

September 1970

AN INFORMATION-PROCESSING MODEL OF OPERANT BEHAVIOR

by

Nils J. Nilsson  
Peter E. Hart

Artificial Intelligence Group

Technical Note 38

## AN INFORMATION-PROCESSING MODEL OF OPERANT BEHAVIOR

by

Nils J. Nilsson

Peter E. Hart

In this note we shall sketch the outline of a simple information-processing model of operant behavior. This model can be viewed as an elaboration of one suggested originally by Deutsch (1960);\* the present exposition, however, uses terms and concepts of computer science. We stress that this model is just one of many alternatives; at various points we shall indicate how certain changes could be made in the model. More research must be done on this and related modeling schemes to accumulate the experience necessary to judge which best explains the empirical data. The first step must be to define one model precisely enough so that it could be implemented, say as a computer program.

### 1. The Major Elements of a Model

An information-processing model of behavior and its acquisition must contain precise definitions for the following elements:

---

\*References are listed alphabetically by author at the end of this note.

- (a) Inputs and Outputs. Any organism interfaces with its environment by means of an input (stimulus) channel and an output (response) channel. An attempt to model the behavior of an organism, therefore, must begin by defining the set of possible stimuli and the set of possible responses.
- (b) Stored Information. In all but the simplest animals, responses are determined not only by the stimulus but also by stored information. This information may consist of traces of previous events and of data "wired into" the animal. A major question for modeling concerns the structure and contents of the information-storage system.
- (c) The Response Computation. The major computation performed by the model is the computation of a response (or a set of simultaneous responses) that depends on the input stimulus and on stored information. Such a computation can be imagined as being performed continuously so that the output is a stream of responses in time.
- (d) Adaptive or Learning Mechanisms. A central goal of our research is to develop models for learning phenomena. By learning we mean here any changes that are made by the model to itself (that is, to the four elements of the model we have just named). Note that this definition includes the possibility of changes in the learning mechanism itself!

In the next few sections we shall describe one example of an information-processing model in terms of these four elements.

## 2. Input/Output

We shall assume that the sensory inputs (at any instant of time) can be represented by a basic set of "sensory predicates." We assume the existence of various sensory analyzing mechanisms that compute these sensory predicates. A representative partial list of example sensory predicates is given in Table I.

If the model animal is as shown in the room in Figure 1, his sensory-analyzing mechanisms at a given instant might, for example, be producing the following sensory predicates:

H(27),Th(13),Te(98.6),S(3),Fv(10 feet),  
 D(15 feet),Li(15 feet),B(-180°),Fo(31),  
 He.

Table I

BASIC SENSORY PREDICATES

A. Internal Stimuli

H(x)	Hunger level has value x.
Th(x)	Thirst level has value x.
Te(x)	Internal temperature is x.
S(x)	Sex drive level has value x.

B. External Stimuli

1. Visual

Fv(x)	Food is in view at range x.
W(x)	Water is in view at range x.
M(x)	Potential mate is in view at range x.
N(x)	Nest is in view at range x.
D(x)	Doorway is in view at range x.
Le(x)	Lever is in view at range x.
Sy(x)	Symbol of shape x is in view.
Li(x)	Light is now on and at range x.

