Landmark Learning: An Illustration of Associative Learning

Andrew G. Barto
Richard S. Sutton
Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01003

Technical Report #81-12

Acknowledgement

This research was supported by the Air Force Office of Scientific Research and the Air Force Avionics Laboratory through Contract No. F33615-77-C-1191 and Contract No. F33615-80-C-1088.

Abstract

In a previous paper we defined the associative search problem and presented a system capable of solving it under certain conditions. In this paper we interpret a spatial learning problem as an associative search task and describe the behavior of an adaptive network capable of solving it. This example shows how naturally the associative search problem can arise and permits the search, association, and generalization properties of the adaptive network to be clearly illustrated.

Landmark Learning: An Illustration of Associative Search

Andrew G. Barto
Richard S. Sutton

In a previous paper (Barto, Sutton, and Brouwer, 1981) we defined the associative search problem and presented a system, called an Associative Search Network (ASN), capable of solving it under certain conditions. An ASN incorporates learning rules that have been carefully designed following Klopf's hypothesis that neurons are goal-seeking systems (Klopf, 1972, 1979, and 1980). Here we present a simple spatial learning problem as an example of the associative search task. The ASN controls locomotion in a spatial environment containing various types of "olfactory" gradients. This interpretation illustrates the task in an intuitively clear form, shows how naturally it can arise, and allows the capabilities of a simple ASN to be clearly described. It was not our intention to either model animal spatial learning behavior or to fully exploit the capabilities of an ASN; rather, we wanted to illustrate its capabilities in as simple a problem as we could construct.

Associative Search

Fig. 1 shows an ASN interacting with an environment E. At each time t, E provides the ASN with a vector $X(t) = (x_1(t),\ldots, x_n(t))$, where each $x_i(t)$ is a positive real number, together with a real valued payoff or reinforcement signal $z(t)$. The ASN produces an output pattern $Y(t) = (y_1(t),\ldots, y_m(t))$, where each $y_i(t) \in \{0, 1\}$. The ASN's action Y is received by E. Each input vector X(t) provides information to the ASN about the sensory situation at time t in which it acts. After performing an action, i.e., after producing an output pattern, the ASN receives (1 time step later) an evaluation from E of the appropriateness of that action for

the situation in which it was made. This evaluation is received by the ASN as the value of a payoff or reinforcement signal z. The evaluation alone is not sufficient to determine whether the preceding action was the best possible in the given context. The associative search task is to learn, for each input vector, to perform the action which maximizes the payoff value. In other words, it must learn to perform the best action in each sensory situation. Different actions can be optimal in different sensory situations. This class of problems is more completely described in Barto et al., 1981, where it is distinguished from the simpler pattern recognition tasks that can be solved by perceptron-like learning rules.

Spatial Learning as Associative Search

If an ASN is viewed as controlling the locomotory behavior of an organism in a spatial environment, then input vectors are associated with places in space and ASN output patterns control movement. We have created a simple spatial environment in which to illustrate this interpretation of the associative search problem and a simple ASN's behavior. Fig. 2 shows a spatial environment consisting of a central landmark (shown as a tree) surrounded by four other landmarks (shown as disks). Thinking of this as an olfactory environment for a simple organism, we let each landmark possess a distinctive "odor" which can be sensed at a distance. Accordingly, to each landmark is associated a spatial distribution, linearly decreasing with distance from the landmark. The distributions extend as far as the large ellipses shown in Fig. 2. The asterisk shows the location of the ASN.

When the ASN is in a partiuclar location, its input pattern is determined by its distance from each of the landmarks. We let the central landmark act as an attractant for the ASN by letting its "odor" be the value of the payoff or reinforcement signal z. The other landmarks are "neutral" in that proximity

to them is not rewarding to the ASN. Input to the ASN therefore consists of five values giving the odor concentrations due to the central "tree" and the north, south, east and west neutral landmarks.

Fig. 3 shows an ASN with five input pathways, labelled vertically on the left according to the landmarks to which they respond. The shaded input pathway N indicates that the ASN is near the north neutral landmark. There are four output pathways labelled horizontally at the bottom as controlling "actions". The manner in which these actions determine locomotion was chosen solely for the sake of simplicity. There is an output element for each compass direction. Each output element produces an output of 0 or 1 at each time step. For example, if N=0, S=1, E=1, and W=0 (as shown by the shaded output elements in Fig. 3), the ASN will move a fixed distance south and east. We use a kind of "reciprocal inhibition" between the north and south elements and between the east and west elements so that at each time step usually only one of each pair of elements outputs a 1. Clearly, we are not attempting to model in any detailed manner the motor control system of an organism (for example, there is no explicit spatial orientation of the ASN).

The arrangement of input and output pathways used in Fig. 3 permits the connection weights to be displayed in convenient form as circles centered on the intersections of input pathways and the vertical output element "dendrites". Positive weights are shown as hollow circles, and negative weights are shown as solid circles. The sizes of the circles indicate the relative magnitudes of the corresponding weights. The uppermost "tree" input is the specialized payoff pathway z which has no associated weights. These connection weights form an associative matrix which is similar to those widely discussed in the literature (e.g., Anderson et al., 1977, Amari, 1977, and Kohonen, 1977) but which gathers information by means of the more complex

closed-loop learning rules to be described below.

The ASN's task in this environment is to 1) find the central landmark by climbing the attractant distribution and 2) associate with each place that action which causes movement toward the central landmark. The first part of this task is a simple hill-climbing problem which does not require long-term memory. The second part is an example of the associative search task. Although the payoff signal is derived from a single spatial distribution (the "odor" of the tree), the optimal action is clearly a function of the ASN's location. For example, if the ASN is south of the central landmark, it is best for it to move north; if it is north of the central landmark, it is best for it to move south. Consequently, the search for the optimal action in each place requires maximization of functions of ASN actions which differ from place to place. (A predictor as discussed in Barto et al., 1981, is not required for this spatial learning task since the functions to be maximized vary smoothly over time.) As a result of solving the second part of this problem, the ASN can proceed directly to the central landmark simply by performing the actions associated with its successive locations. Importantly, this direct approach is possible when the attractant distribution is very noisy, intermittant, or even totally absent (as we demonstrate below).

