishable from each other. The more the agent knows, the smaller these indistinguishable equivalence classes.

In this framework reasoning can be shown to be the concentration of information from a set of locations to a smaller set of locations, and also to be the making explicit of the information contained implicitly in the larger set (Rosenschein and Kaelbling, 1986). Agent states can thus be thought of as being ordered in terms of the amount of

explicit information they contain. The greater the explicit information content, the more precisely the state of the world is known.

The usefulness of knowledge for an agent is, of course, in guiding action towards a goal. In process dynamics terms the agent needs knowledge in order to tell what trajectory to follow. In areas of the state space where trajectories diverge, there is sensitive dependence on the current state, so even small errors in determining a trajectory may mean missing the goal later. In areas where the flow converges, little knowledge is needed: the agent can't help taking the right action.

6 Brief Notes on Other Issues

This section presents brief notes on some conjectures relating agent theory to dynamics. These notes are intended as a basis for discussion about these issues with others both in the areas of dynamics, AI and cognitive science. The notes will be revised and expanded as a result of discussions.

6.1 Learning

Learning consists in shaping the phase-space of the system. For example, goal-directed and aversive behaviour are produced by introducing attractors and repellers in the phase-space. Shaping of the phase-space is produced by changing the computational mapping between environmental inputs, internal state, and agent outputs. The behavioural trajectories of an agent are produced by the iterative computation of this mapping using the current environmental inputs and current internal state. The attractors and repellers can be both implicitly and explicitly represented in the structures that define this computational mapping. It is reasonable to assume that low level learning of the kind produced by conditioning is implicitly represented, while learning procedures from linguistic descriptions is explicitly represented.

6.2 Linking Cognition to Action at Fixed Points

Concepts of dynamics are also applicable to some of the internal processes taking place inside agents. In this section I briefly discuss two proposals: (a) that the process of constructing increasingly abstract representations has fixed points, and (b) that for maximal generality the two processes of cognition and action are best linked to each other at fixed points of the cognition process.

One role of the cognitive and perceptual processes is to obtain representations of the environment. I regard representations as approximations, because they contain less information than the objects they represent. Such approximations are more economical than the original objects, because they abstract away from irrelevant detail. It is often pointed out in the literature of psychology and AI that perceptual processing proceeds through a number of levels in terms of the abstractness or generality of the

representations used. I assume that the iterative progression through such levels of mapping into more general representations has limits which are fixed points. These fixed points are irreducibly abstract concepts. I want to emphasize, that I do not mean irreducibility to properties. If reduction to properties is possible, the properties are themselves represented at the most abstract level. My most abstract conception of a dog has still properties like legs, but the legs are the most abstract kind that I have available.

Action patterns are produced by using a mapping to produce patterns on the "surface" of the agent starting from representations inside. This surface is the set of actuators or transducers of the agent. In terms of directionality, actions are an "inside-out process" in terms of causation from agent to environment.

Cognitive and action patterns are best linked together at increasing levels of condensed representation in order to achieve economy and generality. The most economic way of doing this is to place the linkage at fixed points. The progression from cognition to action (and back to cognition again via the environmental feedback loop) is via the focus of fixed points where the representation from cognition and the representation for action are at the same level: when looked upon from the cognitive point of view, it is an appropriately abstract representation of a situation *as it is* ; when looked upon from the action point of view, the representation is of a situation *as it is to be* as the result of the action, when appropriately expanded. The result of condensation through cognition is a fixed point. This fixed point is then used as the starting point (or parameter value) for expansion into an action pattern. The usefulness of the fixed-point representation is in its economy both in size, and also in its abstractness: many similar situations are conceptually classified together for the purposes of identical or similar action. Generalisation over similar situations is thus obtained, producing economy.

6.3 Problem Solving

Problem solving can be interpreted as a trajectory towards stable equilibria, as has been discussed in relation to goals earlier in this paper. Alternative paths to an attractor are disjunctive solutions, while points on a single path are conjunctive subgoals.

The problem-solving metaphor can possibly be extended to logical inference, as is done in intuitionistic logic. It may be natural to interpret the existence of a proof as convergence to a maximal informational state.