2. Auditory

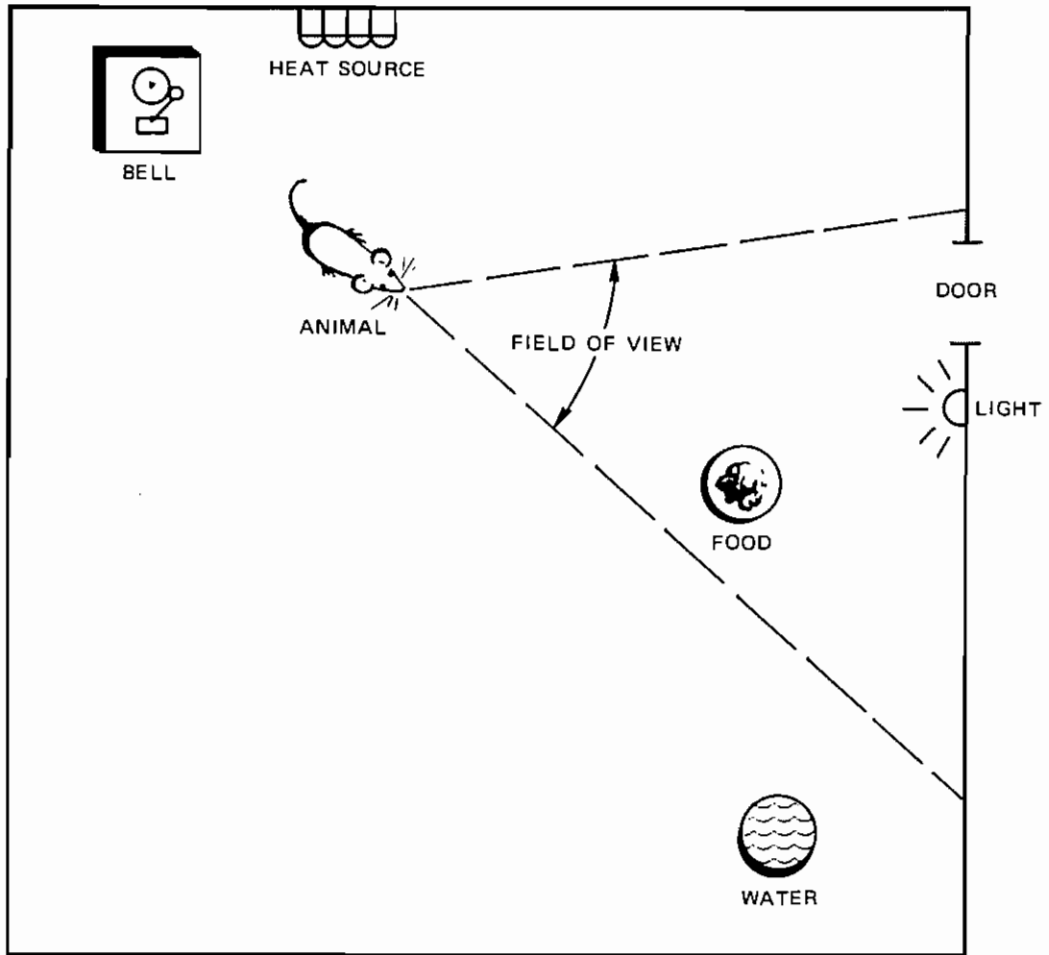
B(x)	Bell is sounding at angle x.
C(x)	Species member is crying at angle x.

3. Olfactory and Taste

Fo(x)	Food smell of intensity x.
Ft(x)	Food taste of intensity x.

4. Somesthetic

P	Pain present.
He	External heat present.
Co	External cold present.
Ta	Pleasing tactile sensation present (e.g., sex, grooming).



TA-710582-22

FIGURE 1 EXAMPLE ENVIRONMENT OF MODEL ANIMAL

Our present model will not be concerned with how these predicates are computed, but it may have provisions (attention mechanisms) for deciding which of them are computed. From the point of view of the rest of the model, the sensed world is known only through those predicates computed at any given instant.

Our responses will be constructed from a basic set of rather high-level motor actions. We must assume that the execution of these is accomplished and monitored by other mechanisms outside the model. The prime task of the model is to compute which subset of the responses is to be executed at any given instant. A representative partial list of basic responses is shown in Table II.

### 3. Stored Information

We are interested in explaining behavior that satisfies certain basic drives of an animal: for example, hunger, thirst, sex, avoidance of pain, etc. Therefore the model must store information describing what actions work toward satisfying these drives from the situations in which the animal ordinarily finds itself. As far as the use of this information is concerned, it is immaterial whether the information was originally built into the model or was obtained by some learning process.

Now a perfectly straightforward, but utterly impractical, method of storing this information is in a table that pairs every imaginable situation with the appropriate response. Response "computation" is then accomplished simply by a table look-up. Because the set of possible situations is the set of all the different outcomes of the sensory predicate computations, the table would be too large to enable such a storage system to be feasible. Fortunately such a table listing all of the different situations is unnecessary because typically it is only some key aspect or property of a situation that determines whether or not a given response is appropriate. For example, a rat may be trained to obtain food by pushing a lever whenever a bell sounds. Certainly all of the other environmental features that are simultaneously present (such as the presence of other rats) define the total situation, but the rat learns that these other features are unimportant in determining the response. Thus, instead of a table of situations, we are led to propose something like a table of important properties of the situations in which the animal may find itself.