The Learning Rule

The ASN presented here uses the same type of learning rule as discussed in Barto et al., 1981. Let $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$ denote the signals at time t from the north, south, east, and west landmarks respectively, and let $z(t)$ denote the signal from the central landmark. Each output element j, $j = 1,\ldots, 4$, has a weight w_{ij} associated with neutral landmark input x_i, $i = 1,\ldots, 4$, and an additional weight w_{0j}. Let $w_{ij}(t)$, $i = 0,\ldots, 4$, denote the values of these weights at time t. Let

$$s_j(t) = w_{0j}(t) + \sum_{i=1}^{4} w_{ij}(t)x_i(t).$$

The output of element j at time t is

$$y_j(t) = \begin{cases} 1 \text{ if } s_j(t) + \text{NOISE}_j(t) > 0 \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

where each NOISE_j, $j = 1,\ldots, 4$, is a mean zero normally distributed random variable (with the same variance for each j).

At each time step, each weight w_{ij}, $i, j = 1,\ldots, 4$, is updated according to the following equation:

$$w_{ij}(t+1) = w_{ij}(t)+c[z(t)-z(t-1)]y(t-1)x_i(t-1). \quad (2)$$

The weights w_{0j} are updated as follows:

$$w_{0j}(t+1) = f[w_{0j}(t)+c_0(z(t)-z(t-1))y(t-1)] \quad (3)$$

where

$$f(x) = \begin{cases} \text{BOUND if } x > \text{BOUND} \\ 0 \text{ if } x < 0 \\ x \text{ otherwise} \end{cases}$$

bounds each w_{0j} to the interval [0, BOUND]. The parameters c and c_0 are positive real numbers determining rates of learning. In all of the simulations described below, c=0.25, c_0=0.5, BOUND=0.005, and the standard deviation of the random variable NOISE_j was 0.01 for $j = 1,\ldots, 4$.

Rule (2) implies that if the firing of an output element in a given place is followed by a movement toward higher attractant concentration z, then the element will become more likely to fire in that place in the future. If firing is followed by a movement toward lower values of z, firing will become less likely in that place. See Barto et al., 1981, for a more de-

tailed discussion of this class of learning rules.*

The weights w_{0j} changing according to (3) permit the ASN to climb the attractant distribution in the absence of landmark information. Rule (3) is similar to (2) applied to a constant signal from a universally present landmark ($x_0(t)=1$ for all t). If c_0 is sufficiently large compared to BOUND (as it was in our simulations), then complete learning will occur in a single trial so that a movement in an up-gradient direction will tend to be followed by a movement in the same direction. This straight line trajectory will tend to continue until it takes the ASN down-gradient. Down-gradient moves will drive w_{0j} to zero so that the random component will dominate. The bound function f is necessary to insure that down-gradient moves can return the weight to zero. The resulting hill-climbing strategy is similar to that used by certain types of bacteria to climb nutrient gradients (Koshland, 1979). Fraenkel and Gunn, 1961, call this strategy klino-kinesis and Selfridge, 1978, calls it "Run and Twiddle" (if things are improving, keep doing what you are doing; if things get worse, do something else).

Learning in a Noiseless Environment

If the attractant concentration can be reliably sensed, then the hill-climbing part of the ASN's task can be accomplished easily. Fig. 4 shows the ASN's trajectory for the case in which there are no neutral landmarks. The central landmark is approached due to the action of (3). Since no associations are formed in this case, that is, since no long-term memory traces are formed, later attempts to climb the same hill will proceed at essentially the same rate as the first attempt.

---

*The rule (2) is identical to that presented in Barto et al., 1981, except that the term $y(t-1)$ is used here instead of $y(t-1)-y(t-2)$. In the previous study, changes in z were attributable to changes in y. Here, y itself determines the change in z because it causes a change in spatial location rather than movement to a particular place.

Fig. 5 illustrates the ASN behavior in the presence of the neutral landmarks. Fig. 5A1 shows the ASN behavior for 35 time steps. Fig. 5A2 shows the state of the ASN as a result of this behavior. Non-zero weights have appeared associated with the north and east landmark input pathways since the ASN has remained in the vicinity of these landmarks (and hence only these pathways were eligible for modification). Since movements north and south were correlated respectively with decreases and increases in the attractant level, weights have formed so that the north and east landmark "odors" inhibit movement north and excite movement south. Weights associated with the east landmark pathway are smaller in magnitude than those for the north landmark since the ASN remained closer to the north landmark. Similarly, the north and east landmark inputs inhibit movement west. Weights for the east output element are too small to be visible since the ASN only infrequently moved east.

Fig. 5A3 shows the results of learning in a vivid form. A vector is shown at each point in a grid covering the entire space. Each vector is the result of computing the values s_j, $j = 1,\ldots, 4$, from the ASN input vector associated with the place at which the vector appears. The resulting 4-tuple is displayed as a vector in the obvious way. The direction of the vector at each location gives the direction of the ASN's most probable first step if it were to start at that location. The vector's magnitude is related to the probability that the ASN will take this step. It is important to note that the attractant distribution of the central landmark is not used to determine the vector fields. The vectors represent information stored in the ASN's memory - not information directly present in the environment. The vectors show how the ASN would tend to move even if the central landmark and its attractant distribution were not present. The generalization capability of

the ASN is clearly shown by the vectors associated with places never visited by the ASN.

