AN APPROACH TO ARTIFICIAL NONSYMBOLIC COGNITION

FRANCO P. PREPARATA
SYLVIAN R. RAY

UNIVERSITY OF ILLINOIS - URBANA, ILLINOIS
AN APPROACH TO ARTIFICIAL NONSYMBOLIC COGNITION

Franco P. Preparata and Sylvian R. Ray

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract No. DAAB 07-67-C-0199; and in part by the Office of Education under contract No. OE C-1-7-071213-4557.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

This document has been approved for public release and sale; its distribution is unlimited.
AN APPROACH TO ARTIFICIAL NONSYMBOLIC COGNITION*
Franco P. Preparata† and Sylvian R. Ray‡
The University of Illinois at Urbana

Abstract

In this paper we describe an approach to the artificial recognition of events of a nonsymbolic nature, such as the bidimensional perspective views of scenes of our everyday world. Scenes are presented as colored pictures and the objective of the cognitive task is the labeling of the interpreted scene objects. The method is based on three major components: i) a preprocessed version of the scene (stimulus), ii) a semantic map and iii) an algorithm which attempts interpretation of the stimulus under the guidance of the semantic map. The algorithm is sequential and proceeds from general to specific, thereby achieving efficient tree-pruning (contextual elimination). Stimulus interpretation is based on attribute-value matching, but classification relies strongly on the accumulated context. Backtracking provisions are available for correction of earlier wrong hypotheses. Experiments are presented and described. The major weakness of the approach is the present lack of a satisfactory theory of abductive inference. Flexibility, generalizability and efficiency appear to be valid merits.

*This work was supported by the Office of Education under contract No. OE C-1-7-071213-4557; additional support was provided by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract No. DAAB-07-67-C-0199.
†With the Coordinated Science Laboratory and the Department of Electrical Engineering; presently on leave at the University of Texas at Austin, Texas.
‡With the Department of Computer Science and the Coordinated Science Laboratory.
1. Introduction and Generalities

The investigation summarized in this paper falls in the general area of artificial intelligence, that is, in our interpretation, the synthesis of automatic systems capable of performing tasks usually considered to be specific of human beings, such as recognizing environments or understanding a written natural language text. These "cognitive" activities can be viewed as the linking of properties of the stimulus being received - be it written text or a direct input from the environment - to a representation of the properties of the past experience of the individual. This representation is what is commonly referred to as the semantic map or cognitive structure.

Although the existence of a semantic map is commonly postulated (see, e.g., [1], Ch. 4), no satisfactory experimental evidence or generally accepted theoretical hypothesis has been offered to characterize its structure. We must note, incidentally that our interest is directed more towards theoretical hypothesis than towards experimental evidence gained from observing living organisms. In fact, we feel that a theory of the organization of a semantic map should be explored for its own merits rather than for its potential similarity to features of the nervous system. For the same reasons, references to psychological intuitions are to be seen more as clues than as proofs to support some selected organizational or algorithmic features.

In order for understanding to occur, it is fair to postulate a universe susceptible of being known and understood. By a negative argument a spatio-temporal universe, such that the properties of its points can be specified at random, cannot
be described in any other way than by the detailed recording of those point properties. In such a universe, compressed descriptions and predictions are not feasible. It appears therefore that a basic prerequisite for learning and understanding is what, for lack of more appropriate words, we shall call a principle of contiguity and permanence (CP-property), that is, the possibility to identify as components of the physical universe ("objects") what appears to be in contiguous domains of space and whose properties are normally constant in time or vary with continuity. This constancy, coupled with the presence of constraints (coarsely identifiable with the physical laws) is conducive to the formation of abstracts, that is, common identifiers of all objects having very similar properties and behaviors. The abstract (or concept, or model) is an extremely rewarding device: not only does it permit summarizing a large variety of equivalent experiences, but it also allows behavioral prediction. Clearly, once the road to abstraction is open, several levels of increasing sophistication are possible.

We shall therefore assume that the universe upon which the automatic system will be called to operate has the property outlined in the previous paragraph, i.e., the CP-property. With this assumption, however, several choices are possible for this universe, ranging from a highly artificial construction to a somewhat simplified version of the world of our everyday experience. For reasons of mental economy the latter choice is more appealing, with the free bonus that evaluation of results is obtained intuitively rather than through some complicated procedure to be developed.

We now return to the discussion of the semantic map. The foregoing analysis indicates that abstract concepts are the basic constituents of the map. Understanding and recognition result from an interaction between the stimulus and
the map, whose formal details, in spite of intuitive illuminations, are still inscrutable.

As hinted before the objective of a cognitive process is to link suitably identified components of the stimulus to concepts in the map, so that the total linking describes a plausible event. We may therefore subdivide the process into two consecutive phases: a deterministic (manipulative) phase and a cognitive phase, which are, respectively, characterized by the absence or presence of interpretation, as we shall explain below.

The deterministic phase, in which stimuli are transformed in order to be more conveniently processed in the subsequent phase, relies upon the "local" properties of the stimulus. The result of this phase is termed the preprocessed stimulus.

The crucial step of the cognitive phase is of an interpretive nature. By interpretation we mean 1) the formulation of hypotheses upon the stimulus-map linkage (i.e., some links are tentatively established) suggested by the evidence provided by the preprocessed stimulus, followed by 2) experiments aimed at testing the hypotheses, until a 3) decision is made as the selection of the hypothesis scoring the highest confidence. The objective to produce the most reasonable interpretation of the stimulus (input) is reminiscent of the decoding problem in statistical communication theory. In both cases "equivalence classes" of inputs are mapped into "representative members": the outstanding difference is that in communication theory, equivalence classes are determined by the

---

1On a psychological basis it may be argued that the process results from the continuous interplay of the two phases: It appears possible, however, to keep them separate when one considers potential ways of automatic implementation.
statistical properties of noise, while in cognitive processes they result from the practical necessity to encompass with a single description a large number of equivalent inputs.

