11

Representational Axes and Temporal Cooperative Processes

John K. Tsotsos

This paper proposes a framework for the integration of time into "high-level" vision. Many characteristics of this framework have much in common with the handling of temporal concepts in low-level vision as well, but these will not be discussed. This framework addresses the integration of representations for temporal concepts with temporal reasoning schemes, temporal grouping, the discrimination and labeling of temporal concepts, and temporal sampling rates. Such issues are not addressed in most other "high-level-vision" methodologies.

Visual perception takes place within a spatio-temporal context, and thus the integration of time into every aspect of processing is crucial. One may draw an analogy here to the use of the term *pragmatics* in research on natural-language understanding. *Syntax* describes the rules for how individual tokens are grouped into larger tokens; *semantics* ascribes meaning to tokens; *pragmatics* relates that meaning to the remainder of the discourse or to the context in which the utterance was found. Most past computer-vision research has dealt with static images—a most unnatural kind of input, since biological visual systems are almost never presented with a single time slice of the visual world out of its spatial and temporal context. The temporal cooperative process that will be described is integrated with a hypothesize-and-test reasoning framework. The structure over which the cooperative process operates changes with time because of the status of interpretation, and iterations are defined by the passing of time for either static images (image contents do not change with time) or time-varying images, where in both cases the images are presented over a period of time. Since for computer vision we must deal with a discretized world, and since time is to be considered, temporal sampling must be an issue.

Is there anything special about time and the processing of time-varying information that does not allow us to treat it simply as a fourth dimension? Although significant effort has been expended on the analysis of time-varying images, this question has never been addressed completely. It is simply for the sake of covenience in processing that most computer-vision research has concentrated on single images, and that most motion research has concentrated on small numbers of images in a time sequence. In most past work on motion analysis and understanding it has been tacitly assumed that the techniques that have been used for analysis of static images apply directly for the time-varying case, and that time is strictly subsequent to spatial analysis. This is not necessarily the case; generally, time must be incorporated into each aspect of processing. The additional constraints provided by the temporal context of a scene are crucial for disambiguation and recognition.

There are three main constraints that time brings to bear on work in motion understanding. First, we cannot stop time. System response is required without unreasonable delay, since the environment continues to change. Yet computation requires a finite period of time. Thus, the processing rate must be sufficiently high to maintain an interpretation of the scene. Second, there is limited storage available. Whereas spatial data are presented in parallel, temporal data are presented serially. This implies that a finite temporal window must be used, and that events over a longer period of time must be sufficiently abstracted that they can be efficiently represented internally. Third, the system must be stable in a noisy environment and must degrade gracefully with increasing noise. Noise can take many forms: quantization noise, data irrelevant to the problem being tackled, and misleading data (including data at the wrong spatial and/or temporal scale). Therefore, some amount of smoothing must be present so that these confounding effects are minimized. This implies that rise and fall times must be chosen accordingly. Similarly, decision-making processes must exhibit procrastination and inertia. They cannot make decisions hastily, and they must temporally integrate results in order to take temporal context into account. A single noise point cannot undo the effect of many samples that exhibit some trend, yet enough samples must have been viewed in order to discover the trend. All this impacts the processing rate for the image sequence.

The problem of time and the interpretation of events in time is not new to psychologists. David Hartley (1749) set forth several propositions pertaining to groups of successive concepts and groups of compound synchronous concepts. He noted that an instance of such a group will raise in the mind expectations for the remaining concepts of the group, whether the concepts occur in a sequence or simultaneously. This is an example of top-down activation of grouping hypotheses. James Mill (1829) elaborated these thoughts and concluded that sensations have a naturally synchronous or successive order. To him, successive order implied order in time whereas synchronous order implied order in space. In addition, successive order implied notions of antecedent and consequent sensations. Grouping received large amounts of attention from the Gestalt psychologists, according to whom grouping principles can be summarized by the terms *proximity, similarity, continuity, symmetry,* and *familiarity* (Wertheimer 1923). Although most studies of such grouping principles were performed mainly by considering the partitioning of a stimulus array in space, each of these has a temporal analogue. For example, the smaller the temporal separation, the greater the tendency to be grouped into a sequence; similarity of temporal primitive; temporal symmetry could refer to oscillatory motions, etc. Unfortunately, for both temporal and spatial versions, these require much elaboration and quantification before they can be immediately applied.

Synchronous order in time was discussed by Gibson (1957) and Hay (1966), who considered the distinctions between physical and optical motions. They, particularly Hay, defined a variety of optical motions as combinations of simultaneous physical motions, thus decomposing complex motions into aggregates of simpler ones. Such decompositions are important notions in my work.

The following experiments point to a strong relationship between expectation and the specialization/generalization of concepts. The experiments of Cooper and Shepard (1973) show the strong positive effect of *a priori* expectations on time for interpretation; those of Bugelski and Alampay (1962) and Palmer (1975) show the effects of generalization of expectation classes. Cooper and Shepard reported that in the identification of letters presented at varying orientations the time taken to identify the letter varied with the amount of rotation (to a maximum value at 180°); this implied that mental rotation and matching were being performed by the visual system, and that if identity and orientation were given before the stimulus the response time was flat across all orientations as long as sufficient time was allowed before the stimulus was presented for expectation formation.

Bugelski and Alampay showed that if a subject is conditioned to expect a given category (or generalization) of stimulus, then the identification time of the stimulus is reduced. They presented stimuli all belonging to the same class of concept (animals); when nonanimal stimuli were presented, the response time increased. This was further examined by Palmer, who also noted the impairment of identification if the context is misleading. (The mechanisms that produce such behavior are not understood.)

Dretske (1981) emphasized temporal integration:

> To understand how certain sets of information are registered, it is important to understand the way a sensory representation may be the result of a temporal summation of signals. To think of the processing of sensory information in static terms, in terms of the kind of information embodied in the stimulus at a particular time, is to completely miss the extent to which our sensory representations depend on an integrative process over time.

The importance of time in sensory perception is given additional credence by the fact that sensory neurons have the ability to sum their input signals not only spatially but also temporally (Kandel and Schwartz 1981). Within the domain of computer vision, the study of motion interpretation has not adequately addressed the issue of temporal integration of results; indeed, much of the work concentrates on the processing of motion information by considering a set of static images. The methodologies developed do not, for the most part, have the ability to combine events into higher-order concepts such as sequential or synchronous events.

The remainder of this paper deals with the representation and organization of temporal concepts, and describes how this organization drives a hypothesize-and-test reasoning process as well as the cooperative process that forms the structure within which the hypothesis response is computed. This research is distinguished from other research on cooperative processes in a number of ways. Schemes such as those of Glazer (1982) and Terzopoulos (1982) use cooperative methods in arriving at a solution to a numerical approximation problem. They use hierarchies of data, but the type of information is uniform; only resolution differs by level of hierarchy. Their problems are posed as numerical ones, and in those cases relaxation methods assist in obtaining a solution. Our case differs in four respects:

- Our information is not uniform but rather different concepts are represented at different levels of the hierarchies.

- There are multiple interacting networks, each organized according to different semantics.

- The data are time-varying (and, more important, the structive over which relaxation is performed is time-varying).

- We are interested in an interpretation task, not an approximation one.

Research such as that of Hummel and Zucker (1980) deals with theoretical foundations underlying relaxation methods. My work should not be considered in such a light. The cooperative process to be described has the qualitative properties I believe are desirable for temporal interpretation, and its performance will be described empirically and in a qualitative fashion through the use of several examples. It should be regarded as an extension of relaxation methods into a domain where they have previously not been used.

There are similarities between the present work and that of Hinton and Sejnowski (1983). They too employ a hypothesize-and-test framework where hypotheses are in one of two states (true and false) and apply a cooperative process to find optimal combinations of hypotheses. Differences exist in that a representation for hypotheses was not presented, nor were specific search mechanisms, and their mathematical analysis (which they based on system energy) is not directly applicable. The work presented in this paper can also be seen as an elaboration and an extension of the vertical and horizontal relaxation processes of Zucker (1978a).

One further major difference exists with past cooperative-computation research. We are dealing with a time-varying data-interpretation task. Past work has shown that, in general, relaxation schemes require large, and potentially very large, numbers of iterations in order to converge to stable solutions. We cannot afford this luxury. In a time-varying context, decisions must be made relatively independent of the number of iterations, so that new data can be considered. Therefore, what is required is a cooperative process that can be characterized in the following way: The only decisions that will be made are those that can be made within a small, fixed number of iterations. We must discover the conditions under which this is possible, for all events of interest, given small amounts of noise in the data. These conditions will be developed in the course of the paper, and will lead to a relationship between image sampling rate and iterations.

In order to describe our scheme—that is, a temporal, high-level vision framework—we must first decide on what we mean by "high level." The approaches of Marr (1982) and Julesz (1980) certainly do not fall within the meaning of high level. On the other hand, work described by Hanson and Riseman (1978), Levine (1978), Ballard et al. (1978), Brooks (1981), O'Rourke and Badler (1980), and Tenenbaum and Barrow (1977) certainly does. What distinguishes these approaches? It is not (as is commonly thought) the use of domain-specific knowledge. The work conducted and motivated by Marr utilized physical constraints of the world, while the VISIONS system of Hanson and Riseman employs knowledge of the appearance of houses; both are forms of knowledge. I believe that the distinction is a deeper one. We can gain insight by looking at some recent distinctions drawn by psychologists between "pre-attentive" and "attentive" vision.

Briefly, the pre-attentive system is a parallel one that can cope with single disjunctive features only, such as the distinction between differently oriented black bars on a white background. (See Treisman and Gelade 1980; Treisman 1982; Treisman and Schmidt 1982.) The attentive system, on the other hand, can handle much greater complexity of visual input. Treisman and co-workers claim that it is a serial system incorporating a focus of attention, and that it thus can deal with the conjunction of features such as color and shape. Moreover, it must play a role in the discrimination of feature conjuncts within a field of conjunctions of similar features. For example, the attentive system must be used to find a green vertical bar in a field of many randomly oriented bars, each a different color. Several discussions on the differences between attentive and pre-attentive vision may be found in the literature; see Julesz and Schumer 1981. Curiously, only static images had been considered in those works. On the other hand, the so-called short-range and long-range motion processes of Braddick (1974) and Anstis (1978) describe perceptual processes that can cope with motions involving small displacements and not requiring form recognition or correspondence (Braddick) and with motions requiring form recognition and larger displacements, necessitating correspondence (Anstis). It would be startling indeed if the static and temporal distinctions drawn in the above works were simply instances of the same process.

The most obvious manifestation (but not the only one) of serial visual processing is visual search. Given a complex image, our gaze typically moves around the image, tracing contours and interesting features until the image has been interpreted to our satisfaction. What could trigger such a search? If one believes that the purpose of vision is to construct some internal representation of the physical world, then one may also hypothesize that, on the basis of pre-attentive vision, a "skeleton" structure is created, which may then, if required, be filled in by the attentive process. This filling in or completion could be driven by the need for completeness of description and disambiguation. For single disjunctive visual features, this skeleton may be complete—it may be compared to Marr's (1982) primal sketch or to Julesz's (1980) textons. Labeling of features of such a skeleton of isolated, disjunctive features proceeds in a "pre-attentive-like" manner, that is, bottom-up and in parallel. Whereas visual search in the static case involves changes of fixation in some manner, search in the dynamic case may involve both search for features that complete the skeleton of objects and temporal search or the generation of expectations in time for completion of the motion skeleton. The motion skeleton may be the result of short-range processes, or what Braddick and Anstis call

Real Motion, whereas Apparent Motion requires the filling in of form-related and correspondence-related information.