These important properties can be expressed by predicates whose truth or falsehood at any instant is determined from the primitive sensory predicates true at that instant. Thus at the instant depicted in Figure 1, the predicate  $(\exists x)(\exists y)[Fv(x) \wedge Li(y)]^*$  can be determined to

---

\*The symbol  $(\exists x)$  means "there exists an x such that," and the symbol  $\wedge$  means "and." Thus our example predicate means "Food is in view (at some range) and a light is in view and on (at some range)."

Table II  
BASIC RESPONSES

A. Internal

Raise and lower blood pressure, heart rate, etc.

B. External

1. "Consummatory" responses: eating, drinking, copulating, sleeping, nest building, gnawing.
2. Anticipating and reflex responses: salivation, fear.
3. Locomotive responses: random exploration. Approach a named stimulus object or location. Withdraw from a named stimulus object or location.
4. Manipulative responses; push lever.
5. Other: rage behavior.

be true from the list of sensory predicates being computed by the sensory analysis systems at that time. We can also, for example, determine that at the same instant the predicate  $[He \wedge (\forall x) [\sim W(x)]]^*$  is true. In fact there may be a very large number of different predicates (some true at one instant and others not). Each predicate represents a different but important aspect of the situation in which the animal may find itself.

For the moment we shall ignore how these predicates come to be synthesized. Formation of knowledge about the important properties of situations is one of the aspects of learning that we propose to study.

Our storage system will represent each of the important properties of situations by a node in a directed graph. (A graph allows us also to store relationships between properties that a simple linear table does not.) Each property node will be labeled by the corresponding predicate that defines the property.

We shall be interested in storage networks containing perhaps hundreds of nodes and thus hundreds of predicates. At any given instant some subset of all of these predicates may be true according to the primitive sensory predicates. The nodes corresponding to these true predicates shall be called start nodes.

Thus at any instant the important properties of a situation (as perceived by the animal) are represented by the set of start nodes, each representing a different property. It may also be the case that two different situations share some of the same properties so that two different situations may have some of the same start nodes.

A subset of all of the nodes is associated with certain basic goals such as the satisfaction of hunger, avoidance of pain, achievement of safety, etc. These nodes shall be called the primary goal nodes. Another set of nodes represents those basic properties of situations that cause the animal pain. These shall be called the primary pain nodes. The general operation of the system will be to seek situations in which the predicates associated with primary goal nodes are true and to avoid situations in which the predicates associated with primary pain nodes are true.

In a directed graph, certain pairs of nodes are connected by directed arcs. Thus, if there is an arc directed from node  $i$  to node  $j$ , we say that node  $i$  is a parent of node  $j$  and that node  $j$  is a successor of node  $i$ . In our model, the arcs represent responses that the organism can make. Suppose that at node  $i$  we have predicate  $P_i$  and that at node  $j$  we have predicate  $P_j$ . Then if an arc  $a_{ij}$  is directed from node  $i$  to node  $j$ , it represents some action that if executed in a situation in which  $P_i$  is true will lead to a situation in which  $P_j$  is true. The

---

\*The symbol  $(\forall x)$  means "for all  $x$ ," and the symbol  $\sim$  means "not." Thus this example predicate means "External heat is present and water is not in view."



directed graph storage system with its nodes and arcs is then a representation of important properties of situations and how they are causally linked together by the various actions in the animal's response repertoire. A simple (if over-idealized) example directed graph is shown in Figure 2.

Since the nodes of the graph represent properties rather than complete situations, it may sometimes be the case that the action represented by an arc directed from one node to another may not reliably produce a situation in which the latter node is a start node. Information about the reliability with which actions result in given successor nodes can be stored as probabilities associated with each arc. That is, associated with arc  $a_{ij}$  (directed from node  $i$  to node  $j$ ) is the probability,  $P_{ij}$ , that the corresponding action will lead to node  $j$  starting with node  $i$ .

In addition to probabilities of actions, we might want to store information about the effort or cost of actions. For example, moving across the room requires more effort than turning one's head. The cost of an action can be stored as a number associated with the corresponding arc.