It is now convenient to treat separately, cognition of symbolic (verbal) and of nonsymbolic (sensorial) stimuli, because there appear to be substantial differences both of an organizational and of algorithmic nature.

A) In the symbolic or linguistic stimulus-map interaction, the stimuli are presented in a form which requires only minor preprocessing, for example, word segmentation. In fact, it is generally recognized that syntactic analysis cannot proceed unambiguously without semantic aid, (see, e.g. [2]), and for this reason, cannot be considered part of the deterministic phase. The difficulty, therefore, appears to be of a two-fold nature: 1) Since inputs may be presented as sentences, the capability to effect syntactic analysis must be available; 2) Words of a language can refer to a much larger domain than that of environmental experience, reaching, for example, the intellectual and emotional spheres. Therefore, the semantic map must incorporate relationships of an extremely subtle nature and this appears to be the most formidable difficulty (notable efforts in this direction are A. R. Quillian's "Teachable Language Comprehender" [3] and current work by R. F. Simmons\(^2\)).

B) In the nonsymbolic interaction the manipulative phase is considerably more complicated. Indeed, consistent with the CP-property, we identify near-uniform portions of the stimulus ("domains") and attempt to link them to map concepts, through the consensus of a carefully selected set of attributes. It

\(^2\)Private communication, April 1970.
is almost superfluous to note the strongly cognitive nature of the stimulus-map links initially established, since a considerable amount of guesswork is required at this preliminary stage. On the other hand, it is conceivable that the necessary semantic map need not be as complex as in the symbolic interaction, due to the simpler nature of the relationships among the concepts involved.

This expected simplification, along with the intent to avoid the apparently extraneous complications of syntactic analysis, guided our choice to focus on the nonsymbolic interaction. The data organization and the procedures described in the sequel are intended for implementation on an automatic processor, specifically - although not necessarily - a general purpose digital computer.

2. Structure for the Semantic Map

As discussed in the previous section, the semantic map should reflect the structure of the universe it is designed to interact with. For this reason, an economical objective suggests that the map complexity be the minimum required to accomplish the cognitive task. This is the rationale of the organization proposed below.

Introducing a further specialization of the choice of nonsymbolic stimuli, we shall consider inputs of a visual nature, such as landscape-like scenes of our everyday world (i.e., their bidimensional perspective view).

The primary constituents of a semantic map are the objects of the universe, and the principal organizing criterion is a binary relation between objects. Therefore the formal model of the map is a directed graph, whose nodes are objects and whose edges describe the relation. This relation - which defies a precise definition - can be thought of as expressing implication, and is
functionally reminiscent of the concept of "mutual information" as encountered in probabilistic information theory (see, e.g., Gallager [4]). Unfortunately, no probability assignment is discernible in our set of objects so that at present the analogy does not seem to be further extensible. On the other hand, it appears satisfactory to accept "implication" as a primitive concept from which probabilistic or logical systems can be constructed.

Two important features of the relation R should now be pointed out. The first is that links in the relation graph do not express mandatory constraints but rather potential implications which are allowed in the universe. This can be rendered explicit by assigning weights, valued in the interval [0,1], to edges of the graph, expressing the strength or plausibility of an implication. The second is the asymmetric nature of R. In view of the nature of the cognitive algorithm (section 4), it is convenient to render R completely asymmetric, i.e., antisymmetric, by using the hierarchical organization of the set of objects, in order to retain only those links of the graph directed from general to specific (such as part-whole relationships). This modified R (reflexive, antisymmetric and transitive) clearly induces a partial ordering, and is referred to as the "context for" relation.

In summary the semantic map is a set of objects \{A\} and a reflexive, antisymmetric, transitive relation R: \{A\}^2 \rightarrow [0,1]. The graph of R is loop-free, i.e., it is the Hasse diagram of a partly ordered set having the whole universe and the nonexisting object as universal bounds [5]. The notations \(A_j \in \mathfrak{S}^{-1}(A_i)\) express the existence of an edge from \(A_i\) to \(A_j\), whose weight, \(r_{ij}\),

\[(1)\]After the inception of the project reported here, an illuminating paper by S. Watanabe [6] eloquently expressed this interesting viewpoint.
is positive. We now present a simple illustrative example.

A fragment of the semantic map is given in Figure 1. Nodes and edges are labeled as previously described. $A_2$ and $A_3$ are details (in the sense of relation $R$) of $A_1$, as well as $A_4$ and $A_5$ are details of $A_3$; we notice that $A_4$ is also a detail of $A_2$. To provide some intuitive significance to the example, we could assign $A_1 = \text{sea}, A_2 = \text{sailboat}, A_3 = \text{ship}, A_4 = \text{hull}$ and so on.

A node of the map is actually recorded as a collection of attribute-value pairs for a judiciously selected set of attributes. This collection of parameters (or collections of parameters, if a certain object can appear in more than one clearly identifiable form) is associated with the symbolic code $A_j$ (the name) to be used for communicating the result of the cognitive process. If more than one collection of attribute values are associated with a given node, it is convenient to assign to each of them a $[0,1]$-valued confidence parameter $\gamma$ expressing the plausibility of the association. If only one collection is present, $\gamma$ is conventionally equal to 1.
The relevant attributes should be selected among the most powerful in discriminating the stimuli of a visual environment. It is convenient to separate the attributes which pertain to individual objects (local attributes) from those which pertain to relations among objects (relational attributes, providing "syntactic information"). It is also worth noting that usually attributes referring to the observation of a sequence of scenes, such as motion attributes, are of the latter type. Since, however, the analysis of sequences of scenes appears only quantitatively more complex than that of a single scene, for the sake of simplicity in this paper we shall deal exclusively with the latter case, and dispense hereafter with the consideration of temporal attributes.

