Fractal (Reconstructive Analogue) Memory

David J. Stucki and Jordan B. Pollack
Laboratory for Artificial Intelligence Research
Department of Computer and Information Science
The Ohio State University
Columbus, OH 43210
stucki@cis.ohio-state.edu
pollack@cis.ohio-state.edu

Abstract

This paper[1] proposes a new approach to mental imagery that has the potential for resolving an old debate. We show that the methods by which fractals emerge from dynamical systems provide a natural computational framework for the relationship between the "deep" representations of long-term visual memory and the "surface" representations of the visual array, a distinction which was proposed by (Kosslyn, 1980). The concept of an iterated function system (IFS) as a highly compressed representation for a complex topological set of points in a metric space (Barnsley, 1988) is embedded in a connectionist model for mental imagery tasks. Two advantages of this approach over previous models are the capability for topological transformations of the images, and the continuity of the deep representations with respect to the surface representations.

The Imagery Debate

The phenomena of mental imagery is widely disputed among cognitive scientists primarily because it occupies a position in the boundary between perception and cognition. On the one hand, mental images seem to be purely symbolic structural descriptions that are independent of any perceptual mechanisms. In this way images are no different from any other knowledge structures and therefore require no special purpose mechanisms, but can be reasoned about and operated on in the traditional propositional fashion. On the other hand, mental images seem to be represented by a special-purpose cognitive architecture that shares components with the visual perceptual system. Under this approach, additional mechanisms must be proposed for inspecting, transforming, and reasoning about images, providing a means for translating between purely symbolic representations and the "visual buffer".

This latter view that mental imagery is performed in an analogue, "pictorial" medium began to regain acceptance in the early 1970s as a result of empirical studies which indicated both that mental imagery belongs to a different modality than language and that there are cognitive tasks in which mental imagery is brought into play when symbolic reasoning and explicit knowledge is insufficient for solving the problem.

As an example of the former, an experiment designed by Lee Brooks (Brooks, 1968), required subjects to imagine a block letter and report whether successive corners were at the extreme top or bottom of the letter. The experiment showed that visually oriented responses (i.e., pointing to the letters Y or N) took longer than verbal responses (saying 'yes' or 'no') implying that the visual response task was interfering with the imagery task. In a similar experiment (also Brooks, 1968), he asked the subjects to report whether successive words in a sentence were nouns. In this case, verbal responses were slower than visually oriented responses. Brooks' conclusion was that mental imagery is distinct from verbal processes, and shares processing resources with the visual perceptual system.

The most famous experiment illustrating the latter was performed by Roger Shepard and Jacqueline Metzler on the mental rotation (Shepard & Metzler, 1971). When presented with pairs of drawings of three dimensional shapes at differing orientations that were either identical or mirror images, the subjects were to report whether the objects had the same shape, independent of any difference in orientation. They found that the response times varied linearly with the difference of angular rotation of the objects, which implied that the subjects were performing a sort of "mental rotation" in order to solve the problem.

Although extensive contributions have been made by many researchers to the theory of analogue imagery (see Finke, 1989; Chandrasekaran & Narayanan, 1990; Tye, 1991), most of its essential qualities have been incorporated into a single framework described by Stephen Kosslyn (Kosslyn, 1981; Kosslyn, 1980). The primary notion in Kosslyn's theory is that the representations of mental images are quasi-pictorial, or "picture-like". This means that in some way the representation preserves some of the topological or spatial properties of the objects being represented, by embedding these relation-

[1]. This research has been partially supported by the Office of Naval Research grant N00014-92-J-1195.

To appear in the Proceedings of the 14th Annual Conference of the Cognitive Science Society, Bloomington, Indiana, July, 1992.

ships in the architectural and functional medium of the representational mechanism. Kosslyn essentially recognizes two structural components and two kinds of processes in any encoding system. The constructs, which he calls the *representation* and the *medium*, are data structures with the medium serving as a host to the representation. For example, when a circular queue is implemented in a computer program as an array, the array is the medium in which the queue is represented. It is also important to note that the word "representation" is being used to denote both the entire encoding system and the data structure for a specific object, the usage being determined from context. The processes which operate within the encoding systems are either for making comparisons between representations, or parts of representations, or for transforming them in various ways (including the generation of new representations in place of old ones).