Search schemes are common components of systems that claim reasoning capabilities. In addition, all such systems exhibit foci of attention that are derived from the "best guesses" for the solution of the problem at hand. However, search in vision can take many forms. In order to conjoin features (such as "red" and "the letter B") into a single percept, search for corresponding features in different portions of processing hierarchies may be required. This, of course, assumes static images. It may be that the feature being searched for has no corresponding instance and thus a visual search—eye motion—must be initiated. (There clearly are other reasons for eye movement as well.) This would be accompanied by establishing expectations as to what the attentive system was looking for, thus biasing the computation. Finally, in the time-varying case, it may be that the corresponding feature is an event that has not yet occurred. In this case, expectations are set up in time, again biasing the computation. These biases may be regarded in one of two ways: as "priming" signals (that is, signals that may facilitate the computation of particular units or concepts) or as manifestations of internal focus of processing attention.

Search for missing discriminatory features will be considered as the main distinction between "high-level" and "low-level" vision. Search for globally consistent results, such as manifested by relaxation schemes, is not included within this distinction, since global consistency must play a role in both levels of vision.

In summary, the major capabilities that an attentive vision system—particularly one addressing time-varying phenomena—must possess are the following:

prototype concept representation, manifested as stereotyped computing units

temporal grouping

temporal expectations

generalization of concepts in relation to expectations

rich search dimensions that enable and distinguish search in image space from search in hypothesis space

spatial and temporal integration of results

generation and maintenance of a focus of attention

an interface to the tokens that may be abstracted from images by early processes.

This paper discusses a framework for the realization of such capabilities. Although it is claimed that attentive vision systems must possess at least the capabilities just summarized, the realization presented here is only one of many possible realizations with the same capabilities. There are no claims on necessity. On the other hand, this framework is indeed sufficient and does satisfy the requirements laid out. All the machinery described in this paper has been implemented as part of the ALVEN expert vision system (Tsotsos 1981a, 1983; Tsotsos et al. 1984), which assesses the performance of the human left ventricle from x-ray image sequences. The experimental work described in this paper was done with that implementation.

Overview

The basic properties of an attentive vision framework involving temporal phenomena were described in the preceding section, and they are manifested in a clear manner in our attentive vision framework. Knowledge organization plays a significant role. The key elements of the framework are the following:

- Four dimensions of knowledge organization, namely IS-A (generalization/specialization), PART-OF (aggregation/decomposition), SIMILARITY, and Temporal Precedence.

- Frames as the prototype knowledge or computing unit, organized along the four dimensions just mentioned. These may be considered as declarative structures, specialized procedures, or some combination of these two. It is not important how they are implemented for the purposes of this framework. The important point is that specialized units are present that interpret specific visual entities, which may be as simple as "change detection" or as sophisticated as "left ventricular systole" and which may involve spatial relations, such as "above" or "inside," and temporal quantities such as velocities or rates of area change. These computing units must have several important properties: they realize when they cannot successfully interpret some visual feature, they can create a data structure (an exception record) describing why they cannot, they can create instance representations of themselves when appropriate, and they can communicate with other units.

- The "leaves" of the PART-OF hierarchy of frames represent the primitive types of features that may be abstracted in a pre-attentive fashion from the images, thus forming the interface between early and attentive processing.

- Hypothesize and test as the basic interpretation paradigm, with hypotheses being activated from the knowledge frames as a result of four interacting search dimensions, namely hypothesis-driven, data-driven, failure-driven, and temporal search. Since hypotheses are derived from prototype knowledge frames, and those frames are organized, hypotheses are also organized in the same fashion.
- A projection mechanism to transduce hypothesis-specific expectations to image-specific ones, and a scheme to recover from inadequate expectations through upward traversal of the IS-A hierarchy, thereby relaxing constraints.
- A cooperative process that integrates results over time and space in order to enforce global hypothesis consistency and determine the best hypotheses, and thus the system focus of attention. The focus exhibits levels of abstraction because of the hypothesis organization. This process should also be able to deal with noisy and incomplete data.

It will be shown that knowledge organization is the driving force behind the interpretation strategy, and that, in addition to the knowledge-structuring properties used in many other knowledge-based systems, the dimensions of knowledge organization have many other important uses, ranging from restricting the temporal sampling rate to supplying the feedback necessary for stability for the cooperative process. These aspects and others will be discussed in detail.

Representation and Organization

A popular form of knowledge representation for "packets" of knowledge is the *frame* (Minsky 1975). Frames may be thought of as primitive computing units, declarative definitional structures, or some combination of the two. Their exact form is not important; it suffices that each is specialized for the computation of some specific visual entity. A version of frames called *classes* is presented in the PSN (Procedural Semantic Networks) formalism of Levesque and Mylopoulos (1979). The remainder of the discussion focuses on knowledge class organization, since organization is of greater importance than the form of the actual knowledge packages.

Class Organization: Generalization, Aggregation, Instantiation

When one is confronted with a large, complex task, "divide and conquer" is an obvious tactic. Task partitioning is crucial; however, arbitrary task subdivision will yield structures that are unwieldy, unnecessarily complex, or inappropriately simple. Furthermore, such structures have poorly defined semantics, lead to inefficient processing, and lack clarity and perspicuity. Within the existing representational repertoire, there exist two common tools for domain subdivision and organization: the IS-A relationship (or generalization/specialization axis) and the PART-OF relationship (or part/whole axis). Brachman (1979, 1982) and Levesque and Mylopoulos (1979) discuss the properties, semantics, and use of these relationships. The IS-A (generalization/specialization) relationship was included in order to control the level of specificity of concepts represented. IS-A provides for economy of representation by representing constraints only once and enforcing strict inheritance of constraints and structural components. It is a natural concept-organization scheme, and it provides a partial ordering of knowledge concepts that is convenient for top-down search strategies. In conjunction with SIMILARITY (another representational construct), IS-A siblings may be implicitly partitioned into discriminatory sets. The PART-OF (aggregation) relationship allows control of the level of resolution represented in knowledge packages and thus control of the knowledge granularity of the knowledge base. It provides for the implementation of a divide-and-conquer representational strategy, and it forms a partial ordering of knowledge concepts that is useful for both top-down and bottom-up search strategies. Concept structure can be represented using slots in a class definition. The slots form an implicit PART-OF relationship with the concept. Representational prototypes (classes) are distinguished from and related to tokens by the INSTANCE-OF relationship. Instances must reflect the structure of the class they are related to; however, partial instances are permitted in association with a set of exception instances, or the exception record, for that class. In addition, a third type of incomplete instance is permitted: the potential instance or hypothesis. It is basically a structure that conforms to the "skeleton" of the generic class, but that may have only a subset of slots filled, and has not achieved a certainty high enough to cause it to be an instance or a partial instance. Details on the precise semantics of IS-A, PART-OF, and INSTANCE-OF may be found in Levesque and Mylopoulos 1979.

Such knowledge organization dimensions have been used in many other knowledge-based vision systems. (See, e.g., Hanson and Riseman 1978; Levine 1978; Sabbah 1981; Brooks 1981; Mackworth and Havens 1983.) Yet, their integration into the interpretation scheme was not completely developed within those works. The knowledge organization was used to structure knowledge and to provide access mechanisms for it. This work will show that knowledge organization can really do much more.

Time

A representation of general temporal concepts is beyond the scope of this work. In fact, many of the details of such representations, such as the calculus presented by Allen (1981), are not relevant. We are concerned with the impact of time-varying concepts on knowledge organization. From this point of view, time-varying aspects impose a partial ordering on elements of a concept; that is, a concept's parts are ordered in time. This ordering involves relationships such as *next, previous, simultaneous,* and *overlap.* Temporal precedence relationships interact with the PART-OF relationship.

Arbitrary groups or sets of events can be represented. If temporal concepts are grouped within some class, then whenever they represent a sequence of events, the particular concept representing the group exhibits a "coarser" temporal resolution than its components. To carry this further: A PART-OF hierarchy of temporal concepts displays levels of temporal resolution.

Description via Comparison: Similarity

Similarity measures that can be used to assist in the selection of other relevant hypotheses on the failure of hypothesis matching are useful to control the growth of hypothesis space. These measures usually relate classes that together make up a discriminatory set (i.e., only one of them can be instantiated at any one time). As such, they relate classes that are at the same level of specificity on the IS-A hierarchy and have the same IS-A parent class. Multiple IS-A parents are permitted as long as each class of the discriminatory set has the same set of IS-A parents. Similarity links are components of the frame scheme of Minsky (1975), and a realization of SIMILARITY links as an exception-handling mechanism based on a representation of the common and differing portions between two classes is presented in Tsotsos et al. 1980. Thus, they are an element of embedded declarative control, and they add a different view of frame representation, thereby enhancing the redundancy of the representation. The three major components of a SIMILARITY link are the list of target classes, the "similarities" expression (the important common portions between the source and target classes—remember that during interpretation the target classes are not active when the SIMILARITY link is being evaluated; thus, in time-dependent reasoning situations the components of the target class that are the same as those in the source class before activation of the

SIMILARITY link can be verified using the similarities expression), and the "differences" expression (the time course of exceptions that would be raised through inter-slot constraints of the source class or in parts of the source class).

Transduction between Domains: Projection

Projection is a transformational link relating representations of the same concept but in differing representational domains. In other words, projection is used to represent hypothesis-to-signal transductions. It is important because this enables the implementation of "priming" signals from "high-level" hypothesis expectations down to image-specific computing units. It is, for example, the relationship between a prototypical object and its actual appearance in an image. The ALVEN system employs such projections in creating predictions for low-level image operators.

The need for expectations and their use in high-level vision is not a new idea. Its importance was emphasized by Mackworth (1978) and Kanade (1980). However, no clear models exist. We believe that rich and well-defined knowledge organizations are a prerequisite for such expectation capabilities.

The generation of expectations is driven by current best hypotheses. Clearly, if the set of best hypotheses does not include the correct one at some point then the expectations produced will not be verified by the data. A mechanism that utilizes the hypothesis level of specificity in recovering gracefully from such incorrect expectations is required, and it is here that the above-mentioned link between specialization of concepts and expectations is used. Recovery from incorrect predictions involves upward movement along the IS-A hierarchy of hypotheses. This has the effect of relaxing the constraints that generated the original incorrect prediction and allowing for the creation of a more general prediction as the next plan.

Down (1983) presented examples of prediction specifications for simple shapes (such as points, lines, arcs, and circles) and aggregations of these shapes into more complex forms, as well as the methodology for their use. Predictions are classes in their own right and are thus treated as are all other classes. They are a version of limited planning capability.

Attentive Vision and Search

As described earlier, attentive vision can be characterized by several basic characteristics. One of these characteristics, perhaps the most important, is the presence of rich dimensions of search that allow for an interface between data tokens and the interpretation process, distinguish between search in the image and search in hypothesis space, and enable model-driven, goal-driven, and data-driven interpretation. This section addresses the search schemes in our framework and the basic processing cycle within which they operate.

Hypothesize-and-test is the basic recognition paradigm. However, activation of hypotheses proceeds along each of four dimensions concurrently, and hypotheses are considered in parallel rather than sequentially. These dimensions are the same class-organization axes that have been described above. Hypothesis activation is a cyclic process, beginning with data-driven activation and then alternating with goal-driven, model-driven, temporally directed, and failure-directed activations. For a given set of input data, in a single time slice, activation is terminated when none of the four activation mechanisms can identify an unactivated viable hypothesis. Termination is guaranteed by the finite size of the knowledge and the explicit prevention of reactivation of already active hypotheses. Furthermore, the activation of one hypothesis has implications for other hypotheses. Because of the multidimensional nature of hypothesis activation, the "focus" of the system also exhibits levels of attention. In its examination, the focus can be stated according to the desired level of specificity or resolution (the two are related), discrimination set, or temporal slice.

Each newly activated hypothesis is recorded in a structure that is similar to the class whose instance it has hypothesized. This structure includes the class slots awaiting fillers, the relationships that the hypothesis has with other hypotheses (its "conceptual adjacency"), and an initial certainty value determined by the activating hypothesis.