With reference to the local attributes we feel that color, shape, size and orientation form an adequate set. The collections of attribute values assigned to each map node are intended as nominal and represent averages over a population of samples (in conformity with the nature of abstracts). The scale used for each attribute deserves some further comment. Colors are classified into several nuances, which are chosen more densely where it is felt that a more subtle discrimination is required. Shape is grossly provided as an aspect ratio of the smallest rectangle circumscribing the domain: if orientation is significant, it is given by the slope of the major side. Size is provided in conventionally chosen units.

The main function of relational attributes is to express physical constraints of the universe under consideration. The most outstanding such constraint in our universe appears to be "vertical ordering", to which we intentionally limit our selection, although several other constraints may be significant. Therefore between any two map nodes $A_i$ and $A_j$ we can imagine an edge whose $[0,1]$ valued
weight $v_{ij}$ expresses the plausibility that $A_i$ be above $A_j$. Clearly the relation expressed by $v_{ij}$ is reflexive, antisymmetric and transitive, i.e., it induces a partial ordering. The ensuing graph is therefore loop-free (called V-diagram).

We note, incidentally, that the V-diagram is expectedly very sparse since for the large majority of pairs the relation of vertical ordering does not apply.

In summary, the semantic map is organized as a primary graph (the R-diagram) whose nodes represent abstracts of objects and edges depict the binary relation "context for"; nodes are associated with collections of attribute-value pairs. Added to the R-diagram is the V-diagram, that is, the graph of the vertical ordering relation. The constraints of the universe are reflected by the confidence parameters associated both with edges and with the collections of attribute values (in principle, each node can be thought of as being connected to a much larger multitude of nodes. Zero confidence parameters make implausible relations disappear from the graph).

The detailed structure of the semantic map is impressed into the system at its inception, according to our best choice. Presently we shall not consider the possibility of "automatic learning" of the map. Although procedures and criteria for the incremental augmentation of the map appear feasible within the proposed framework, they will not be discussed in this paper and will be the subject of later reports.

3. The Structure of the Preprocessed Stimulus

In the light of the map organization proposed in the previous section, the objective of the cognitive phase can be rephrased as the activation of subsets of the nodes of the semantic map. By activation of a node we intend an affirmative
decision as to the recognition in the stimulus of the object represented by the node. It is clear from the nature of the relation R that the activated nodes must form a connected subgraph of the map. We shall say that the graph of the activated nodes interprets the scene.

As noted in the Introduction of this paper, near-uniform portions of the input scene are identified as "domains" to be cognitively linked to nodes of the bidimensional input scene (presumably rectangular) has been sub-divided by a uniform grid. The fineness of this grid is of considerable importance, since it determines the level of detail resolution, which directly affects the ability to perform the cognitive task. Each elementary square as determined by the grid is labeled according to its prevalent color. Domains are then formed (grown) by grouping together contiguous squares of similar color. Once a domain, $D_j$, is formed, a set of attribute-values is associated with it, i.e., color, shape, size, orientation and vertical position in the scene. The first four attributes have been discussed in Section 2: the only difference in this instance is that values are "measured" and not "nominal". The fifth attribute is self-explanatory and is the basis for the interaction with the V-diagram.

It is intuitively plausible, however, that the "local" information represented by the readily derivable attribute collections of scene domains is not completely adequate for recognition and must be supplemented by additional information of a syntactic nature. The following discussion is meant to give shape and substance to this essential but ill-defined requirement.

In order to facilitate the cognitive task it would be desirable that the product of the manipulative phase, i.e., the pre-processed stimulus, be strongly suggestive of the subgraph of the map yielding the most plausible inter-
To clarify this crucial point, we begin with a highly idealized simple example. Suppose that the scene of figure 2a be offered to the cognitive system. This scene is a mountain landscape with a house by a lake (The function of the symbolic designators of scene domains, \( D_j \), will be apparent soon). The pertinent section of the semantic map is illustrated in figure 2b, where the subgraph of interest, embedded in the entire map, is shown with solid lines (furthermore, node names have been added for explanatory purposes). Assume now that Scene domains \( D_1, \ldots, D_7 \) are partially ordered according to the relation of domain inclusion (\( D_0 \) being conventionally the entire scene). Then it is easily recognized that the resulting diagram is isomorphic to the map subgraph of figure 2b. In this case the cognitive task would reduce to a recall operation, i.e., a search in the map for a subgraph, isomorphic to the inclusion diagram of the input, once (input domain) - (map node) pairs have been identified on the basis of their attributes.

This simple example highlights two very important points, one positive and one negative. The positive point is that a readily derivable geometric relation

![Figure 2a - A simple scene](image)

![Figure 2b - The semantic map subgraph interpreting the scene of Figure 2a.](image)
among input domains is a powerful vehicle to the pertinent map subgraph. The negative point is that to expect isomorphism between the relevant map subgraph and some diagram describing a geometric relation of scene domains is certainly naive. To substantiate the quasi-obviousness of this statement, consider figure 2c. This scene is semantically equivalent to the scene of figure 2a, in the sense that they are both interpreted by the semantic map of figure 2b. The inclusion diagram of the scene (figure 2d), however, does not even resemble the map subgraph interpreting it.

Figures 2c,d,e. A scene analogous to scene 2a and the corresponding diagrams for inclusion and E.