As in the distinction between sound and meaning in language, there is a distinction between image and meaning in mental imagery. Indeed, Kosslyn borrowed the linguists' terms *surface* and *deep* representations to capture this dichotomy. Surface representations are the analogue, pictorial images which we attribute to the "mind's eye". Deep representations are long term (symbolic) memories that can be used to "display" images on the *visual buffer*, which serves as the viewing screen for the mind's eye. The visual buffer is analogous to the memory a computer uses to store a bitmap for its display monitor. This memory serves as a *functional* coordinate space, since while the mapping from the bits to the pixels must preserve the coordinates—which must also be respected by any processes accessing the structure—there is no constraint that the individual bits be physically contiguous.

Although the nature of the surface representations has been specified in great detail in Kosslyn's model, the details of the deep representations have not been as well developed. The theory suggests two types of deep representation for visual images, called literal and propositional. The literal representations are intended to consist of information about what an object looked like, without any reference to coordinate spaces, but Kosslyn has been unable to formulate an appropriate representation and medium that is relevant to the theory:

> "We have not as yet made any strong claims about the precise format of the underlying literal encodings." (Kosslyn, 1981)

The propositional encodings are simply assertions used to describe the properties and features of an object, which presumably can be manipulated by mechanized logic. These representations are governed by syntax driven rules for interpretation and manipulation that are independent of the semantics of their values. The interpretation of a representation is based on the truth-value assigned to it under these rules.

Zenon Pylyshyn has been the most vocal opponent of the analogue approach to mental imagery (Pylyshyn, 1981; Pylyshyn, 1973). However, rather than suggesting an alternative, new theory, Pylyshyn questions the necessity of abandoning the traditional theory of cognitive processing as a physical symbol system of functional architectures:

> "In my view, however, the central theoretical question in this controversy is whether the explanation of certain imagery phenomena requires that we postulate special types of processes or mechanisms, such as ones commonly referred to by the term *analogue*.… whether certain aspects of cognition, generally (though not exclusively) associated with imagery, ought to be viewed as governed by tacit knowledge…or whether they should be viewed as intrinsic properties of certain representational media or of certain mechanisms that are not alterable in nomologically arbitrary ways by tacit knowledge." (Pylyshyn, 1981)

In order to substantiate his feeling that the answer to the first question is 'no,' Pylyshyn attacks the analogue position on two fronts.

The first attack consists of several specific criticisms of the Kosslyn model. He claims that a theory can not serve as a principled or constrained account of mental imagery if it is not substantive (explanatory), or if it is *ad hoc*, or if it has too many degrees of freedom (free parameters). He asserts that the analogue theory fails on all three counts. He does not claim that the analogue position is wrong in principle, but simply that none of the theories advanced so far have satisfied the conditions necessary for a principled account. The crux of the failure of Kosslyn's model, as he sees it is expressed in the following quote from (Pylyshyn, 1981):

> "Cognitive principles such as those invoked by [Kosslyn] would only be theoretically substantive (i.e., explanatory) if they specified (a) how it was possible to have formal operations that had the desired semantic generalization as their consequence—that is, how one could arrange a formal representation and operations upon it so that small steps in the formal representations corresponded to small steps in the represented domain—and (b) why these particular operations, rather than some other ones that could also accomplish the task, should be used…"

The second front on which Pylyshyn attempts to undermine analogue imagery systems is parsimony. Even the phrasing of the question quoted at the beginning of this section betrays his belief that as long as a propositional attitude can account for all of the empirical evidence on mental imagery, that to advance a theory requiring specialized mechanisms violates Occam's Razor. After all, Pylyshyn might say, if physical symbol systems have succeeded in explaining this much of cognition already, then the simplest thing would be if they could do the whole job.

To provide an example of the error that Pylyshyn perceives in the analogue viewpoint, he briefly discusses the scanning experiments Kosslyn performed with his colleagues in the 1970's (see Kosslyn, 1980). The results of these experiments showed that the further away from the current point of focus in an image a target object was, the longer it took to refocus on the target. Pylyshyn admits that this is clear evidence that inter-object distances are represented in mental images. However, he balks at the conclusion that this implies that the images have spatial extent. He argues here for a distinction between "having" and "representing" dimension or size. Once this distinction is recognized, he insists, the argument of the previous paragraph becomes perfectly natural.