The "test" part of hypothesize-and-test is accomplished by the evaluation of constraints specified in the knowledge classes. The matching result of a hypothesis for the purpose of hypothesis ranking is summarized as either success or failure. Matching is defined as successful if each hypothesis component that is expected to be present at the time of matching has a corresponding active hypothesis that matches successfully and each slot and inter-slot constraint within the hypothesis evaluates to true. Matching is defined as unsuccessful if any slot or inter-slot constraint evaluates to false, or if any expected hypothesis component fails matching or does not exist and cannot be found through any mechanism.

Goal-Directed and Model-Directed Search

Top-down traversal of an IS-A hierarchy (that is, moving downward when concepts are verified) implies a constrained form of hypothesize-and-test

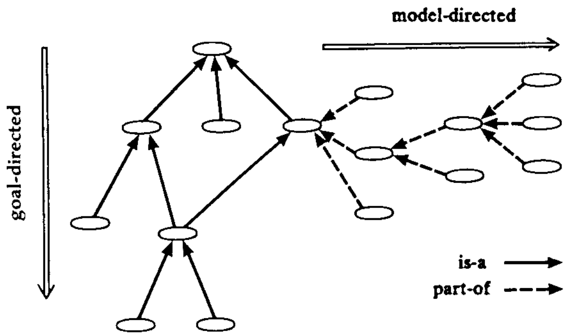


Figure 1
The directions of search for goal-directed and model-directed search processes.

for more specialized concepts. Similarly, top-down traversal of the PART-OF hierarchy implies a constrained form of hypothesize-and-test for components of classes that reflect greater resolution of detail. This search dimension along either representational axis is success-driven (figure 1).

A successful match (distinguished from "instantiation") of an IS-A parent concept implies that perhaps one of its IS-A children applies; a successful match of an IS-A child implies that its parents should also be true. Multiple IS-A children can be activated by a parent, but a more efficient scheme would be to activate one of the children if all children form a mutually exclusive set, or one child from each of several such sets, and then allow lateral search to take over. This selection may be guided by meta-knowledge associated with the IS-A parent hypothesis class. The lateral search mechanism will then determine how many IS-A children in a discriminatory set are viable possibilities. Note that hypotheses are activated for each class in a particular IS-A branch as the hierarchy is being traversed, and thus tokens will be created for each on instantiation. The activation of a hypothesis implies activation of all its PART-OF components as hypotheses as well. Cycles are avoided since at most one hypothesis for a particular class can exist for each time interval and set of structural components.

In the case of top-down PART-OF hierarchy traversal, activation of a hypothesis forces activation of hypotheses corresponding to each of its components, i.e., slots. The implication is that all slots must be filled in order for the parent hypothesis to be instantiated. Slots may have a temporal ordering, a feature handled by the temporal search mechanism interacting with this one. The search is therefore for all components of a class, increasing the resolution of the class definition.

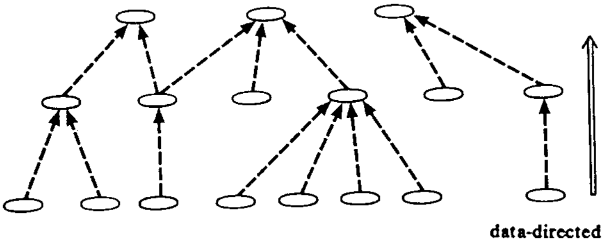


Figure 2
The direction of search for the data-directed search process.

Data-Directed Search

The PART-OF hierarchy can also be traversed bottom-up in aggregation mode (figure 2). Bottom-up traversal implies a form of hypothesize-and-test where hypotheses activate other hypotheses that may have them as components, i.e., data-directed or event-driven search. This form of search has important implications for the definition of the knowledge base. The leaves of the PART-OF hierarchy are required to represent the types of tokens that can be abstracted from images, thus interfacing attentive with early vision processing components. This definition of the interface is independent of the number or form of "intrinsic" images computed during early vision.

This form of search also is success-driven. A successful match of a hypothesis implies that it may be a component of a larger grouping of hypotheses, and thus each possible PART-OF parent hypothesis is activated. Guidance for limiting the number of activations can be obtained from relevant meta-knowledge associated with the activating hypothesis class. Activation of hypotheses in this direction implies activation of all IS-A ancestors of new hypotheses. Arbitrary hypothesis groupings can be accomplished, but specific groupings can be recognized only if defined as a class.

Lateral Failure-Directed Search

Lateral search is a very different process than the previous two, since it is failure-driven. The search is along the SIMILARITY dimension (figure 3) and depends on the exception record of a particular hypothesis. Typically, several SIMILARITY links will be activated for a given hypothesis, and the resultant set of hypotheses is considered as a discriminatory set (i.e., at most one of them may be correct). Discriminatory sets are not allowed to

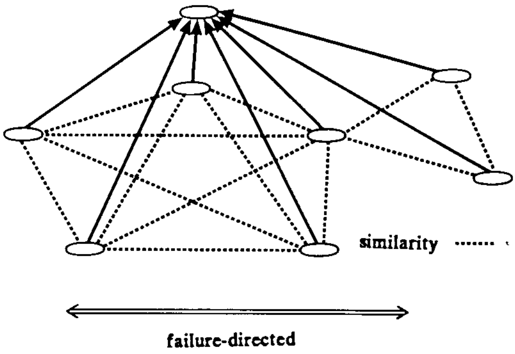


Figure 3
The directions of search for the failure-directed search process.

intersect. SIMILARITY interacts with the PART-OF relationship in that exceptions raised that specify missing slot tokens are handled by the hypothesis's PART-OF parent. Of course, the source and target classes of those links are at a different level of resolution.

Two situations may arise that require special consideration. The first occurs when the failing hypothesis has no PART-OF parent, as may occur during the first stage of data-driven search. The second may occur when no similarity link can be found that handles some of the exceptions raised. The goal of the exception-handling mechanism is to use all raised exceptions in some way through the similarity links. In the former case, when the failing hypothesis has no PART-OF parent, the similarity links found within its own structure are tried. In the latter case, when no similarity link can be found that can handle certain exceptions, the IS-A inheritance mechanism plays a role. If no similarity link can be found within the hypothesis itself or within a PART-OF parent, this means that the exhibited phenomena disagree with the hypothesized ones in a major way. For example, suppose that a hypothesis defining a particular type of "contract" motion was under consideration. The immediately avilable similarity links may be set up handle differences in rate of contraction. They would not point to appropriate hypotheses if the motion simply ceased. Therefore, similarity links are inherited from IS-A ancestors by either the PART-OF parent or the hypothesis itself as necessary. The IS-A ancestors of each hypothesis are also active hypotheses. The end result would be that the exception causes new hypotheses to be activated at some higher level of generalization, rather than at the same level. In this way traversal back up the IS-A hierarchy can be accomplished. This is not necessarily a form of

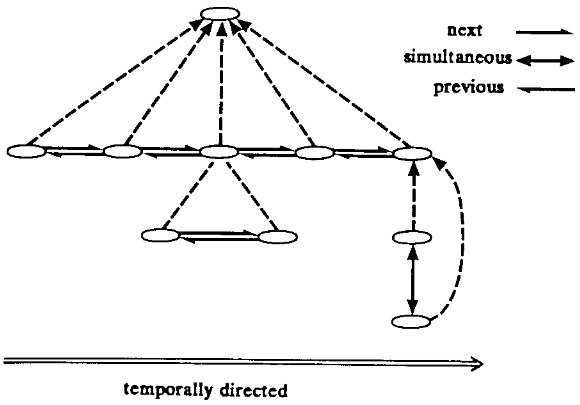


Figure 4
The direction of search for the temporally directed search process.

backtracking. It will indeed be backtracking in situations where wrong hypotheses are activated. However, since the context is time-varying, this mechanism also allows fast reaction to changes in data from hypotheses that are no longer viable because their expectations in time are no longer exhibited in the data.

Temporally Directed Search

Temporal search, a special case of hypothesis-driven search along the PART-OF dimension, is relevant whenever a class has an IS-A relationship with the SEQUENCE class. This is shown in Figure 4. Elements of a sequence may be compound events, such as other sequences, simultaneous events, or overlapping events. In a sequence, each element has a PART-OF relationship with the event class. Thus, on activation of the class, it is meaningless to activate all parts, as stated above, at the same time. Activation of the parts occurs only when their particular temporal specifications are satisfied. This form of search can take place only when temporal ranges are known. Arbitrary forms of temporal grouping can otherwise occur in a data-driven fashion, and specific groupings, if labeled by the creation of a class, can be recognized from them. (Causal or existential dependencies, a special case of temporal search, are not discussed here.)

The Search Cycle

The four search dimensions described must be coordinated in order to achieve the desired results, namely that each is used when appropriate.

Decisions on which dimension to use are governed by the following:

- There is a drive to instantiate the most specific classes for each data item at each time interval. Therefore, as soon as a successful match is obtained for a hypothesis, activate the next most specialized hypothesis for that data grouping.
- Instantiation implies a drive for completeness of description involving all components of a class description.
- There is a need to achieve instantiation in as few time samples as possible. Thus, as soon as an unsuccessful match is obtained, activate the relevant alternate hypotheses for that data grouping.
- Acquisition of new data items, regardless of whether they be acquired as the basic sampled data or as data found by specialized procedures initiated by hypotheses, necessitates data-driven search.
- Specific data items or groupings of data items must be considered individually with respect to the appropriate search scheme at specific points in time.
- Activation of a hypothesis by any method implies activation of all IS-A ancestors (perhaps several levels) and each direct PART-OF component (unless already active).

These basic rules ensure that the search schemes are indeed mutually complementary and are used only when appropriate.

Figure 5 presents the coordination of the different search modes within the processing cycle.

Hypothesis Structure and Its Properties

Hypotheses, like generic knowledge classes, are organized in specific ways. Connections among hypotheses are referred to as *conceptual adjacencies.* If a knowledge organization relation (IS-A, PART-OF, SIMILARITY, Temporal Precedence) exists between two classes and hypotheses are active for those two classes such that one hypothesis was activated by the other, then the hypotheses also have that same relation. The conceptual adjacency is one of the major components of hypothesis ranking, since it specifies what kinds of global and local consistencies play a role for a given hypothesis. In fact, the certainty updating scheme only uses information about conceptual adjacency and hypothesis matching. The set of conceptual adjacencies for a given hypothesis varies with time, as do its matching characteristics.

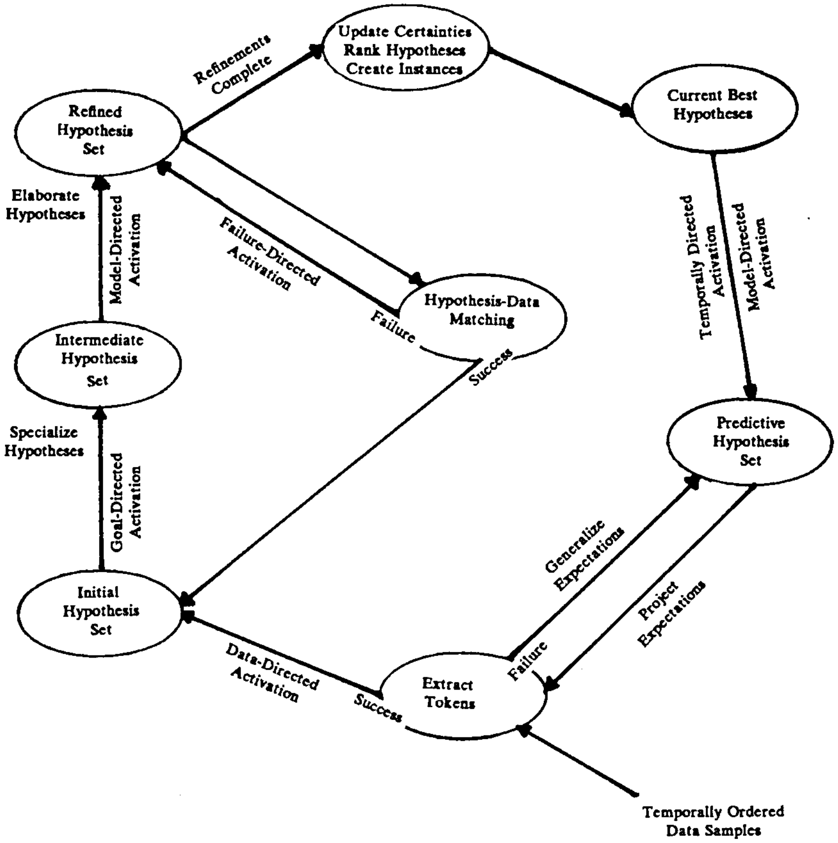


Figure 5
An attentive control strategy.