Fig. 5B shows how the ASN behaves for about 800 time steps. It climbs the attractant distribution and remains in the vicinity of the central landmark (Fig. 5B1). The resultant associative matrix values (Fig. 5B2) show that the north landmark signal inhibits the north output element and excites the south output element. Consequently, when the ASN is in the vicinity of the north landmark, it will tend to move south. Similarly, a strong signal from the south landmark will cause the ASN to move north. The weights associated with the east and west landmarks similarly affect the east and west output elements. The resultant movement tendencies are shown as a vector field in Fig. 5B3. This form of learning is not dependent on the central location of the attracting landmark. Fig. 6 shows a vector field determined from the contents of the ASN's memory after about 800 time steps of learning with the attracting landmark located off-center. The importance of this illustration is that it shows that the learning rule is capable of not only determining the correct signs for the weights but also their correct magnitudes.

The information stored in the association matrix formed during exploration of this spatial environment can be used by the ASN to guide movement even in the absence of the attractant gradient. In Fig. 7 is shown the behavior of the ASN after learning by exploration of the environment with the attractant landmark in the center. The central landmark and its attractant distribution have been removed from the environment, and the ASN starts at a place it has never before visited. The ASN takes a direct route to the former location of the central landmark. This occurs because the context vector associated with each place "keys out" the appropriate action. The ASN remains near the central landmark's former location.

Re-learning in a Modified Environment

Here we illustrate how the ASN can reorganize its associative matrix due to changes in its environment. We allowed the ASN to learn in the original environment (Fig. 2) until it was able to associate the best movement with each place. We then interchanged the east and west landmarks. Fig. 8A shows the vector field resulting from evaluating the ASN's associative matrix in the altered environment. The central landmark location is now a saddle point rather than a stable focus. Starting from a central position, the ASN is "misled" by its sensory information and follows the vector field away from the central landmark (Fig. 8B1). Since this movement is down the attractant gradient, the ASN alters the weights to the east and west output elements from the east neutral landmark input (which now responds to the landmark to the west). This re-learning results in the network of Fig. 8B2 and the vector field of Fig. 8B3. A similar excursion to the east modifies the weights associated with the west neutral input which now responds to the landmark to the east (Fig. 8C). If the attractant distribution had been absent, no re-learning would have occurred.

Learning in a Noisy Environment

Climbing a hill as large and reliably sensed as the attractant distribution of the preceding illustrations is not a difficult task. When the attractant concentration can be sensed only in the presence of noise, the task becomes more difficult and more interesting. The sensitivity of the ASN to neutral context information permits it to improve its performance in climbing a noisy hill with repeated attempts.*

---

*Although we do not illustrate it here, we would expect that context information would also facilitate the more difficult problem of higher dimensional search.

Fig. 9A shows the ASN performance, starting with all weights zero, as it climbs the attractant concentration corrupted by additive noise. The noise is normally distributed with a standard deviation of 0.02. Comparing Fig. 9A with Fig. 4 or Fig. 5B1 shows that hill-climbing performance is significantly degraded. After sufficient experience with the noisy attractant concentration (1107 time steps), the ASN uses neutral landmark guidance to directly approach the goal even with the same noise level in the attractant concentration (Fig. 9B).

There are other means for improving hill-climbing performance in the presence of noise such as direct low-pass temporal filtering of the attractant signal as it is received by the ASN over time. We have not optimized hill-climbing behavior of the ASN in the absence of landmark guidance. Consequently, Fig. 9 does not compare landmark guided hill-climbing with the best hill-climbing behavior that can be accomplished without landmark guidance. What is important in this comparison, however, is that the association of neutral context information during a search permits the system to improve its performance with repeated attempts to approach a goal in the same or similar environments. Even the most highly tuned pure hill-climbing strategy does not learn from its experience in this manner. This example illustrates that the exploitation of neutral sensory information can provide significant adaptive advantages if the same or similar search problems occur repeatedly.

A Remark on Linearity

The associative search problem posed by the spatial environment of Fig. 2 is simple enough to be solvable by an ASN capable of making only linear associations. The influences of the neutral landmarks merely superimpose to form the desired control surface. If this were not the case, the ASN which we have described would not be able to form a stable mapping. Due to its linearity, it is not able to represent arbitrary patterns of location-action

associations; that is, only certain types of vector fields can be learned.

In our current research, we are investigating two methods for extending the ASN's capabilities to include nonlinear associations. The first relies on the observation that more varied associations can be formed as the number of landmarks increases. If, for example, there were a landmark at each spatial location, then a linear ASN could learn arbitrary location-action associations (this would be similar to the approach taken in the BOXES system of Michie and Chambers, 1968). This suggests that it would be useful for a system to effectively "create" landmarks where needed in order to refine its representation of space. Such a landmark, which we call a "virtual landmark", would be created by the formation of an appropriate nonlinear combination of the sensory signals provided by the real landmarks.

Another approach to nonlinearity is related to the "Patchwork Map" theory described by Kuipers, 1977. Here, the system's knowledge of space would consist of several different associative mappings appropriate for guiding locomotion in different regions of space. The system would need to develop nonlinear switching circuits for accessing the correct associative structure when entering each region. Both of these approaches to nonlinear learning are applicable to a wide range of spatial and non-spatial problems. We are finding that the simple spatial interpretation described in this article provides a concrete and generalizable framework for approaching these very difficult and general problems.

Conclusion

We have illustrated the behavior of an ASN in a simple spatial learning task. The spatial problem provides a vivid way to demonstrate the search, association, and generalization capabilities of an ASN. Although we have illustrated these capabilities in an extremely simple form, it should be realized

that the methods employed have much wider applicability. The spatial learning problem is an example of a wide class of problems, some of which require paths to be learned through spaces which do not necessarily represent physical space. For example, the space may be the state-space of a dynamical system in which case the vector fields developed represent hypothesized system dynamics. Associative learning capabilities provide a simple means whereby experience in attempting to solve a problem can be accumulated and used to drastically improve performance in similar problems. The necessity for explicit search is minimized by storing in long-term memory the information gained in previous searches.

Finally, we wish to comment on the simplicity of the ASN illustrated. It consists of just four adaptive elements acting in parallel. Since the adaptive elements themselves embody fairly sophisticated learning rules, utilizing both short-term and long-term memory, we did not need to construct a special purpose network to perform the landmark learning tasks which we have presented. The behavior illustrated is a very natural consequence of a set of elements operating according to a carefully designed closed-loop learning rule.