The Dynamical Systems Road to Parsimony

The discovery within the last thirty years that non-linear dynamical systems are capable of exhibiting deterministic but unpredictable behavior and of generating fascinating images has sent a reverberating ripple through the physical sciences that caught the attention of world (Gleick, 1987). As tools have begun to emerge over the last three decades for analyzing, controlling and understanding non-linear systems, the taboo against these non-linearities has diminished, opening the doors for a broader class of the sciences to assimilate dynamical modelling into their theories.

As constituents of the physical universe, it is obvious that brains are subject to physical law and the passage of time, but it was not clear until very recently that there was anything to be gained by viewing cognition, and its various elements, from the physical perspective. For example, the earliest use of dynamical systems as explanatory tools for cognitive functions came from neuroscientific research, such as the work of Walter Freeman on EEGs of the olfactory bulb (Freeman, 1979; Skarda & Freeman, 1987). In the last decade dynamical systems have been applied to coordinated behavior (Kelso & Scholz, 1985; Jordan, 1986), decision processes (Usher & Zakay, 1990), language acquisition (Pollack, 1991; van Gert, 1991) and several other aspects of cognitive and perceptual processing.

The earliest inspirations for our current model can be traced back to a research plan presented in (Pollack, 1989) in which it was proposed that it is within the intersection of AI, Neural Networks, Fractal Geometry and Chaotic Dynamical Systems that various conundrums for cognitive science will be resolved. In this work the relationship between fractals and memories was proposed:

> "Consider something like the Mandelbrot set as the basis for a reconstructive memory. Rather than storing all pictures, one merely has to store the 'pointer' to a picture, and, with the help of a simple function and large computer, the picture can be retrieved.…"

A reconstructive memory based upon fractals will require a solution to the "fractal inversion" problem: given a picture within the generative range of some dynamical system, determine the precise parameters that would cause the dynamical system to generate it. Although a very hard problem in general, a mathematician claims to have solved it using the techniques of "Iterated Function Systems" (IFSs) (Barnsley, 1988)[2]. In this approach, the 'pointer' referred to in the above quote would be a single point in a multi-dimensional space of IFS parameters. Although IFSs have primarily received attention for their compact representation of visually complex two dimensional sets (fractals), Barnsley's results can be extended to more classical Euclidean sets, or even to three dimensions. This framework provides a strong mathematical and parsimonious foundation for our contribution to the imagery debate.

Fractal Memory (FRAME)

We have been developing a prototype reconstructive memory system based on fractals, called FRAME. Our encoding system for images is derived from Barnsley's work on IFSs and is commensurate with Kosslyn's dichotomy of deep and surface levels of description. The deep representation of an image is a small set of contractive affine transforms (i.e., linear functions, each of which maps the domain to one of its subsets) over a metric space (e.g. the Euclidean plane). The surface representation is the *attractor* (i.e., fixed point) of the functional union of this set, and can be constructed from the trajectory of a single point in the metric space through random selection and application of the transforms. We have shown that our sequential cascaded networks (SCNs), which are mathematically similar to IFSs, will exhibit state trajectories with complex, fractal properties when randomly stimulated, indicating that a simple neural model can instantiate the mathematical theory of iterated functions (Pollack, 1991; for another approach, see Stark, 1991).

While the images from IFSs are generally thought of as the result of a random infinite sequential process, this iterative process is extremely amenable to massive parallellization, producing rapid visual image reconstruction and even animation from the deep codes. This follows from the fact that the image is the fixed-point attractor of the IFS over the whole space. In other words, since every point in the space follows a trajectory under the IFS that approaches the attractor at an exponential rate, a large number of processors running the same IFS (with different random sequences and initial conditions) in parallel will produce almost instantaneously a surface representation of the image that captures its gross structure. Of course, finer detail will emerge over time.

2. This claim remains unverified, since his solution is being treated as proprietary by his corporation, which is selling digital image data compression systems.