Temporal order satisfaction plays a role in the determination of component grouping strength, but is not the only actor. Each hypothesis has a self-contribution as well, which is based on the matching result of internal constraints.

Basically, hypotheses that are connected by conceptual adjacencies that imply consistency support one another, and those linked by adjacencies that imply inconsistency compete with one another by inducing inhibition. The IS-A relationship is in the former group; the SIMILARITY relationship is in the latter group. The *focus* of the system is defined as the set of hypotheses whose elements are the highest-ranked hypothesis at each level of specificity for each set of structural components being considered in the given time slice. Because of the slow change of certainties inherent in relaxation schemes, this focus exhibits inertia, or procrastination; i.e., it does not alter dramatically between certainty updates. It is a nonlinear scheme. Both global and local consistency are enforced through the contributions of hypotheses to one another via their conceptual adjacencies.

For the cooperative process, there are four major components that contribute to the certainty for each hypothesis: contributions from more general hypotheses along IS-A; contributions from competing hypotheses along SIMILARITY; contributions from component hypotheses along PART-OF; and the grouping strength of those components due to the satisfaction of temporal ordering considerations, spatial constraints or other component interrelationships, and matching results. Each of these is relevant only among active hypotheses.

Hypothesis Structure Consistency

It is important to define the notion of "consistency" in such a network of hypotheses. Within our framework, there are really three views of consistency that are considered:

global consistency (Are all the instances resulting from interpretation related to one another in reasonable ways?)

internal consistency (Does each instance have sufficient support from the data elements that directly constitute it?)

competition consistency (Do the competing hypotheses for each instance "believe" that the correct hypothesis was instantiated?)

These three considerations will be assumed to have equal importance.

Global consistency is required among all instances along their IS-A relationships. The semantics of IS-A imply strict inheritance, so that if a hypothesis is instantiated, then all its IS-A ancestors must also be instantiated. Thus, along the IS-A dimension, hypothesis responses are related by

$$R(h,t) \leq \min_{j \in N_{\text{IS-A}}(h,t)} R(j,t),$$

where $N_{\text{IS-A}}(h,t)$ is the set of IS-A ancestors for hypothesis h at time t, the hypotheses j are elements of that set, and $R(h,t)$ is the response of hypothesis h at time t.

Internal consistency, or sufficient internal support, is reflected by specific mechanisms within the updating rule. Unfortunately, no characterization is possible, and instances are created when their total support from all sources causes their certainty to achieve a threshold.

Competition consistency is defined along the SIMILARITY dimension. Each competing group has the same set of direct IS-A ancestors by definition (because otherwise the competition would not be meaningful). Moreover, the hypotheses in a competing group are mutually exclusive, so only one can be instantiated. Therefore, responses of a discriminatory group are related by

$$\sum_{h \in N_{\mathrm{SIM}}(*,t)} R(h,t) = \min_{j \in N_{\mathrm{IS\text{-}A}}(*,t)} R(j,t),$$

where $N_{\mathrm{SIM}}(*, t)$ is the set of competing hypotheses at time t and where $N_{\mathrm{IS\text{-}A}}(*, t)$ is the set of IS-A ancestors for that group. If the IS-A ancestors have been instantiated, then the right-hand side reduces to 1.0, which is the same as in standard relaxation. Consistency is enforced through the response-normalization process, and these relationships will appear in the updating rule to be presented below.

Compatibilities for the Cooperative Process

The conceptual adjacency relations manifest themselves as compatibilities in the updating of hypothesis certainty. In standard relaxation (Zucker et al. 1977), compatibility factors form the "model" (the view of consistency that the RLP has). In our scheme, each organizational relation has an associated compatibility. Each has a very intuitive meaning, and they are all listed below.

- self-compatibility: If a hypothesis succeeds—that is, if its internal constraints match the data successfully (the "glue" or "grouping strength" that binds its parts together)—it supports itself. If it fails, it inhibits itself. A failing hypothesis is not deleted from consideration on match failure alone.
- PART-OF compatibility: If a hypothesis has parts (features that can be observed), it receives a positive contribution from each part. A hypothesis cannot inhibit a part, because that part must be allowed to participate in other groupings in an unfettered manner.
- IS-A compatibility: Let hypothesis h_1 be IS-A related to h_2, i.e., h_1 IS-A h_2. When updating h_1, if both h_1 and h_2 match successfully, h_2 supports h_1. If h_1 succeeds and h_2 fails, h_2 inhibits h_1. Failure of h_1 has no effect on h_2.
- SIMILARITY compatibility: Hypotheses related by SIMILARITY are competitors—only one out of such a discriminatory set can be instantiated.

Let h_1 be related via a SIMILARITY link to h_2. When updating h_1, if h_2 fails, h_2 supports h_1, and if h_2 succeeds, h_2 inhibits h_1.

- temporal compatibility: Let the temporal relation between hypotheses h_1, h_2, and h_3 be

$h_1 \leftarrow$ previous $-$ $h_2 \leftarrow$ previous $-$ h_3.

Also let h_{sequence} be the hypothesis that represents this sequence, so that each of h_1, h_2, and h_3 is PART-OF h_{sequence}. When updating h_{sequence}, each of h_1, h_2, and h_3 would contribute to h_{sequence} when they appear through PART-OF. If no special mechanism were present for temporal order, the PART-OF contribution alone would not provide a discriminatory effect if the order were wrong. Therefore, there is a bonus contribution to h_{sequence} during the event h_2 due to the last nonzero response of h_1 if it appeared in the correct order, and likewise if h_3 follows h_2. This contribution decays exponentially; thus, the further in the past it happened, the smaller the force of the "glue" that groups these elements together. This will be termed *"previous" support*. Similarly, h_{sequence} receives a bonus inhibition due to h_3 that does not decay with time if h_3 appears before h_2 and this is the "next" inhibition. The same occurs if h_2 occurs before h_1. Clearly, "next" and "previous" are reciprocal relations except at the ends of the sequence.

Compatibilities are set to values between 1.0 and -1.0, where 1.0 means that the hypotheses are strongly compatible, 0.0 means that they are independent, and -1.0 means that they are strongly incompatible. They will appear in the form $1/k_{i,j}$ in the remainder of the discussion, where the absolute value $|k_{i,j}| \geq 1.0$, so that $-1.0 \leq 1/k_{i,j} \leq 1.0$. Here "$i$, j" corresponds to the type of compatibility between hypotheses i and j and can be of the following types:

self-compatibility: k_{self}

SIMILARITY compatibility: k_{SIM}

IS-A compatibility: $k_{\text{IS-A}}$

PART-OF compatibility: $k_{\text{PART-OF}}$

temporal compatibility: "previous" support is embodied in k_{prev}

temporal compatibility: "next" inhibition is given by k_{next}.

Each compatibility value is positive unless explicitly prefixed with a minus sign. A set of empirically derived inequalities that constrain the values of each of these compatibilities will be presented.

Initial Certainties of Hypotheses

The activation of hypotheses is the first step of processing. Once an initial set has been activated, matching is performed for each of those active hypotheses. Then, depending on those results, other hypotheses may be activated. On activation, each hypothesis receives an initial certainty, and this certainty is updated after all hypothesis activation and matching has been completed for that time interval.

Hypotheses are activated via the search mechanisms outlined above. Each hypothesis has an associated structure that conforms to the generic class to which it is related. In addition, each is assigned an initial certainty depending on how it was activated and ensuring that consistency relationships are maintained. Let the activated hypothesis be h, a single activating hypothesis be h_a, a set of simultaneously activating hypotheses be H_a, the activation time be t_0, and the hypothesis response be R. Initial certainties are assigned depending on the following activation types.

- Data-driven activation along PART-OF. The parts of a class, from which a hypothesis is derived, are represented by the set $N_{\text{PART-OF}}(h, *)$ and each of those parts can activate the hypothesis. (Of course, there can be only a single activator as well.) Thus, $H_a \subset N_{\text{PART-OF}}(h, *)$, and the initial certainty is given by

$$R(h, t_0) = \frac{\sum_{j \in H_a} R(j, t_0)}{|H_a|}.$$

- Hypothesis-driven activation along PART-OF: $R(h, t_0) = R(h_a, t_0)$.

- Hypothesis-driven activation along IS-A where several hypotheses in the set $H_a \subset N_{\text{IS-A}}(h, *)$ may participate in the activation: $R(h, t_0) = \min_{j \in H_a} R(j, t_0)$.

- Temporally driven activation. This is a special case of hypothesis-driven activation along PART-OF, and thus the initial certainty is computed in the same manner.

- Failure-driven activation along SIMILARITY: When a competitive set N_{SIM} first comes into being at time t_0, it does so by hypothesis-driven activation along IS-A, so that each hypothesis receives an equal share of the minimum certainty of their activating IS-A ancestors $H_a \subset N_{\text{IS-A}}(h, *)$:

$$\frac{\min_{j \in H_a} R(j, t_0)}{|N_{\text{SIM}}|}.$$

Suppose there is currently a set of competitors $N_{\mathrm{SIM}}(*, t_0 - 1)$ and K new hypotheses are added to that set at time t_0. The already existing hypotheses donate half of their response to a pool, and then that pool is shared equally over the new and old hypotheses. Thus, the initial certainty of each of those K new hypotheses is given by

$$\frac{\sum_{j \in N_{\mathrm{SIM}}(*, t_0 - 1)} R(j, t_0 - 1)/2}{|N_{\mathrm{SIM}}(*, t_0 - 1)| + K}$$

and each of the hypotheses h in $N_{\mathrm{SIM}}(*, t_0 - 1)$ has its certainty adjusted to

$$\frac{R(h, t_0 - 1)}{2} + \frac{\sum_{j \in N_{\mathrm{SIM}}(*, t_0 - 1)} R(j, t_0 - 1)/2}{|N_{\mathrm{SIM}}(*, t_0 - 1)| + K}.$$

These certainties are the adjusted ones before updating is done for time t_0. This sharing scheme maintains that the sum of the certainties in the competing set satisfies the definition of IS-A consistency. The design of the certainty sharing was motivated by the fact that the addition of a competing hypothesis must not undo the accumulated results of the activating hypothesis's matching history, i.e., hypothesis's matching inertia. It is an assignment that preserves hypothesis relative ranking.