Hence the very significant conclusion that geometric relations are powerful clues to contextual conditioning, i.e., the semantic relation R, but none of them can be used in a rigidly constraining manner. Rather, the matching of the geometric (physical) structure to the semantic structure is "soft" in nature, that is, the geometric structure (syntax) is by no means sufficient to rigidly characterize events of the universe. Failure to recognize this property and reliance on graph
isomorphisms appear to limit the applicability of some proposed schemes (see, e.g., A.S.P.[7]) to formalized data bases, i.e., non-cognitive. It must be pointed out that the discussed relatedness of semantic and geometric structures is simply a manifestation of that CP-property which we have assumed as characterizing our universe.

In the light of the foregoing discussion, we propose to describe the geometric structure of the domains of the input scene by means of a relation E, intended to give quantitative measure to the CP-property. Specifically, E is a mapping and $e_{ij}$ can be verbally expressed as "$D_i$ envelopes $D_j$ with degree $e_{ij}$". The operational definition of $e_{ij}$ is certainly important, in view both of its effectiveness in the overall process and of the ease of its implementation, but presumably not critical. Indeed, a reasonable choice could be that $e_{ij}$ measures the normalized intersection of the convex hull of $D_j$ with $D_i$. Another possible choice for $e_{ij}$ is the normalized angle subtended by $D_j$ at the center of gravity of $D_i$. Once this choice of E has been made, a graph describing the scene is obtained. This will be referred to as the E-diagram of the scene and domains $D_j$'s will be termed scene-nodes. We notice that, in general, E is reflexive but neither antisymmetric nor transitive, i.e., E fails to generate a partly ordered set and the E-diagram may contain loops. As an example, the E-diagram of scene 2c is illustrated in figure 2e. Its resemblance to the semantic subgraph is more apparent: it coincides with it if, say, one eliminates all edges whose weights are less than 0.3.

In spite of this interesting result, let us dispel the impression that this example may be formalized into a cognitive algorithm. In fact, in the next
section we shall describe the necessarily more sophisticated procedures aimed at performing the linking between the scene (E-diagram) and the map (R-diagram).

4. An Algorithm for Cognition

It is evident that the fundamental difficulty of the cognitive task is the "softness" of the matching between stimulus and map. This softness arises—as noted in previous sections—from the fuzzy match of scene node and map node, the vague relatedness of contextual conditioning to physical proximity, the fiduciary character of the map structure. In other words, input domains yield ambiguous interpretations, which have to be resolved on the basis of semantic consistency with the aid of contextual redundancies. But this difficulty, which is of a philosophical nature, is by no means the only one. Another difficulty is the efficient implementation of this soft match. The remark is not simply engineering-motivated and therefore ignorable in principle. Due to memory size and processing time limitations, different implementation strategies could reasonably make the difference between feasibility and unfeasibility.

To substantiate this point, we offer a negative argument. Assume that the scene domains are considered individually and interpreted on the basis of local attributes alone. Then each domain may be associated ("labeled") with several map nodes, with varying confidence values, that is, each domain may receive more than one plausible interpretation. These interpretations of individual domains are then used to arrive at a global interpretation of the scene, for example, by taking all possible combinations of labels and seeking some form of consensus in the semantic map. Even with the enormously simplifying assumption of isomorphism between semantic structure and an adequately selected geometric structure, the
difficulty encountered by this method would be computational, since combinatorial explosion of the cases to be examined would occur once the map complexity goes beyond a rudimentary level. The illustrated shortcoming is a consequence of the strategy adopted, (bottom-to-top) which arrives at the general context by starting from the minute details, at the whole object by piecing together its component parts.

Instead, the algorithm we propose embodies the reverse strategy (top-to-bottom), i.e., the hierarchical structure of the universe is traversed in the opposite direction, as will be apparent from the description to follow. To provide some helpful motivation, however, suppose that D₁ of the scene has been interpreted as node Aᵢ of the map. Then, interpretations of the "details" of D₁ (in the sense of relation E, Section 3) will be sought not in the entire map diagram but rather among the descendants of Aᵢ. This strategy results in two major advantages:

1) A substantial reduction of the amount of memorized data to be searched. We have a form of "tree-pruning," which in our scheme, corresponds to early discard of those paths in the semantic graph which are unlikely to yield plausible interpretations.* We shall return to this point more extensively in the sequel.

2) Feature extraction, so crucial in pattern recognition (see, e.g., [8]), becomes much less critical. Indeed the difficulty of extracting useful attributes possessing adequate discriminative power grows with the universe size.

* We wish to point out that psychological intuition tends to suggest that some analogous mechanism acts in human cognitive processes, in the sense that context identification tends to prevent the generation of ambiguities, as opposed to resolving ambiguities.
Since our algorithm is aimed at "local" classification in a small subuniverse, it is quite plausible that nonoptimally selected attributes will be completely adequate.

The objective of the algorithm can be stated as the organization of the set of scene nodes into a tree each path of which is isomorphic to a path of the map diagram. This is achieved by progressive alteration of the E-diagram, as the discussion in Section 3 strongly suggests. Specifically, the algorithm attempts to perform this task by establishing interpretive links and by selectively removing edges of the E-diagram, and terminates upon exhaustion of scene nodes.

We now introduce some nomenclature.