IFS	*a*	*b*	*c*	*d*	*e*	*f*	*p*
Sierpinski's Gasket	0.5	0.0	0.0	0.5	0.125	0.125	0.333
	0.5	0.0	0.0	0.5	0.25	0.375	0.333
	0.5	0.0	0.0	0.5	0.375	0.125	0.333
A Dragon	0.5	-0.4	0.5	0.5	0.429	0.143	0.5
	0.5	-0.4	0.5	0.5	0.5	-0.214	0.5
Conch Spiral	-0.24	-0.0825	0.125	-0.25	0.428	0.868	0.1
	0.925	-0.225	0.266	0.925	0.15	-0.11	0.9

Figure 1: *Table of coefficients for the affine transformations of three IFSs approximated by FRAME networks*

As mentioned above, one of the most difficult problems is finding the deep representation for a given image. In FRAME, *the fractal inversion problem is refashioned as a neural network learning problem.* While we have not completely solved it, and acknowledge the need for large amounts of domain knowledge, we believe that the emergent complexity of simple nonlinear dynamical systems will provide a computationally feasible solution. The following model is an initial confirmation of this hypothesis.

As a first step towards this goal we used a network with the SCN architecture, which corresponds to the IFS structure. This network models an IFS by tracing the orbit of a point in the unit square under a probabilistically weighted sequence of transformations that are selected by the inputs to the net. The outputs of the network are recurrently connected back to the inputs, to simulate the iterative nature of the IFS. The network is trained in a supervised environment, in which the "teacher" knows the IFS for the image which is being learned. Since the teacher provides target outputs during training, the recurrent connections are only used during performance of the network.

A training set was randomly generated by forming triplets consisting of a point lying on the fractal attractor, an index, and the image of the point under the indexed transformation of the IFS for that attractor. The three sets on which our training sets are based are typical fractal images. These images, while lacking the geometric simplicity of more traditional experimental stimuli, nevertheless mirror the complexity of the stimuli found in nature. The coefficients for the three IFSs we trained on are displayed in the above table, in which each row represents a single affine transformation of the form

$$f(x, y) = (ax + by + e, cx + dy + f) ,$$

and the value of p is the probabilistic weight of the transformation in the reconstruction algorithm. The transformation indices were represented as 1-in-N encodings presented as input to the network, while the output and state of the network represented x-y coordinate values in the unit square. Since the network's task during training is to induce the invariant mathematical relationships in the training set, the training set needs to be large enough to eliminate any bias. On the other hand, a large training set imposes too many simultaneous constraints when using epoch learning. To balance these considerations, the network was only trained on a small subset of the entire training set during each epoch, similar to the independent method of (Cottrell & Tsung, 1991). In order to solve the bias problem, a new subset was randomly chosen at increasing intervals. The network was trained 100 epochs beyond convergence (Pollack, 1991), with the

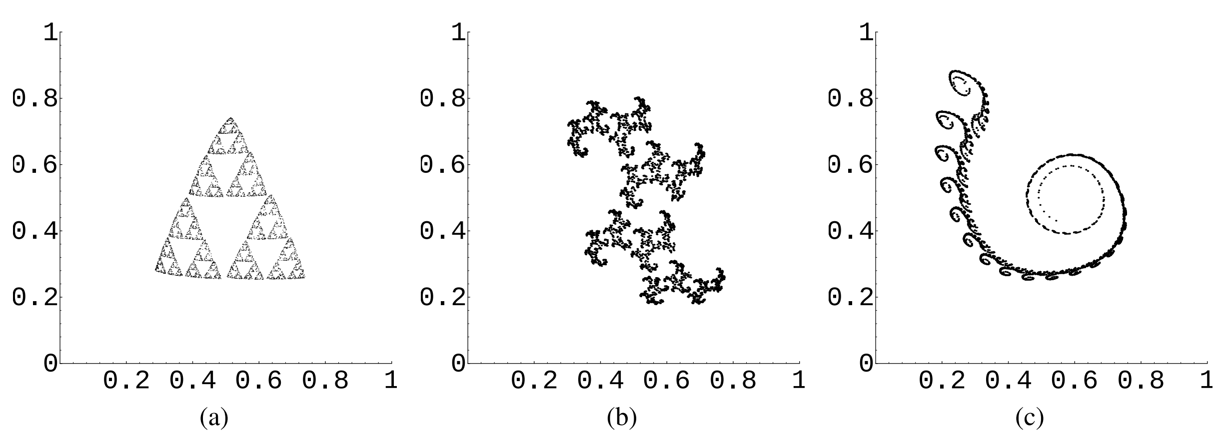

Figure 2: *Reconstructed images generated from FRAME networks that learned the "deep representations" of (a) Sierpinski's Gasket, (b) a Fractal Dragon, and (c) a Conch Spiral.*