Axioms could be interpreted as fixed points since they require no further proof process.

Reasoning is seen as finding a path from the point corresponding to a formula to a fixed point corresponding to an axiom. The path is the proof process, consisting of a sequence of transformations of the formula according to

derivation rules. The derivation rules correspond to reasoning actions of an agent.

Alternatively, formulas to be proved can be regarded as problems to be solved, with constituent subformulas as subgoals. The informational state of the agent increases as more and more constituents are proved. Such an interpretation has been used by Kripke (1965) to provide a semantics for intuitionistic logic. It will be of interest to analyse the dynamics of changes in informational states along the lines suggested in this paper.

6.4 Process Dynamics as an Implementation Strategy

The theoretical framework proposed represents agents as processes (or bundles of processes) and constraints acting on these processes, producing the appropriate dynamics. This suggests that the implementation should be in the form of machines recursively composed of parts. Machines can generally be described as bundles of processes, with the machine structure acting as constraints on the processes.

Rosenschein and Kaelbling (1986) has offered a detailed formalisation of this concept by modelling machines as a pair of processes (input and output), subject to behavioural constraints. The constraints impose a structuring on the state-space of the output process that can occur in the machine, producing what we call a dynamics for that process. Having a dynamics *means* being constrained, so that not all possible state trajectories are possible, and those that are possible must take a shape that satisfies the constrains.

Convenient high-level languages need to be developed for expressing the constraint system in order to facilitate the design of artificial agents. Rosenschein and Kaelbling (1986, 1989) have developed one example of such a language, REX, and its extension to the declarative specification of goals, GAPPS. This language has been used in robotics by Rosenschein and by the author in developing agents as intelligent interfaces to computer software.

7 Conclusions

This paper has taken the first steps in providing an interpretation of agent theoretic concepts in terms of process dynamics concepts. If successful, this approach could have the double advantage that it provides both a mathematical theory which is already known to be successful in other application domains like physics and biology, and it also gives a hint at implementation strategy.

8 References

Abraham, R.H. and Shaw, C.D. (1981) Dynamics, the Geometry of Behavior, Parts 1-4. Santa Cruz, CA: Aerial Press.

- Cvitanovic, P. (ed) (1984) *Universality in Chaos*. Bristol: Adam Hilger.
- Devaney, R.L. (1986) *An Introduction to Chaotic Dynamical Systems*. Menlo Park, CA: Benjamin/Cummings.
- Georgeff, M. and Lansky, A. (eds) (1986) *Proceedings of the Conference on Reasoning about Actions and Plans*. Los Altos: Kaufmann.
- Kiss, G.R. (1988) *Some aspects of agent theory*. Project Report HLD/WP/OU/GRK/24.
- Kiss, G.R. and Brayshaw, M. (1989) *Agent architecture and implementation strategy*. Project Report HLD/WP/OU/GRK,MB/25.
- Kripke, S. (1965) *Semantical analysis of intuitionistic logic I*. In: J.N. Crossley and M.A.E. Dummett (eds) *Formal Systems and Recursive Functions*.
- Rosenschein, S. (1985) *Formal theories of knowledge in AI and robotics*. *New Generation Computing*, 3, 345-357. Also SRI International Technical Note 362.
- Rosenschein, S. and Kaelbling, L. P. (1986) *The synthesis of digital machines with provable epistemic properties*. In: Halpern, J. (ed) *Theoretical aspects of reasoning about knowledge*. Proc. of the 1986 Conference. Los Altos, CA: Morgan Kaufman.
- Rosenschein, S. and Kaelbling, L. (1989) *Integrating planning and reactive control*. In: Proc. NASA/JPL Space Telerobotics Conference, Pasadena, CA. January, 1989.
- Rosenschein, S. (1989) *Synthesizing information-tracking automata from environment descriptions*. In: Proc. Toronto Conference on Knowledge Representation, 1989.
- Thompson, J.M.T. and Stewart, H.B. (1986) *Nonlinear Dynamics and Chaos*. Chichester: Wiley.