References

Amari, S.: Neural theory of associaiton and concept-formation. Biol. Cybernetics 27, 175-185 (1977)

Anderson, J.A., Silverstein, J.W., Ritz, S.A., Jones, R.S.: Distinctive features, categorical perception, and probability learning; some applications of a neural model. Psychological Rev. 85, 413-451 (1977)

Barto, A.G., Sutton, R.S., Brouwer, P.S.: Associative search network: a reinforcement learning associative memory. Biol. Cybernetics, to be published (1981)

Fraenkel, G.S., Gunn, D.L.: The orientation of animals: kineses, taxes and compass reactions. New York: Dover 1961

Klopf, A.H.: Brain function and adaptive systems - A heterostatic theory. Air Force Cambridge Research Laboratories research report AFCRL-72-0164, Bedford, MA. (1972)(AD742259). (A summary in: Proceedings of the International Conference on Systems, Man and Cybernetics, IEEE Systems, Man and Cybernetics Society, Dallas, Texas, 1974)

Klopf, A.H.: Goal-seeking systems from goal-seeking components: implications for AI. The Cognition and Brain Theory Newsletter, Vol. III, No. 2 (1979)

Klopf, A.H.: The hedonistic neuron: A theory of memory, learning and intelligence. Washington, D.C.: Hemisphere, to be published, 1980

Kohonen, T.: Associative memory: A system theoretic approach. Berlin: Springer-Verlag 1977

Koshland, D.E. Jr.: A model regulatory system: bacterial chemotaxis. Physiological Rev. 59, 811-862 (1979)

Kuipers, B.J.: Representing knowledge of large-scale space. M.I.T. Artificial Intelligence Laboratory report AI-TR-418. Cambridge, MA 1977

Michie, D., Chambers, R.A.: BOXES: an experiment in adaptive control. In: Machine Intelligence 2, pp. 137-152. Dale, E., Michie, D., eds. Edinburgh: Oliver and Boyd 1968

Selfridge, O.G.: Tracking and trailing: adaptation in movement strategies. Unpublished draft (1978).

Figure Captions

Figure 1. An ASN interacting with an environment E. The ASN receives input signals $x_1, \ldots, x_n$ and a payoff or reinforcement signal z from E and transmitts actions to E via the output signals $y_1, \ldots, y_m$.

Figure 2. A spatial environment consisting of a central landmark (shown as a tree) surrounded by four other landmarks (shown as disks). Each landmark possesses a distinctive "odor" which can be sensed at a distance. Odor distributions decrease linearly from their associated landmarks and become undetectable at ellipses. The asterisk shows the location of the ASN.

Figure 3. The ASN controlling locomotion in the spatial environment. The five input pathways are labelled vertically on the left according to the landmarks to which they respond. The shaded input pathway N indicates that the ASN is near the north neutral landmark. The four output pathways controlling actions are labelled horizontally at the bottom according to the direction of movement they cause. The shaded output elements indicate that a southeast movement is being made. The associative matrix weights are displayed as circles centered on the intersections of the horizontal input pathways and vertical output pathways. Positive weights are shown as hollow circles, and negative weights are shown as solid circles.

Figure 4. The ASN's path is shown as it climbs the attractant gradient in the absence of landmark guidance. No long-term memory traces are formed, and later attempts to climb the same gradient will proceed at essentially the same rate.

Figure 5. ASN behavior in the presence of neutral landmarks. A1) ASN behavior for 35 time steps. A2) The state of the ASN as a result of the experience shown in A1. The north and east landmark "odors" have come to inhibit movement north and excite movement south since, in the vicinity of the

north and east landmarks, movement north was correlated with attractant level decreases and movement south was correlated with attractant level increases. The north and east odors also enhance movement west since movement in this direction was also correlated with increases in attractant levels. A3) A vector field representation of the ASN state shown in A2. The direction of the vector at each location gives the direction of the ASN's most probable first step if it were to start at that location. These vectors represent the contents of the associative memory and thus show how the ASN would move even in the absence of the attractant distribution. A simple form of generalization is shown by the existence of vectors at places never visited by the ASN. B1) ASN behavior for about 800 time steps. It climbs the attractant gradient and remains in the vicinity of the central landmark. B2) The state of the ASN after about 800 time steps shows that proximity to the north landmark will make the ASN move south, proximity to the south landmark will make it move north, and similarly for the east and west landmarks. B3) A vector field representation of the ASN state shown in B2. Again, the vectors show how the ASN would tend to move even in the absence of the attractant gradient.

Figure 6. A vector field representation of the ASN's state after about 800 time steps in an environment with the attractant landmark located off-center. The learning rule is capable of determining the correct magnitudes for the weights in addition to the correct signs.

Figure 7. Use of long-term memory. With the ASN state as shown in Fig. 5B2 and the central landmark and its attractant gradient removed, the ASN takes a direct route to the central landmark's former position from a place it has never before visited. Stimulus patterns associated with successive positions "key-out" the appropriate actions.

Figure 8. Re-learning in a modified environment. After learning in the original environment (Fig. 2) we interchanged the east and west landmarks. Now the landmark to the west causes activity in the input pathway labelled east, and the landmark to the east causes activity in the input pathway labelled west. A) The vector field resulting from evaluating the ASN's state in the altered environment. The east and west landmarks now provide misleading information. B) Relearning from a western excursion. B1) Starting from the central position, the ASN is "misled" by its sensory information and goes away from the central landmark. Since this movement is down the attractant gradient, the ASN alters its weights and relearns as it climbs the attractant gradient back to the center. B2) The ASN state after the excursion west shown in B1. The influence of the input pathway from the east landmark has reversed so that proximity to the east landmark (now to the west) causes the ASN to move east rather than west. B3) The vector field representation of the ASN's memory contents after the excursion west shown in B1. C) Having experienced a western excursion and appropriately modifying its memory contents, the ASN is similarly misled by the information provided by the other re-located landmark. C1) The spatial path of an eastern excursion. C2) The ASN state after the eastern excursion. The influences from the input pathways have been reversed. C3) The vector field representation of the ASN's memory contents shows that the appropriate reorganization has taken place.