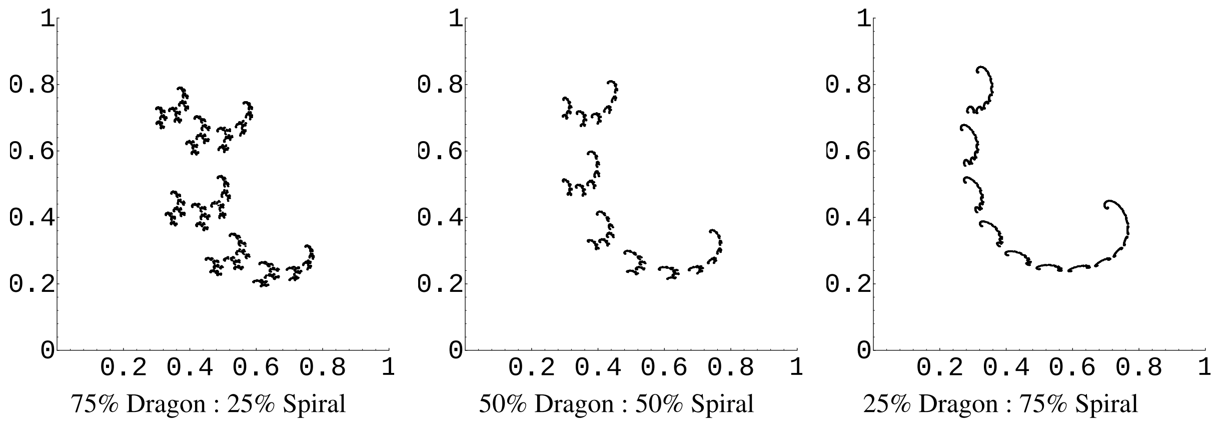

Figure 3: *Reconstructed images from three intermediate stages in the smooth interpolation of the deep network representations for the Fractal Dragon and the Conch Spiral.*

error threshold typically set to a value in the range [0.01, 0.05] and the overtraining usually distributed among several training sets, due to the cycling algorithm. To retrieve the image, the network is started with some random initial point, whose trajectory is plotted for 5000 iterations. The first 50 points in the trajectory (transients) are dropped, allowing it to approach the attractor. The performances of the networks trained on the IFSs from the table are pictured on the previous page. These results show the representational efficacy of this architecture and learnability of the set of transformations of an IFS.

One of the exciting aspects of our fractal representation for images is that *the surface representations vary continuously with the deep representations*. In other words, IFS codes meet Pylyshyn's challenge of small changes in one representation effecting small changes in the other. Furthermore, it varies in a predictable fashion. The affine nature of the transformations of an IFS allow us to decompose a transform into its primitive components: translation, rotation, and scaling. Thus the equation $f(x, y) = (ax + by + e, cx + dy + f)$ becomes

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (r\cos\theta)\, x + (-s\sin\varphi)\, y + e \\ (r\sin\theta)\, x + (s\cos\varphi)\, y + f \end{pmatrix}$$

where r and s are scaling factors, e and f are translational parameters, and θ and φ are angular rotation. Therefore, any of these transformations of the image can be accomplished by simply manipulating the appropriate parameters in the IFS.

A surprising capability of our model is that *one image can be continuously deformed into another image* merely by linearly interpolating between their deep codes. This technique can be implemented in our cascaded back propagation model by adding an additional cascaded layer to the network that chooses which code to use. The deep representations for the images are loaded into the upper network from long term memory as parallel slices of its weight matrix. The inputs to this network then select a linear combination of these representations as the weights of the middle network, producing an interpolated image (trajectory) on the lower net. Refer to (Pollack, 1987) for architectural details. This is illustrated by the "snapshots" of an interpolation between the spiral and the dragon shown above. By increasing the number of snapshots—by reducing the grainsize of a discrete interpolation—it is possible to produce an animation of the deformation process.