Certainty Updating in Time

A variant of the temporal relaxation rule introduced in Tsotsos et al. 1980 will be used for hypothesis certainty updating. Basically, a neighborhood whose members change with time is responsible for the contribution part of the update. The hypothesis must reflect the same IS-A and PART-OF relationships with other hypotheses as does the generic class of which it may be an instance, with classes the generic class is related to. However, the temporal and SIMILARITY relationships of the hypothesis may only be a subset of those in the class. It will not, for example, always be the case that the discrimination will take place among all possible choices, nor will it always be the case that the correct temporal sequence of events will occur. Figure 6 shows a typical set of relationships among generic classes and hypotheses during the recognition process. This neighborhood, derived from the conceptual adjacencies described above, may be thought of as a "conceptual receptive field" for the hypothesis, because changes in response in any of the neighborhood members will result in changes in the hypothesis itself. Although the numbers of contributors may vary widely for given hypotheses, this variation has no adverse effect on the certainty

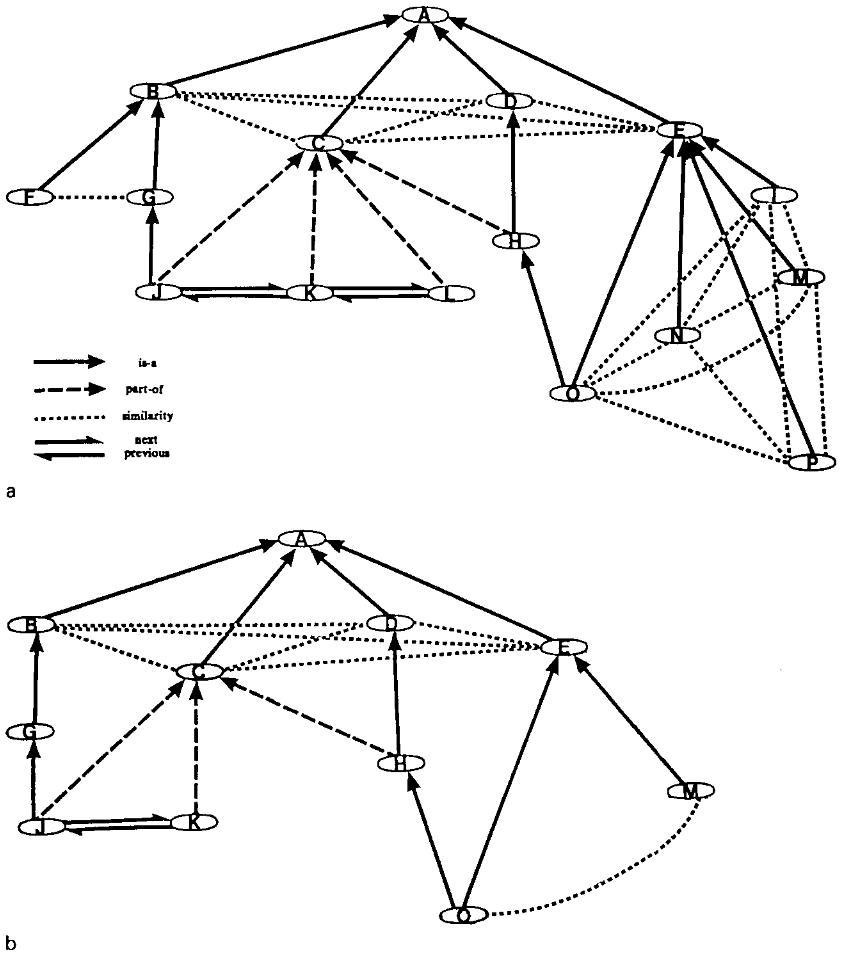


Figure 6
The relationship between generic knowledge (a) and a representative hypothesis structure (b) that may be created from it.

updating, since the magnitude of each contribution is weighted by the number of contributors where necessary.

The rule is now presented. The response of hypothesis h at time $t + 1$ is defined by

$$R(h, t + 1) = \frac{R(h, t)\left(\sum_{j \in N(h,t)} w(h, j, t) \frac{R(j, t)}{k_{h,j}(t)}\right) \min_{a \in N_{\text{IS-A}}(h,t)} R(a, t + 1)}{\sum_{m \in N_{\text{SIM}}(h,t)} \left(R(m, t) \sum_{n \in N(m,t)} w(m, n, t) \frac{R(n, t)}{k_{m,n}(t)}\right)},$$

where $R(h, t)$ is the hypothesis response (or certainty) at time t and is restricted to the range from 0.0 to 1.0, and where $N(h, t)$ is the set of all hypotheses that are neighbors to h at time t and is the union of the following five sets:

$N_{\text{SIM}}(h, t)$, the set of all hypotheses that are neighbors to h through a SIMILARITY connection at time t (including h),

$N_{\text{IS-A}}(h, t)$, the set of all hypotheses that are neighbors to h through an IS-A connection at time t,

$N_{\text{PART-OF}}(h, t)$, the set of all hypotheses that are neighbors to h through a PART-OF connection at time t,

$N_{\text{previous}}(h, t)$, the set of all hypotheses that are neighbors to h through the temporal sequence connection "previous," and

$N_{\text{next}}(h, t)$, the set of all hypotheses that are neighbors to h through the temporal sequence connection "next."

Also in the above equation, $w(h, j, t)$ is the weight of the contribution by hypothesis j to hypothesis h at time t in relation to the other contributions. The sum of the weights over the set $N_{\text{SIM}}(h, t) \cup N_{\text{IS-A}}(h, t) \cup N_{\text{previous}}(t) \cup N_{\text{next}}(t)$ must be 1.0, and the sum of the weights over the set $N_{\text{PART-OF}}(h, t)$ must be 1.0 for convergence purposes. Since the hypothesis structure varies with time, so clearly do the assignments of weights. Furthermore, $k_{h,j}(t)$ is the compatibility between hypotheses h and j at time t and is determined by the type of relationship between the two hypotheses. There are six types, as described above, and the hypothesis matching result at time t determines whether the value is positive or negative for certain ones.

The contribution portion of this rule, through judicious choices of weighting factors, is restricted to the range $0.0 \leq$ contribution ≤ 2.0, as it is in Zucker et al. 1977. However, normalization takes place only among hypotheses that are comparable, that is, elements of the same discriminatory set. It would be meaningless to try to normalize between levels of abstraction. Moreover, because of the definition of IS-A consistency, the sum of the responses over a discriminatory set is not normalized to 1.0 necessarily, but rather to the minimum updated response of the IS-A ancestors of the hypothesis, thus the last term of the numerator. This implies that the normalization process must be done in a strict order, from general hypotheses to specific ones along the IS-A relationship. If there is no discriminatory set the denominator is set to 1.0, and if there are no IS-A ancestors that term is also 1.0.

If we expand the contribution portion of the rule, the values of the weights will become apparent.

$$\sum_{j \in N(h,t)} w(h,j,t)\frac{R(j,t)}{k_{h,j}(t)}$$

expands out to the sum of the following terms:

- self-contribution: $w(h,h,t)\dfrac{R(h,t)}{k_{\text{self}}(t)}$.

- SIMILARITY contribution: $\displaystyle\sum_{j \in N_{\text{SIM}}(h,t), j \neq h} w(h,j,t)\frac{R(j,t)}{k_{\text{SIM}}(t)}$.

- IS-A contribution: $\displaystyle\sum_{j \in N_{\text{IS-A}}(h,t)} w(h,j,t)\frac{R(j,t)}{k_{\text{IS-A}}(t)}$.

- PART-OF contribution: $\displaystyle\sum_{j \in N_{\text{PART-OF}}(h,t)} w(h,j,t)\frac{WR(j,t)}{k_{\text{PART-OF}}(t)}$,

where $WR(j,t)$ is the weighted response of the PART-OF subtree rooted at hypothesis j.

- previous contribution: $\displaystyle\sum_{j \in N_{\text{previous}}(h,t)} w(h,j,t)R(j,t_{\text{last}})e^{(-t+t_{\text{last}})/k_{\text{prev}}}$.

- next contribution: $\displaystyle\sum_{j \in N_{\text{next}}(h,t)} w(h,j,t)\frac{R(j,t_{\text{last}})}{k_{\text{next}}}$.

The following are the weight assignments.

- for PART-OF contributors: $w(h,j,t) = \dfrac{1}{|N_{\text{PART-OF}}(h,t)|}$,

so that $\sum_{j \in N_{\text{PART-OF}}(h,t)} w(h,j,t) = 1.0$. Of course, the number of PART-OF contributors varies with time.

- The sum of the weights of all the remaining contributors, including the self contribution, must be 1.0. The weights, therefore, are not fixed but vary with time depending on which contributors are present. The remaining major contributions reflect the hypothesis's internal consistency, its global consistency, and its consistency as viewed by its competitors. Each of these three contributions is weighted equally. In the case that all are present, the following weights are assigned: For IS-A contributors, $w(h, j, t) = 1/3|N_{\text{IS-A}}(h, t)|$. (The number of IS-A contributors changes with time.) For SIMILARITY contributors, $w(h, j, t) = \frac{1}{3}$. (The weighted sum is not required here, since the sum of responses may be at most 1.0, and this is enforced through normalization.) For grouping strength, self-contribution, previous support, and next inhibition are equally weighted, and when all are present $w(h, j, t) = \frac{1}{9}$ so that the sum of the weights is $\frac{1}{3}$. If only the self-contribution is present, then it is weighted by the entire $\frac{1}{3}$ amount. If the self-contribution and (say) the next inhibition are present, each is weighted by $\frac{1}{6}$.

The PART-OF contribution is always present and is positive. The weights on any other contributions may vary, but the sum of those weights must always be 1.0.

Result Propagation through the Network

Since iterations are related to time samples, it is important to address the problem of result propagation through the network of hypotheses. In standard RLPs, results propagate to neighboring processes, as a result of several iterations, and the field of influence of a given process is determined directly by the number of iterations; the greater the number of iterations, the larger the field. In our case, all results must propagate to any other processes that may need them during a small fixed number of iterations, because for the next iteration new data will be presented to the system. This does not mean that a consistent solution is also found within that same small number of iterations; iterations are required for the temporal integration of results.

The results that must be communicated are of two types: changes in certainty and changes in hypothesis matching state. The hypotheses are organized in the same fashion as the generic knowledge classes from which they are derived, namely along the IS-A, PART-OF, SIMILARITY, and Temporal Precedence dimensions. The first three of these (Temporal Precedence is a special case of PART-OF) enable results to propagate as desired. SIMILARITY networks pose no problem, since each active hypothesis is directly connected to each other active hypothesis.

PART-OF also poses no problem, because the PART-OF contribution is computed as a weighted sum of the entire subtree rooted at the contributing hypothesis. This, in addition, imposes an ordering on the computation of updated certainty values.

The role of the IS-A dimension requires more elaboration. Because of inheritance, match results are conveyed down the IS-A hierarchy implicitly. There is no upward communication along this dimension. Because of the strict order of normalization (downward along IS-A), changes in an IS-A ancestor's certainty will result in changes in the normalization factor for the IS-A child. Increases in ancestor certainty will allow the child's certainty to increases, and vice versa. Communication along all relevant branches of the hypothesis network, both of match results and of certainty changes, is thus guaranteed by the nature of the knowledge organization and definitions of network consistency.

Nonlinearity and Feedback

A linear system has response that is computed according to current input and current state, independent of the rate at which response is changing. The updating rule presented above is nonlinear. Because of the multiplicative nature of the rule's numerator, the change in response is greater with increasing previous response, and with increasing IS-A parent response. The question that must be asked here is: "Are the nonlinearities inherent in the theory, or do they appear as a result of the particular realization of the theory?" This will be only partially addressed here. However, owing to other aspects of the theory, the effect of the nonlinearities is minimal.

One important aspect of the updating rule is the normalization of response. This is necessary because of the definitions of network consistency, and is thus an inherent nonlinearity of the theoretical foundations of this framework. Moreover, from a practical point of view, normalization is necessary in order to ensure that responses do not grow unchecked, something that would occur even with a linear rule. A second source of nonlinearity is due to the first term of the numerator, the multiplication by the response of the hypothesis under consideration. This is present in the original RLP model of Zucker et al. 1977, and is a carryover from there. It is an implementation-dependent nonlinearity, and no attempts were made to reformulate this in a linear manner.

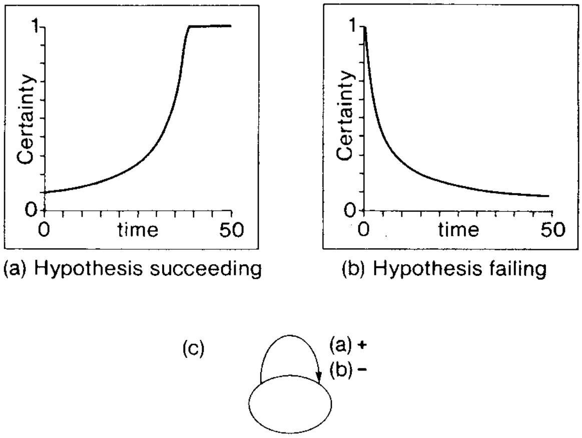
Nonlinear, time-varying systems in open loop configuration are rather difficult to characterize and control. The incorporation of feedback, on the other hand, has definite advantages. Feedback reduces sensitivity of response to parameter variations (in this case, compatibility values), reduces the effect of noise disturbances, and makes the response of nonlinear elements more linear.