Interpretive links are established between scene nodes \( \{D\} \) and map nodes \( \{A\} \). The link between \( D_i \) and \( A_j \) has a \([0,1]\)-valued weight \( w_{ij} \), which expresses the confidence of this node-to-node mapping. This mapping is symbolically denoted by \( \mathcal{L} \), i.e., if \( D_i \) is linked to \( A_j \), we have \( A_j = \mathcal{L}(D_i) \) and \( D_i = \mathcal{L}^{-1}(A_j) \). Clearly, \( \mathcal{L} \) is single-valued, but not \( \mathcal{L}^{-1} \). \( \mathcal{D}(D_i) \) denotes the set of domains \( D_j \) such that \( e_{ij} \geq \theta \) (e.g., \( \theta = 0.4 \)); \( \mathcal{J}(A_k) \) the set of descendants of map node \( A_k \); \( \mathcal{L}(D_i) \) denotes the level of \( D_i \). A subset, denoted \( \{A\} \) of map nodes are termed depth-1 map nodes. For an arbitrary ordered pair \( \{D_i, A_j\} \), the \([0,1]\)-valued correlation \( c_{ij} \) between \( D_i \) and \( A_j \) is defined as

\[
(1) \quad c_{ij} = 1 - \frac{\sqrt{\sum x_h(D_i) - x_h(A_j)^2}}{\sqrt{\sum x_h^2(D_i) + \sum x_h^2(A_j)}}
\]

where \( x_h(D_i) \) and \( x_h(A_j) \) are the values of attribute \( x_h \) for \( D_i \) and \( A_j \), respectively (the given choice for \( c_{ij} \), which utilizes the relative distance of \( D_i \).
and $A_j$ in a Euclidean attribute space, does not rest on any specific philosophical grounds—rather, it has been selected for its ease of computation and because it satisfies the intuitive prerequisites of being 1 when attribute vectors of $D_i$ and $A_j$ coincide and of approaching 0 for increasing distance). The algorithm will refer at the beginning of each recursion to three current sets of domains: $R$ is the set of domains selected for processing; $E \subseteq R$ is the set of the selected domains which can be satisfactorily interpreted; $\mathcal{D}$ is the set of the currently interpreted domains (and, at each recursion $\mathcal{D}$ is augmented according to the rule $\mathcal{D} \leftarrow \mathcal{E}$).

The main difference between the $k$-th recursion and the initial one resides in the formation of the set $\mathcal{R}$. At the $k$-th recursion, $\mathcal{R}$ is given by $\cup \mathcal{B}(D_i)$ over all $D_i$'s for which $\mathcal{E}(D_i) = k-1$ and $D_i \in \mathcal{E}$ (that is, informally, $\mathcal{R}$ consists of the domains geometrically related to domains which were satisfactorily interpreted in the previous recursion). In the initial recursion, as the preceding discussion suggests, in order to take better advantage of the tree-pruning capabilities of the algorithm, it is desirable to identify the scene nodes corresponding to map nodes of depth 1. This is no simple task; non-enclosedness and domain size, however, appear to be adequate clues for this purpose.

We now examine in some detail the main concept of the cognitive procedure. We may say, informally, that is proceeds towards levels of increasing detail, and for this reason is referred to as the forward subroutine. Its flow-diagram is schematically exhibited in Fig. 3.

---

1 Errors due to the application of this criterion are susceptible to correction at later stages.
Fig. 3. Flow-diagram of the forward subroutine.

The Forward Subroutine

Step 1. Set \( k = 1, \mathcal{R} = \emptyset, \mathcal{E} = \emptyset, \mathcal{D} = \emptyset \).

Step 2. Form \( \mathcal{R} \). \( D_i \in \mathcal{R} \) if area \( D_j \geq \alpha \) and \( e_j < \varepsilon \) for every \( s \), where \( 0 < \alpha < 1 \), and \( 0 < \varepsilon < 1 \) are appropriate constants. (By so doing we identify large domains not substantially enclosed by other domains).

Step 3. Classify \( D_i \in \mathcal{R} \). For \( D_i \in \mathcal{R} \), let \( D_i \in \mathcal{E}(D_s) \) \((k = 1, 2, \ldots)\), with \( e_{s_1, i} > e_{s_2, i} \geq \ldots \). Select \( e_{s_1, i} \) as \( \max e_{s_j, i} \) where \( A_j \in \mathcal{E}(D_s) \) for the smallest \( k \) for which \( e_{s_1, i} \geq \theta_1 \) (\( \theta_1 \) an appropriate constant). (By so doing, we try to classify a likely geometrical detail of \( D_s \) among the plausible semantic details of \( A_j \), the interpretation of \( D_s \)).

Step 4. Verify vertical ordering and form \( \mathcal{E} \). For \( D_i \in \mathcal{R} \) and \( e_{s_1, i} \geq \theta_1 \) (\( \theta_1 \) an appropriate constant), determine all \( A_k 's \) so that \( v_{km_i} > 0 \). Obtain the set of domains \( \mathcal{K}_i \subseteq \mathcal{D} \cup \mathcal{E} \) defined as

\[
\mathcal{K}_i = \left( \bigcup_{k: v_{km_i} > 0} \mathcal{E}^{-1}(A_k) \right)
\]
If \( K_1 = \emptyset \), set \( c_{im_1} \rightarrow w_{im_1} \); if \( K_1 \neq \emptyset \), for \( D \in K_1 \), if \( D \) is above \( D_i \) in the scene, set \( c_{im_1} \rightarrow w_{im_1} \), if \( D \) is below \( D_i \) \((D = \mathcal{L}(A_k))\) and \( c_{im_1} \rightarrow \gamma \cdot \max k v_{km_i} \geq \theta_1 \), set \( c_{im_1} \rightarrow \gamma \cdot \max k v_{km_i} \rightarrow w_{im_1} \) \((\gamma \) an appropriate constant). (By so doing we retain the domains whose physical positionings are consistent with the vertical ordering constraint. See figure 4 for an illustration). Assign \( D_i \) to \( \mathcal{E} \) if \( w_{im} \geq \theta_1 \).