Conclusions

As part of a general theory of reconstructive memory based on fractal inversion, we demonstrate a neural network that is capable of approximating the mathematical theory of iterated function systems. This model addresses the imagery debate in that the weights of the network serve as the deep representation of an image, which is reconstructed using fractal analogue techniques. We believe this substantially addresses the foundational needs of theories of mental imagery, such as Kosslyn's. This model also circumvents Pylyshyn's substantivity criticism of Kosslyn's work. The mathematics in which our system is embedded guarantees that the surface representation of an image will vary continuously with the parameters in the deep representation. In other words, small changes in one will correspond to small changes in the other. This leads to the simple ability to rotate, zoom, and translate mental images by operating on the compact codes. It also leads to a prediction for a cognitive ability to smoothly deform one image to another, which is clearly a task of the "imagination," and which has not been accounted for by any other theories that we are aware of to date.

There are two areas which are incomplete in our fractal memory model. We have begun to solve the problem of inducing deep representations from surface images by constraining the learning task to a supervised environment. However the general problem, which is much more difficult, will be the focus of further research. A second area of further work is in the design of a memory system, containing many deep codes, that serves as a mechanism for the recognition task. If the first problem is solved, recognition of visual images can be done rather quickly using nearest-neighbor techniques.

The complexity attributed to objects in the world is at best dependent on the modelling tools and interpretations that are available. Fractal geometry's first lesson was that the apparent complexity of nature (e.g., the shape of a tree or coastline, the branching structure of rivers and lungs) simply reflected an unsuitable mathematical formalism. Perhaps its second lesson will speak more to the imagination.

Acknowledgments

The authors would like to thank Peter Angeline, John Kolen, and Greg Saunders for criticizing earlier drafts of this paper. Special thanks also goes to John Kolen for many long discussions and suggestions throughout the course of this research.

References

Barnsley, M. (1988). *Fractals Everywhere.* Academic Press.

Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22, 349-368.

Chandrasekaran, B. & Narayanan, H. (1990). Towards a theory of commonsense visual reasoning. In K. V. Nori and C. E. Veni Madhavan, (Eds.), *Foundations of Software Technology and Theoretical Computer Science: Proceedings of the Tenth Conference, Lecture Notes in Computer Science, 472.* Springer-Verlag.

Cottrell, G. W. & Tsung, F. S. (1991). Learning simple arithmetic procedures. In John Barnden and Jordan Pollack, (Eds.), *High-level Connectionist Models*, Norwood, N. J.: Ablex.

Finke, R. A. (1989). *Principles of Mental Imagery.* The MIT Press.

Freeman, W. J. (1979). Nonlinear dynamics of paleocortex manifested in the olfactory EEG. *Biological Cybernetics, 35*, 21-34.

Gleick, J. (1987). *Chaos: Making a New Science.* Viking.

Jordan, M. I. (1986). Serial order: A parallel distributed processing approach. ICS Report, La Jolla, CA.

Kelso, J. A. S. & Scholz, J. P. (1985). Cooperative Phenomena in Biological Motion. In Haken, (Ed.), *Complex Systems: Operational Principles in Neurobiology, Physical Systems, and Computers*. Springer-Verlag.

Kosslyn, S. M. (1980). *Image and Mind.* Harvard University Press.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: a theory. *Psychological Review, 88*, 46-66.

Pollack, J. B. (1987). Cascaded back propagation on dynamic connectionist networks. In *Proceedings of the Ninth Conference of the Cognitive Science Society.* Seattle, 391-404.

Pollack, J. B. (1989). Implications of Recursive Distributed Representations. In David S. Touretzky, (Ed.), *Advances in Neural Information Processing Systems*. Los Gatos, California: Morgan Kaufmann.

Pollack, J. B. (1991). The induction of dynamical recognizers. *Machine Learning, 7*, 227-252.

Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*, 1-24.

Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16-45.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.

Skarda, C. A. & Freeman, J. (1990). How brains make chaos in order to make sense of the world. In James A Anderson, Andras Pellionisz & Edward Rosenfeld, (Eds.), *Neurocomputing 2*. The MIT Press.

Stark, J. (1991). Iterated function systems as neural networks. *Neural Networks*, 4, 679-690.

Tye, M. (1991). *The Imagery Debate.* The MIT Press.

Usher, M. & Zakay, D. (1990). Multiattribute Decision Processes by Neural Networks. In *Advances in Neural Information Processing Systems 3*. Morgan Kaufmann.

van Gert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*, 1,3-53.