The essential property of feedback is its iterative comparison of the current state with the desired state in such a way that results of comparison can be used to correct the system toward the desired state (Zucker 1978b). The desired state in a recognition task is unknown, yet global consistency is sufficient to ensure that a correct interpretation can be obtained. Feedback is inherent in this framework in two ways: exception recording and handling via the similarity mechanism, and levels of IS-A abstraction and downward communication between them. It is evident from the experimental results that the resulting scheme is well behaved, yet analytic proof is elusive.

Performance of the Temporal Cooperative Process

Let us take a simple situation and see how this updating rule performs. Figure 7 presents a single hypothesis, totally disconnected from any other hypotheses, whose contributions come from its own matching success or failure and from its PART-OF elements.

The PART-OF contribution, for purposes of the first few examples, is 1.0. There is no need for normalization, since there is no discriminatory set.



(a) Hypothesis succeeding

(b) Hypothesis failing

Figure 7
Certainty changes with time for a single hypothesis with self-contribution and PART-OF contributions.

On the other hand, this forces the need for a careful analysis. Normalization forces all responses to behave, that is, to stay between 0.0 and 1.0. Without it, response may be unbounded. Figure 7c shows the hypothesis configuration, while figures 7a and 7b show the certainty over time if the hypothesis always succeeds or always fails in matching, respectively. The updating rule in this case reduces to

$$R(1, t) = R(1, t - 1) + \frac{R(1, t - 1)^2}{k_{\text{self}}}$$

for figure 7a and to

$$R(1, t) = R(1, t - 1) - \frac{R(1, t - 1)^2}{k_{\text{self}}}$$

for figure 7b. Approximating each of these with an ordinary differential equation and solving, the response functions become, respectively,

$$R(1, t) = \frac{R(1, 0)}{1 - R(1, 0)*t/k_{\text{self}}}$$

and

$$R(1, t) = \frac{R(1, 0)}{1 + R(1, 0)*t/k_{\text{self}}}.$$

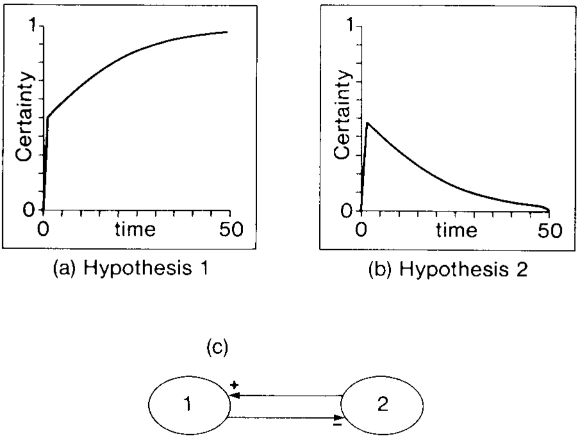
In the constant failure case, the response smoothly tends toward 0 with increasing time and the speed of decrease can be controlled by adjusting k_{self}. For the constant success case, as is intuitively clear, there is no such nice property. Indeed, the response will achieve 1.0 at time

$$\frac{k_{\text{self}}*(1 - R(1, 0))}{R(1, 0)}$$

and keep on increasing. Clearly, the smaller k_{self} is, the faster this occurs. We will want decisions on interpretation to occur as close to the end time of an event as possible, but most definitely not after the event has ended. Therefore,

$$k_{\text{self}} \leq \text{minimum event duration}* \frac{R(1, 0)}{1 - R(1, 0)},$$

where the minimum event duration has units of "temporal measurements." A temporal measurement is taken "between" two data samples. For simplicity we will ignore the final term due to initial hypothesis response.



(a) Hypothesis 1

(b) Hypothesis 2

(c)

Figure 8
Certainty changes with time for a pair of competing hypotheses with SIMILARITY and PART-OF contributions.

It will be apparent that this does not ill affect the remainder of the analysis, since a structure with an isolated hypothesis can never occur. We will require that the number of iterations (or the number of temporal measurements) required in order to "recognize" a particular concept always be less than or equal to k_{self}. In other words, we equate iterations with temporal measurements.

Hypotheses are never alone; they have neighbors. The first neighbor that will be investigated is the one connected via SIMILARITY. Figure 8 presents hypotheses where both PART-OF and SIMILARITY contributions are present, but not self-contributions. Hypothesis 1 always succeeds; hypothesis 2 always fails. The updating rule for this situation reduces to the pair

$$R(1,t) = R(1,t-1) + \frac{R(1,t-1)*R(2,t-1)}{k_{\text{sim}}},$$

$$R(2,t) = R(2,t-1) - \frac{R(2,t-1)*R(1,t-1)}{k_{\text{sim}}}.$$

In this case the normalization factor has a value of 1.0. The initial values $R(1,0)$ and $R(2,0)$ sum to 1.0, and it is clear from the updating rule that the denominator of the updating rule will always be $R(1,t-1) + R(2,t-1)$, or 1.0. By the same approach as in the preceding example, the response curves can be approximated by

$$R(1, t) = \frac{1}{\dfrac{1 - R(1, 0)}{R(1, 0)} * e^{-t/k_{\text{SIM}}} + 1},$$

$$R(2, t) = \frac{1}{\dfrac{1 - R(2, 0)}{R(2, 0)} * e^{t/k_{\text{SIM}}} + 1}.$$

In both cases, desirable properties are exhibited. The response for hypothesis 1 tends to 1 with increasing time, and the rate can be controlled by adjusting k_{SIM}. The response for hypothesis 2 tends to 0 with increasing time, and again the rate is set by k_{SIM}.

Let us now join these two examples and investigate the situation where there is both self-contribution and SIMILARITY contribution. This is portrayed in figure 9.

Two hypotheses are present: that hypothesis 1 always succeeds and that hypothesis 2 always fails. The updating rule for this situation is

$$R(1, t) = \frac{R(1, t-1) + \dfrac{R(1, t-1)^2}{2k_{\text{self}}} + \dfrac{R(1, t-1)*R(2, t-1)}{2k_{\text{SIM}}}}{\text{NORM}},$$

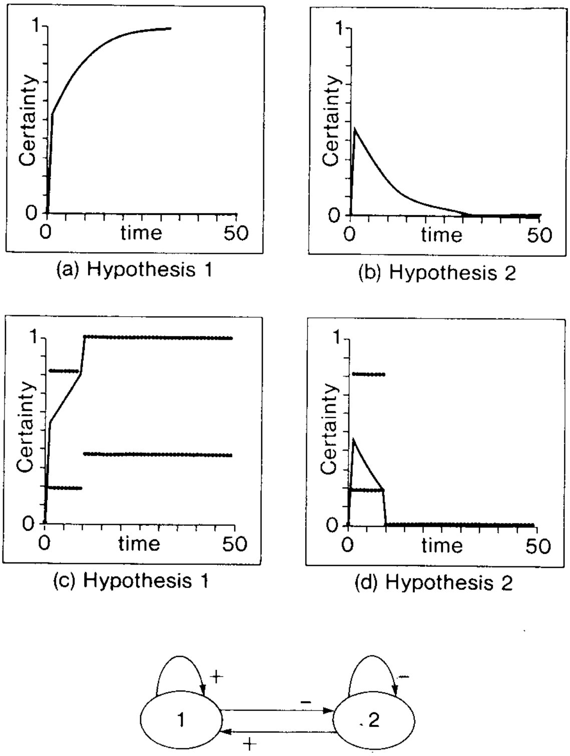
$$R(2, t) = \frac{R(2, t-1) - \dfrac{R(2, t-1)^2}{2k_{\text{self}}} - \dfrac{R(2, t-1)*R(1, t-1)}{2k_{\text{SIM}}}}{\text{NORM}},$$

where

$$\text{NORM} = 1 + \frac{R(1, t-1)^2}{2k_{\text{self}}} - \frac{R(2, t-1)^2}{2k_{\text{self}}}.$$

This is not easy to deal with analytically, and thus an approximation will not be presented. However, figures 9a and 9b show the typical response profiles that such a configuration produces. It appears as if these results are well behaved, and indeed through experimentation it was found that this is the case.

All response computations are subject to lower and upper thresholds for hypothesis deletion and instantiation, respectively. The curves shown in the preceding figures have a passing resemblance to exponential functions. A common notion for exponentials is the time constant, or the amount of time required for the function to reach $1/e$ of its initial value if decreasing



(a) Hypothesis 1

(b) Hypothesis 2

(c) Hypothesis 1

(d) Hypothesis 2

Figure 9

Two competing hypotheses with SIMILARITY, PART-OF, and self-contributions. Diagrams c and d show the effect of applying the dynamic threshold mechanism for hypothesis instantiation and deletion. Upper (instantiation) and lower (deletion) thresholds are shown as dotted lines. The change in threshold values with time is apparent.

or $(1 - 1/e)$ of its final value from an initial value of 0.0 if increasing. Using the initial values of hypotheses in competing sets defined earlier to compute the height of the exponential, we set the instantiation threshold at

$$\left(1 - \frac{1}{|N_{\mathrm{SIM}}|}\right)*(1 - e^{-1}) + \frac{1}{|N_{\mathrm{SIM}}|}.$$

$|N_{\mathrm{SIM}}|$ is the number of competing hypotheses in the discriminatory set, and this relationship is derived under the assumption that all hypotheses start at equal certainties (which is the case when they are all activated at the same time.) The corresponding deletion threshold is set at

$$\frac{e^{-1}}{|N_{\mathrm{SIM}}|}.$$

The effects of applying these thresholds to the previous case are shown in figures 9c and 9d. The amount of time required to reach the instantiation threshold is the response time of the system. With contributions other than the self-contributions, this response time is significantly shorter than with self-contribution alone. For the remainder of this discussion, a given hypothesis will have achieved *convergence* (or the reaching of a decision) when the hypothesis's response achieves the instantiation threshold. It will have achieved *useful convergence* if the number of iterations (or time samples) is less than or equal to the minimum duration of the event represented by that hypothesis.