Step 5. Revise E-diagram. For \( D_1 \in \mathcal{E} \), \( D_1 \in \mathcal{E}(D_s) \), if \( w_{im} \geq \theta_2 \) \((\theta_2 \) an appropriate constant) assign \( D_1 \) to \( \mathcal{E} \) and set \( \mathcal{L}(D_1) = k \); moreover, set \( e_{is} = 0 \) and \( e_{ti} = e_{it} = 0 \) for \( s \neq t \). If \( w_{im} < \theta_2 \), set \( e_{is} = e_{si} = 0 \). (Thus the E-diagram reflects the obtained interpretations).

Step 6. Compute global confidence. For \( D_1 \in \mathcal{E} \), \( D_1 \in \mathcal{E}(D_s) \) and \( A_r = \mathcal{L}(D_s) \), compute

\[
\alpha_s = (\Sigma(e_{si} + r_{rm_i})w_{im}) / (\Sigma(e_{si} + r_{rm_i}))
\]
where the summations run over all $i$ for which $D_i \in \mathcal{D}(D_s)$. Note, that the definition of $\alpha_s$ takes into account both domains which are geometrically closely related to $D_s$ (through $e_s$) and concepts which are semantically closely related to $A_r$ (through $r_{rm}$).

**Step 7. Update confidence values.** For $D_s$ and $A_r$ as in step 6, $w_{sr}$ is updated as $f(\alpha_s, w_{sr}) \rightarrow w_{sr}$, where $f(\alpha, w)$ is an appropriate function \(^{(1)}\). This updating is then iteratively performed on all domains $D_s \in \mathcal{D}$, $\mathcal{E}(D_s) = k$ for $h = k-2, k-3, \ldots, 1$, as follows: For $\mathcal{E}(D_s) = h$ ($h \leq k-2$), and $A_r = \mathcal{E}(D_s)$, $w_{sr}$ is updated as $f(\alpha_s, w_{sr}) \rightarrow w_{sr}$, where

$$\alpha_s = \sum w_{im} \frac{r_{rm}}{\sum r_{rm}}$$

where the summations run over all $i$ for which $D_i \in \mathcal{D}(D_s)$ (notice the dominant role played by the semantic links).

This concludes the description of the forward subroutine. A pictorial illustration of its operation is given in figure 5, where the pertinent portions of the E-diagram and the map are exhibited.

---

\(^{(1)}\) The exact form of the functions $f(\alpha, w)$ is presumably not critical, once some prerequisites of an intuitive character are met, such as

1. $0 \leq f(\alpha, w) \leq 1$.
2. $f(0, w) = 0$, $f(1, w) = 1$, $f(w, w) = w$.
3. $\left| \frac{\partial f(\alpha, w)}{\partial \alpha} \right|_{\alpha=w} = \text{const} = \mu < 1$,

which reflect the updating effect of $\alpha$, and the stabilizing effect of the parameter $\mu < 1$. For example, the polynomial $a(w)\alpha^3 + b(w)\alpha^2 + c(w)\alpha$ with $a(w) = -\frac{\mu}{w(2-w)}$, $b(w) = \frac{\mu}{2-w}$, $c = 1 - \frac{\mu}{2-w}$ might be an adequate choice for $f(\alpha, w)$. 
At this point we evaluate the results of the k-th recursion. These are satisfactory if after the updating step 7, no $w_{ij}$ for $D_i \in \mathcal{D}$ has dropped below $\theta_2$ or if, as a result of the removal of edges from the E-diagram performed in step 5, the latter is still connected. The contrary events indicate that errors have been made at some previous recursion and we obtain a subset $\mathcal{K} \subset \mathcal{D}$ of domains which must be re-interpreted. This task is accomplished by the test subroutine described below (see figure 6 for its simplified flow-diagram):

**Figure 5. An illustration of the forward subroutine**

**Figure 6. Flow diagram of the test subroutine**
Step 8. Test \( w_{sj} \). For \( D_s \in \mathcal{D} \), if \( w_{sj} < \theta_3 \), assign \( D_s \in \mathcal{K} \).

Step 9. Test for connectedness. Let \( E' \) be a disconnected portion of the E-diagram having zero intersection with \( \mathcal{D} \). For \( D_i \in E' \), assign \( D_i \) to \( \mathcal{K} \) if \( e_{ri} \geq \theta_4 (0 < \theta_4 < 1 \), an appropriate constant). (By so doing we will restore in step 13 those links in the E-diagram which may potentially lead to successful interpretations of \( D_i \)).

Step 10. Test \( \mathcal{K} \). If \( \mathcal{K} = \emptyset \), increment \( k \) and go to step 11. If \( \mathcal{K} \neq \emptyset \), go to step 12 (enter backward subroutine).

Step 11. Initiate recursion. The set \( \mathcal{R} \) is given by \( \mathcal{R} = \bigcup \mathcal{S}(D_s) \), over all \( D_s \) for which: \( \mathcal{S}(D_s) = k \), \( D_s \in \mathcal{E} \) and \( w_{sm} \geq \theta_2 \) (see step 5). Return to step 3.

With step 11 a recursion cycle is completed. We now want to explicitly evidence the interpretive nature of the forward subroutine by relating it to our previous discussion of interpretation (see Introduction). Clearly steps 3 and 4 assume as a premise the hypothesis that \( D_s \) is associated with \( \mathcal{S}(D_s) \): these steps perform experiments aimed at testing the validity of this hypothesis. The result of these experiments is used to update the confidence with which the hypothesis is held (Steps 6 and 7). Concurrently these experiments lead to the formulation of new hypotheses, expressed by new tentative node-to-node associations (Step 11) and deletions of links of the E-diagram (Step 5). We also note that, as a result of step 5, the interconnection of members of \( \mathcal{S} \) is a tree.