Figure 9. Learning in a noisy environment. A) ASN behavior, starting with all weights zero, as it climbs the attractant gradient corrupted by additive noise. Hill-climbing performance is significantly degraded (cf. Fig. 4 or Fig. 5B1). B) After sufficient experience with the noisy attractant gradient (1107 time steps), the ASN uses neutral landmark guidance to directly approach the goal even with the same noise level in the attractant gradient. Previous experience in the same or similar environments can be used to improve performance.

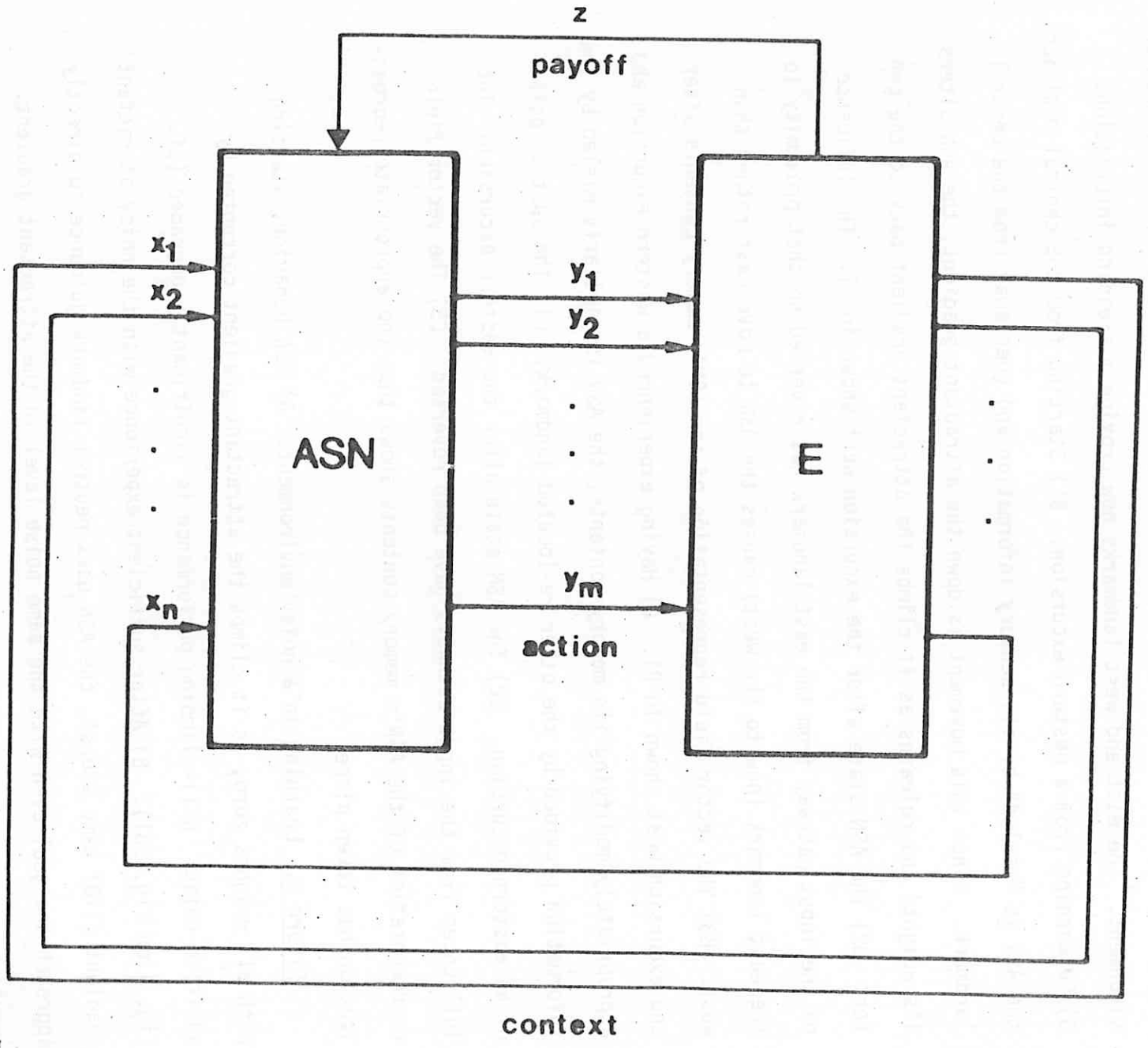

Fig. 1

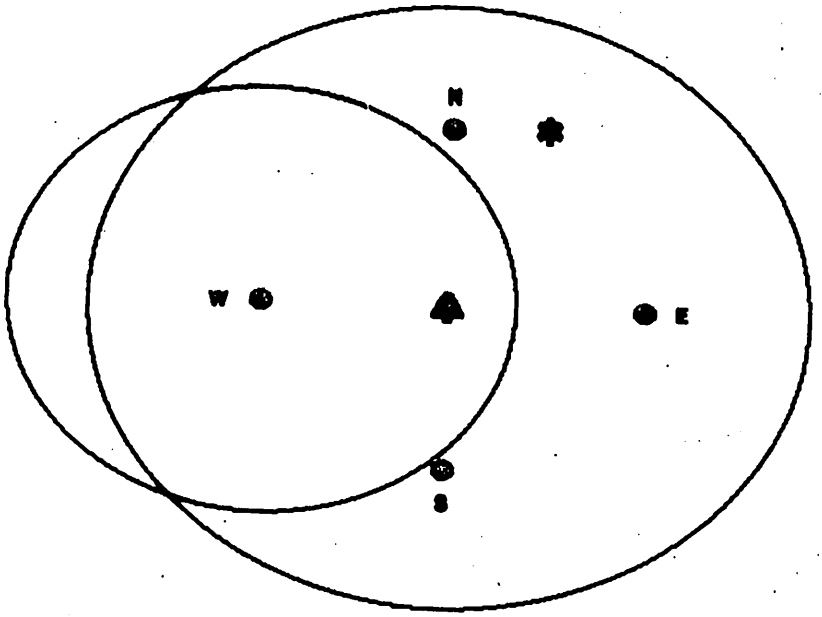

Fig. 2

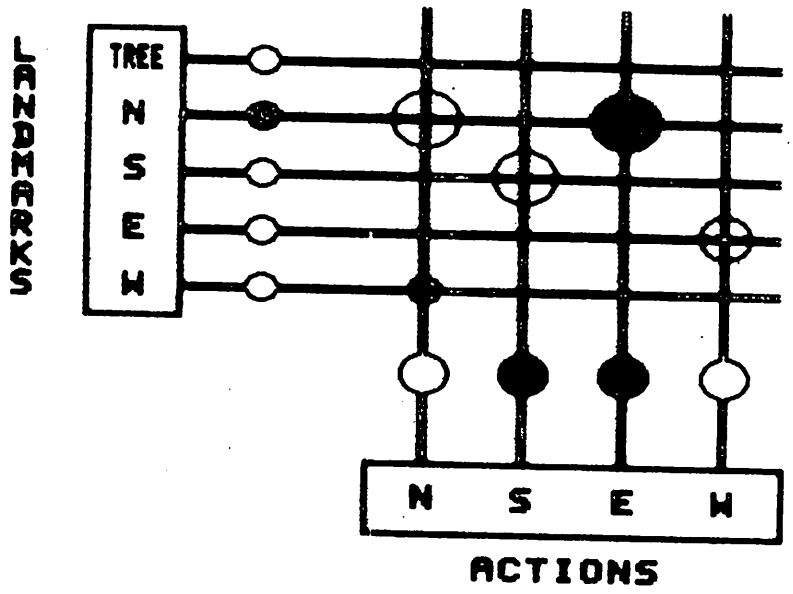

Fig. 3

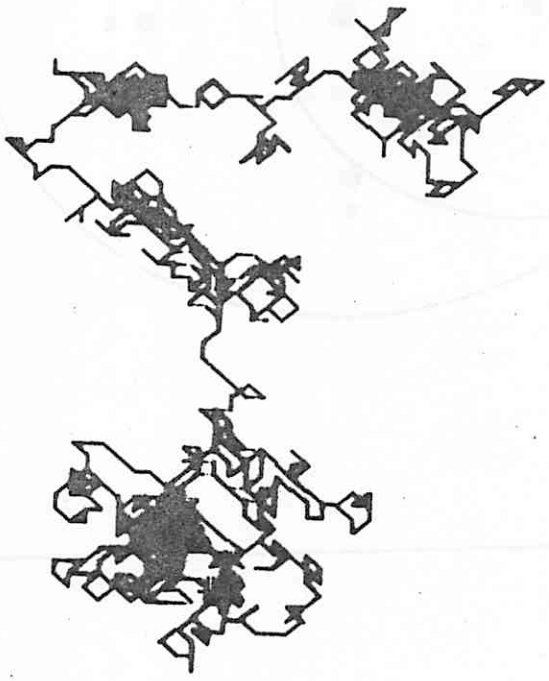

Fig. 4

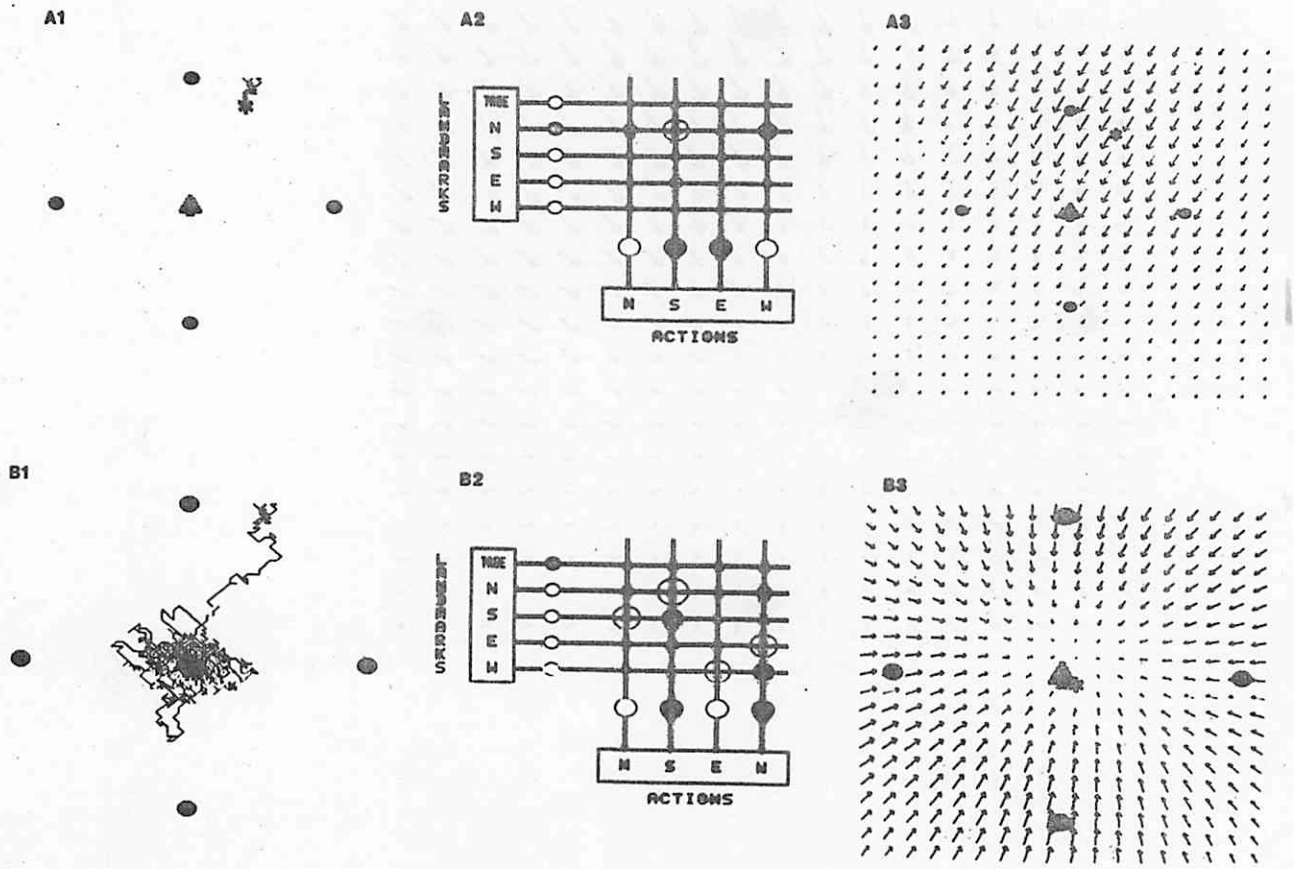
A1
A2
TABLE
N
S
E