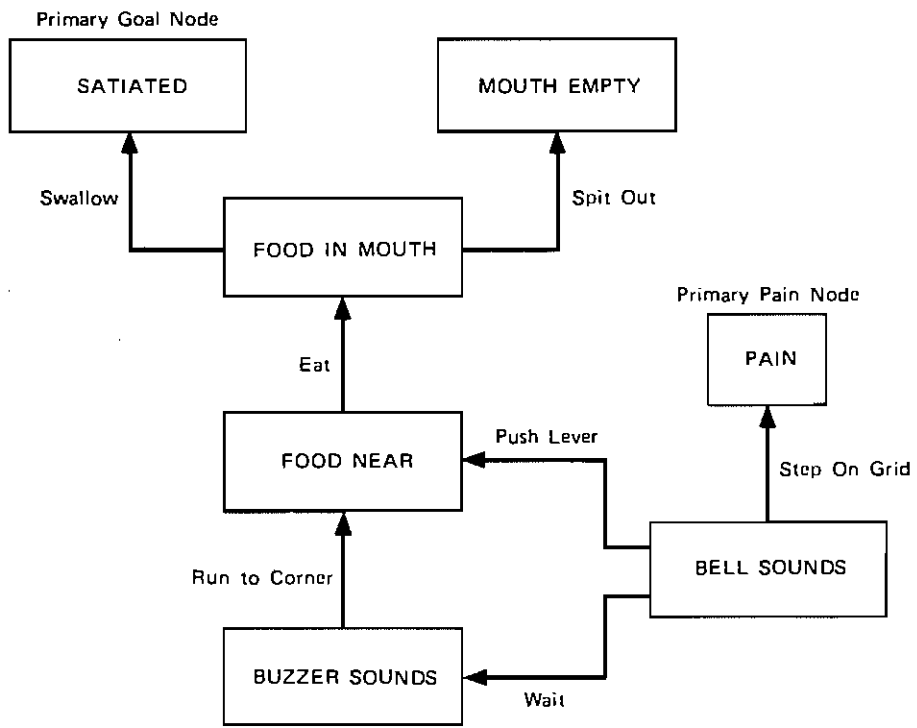
To summarize the form of our information-storage system, we have:

- (a) A set of primary goal nodes
- (b) A set of primary pain nodes
- (c) A set of nodes labeled by predicates. (The subset of these nodes having predicates that are true in light of the current sensory predicates are called start nodes.)
- (d) A set of arcs, labeled by actions, directed from one node to another node. (The arcs also have probabilities and costs associated with them.)

Having established this form of memory organization, we can now discuss various techniques for computing responses and for incorporating learning.

#### 4. The Response Computation

We shall now discuss the problem of how to compute a response (action) given the information-storage structure discussed above and a set of sensory predicates that describe (more or less) the present state of the world. We assume first that the model continuously computes the relative desirability (and abhorrence) of the primary goal (and pain) nodes. Thus at a particular instant for a hungry animal these might be given by the list: (need for food: 92; need for water: 80; need for sex: 17; need to avoid electric shock: -20; need for warmth: 42; etc.) Such a list assigns a number, called the motivation value, to each primary goal and pain node.



TA-710582-23

FIGURE 2 SIMPLE DIRECTED GRAPH FOR INFORMATION STORAGE

Loosely speaking, we want the operation of the model to be such that its responses lead to primary goal nodes with large positive motivation values and avoid primary pain nodes with large negative motivation values. Following this general idea there are several specific possibilities. A more precise discussion of some of these requires some definitions.

First we define a path of length k in a graph to be a sequence of nodes  $n_{i1}, n_{i2}, \dots, n_{ik}$  with each  $n_{ij}$  a successor of  $n_{i,j-1}$  for  $j = 2, \dots, k$ . The cost of a path is then the sum of the costs of all of the arcs on the path. The motivation value of a path is the sum of all of the motivation values of any primary goal or pain nodes on the path. For example, suppose  $n_1, n_2, n_3, n_4$  is a path from node  $n_1$  to node  $n_4$ , and suppose that node  $n_4$  is the only primary goal node on the path. Then if there are no primary pain nodes in the path, the motivation value of the path is just the motivation value of node  $n_4$ , and the cost of the path is  $C_{12} + C_{23} + C_{34}$ .