We must now develop a remedial action to be taken when, in Step 10, \( \mathcal{K} \neq \emptyset \). This task is accomplished by the backward subroutine described below in some detail (see figure 7 for a flow diagram).
Step 12. For \( D_i \in \mathcal{K} \) determine \( m = \min_{i} \mathcal{L}(D_i) \), and replace \( k \) with \( m \) (clearly \( m \leq k \)).

Step 13. Remove from \( \mathcal{D} \) each \( D_i \) such that \( \mathcal{L}(D_i) \geq m \) and restore the pertinent connections in the E-diagram.

Step 14. For each \( D_i \in \mathcal{K} \), penalize future consideration of the previously selected interpretations and go to step 11.

Figure 7. Flow diagram of the backward subroutine

Figure 8. Schematic flow-diagram of the cognitive algorithm
This concludes the rather sketchy presentation of the cognitive algorithm whose overall flow-diagram is given in Figure 8. Important details have been omitted, such as the control of acceptance thresholds $\theta_1, \theta_2, \theta_3, \theta_4$ during backtracking (loosening and tightening of the thresholds). The top-to-bottom, or general-to-specific character of the procedure should now be quite apparent. We should remark that efficiency is not the only characteristic feature of a top-to-bottom or tree-pruning strategy. A subgraph of the map expresses an interdependency among the nodes it contains, and this interdependency is conceivably stronger between directly connected nodes than for larger node distance. In summary this interdependency can be exploited in two reciprocal ways:

1) In curtailing vastly the number of map nodes to be compared with the scene node being examined, by restricting them to the immediate descendants of a map node (tree-pruning).

2) In utilizing the confidence value with which a given map node is being interpreted to affect the confidence values of the antecedents of this node (sequential cognition).

It is appropriate to mention the decided analogy of the described cognitive algorithm to sequential decoding algorithms of recurrent codes as are encountered in communication engineering (see [4], Chapter 6). The analogies between cognitive processes and statistical communication theory are indeed far reaching and need not be overemphasized. We simply want to mention the possibility of a cognitive algorithm mirroring the maximum-likelihood decoding method of convolutional codes due to Viterbi [8]. Viterbi's method does not require backtracking, but depends heavily upon knowledge of the "memory" of the encoding
process. The corresponding concept in the semantic map would be the typical length of a path in the semantic map between essentially unrelated concepts. This analogy is not further analyzed here, but is suggested as a possible promising alternative.

5. Sample Experiments

In this section we illustrate a simplified version of the previously described method with two experiments, which are part of a series of test runs. The algorithm was programmed in PL/1 and run on the IBM 360-75 computer.

The adopted semantic map contains only 24 nodes, and is represented as a matrix in Figure 9. Only the 12 nonzero rows of this matrix are represented, corresponding to \( A_i \) such that \( \mathcal{F}(A_i) \neq \emptyset \).

| Sky   | Water | Mountain | Field | Plane | Cloud | Sun   | Island | Sand | Rock | House | Tree | Road | Animal | Pond | Window | Roof | Leaves | Trunk | Vehicle | Boat | Hull | Sail | Ice |
|-------|-------|----------|-------|-------|-------|-------|--------|------|------|-------|------|------|--------|------|--------|------|-------|-------|-------|------|
| .7    | .9    | .9       | 1.0   | .9    | .9    | .7    | 1.0    | .9   | .9   | .9    | .9   | .9   | .9     | .9   | .9     | .9   | .9    | .9    | .9    | .9   |

Figure 9. Matrix description of the R-diagram

Nodes marked with an asterisk are depth-1 nodes. Analogous information regarding the V-diagram is omitted, as well as attribute specification. We mention however
that, in this pilot experiment, the only used attributes are colors. The critical thresholds for the algorithm were chosen as follows: \( \alpha = 0, \epsilon = 0.4, \theta_1 = 0.5, \theta_2 = 0.5, \theta_3 = 0.5, \theta_4 = 0.3, \gamma = 0.5 \). Input data are supplied to the system in the following manner. A 16 x 16 grid is superimposed on the scene and for each of the resulting 256 squares the prevalent color is judged as one of eight possible choices.

The first example is highly artificial and was used essentially as a code check. We report it here because it lends itself, for its simplicity, to a detailed description without unnecessarily burdening the reader. The starting pictorial data is represented in figure 10 and, in its rudimentary simplicity,

![Figure 10. A scene to be recognized](image)

is self-explanatory. Domains are labelled with integers for easy reference. In figure 11a we illustrate the initial E-diagram as obtained at the end of the preprocessing stage (edges weighted less than 0.1 have been omitted). Figure 11b shows the modified E-diagram at the end of the first recursion (see p. 22) with the labels and corresponding confidence values for the interpreted domains. The final E-diagram is given in figure 11c and shows correct interpretations with
high confidence values. No resort to the backward subroutine occurred during the execution of the program.

Figure 11. Successive modifications of the E-diagram during the execution of the algorithm.

The natural scene corresponding to the second example is given in Fig. 12 with regions indicated. Analysis of the scene results in the E-diagram of Fig. 13. After the first level regions have been assigned, the second level regions are assigned with the indicated confidence measures. Thereafter, the vertical check finds that region 9 bears a relatively implausible relation to region 10 causing the confidence of region 9 to be reduced to .59. The subsequent confidence updating alters the first level values to those marked "u". Since no confidence values are less than 0.5 (=0.1) and all regions are joined in the E-diagram, the program halts at this point.