Using the above definition of convergence, I now elaborate on the experimentation performed for the current configuration. Figure 10 (a solid family of curves) shows the empirical relationships found between k_{self} and k_{SIM} for varying values of $|N_{\mathrm{SIM}}|$. It was found that in order to ensure the required behavior, namely convergence in no more than k_{self} iterations, k_{self} must be greater than or equal to k_{SIM}. More precisely, the following was found to be the relation among k_{self}, k_{SIM}, and the number of competitors in the set N_{SIM}:

$$k_{\mathrm{SIM}} = 1.0 \quad \text{for } 2|N_{\mathrm{SIM}}| \le k_{\mathrm{self}} < \frac{7|N_{\mathrm{SIM}}|}{3},$$

$$k_{\mathrm{SIM}} \le \frac{3(k_{\mathrm{self}} - |N_{\mathrm{SIM}}|)}{4|N_{\mathrm{SIM}}|} \quad \text{for } k_{\mathrm{self}} \ge \frac{7|N_{\mathrm{SIM}}|}{3}.$$

These relationships are conservative approximations to the family of curves presented in figure 10. Spot checks for values of $|N_{\mathrm{SIM}}|$ not on this graph were tested with satisfactory results. There is no convergence within k_{self}

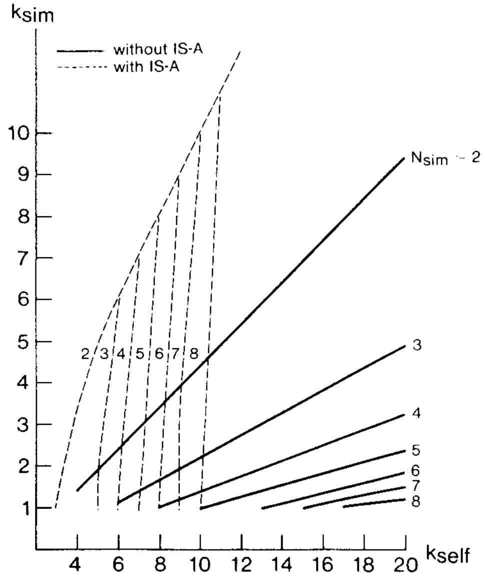


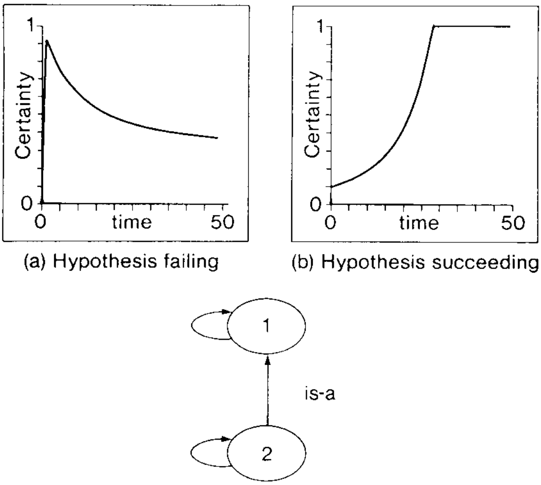
Figure 10
The empirical relationship between the self-contribution compatibility k_{self} and the SIMILARITY compatibility k_{SIM} for varying numbers of competing hypotheses. The solid curves show the relationship without the IS-A compatibility; the dotted curves show the effect with IS-A.

time units (iterations) for values of $k_{\text{self}} < 2|N_{\text{SIM}}|$. The fastest convergence occurs when $k_{\text{SIM}} = 1.0$ (approximately $2|N_{\text{SIM}}|$ time units).

The next compatibility type we will discuss in IS-A compatibility. A hypothesis configuration with IS-A comtribution is shown in figure 11. IS-A contributions may be considered as a "high-level bias" mechanism. If the biased hypothesis is succeeding, it should speed up its increase in response. Indeed, for an IS-A hypothesis with response 1.0 the updating rule reduces to

$$R(1,t) = R(1,t-1) + \frac{R(1,t-1)^2}{2k_{\text{self}}} + \frac{R(1,t-1)*R(2,t-1)}{2k_{\text{IS-A}}},$$

where hypothesis 1 IS-A hypothesis 2. The normalization factor is 1 since there are no competitors, and it is assumed that the IS-A parent hypothesis has response 1.0. This will cause the response of hypothesis 1 to increase faster than the hypothesis with self-contribution alone, and the degree of speedup can be controlled by $k_{\text{IS-A}}$. If the hypothesis fails, there is no IS-A contribution. Thus, failing hypotheses are allowed to decay on the basis



(a) Hypothesis failing

(b) Hypothesis succeeding

Figure 11
Certainty changes with time for two hypotheses related by IS-A with IS-A, PART-OF, and self-contributions present.

of their self-contribution, while succeeding hypotheses are given an extra boost.

An important question, however, is: How much faster is the increase in response? In the case where self, SIMILARITY, and IS-A contributions are present, experiments reveal the results shown in figure 10 (the dashed family of curves) for the empirical relationship between k_{self} and k_{SIM} for varying values of $|N_{\text{SIM}}|$ and where the largest value of $k_{\text{IS-A}}$ that allowed useful convergence was used. This largest value and its relationship with k_{self} are shown in figure 12. Without IS-A useful convergence could not be achieved for $k_{\text{self}} < 2|N_{\text{SIM}}|$, but with IS-A it is possible for $k_{\text{self}} \geq |N_{\text{SIM}}| + 2$ with $k_{\text{SIM}} = 1.0$ and $k_{\text{IS-A}} = 1.0$ in time approximately $|N_{\text{SIM}}| + 1$. For $k_{\text{self}} \geq |N_{\text{SIM}}| + 3$, we have $k_{\text{SIM}} \leq k_{\text{self}}$ and $k_{\text{IS-A}} \leq k_{\text{self}}/(|N_{\text{SIM}}| + 1)$. (This is a very conservative estimate when $|N_{\text{SIM}}| = 2$.) Of course, $k_{\text{IS-A}} = 1.0$ will ensure the fastest decision.

Two final considerations will complete the characterization of the above three types of compatibilities. For the case where h_1 IS-A h_2, the values of the compatibilities should ensure that if h_1 succeeds and h_2 fails then the response of h_1 decays. This will be true if

$$\frac{R(h_1, t)^2}{k_{\text{self}}} - \frac{R(h_1, t)R(h_2, t)}{k_{\text{IS-A}}} < 0.$$

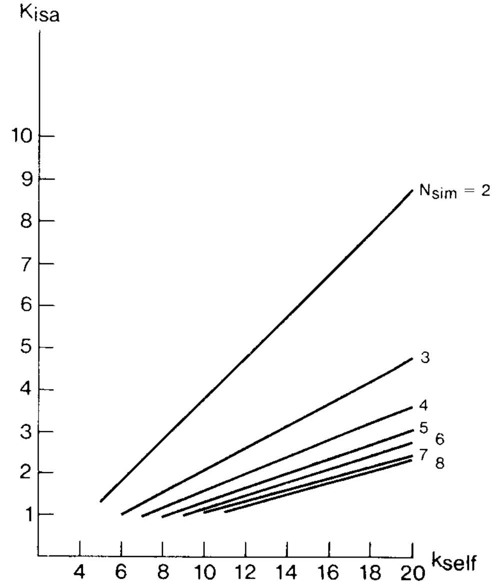


Figure 12
The empirical relationship between the largest value of the IS-A compatibility that enables useful convergence and the corresponding values of the self-contribution compatibility for varying numbers of competing hypotheses.

Since by IS-A consistency $R(h_2, t) \geq R(h_1, t)$, necessarily $k_{\text{self}} > k_{\text{IS-A}}$. If a third hypothesis h_3 is added such that h_3 IS-A h_2 and h_1 SIMILARITY h_3, and if h_1 and h_2 succeed and h_3 fails, we want to ensure that h_3 decays more rapidly than without the IS-A relation. This will be the case if

$$\frac{-R(h_3, t)}{k_{\text{SIM}}} + \frac{2R(h_2, t)}{k_{\text{IS-A}}} > \frac{R(h_1, t)}{k_{\text{self}}}.$$

This comes from the differences in the denominators of the rule between the two cases. Since $R(h_2, t) \geq R(h_3, t)$, and $k_{\text{self}} > k_{\text{SIM}}$ as described previously, then necessarily $k_{\text{SIM}} > k_{\text{IS-A}}$. The result is that $k_{\text{self}} > k_{\text{SIM}} > k_{\text{IS-A}}$, with specific values being set using the relationships derived experimentally.

For the previous discussion, it was assumed that the PART-OF contribution was always 1.0. In general, this is clearly not the case. Experiments were conducted on the full range of acceptable values for the compatibilities discussed above, for sets of competitors that had all manner of varying PART-OF contributions, and with $k_{\text{PART-OF}} = 1.0$. The longer the event duration (that is, the larger the value of k_{self}), the smaller the difference of

PART-OF contribution that could be present if a proper decision was to be made. (Roughly speaking, for short events the difference required was about 50 percent; for longer events it could be as little as 20 percent.) Large values of k_{self} coupled with relatively small values of k_{SIM} and $k_{\text{IS-A}}$ were preferred so that the the number of iterations (time samples) would be large enough to accumulate information over the entire duration of the event. No additional restrictions on the values of the other compatibilities were necessary, and the ranges for convergence defined above still held. Therefore, the final ordering of compatibility values was

$$k_{\text{self}} > k_{\text{SIM}} > k_{\text{IS-A}} \geq k_{\text{PART-OF}} = 1.0,$$

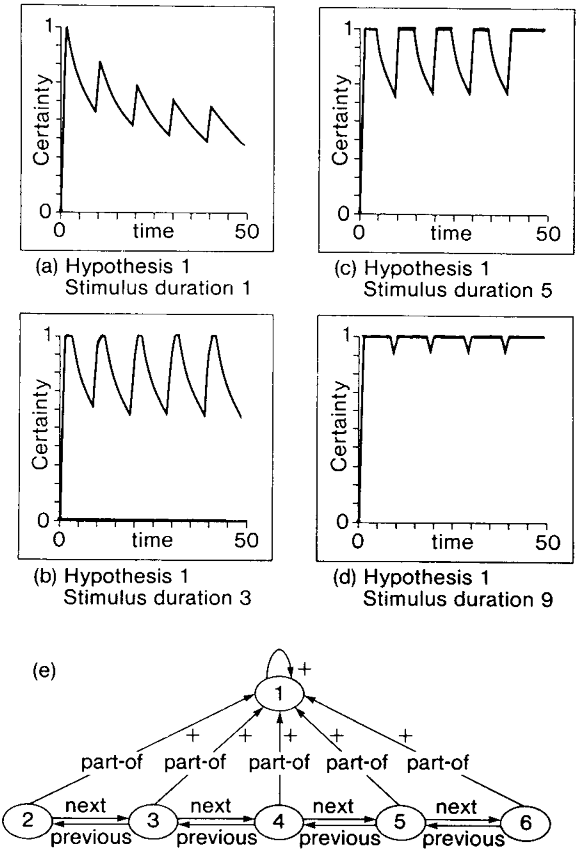
with further restrictions on k_{self} and $k_{\text{IS-A}}$ as defined above.

The next example is a simple sequence group that embodies the PART-OF contribution as well as the support and inhibition available for temporal grouping (figure 13). The hypothesis representing the sequence is labeled 1. The stimuli, numbered 2 through 6, have a high response of a given duration beginning at a specific time and a zero response elsewhere. The response of group 1 decays rather quickly but recovers when a new stimulus of the sequence appears. In figure 13a the stimuli are on for one time unit and off for nine. During those nine time units, the sequence hypothesis self-inhibits because there are no component parts to support it and cause it to match successfully. The sequence in figure 13a is never very sure of itself. In figure 13b, each stimulus is on for three time units and off for seven. In this case, there is some consistency to the response, and as the stimulus-on period increases to five in figure 13c and nine in figure 13d the sequence becomes more consistent with time.

There are three mechanisms at work in this example: self-contribution, PART-OF contribution, and PREVIOUS contribution. In this example, the stimuli were moving in the correct direction. A PREVIOUS contribution (always positive) appears only if the hypothesis IS-A SEQUENCE. NEXT contributions are always negative.

The value of the PREVIOUS compatibility is set depending on the temporal separation of events for which it is desired to cause a strong "gluing" effect. This can be accomplished if k_{prev} = maximum temporal separation in time units (or iterations). Therefore, for temporal separations between events in a sequence less than this value the compatibility will be greater than $1/e$; if the separation is greater, the compatibility will be less than $1/e$.