W
N S E W
ACTIONS
A3
B1
B2
TABLE
N
S
E
W
N S E W
ACTIONS
B3

Fig. 5

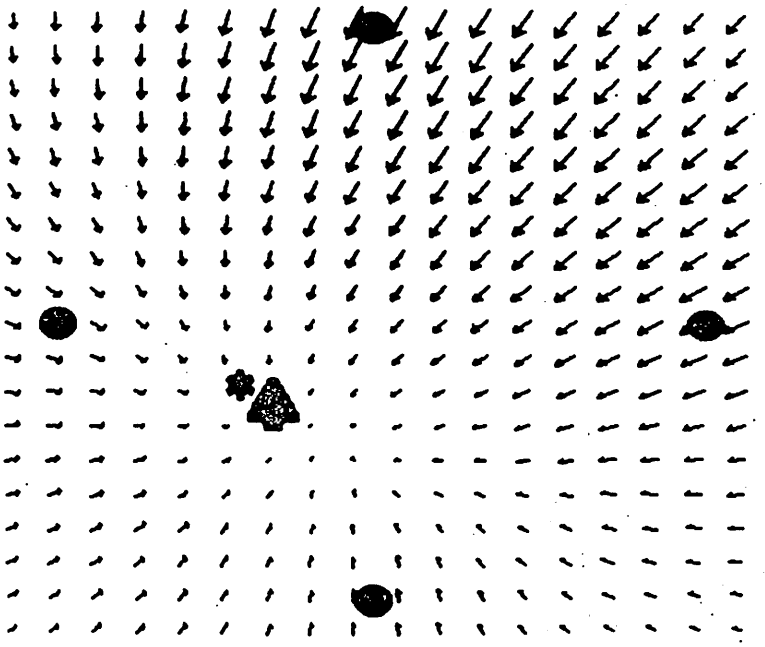

Fig. 6

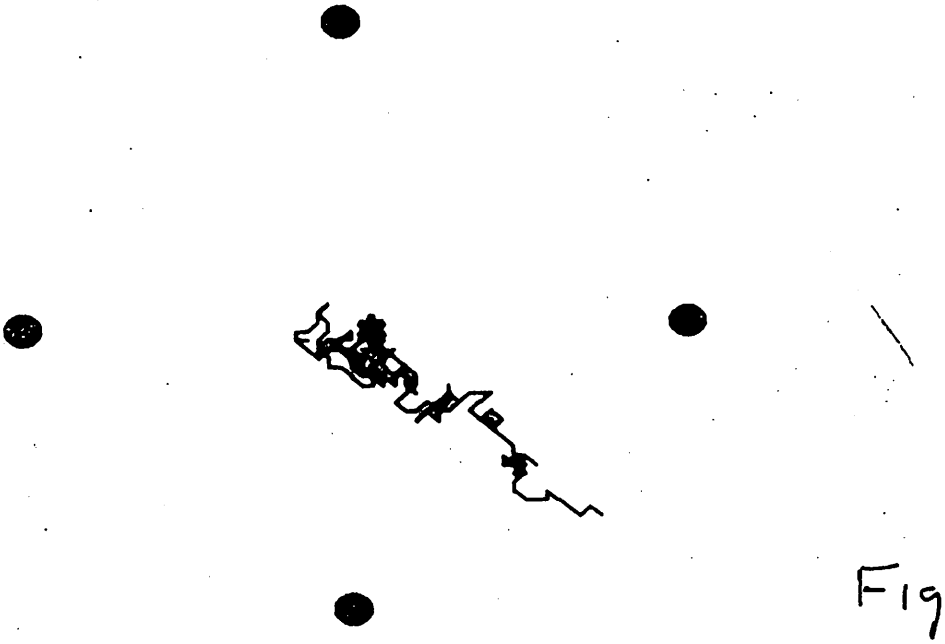

Fig 7

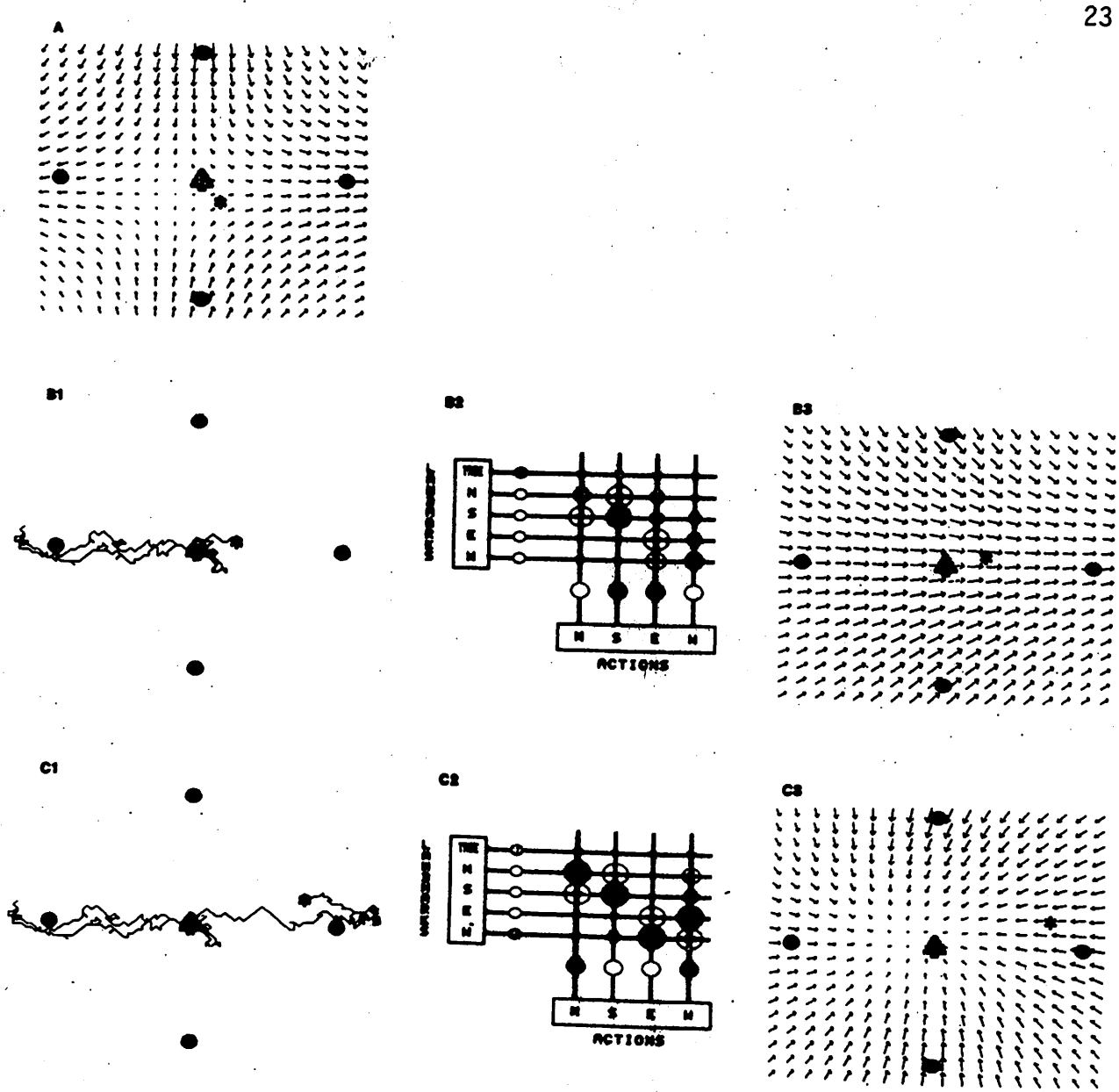

Fig. 8

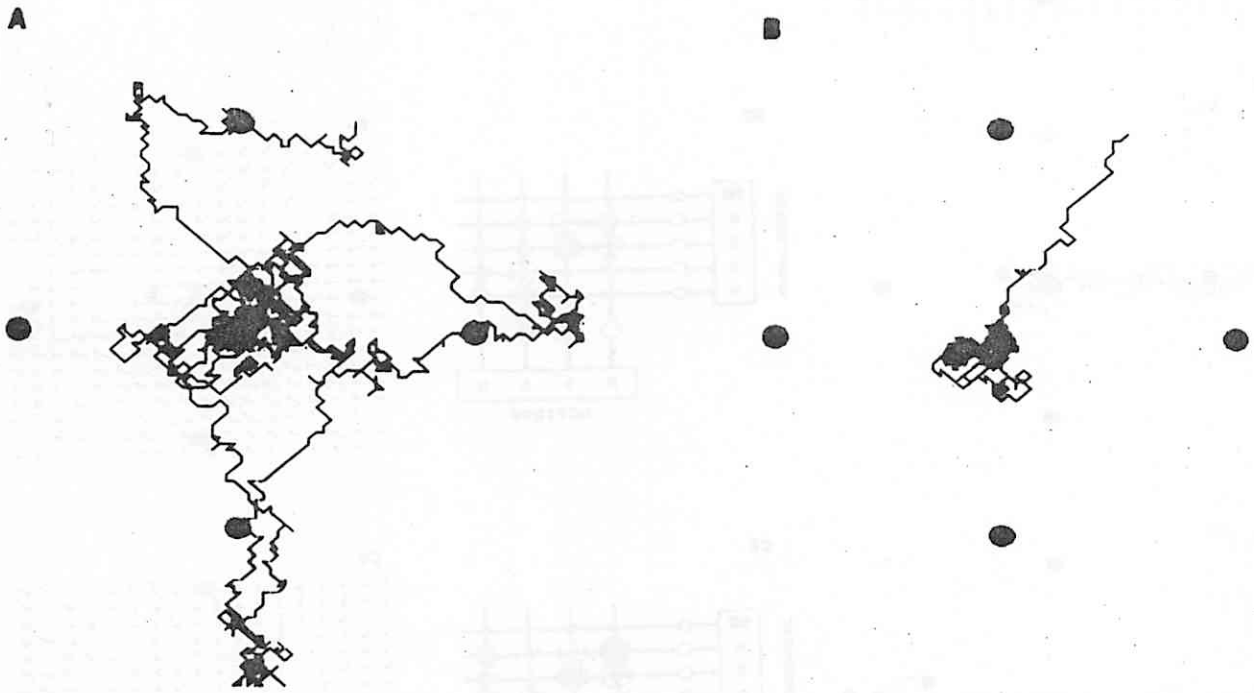

Fig. 9