Two instructive errors occur in the result (1) the hypothesis that region 10 is a mountain (actually, a rooftop) and (2) the hypothesis that region 12 is snow (actually, a white carport roof). The first case would be reduced in plausibility if any data providing depth (distance) or texture were
provided. As is, only the vertical relationship data questions the plausibility that region 10 is a mountain. Error 2, in which region 12 = snow is more interesting. On the basis of the limited semantic model of the program, this analysis is quite reasonable. But what extensions of the model would make possible the plausible conclusion that region 10 is a roof top? With sufficient resolution, possibly the structure of the associated house could be deduced. We feel that it is more immediate to deduce the temperateness of the scene, a visually imperceivable property, on the evidence of green foliage, for example. This observation then feeds back to the perceivable domain, reducing the plausibility of snow.

The introduction of imperceivable properties and their interplay with the perceivables is certainly among the more interesting directions which should be pursued.

Fig. 12 Second Example Scene
6. A Critique

Some aspects of the proposed model for a cognitive process are not entirely satisfactory and will now be critically analyzed.

The major weakness is the absence of a general theory of abductive inference, i.e., a coherent manipulative formalism for the computation of the plausibility or global confidence of complex structures of objects, based on a quantitative definition of their "local" pairwise implications. We have tried to cope with this difficulty by empirically quantifying the map and selecting the algorithmic rules. This position, however, is only temporarily defensible and an acceptable formal system is needed, analogous to a probability system, for situations in which a priori probabilities cannot be defined.

Similarly, it would be desirable to have a uniform way to handle the various parameters upon which the stimulus-map linking is based. For example, the correlation of attribute-value pairs and the verification of vertical ordering should be combined in a theoretically sounder fashion than is presently done. This not only would confer a better philosophical appeal to the proposed algorithm, but would simplify implementation by avoiding "ad hoc" subroutines.
Also, the present capabilities of the system are contingent on the fact that the stimulus is presented with sufficient attribute richness, i.e., with colors. Colors appear to possess powerful discrimination capabilities. Their removal, i.e., the acceptance of black-and-white stimuli, will presumably require - for the same degree of success - both a more complex preprocessing and a more sophisticated map.

Finally the algorithm is incapable of associating as belonging to the same object, portions which are apparently separated by the interposition of another object. This, however, represents a moderate difficulty. The removal of the mentioned shortcomings is the subject of continuing investigation.

As closing remarks, we note that the recognizing abilities of the system depend upon the adopted semantic map. In other words, simple replacing of the semantic map enables the system to operate on a completely different universe. Moreover, the complexity of processing depends essentially upon the average number of descendants of map concepts (i.e., upon the size of sets within which classification must be performed). Therefore substantial growth of the semantic map is permissible with a very moderate growth of the processing complexity. This is the single most interesting feature of the described algorithm, which points to its generalizability to universes of increasing complexity.

Acknowledgement

The authors gratefully acknowledge the invaluable cooperation of Mr. W. T. Bielby, who implemented the programming of the cognitive algorithm with enthusiasm and great ability. The interaction with their colleagues and students, at Illinois and elsewhere, also greatly contributed to the present work; specifically, they like to remember the stimulating discussions they had with H. Maturana, A. Grasselli, R. F. Simmons, H. Von Foerster, and R. T. Chien.
References


Distribution List (cont'd.)

Commanding General
U.S. Army Strategic Communications Command
Attn: HC-030-SAC
Fort McPherson, Arizona 85643

Commander
U.S. Naval Weapons Center
Attn: Technical Library
China Lake, California 93555

Library (Code 2124)
Technical Report Section
Naval Postgraduate School
Monterey, California 93940

Glen A. Myers (Code 526H)
Attn: Professor of Elec Eng
Naval Postgraduate School
Monterey, California 93940

Dr Leo Young
Stanford Research Institute
Menlo Park, California 94025

Lenkurt Electric Co., Inc.
1105 County Road
San Carlos, California 94070
Attn: Mr E.K. Peterson

Director
Electronic Sciences Laboratory
University at California
Berkeley, California 94720

ADDENDUM
Dr Joel Trimble, Code 437
Information Systems Branch
Office of Naval Research
Department of the Army
Washington, D.C. 20360

U.S. Naval Oceanographic Office
Attn: M. Rogofsky, Librarian (CODE 640)
Washington, D.C. 20390

CORRECTION
Director, Electronic Programs
Attn: CODE 427
Office of Naval Research
Department of the Navy
Washington, D.C. 20360

DELETE
Technical Director
U.S. Army Limited War Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

REPLACE WITH
Technical Director
U.S. Army Land Warfare Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

DELETE
USAF European Office of Aerospace Research
APO New York 09667

REPLACE WITH
European Office of Aerospace Research
Technical Information Office
Box 14, FPO New York 09510

DELETE
Dr John R. Raggazzini, Dean
School of Engineering and Science
New York University
University Heights
Bronx, New York 10453

REPLACE WITH
New York University
Engineering Library
Bronx, New York 10453
In this paper we describe an approach to the artificial recognition of events of a nonsymbolic nature, such as the bidimensional perspective views of scenes of our everyday world. Scenes are presented as colored pictures and the objective of the cognitive task is the labeling of the interpreted scene objects. The method is based on three major components: i) a preprocessed version of the scene (stimulus), ii) a semantic map and iii) an algorithm which attempts interpretation of the stimulus under the guidance of the semantic map. The algorithm is sequential and proceeds from general to specific, thereby achieving efficient tree-pruning (contextual elimination). Stimulus interpretation is based on attribute-value matching, but classification relies strongly on the accumulated context. Backtracking provisions are available for correction of earlier wrong hypotheses. Experiments are presented and described. The major weakness of the approach is the present lack of a satisfactory theory of abductive inference. Flexibility, generalizability and efficiency appear to be valid merits.
<table>
<thead>
<tr>
<th>KEY WORDS</th>
<th>LINK A</th>
<th>LINK B</th>
<th>LINK C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern Recognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scene Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonlinguistic Cognition</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>