Next, we define the benefit/cost ratio of a path to be the motivation value of the path divided by the cost of the path. An optimal path between two nodes is then that path with the largest benefit/cost ratio. Let us denote by  $R(i,j)$  the benefit/cost ratio of an optimal path between node  $i$  and node  $j$ . [If no path exists at all between  $i$  and  $j$ , then  $R(i,j)$  is undefined.] Typically we shall be interested in finding that optimal path between any member of a set of start nodes and any member of a set of primary goal nodes that maximizes the value of  $R(i,j)$ , where  $i$  and  $j$  range over the start node set and goal node set, respectively.

If the arcs of our graph all had unit probabilities (that is, every action had a certain and known result), then we might want to execute that action corresponding to the first arc on the best optimal path between the set of start nodes and the set of primary goal nodes. The occurrence of nonunit probabilities on arcs complicates things somewhat in that we must then compute actions that tend to maximize average benefit/cost ratios. To simplify the following exposition, we shall assume that we have only unit (or zero) probabilities. The extension to the general case is straightforward, although it does present some computational difficulties.

Various other computational strategies come readily to mind. We will quickly summarize some of the main alternative ideas. One, based on Deutsch's model, might involve backward search from that primary goal node having the largest positive motivation. The backward search finds the least costly path to any start node, and the first action out of the start node on this path is excited. (Perhaps a similar backward search is conducted simultaneously from that primary pain node having the largest negative motivation. The first action out of the start node on the least costly path to a start node is inhibited.) Such a strategy leads naturally to separate excitatory and inhibitory systems, a separation supported by some biological evidence (Stein, 1964).

Another computational strategy involves the "propagation" of motivation values from the primary goal and pain nodes outward to the rest of the nodes in the system. The strength of the propagation is influenced by the probabilities and costs associated with the arcs. A set of responses can then be computed as follows: Any action linking any start node with a successor node having a large positive (induced) motivation value tends to be excited (perhaps in proportion to that motivation value). Similarly, any action linking any start node with a successor node having a large negative motivation value tends to be inhibited. When two conflicting actions are simultaneously excited (or when the same action is simultaneously excited and inhibited), we assume that some method for resolving the conflict exists. Otherwise, there is no reason why several responses cannot be simultaneously excited and others inhibited if such activity works toward primary goals and avoids primary pains.

We will need to explore and precisely define several of these possibilities before we will be able to comment more critically on their comparative suitability as elements of a model.

## 5. Learning Mechanisms

Previously we stated that possible learning mechanisms include any methods for changing any element of the model, including changing the learning mechanism itself. Here we shall discuss primarily those issues surrounding changes to the directed-graph storage system.

The obvious changes that can be made involve adding and deleting nodes and adding and deleting arcs. For example, the following simple technique for adding arcs is based on a suggestion by Deutsch: Construct arcs directed to each of the set of start nodes from each of the set of nodes that were start nodes at the immediately preceding instant. Label each of these arcs by the response just made. Obviously this strategy works toward building a model of the causal relationships inherent in the animal's world.

A somewhat more complex technique involves the modification of the probabilities associated with the arcs. If  $S_i$  is the set of start nodes at one time instant, and if  $S_{i+1}$  is the set of start nodes at the next time instant after making a response  $a$ , we increment the probabilities associated with the arcs labeled by the response  $a$  linking nodes in  $S_i$  with those in  $S_{i+1}$ . (If any of these arcs are absent, they must first be created. At creation, the arcs may initially have low probability values associated with them.) Similarly we decrement the probabilities of these arcs labeled by action  $a$  that are directed from nodes in  $S_i$  to nodes not in  $S_{i+1}$ . To accord with well-established psychological evidence, learning effects (perhaps the sizes of probability increments) might be made proportional to the absolute difference of the (induced) motivation values of the respective nodes in  $S_i$  and  $S_{i+1}$ . Thus, for example, an action that results in traversing from a node of low motivation value directly to one of high motivation value is

highly significant and is more likely to be learned than an action that does little to change the motivation value. The cost values associated with each arc can be adjusted in a similar way so that they reflect the actual effort involved in the various actions.

There are other types of changes that can be made to the directed-graph structure as a result of learning. First, when the same sequence of actions recurs several times, the sequence ought to be consolidated as a single macro-action. This macro-action can then be represented in the directed-graph structure by a single arc connecting the initial node to the terminal node. The synthesis of macro-actions allows the animal to master much longer chains of behavior.

Another change to the directed graph involves the splitting and agglomeration of nodes. A single node perhaps should be split into two or more nodes whenever the same action labels two or more different arcs directed outward from that node; for in this case, the same action taken supposedly in the same situation leads to two or more different consequent situations. Obviously in this case what we thought was just one situation was really two or more. When a single node is split, each must have its own predicate, and these must be proposed for each of the resulting nodes after splitting. (The disjunction of the separate predicates ought to be equivalent to the single predicate associated with the single node before splitting.) We have yet to work out precise ideas for mechanisms to propose new predicates, but obviously such mechanisms will play an important role in the model.

Similarly, certain nodes that repeatedly are start nodes simultaneously should perhaps be consolidated into a single node in order to conserve storage space and computation time.

There are other types of learning also involving changes to the sensory predicates and to the action computation mechanism, but these kinds of changes are subtler and perhaps ought to be studied later, after experience is gained with the more obvious learning strategies.

## 6. Miscellaneous Considerations

We desire that our model possess two additional features, but we have not yet worked out their details. First is some attention-focusing mechanism that directs the computation of sensory predicates. We have in mind mechanisms of the sort that will give special priority to computing whether or not food is present when the animal is hungry, etc. Perhaps decisions to compute or not to compute sensory predicates can be made one of the actions in the response repertoire of the system.

Another feature that our system must have is the ability to make a response in novel situations not covered by the directed-graph storage network (e.g., in situations in which the set of start nodes is empty). One simple-minded approach is to make some random response in this case, but we must still provide a learning mechanism for creating a new node to cover this new situation from which a random action was

taken. (It may have led to a start node with high motivation value!) When a new node is thus created, we must associate with it a predicate. The predicate certainly should be one composed out of the sensory predicates that were computed in the new situation, but in addition our model must have a means for hypothesizing the important property of that situation.

Although the strategy of generating random responses may be an adequate first suggestion, we would hope that research would indicate ways in which the extent of random variation could be lessened. One possibility is as follows: Suppose the system is in a situation in which the set of start nodes is empty. Instead of making a random response, we recompute the set of sensory predicates now with slightly weaker criteria. (Eventually, of course, this tactic will cause the system to hallucinate.) The result of these recomputations should be to expand the set of start nodes so that some action can be selected. An action computed in this way should often be more relevant to the novel situation than would an action selected completely randomly. If this action leads to a node with high motivation value, then again we must have a mechanism for creating a new node and its predicate representing an important property of the novel situation.

## 7. Summary of our Preliminary Work

To summarize, our work to date has identified several distinct although interacting model components, and has also suggested solutions to the problems of constructing these components. We have reviewed below the basic model components together with a simplified summary of how these components might be modeled.

### a. Model Input and Output

We have assumed that impinging stimuli can be represented by a set of basic sensory predicates. The sensory predicates indicate such things as whether a bell is ringing, whether food is in view, and so forth. Responses would be modeled by a rather high-level action such as eating, walking (either randomly or to a named object), and so forth.

### b. Information Storage

We have suggested that information about the effects of actions be stored as a directed graph. The nodes of the graph represent properties of the stimulus field; the arcs between nodes are labeled with actions that are likely to transform one stimulus situation (represented by the node at the tail of the arc) to a new stimulus situation (represented by the node at the head of the arc). The directed graph also contains information about "primary rewards and punishments."

### c. The Response Computation

The basic response computation has the form of a search of the above-mentioned graph. The purpose of the search is to

find a path (or paths) from the node(s) representing the initial stimulus situation to a node (or nodes) representing a primary reward. The path also must avoid nodes representing primary punishments. Once a path is found, the action labeling the first arc on the path is performed.

d. Learning Mechanisms

The effect of learning is to modify the graph storage mechanism. If experience shows that a certain action, when taken in a certain stimulus situation, leads to some particular new stimulus situation, this experience is reflected by the addition of new arcs and nodes to the graph.

## REFERENCES

Deutsch, J. A. (1960), The Structural Basis of Behavior, University of Chicago Press, Chicago, Illinois.

Stein, L. (1964), "Reciprocal Action of Reward and Punishment Mechanisms," in The Role of Pleasure in Behavior, R. Heath, ed., Harper and Row, New York, New York.