The situation when the correct direction is observed causes faster response time for the sequence hypothesis due to the PREVIOUS contribution than



(a) Hypothesis 1
Stimulus duration 1

(c) Hypothesis 1
Stimulus duration 5

(b) Hypothesis 1
Stimulus duration 3

(d) Hypothesis 1
Stimulus duration 9

Figure 13

Certainty changes with time for a simple sequence grouping of hypotheses, with varying durations for the hypotheses 2, 3, 4, 5, and 6. PART-OF, PREVIOUS, and self-contributions play a role.

without it. However, in cases where the data are not in the correct order the setting of compatibilities requires more care. The contributions from previous or next events are taken into account each time the contribution from any single event is computed. Only the previous or next events of that single event are considered. The certainty of the previous or next contribution is determined using its most recent certainty value, denoted by t_{last} below. In the case where both the previous elements and next elements are present, there are two competing contributions, weighted equally, to the updating of the sequence hypothesis (that are different from previous situations):

$$R(h_2, t_{\text{last}}) * e^{(-t+t_{\text{last}})/k_{\text{prev}}} - \frac{R(h_4, t_{\text{last}})}{k_{\text{next}}}.$$

The sequence constraints of the hypothesis are violated, so the hypothesis itself is self-inhibiting. It is important that the positive term due to the previous event not outweigh the negative term due to the next event. Indeed, the desired result is that the decay is accelerated because of the next effect, and this would be ensured if the negative term were larger in magnitude than the positive term of the above contributions. Therefore, if both next and previous events were of the same strength, and in the worst case where t_{last} is just the previous time interval rather than farther back in time, the following relation should hold in order to ensure this:

$$k_{\text{next}} < e^{1/k_{\text{prev}}}.$$

The situations presented above make it possible to derive empirical constraints on the values of the compatibilities and to relate those values to characteristics of the temporal domain being considered.

A Brief Look at Some More Complex Examples

The remaining several examples are too complex to permit any specific quantitative analysis. They do, however, represent common situations in low-level and high-level vision. They are presented to demonstrate that the machinery presented does indeed function in complex situations as long as the guidelines for the setting of compatibility constants are obeyed. Figure 14 illustrates a rather common situation, that of orientation selection, with static data. Assume that all eight orientations are considered here, even though pictorially is it difficult to portray both orientations along a single line.

The stimulus array is 5 × 5; the darker the element, the higher the

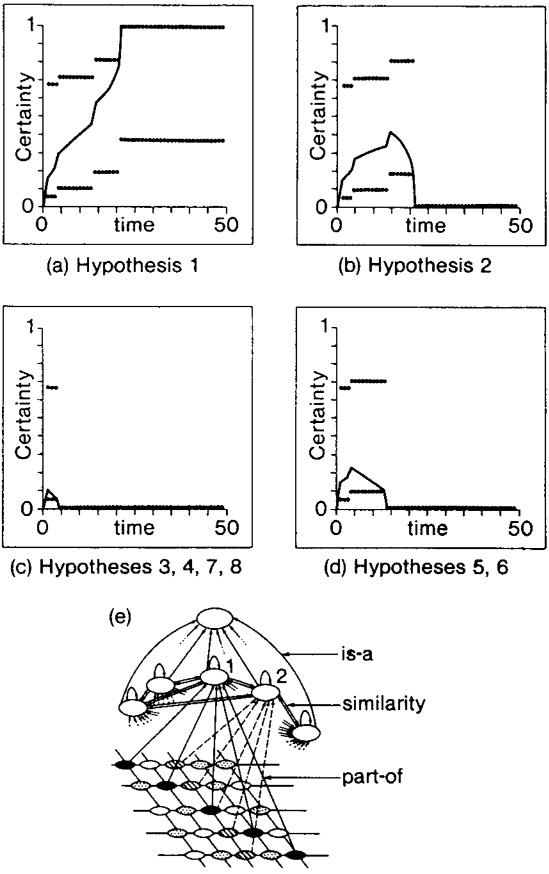


Figure 14

An orientation-selection hypothesis structure, with stimuli on a grid, of varying strengths (coded by shading), and the resulting certainty-vs.-time profiles for each of the eight orientation hypotheses. As the hypothesis number increases, the orientation it represents rotates 45° clockwise, with the orientations for hypotheses 1 and 2 shown.

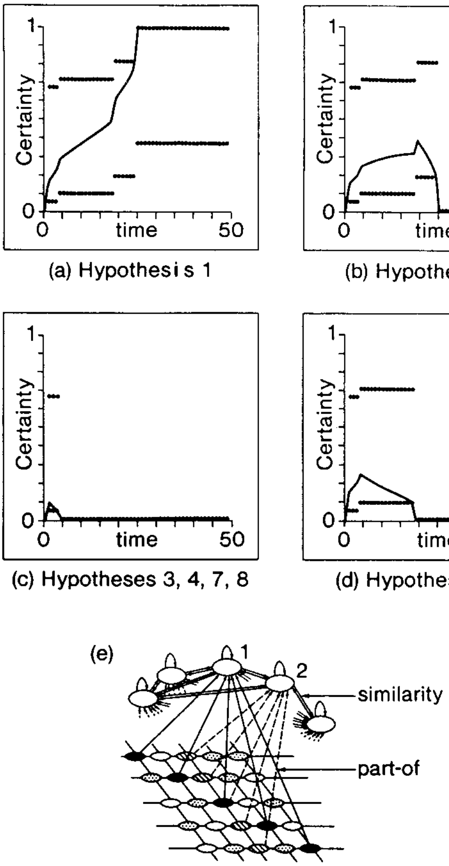
response value. (The black ones have response 1.0, the striped ones 0.5, the white ones 0.0, and the remainder 0.2.) Eight hypotheses are in competition here, with hypothesis 1 representing the dark elements (which form a consistent orientation grouping going from the top left of the array to the bottom right) and with hypothesis 2 representing the striped elements (which form a consistent orientation grouping from the top to the bottom of the array). The 180° opposite orientation to these, we assume, is inconsistent. Both hypotheses 1 and 2 succeed in their matching, since they are looking at elements that semantically form an orientation group. The other hypotheses fail. The result for hypothesis 1 is shown in figure 14a, and that for hypothesis 2 in figure 14b. The results for hypotheses 3, 4, 7, and 8 are the same and are shown in figure 14c. The results for hypotheses 5 and 6, which also are the same, are shown in figure 14d. The clear winner is hypothesis 1 after about 25 time units (iterations). The other hypotheses are never even close to the instantiation threshold. The threshold dependency on the number of competitors is very clear. It jumps when hypotheses 3, 4, 7 and 8 are deleted, when hypotheses 5 and 6 are deleted, and again when hypothesis 2 is deleted.

Figure 15 shows exactly the same setup, except that there is an IS-A constraint that affects each of the eight competing orientation hypotheses.

Basically, it is a hypothesis that biases the situation—it communicates to each hypothesis that there is indeed an orientation element in the stimulus array. It does not identify which one, however. The successful hypotheses accept this bias, while the failing ones reject it. The effect is that the only change from the previous case is that the instantiation of the correct hypothesis occurs sooner, by about five time units. Response time is decreased by the addition of top-down biases.

Using the same stimulus array but adding the time dimension, figure 16 portrays a direction-selection experiment.

Each stimulus is on for eight time units and off for two. The correct ordering for hypothesis 1 is from the top left corner to the bottom right corner. For hypothesis 2, each of the stimuli has the same characteristics as for hypothesis 1 except for strength, and the correct order is from top to bottom. In fact, all hypotheses share an event: the middle one. Hypotheses 1 and 2 succeed in matching; the remainder do not. Each stimulus, when it becomes active, activates of reactivates each higher-order hypothesis that it is PART-OF. The life span of hypothesis 5, whose elements are as strong as its correctly matching counterpart 1 but whose ordering is wrong, is rather short. In fact, as soon as the inhibition due to the reversal in order comes into play at the start of the second stimulus, it is quickly removed.



(a) Hypothesi s 1

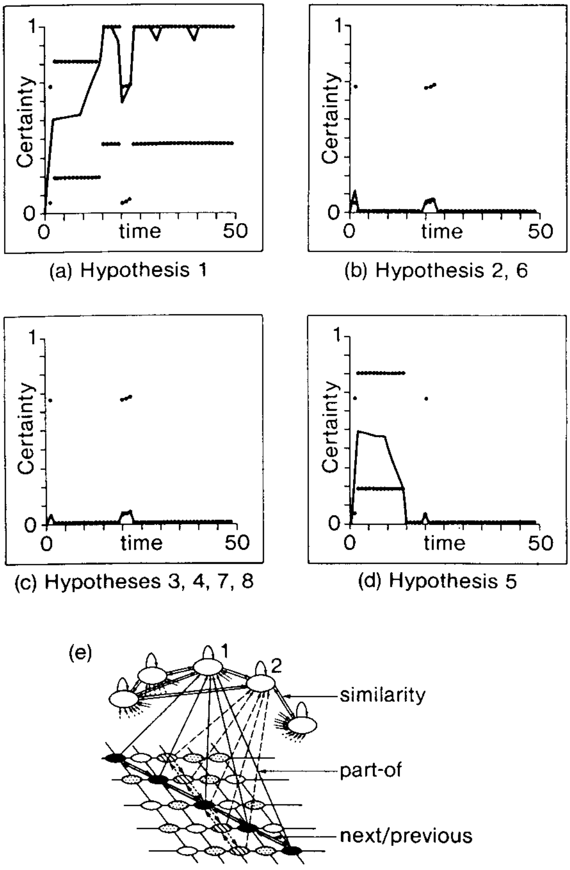
(b) Hypothesis 2

(c) Hypotheses 3, 4, 7, 8

(d) Hypotheses 5, 6

Figure 15

An orientation-selection hypothesis structure, with stimuli on a grid, of varying strengths (coded by shading), and the resulting certainty-vs.-time profiles for each of the eight orientation hypotheses. As the hypothesis number increases, the orientation it represents rotates 45° clockwise, with the orientations for hypotheses 1 and 2 shown. This structure has an IS-A relationship for each of the orientation hypotheses that acts as a top-down bias. The resulting speedup in time to convergence can be seen.



(a) Hypothesis 1

(b) Hypothesis 2, 6

(c) Hypotheses 3, 4, 7, 8

(d) Hypothesis 5

Figure 16

A direction-selection hypothesis structure, with stimuli on a grid, of varying strengths (coded by shading), and the resulting certainty-vs.-time profiles for each of the eight direction hypotheses. As the hypothesis number increases, the direction it represents rotates 45° clockwise, with the directions for hypotheses 1 and 2 shown. For example, hypothesis 5 represents the exact opposite direction to hypothesis 1. The temporal precedence relations among the grid stimuli are shown. Each stimulus is allowed to (re-)activate its PART-OF parent hypothesis.

Hypothesis 1 is again the clear winner. During the stimulus-off period, however, there is self-inhibition. The total decrease in response is, of course, a function of the duration of the off time as well as the decay constant.

Figure 17 shows the results of the same direction-selection task, with an added IS-A constraint. The effect of the IS-A constraint should be the same as was shown for the orientation-selection case. Indeed, the slope of initial increase of response for hypothesis 1 is slightly higher, but the time for instantiation could not be shorter since the second stimulus must appear before discrimination can occur. The decays are compensated slightly. Figure 18 shows the results from the same structure, but with a different definition of SEQUENCE; only the first stimulus of a sequence can activate the higher-order SEQUENCE unit. This is included for purposes of comparison with the previous example. The definition of SEQUENCE used here leads to a more stable response curve for hypothesis 1. It is clear that one may experiment with a variety of schemes. With these more complex examples, in each case the structure over which the cooperative process operates changes with time; however, even more complex situations arise in a real application.

To summarize: The temporal cooperative process displays qualitative results in both static and dynamic situations that are both desirable and consistent. Guidelines were presented for the setting of compatibility values that would achieve such results. The process is remarkably insensitive to actual compatibility values so long as the guidelines are obeyed. Since iterations are considered in time and have the same meaning in either static or dynamic situations, we can begin to relate system response time in static cases using the same terms as for dynamic cases.

Noise Effects

Up to this point, the effect of noisy data has not been addressed. Experiments conducted with randomly generated, normally distributed, noisy matching data are described in detail in Tsotsos 1981b. Basically, a set of competing hypotheses are created, with no semantics, and a match result over time is generated for each. For example, for two competing hypotheses where one is correct, the other is false, and the data are perfect, true matching data would be generated for the former and false data for the latter. With the addition of 10 percent noise, the matching data for the correct hypothesis would be true only for 90 percent of the samples generated; they would be false for only 90 percent of the samples

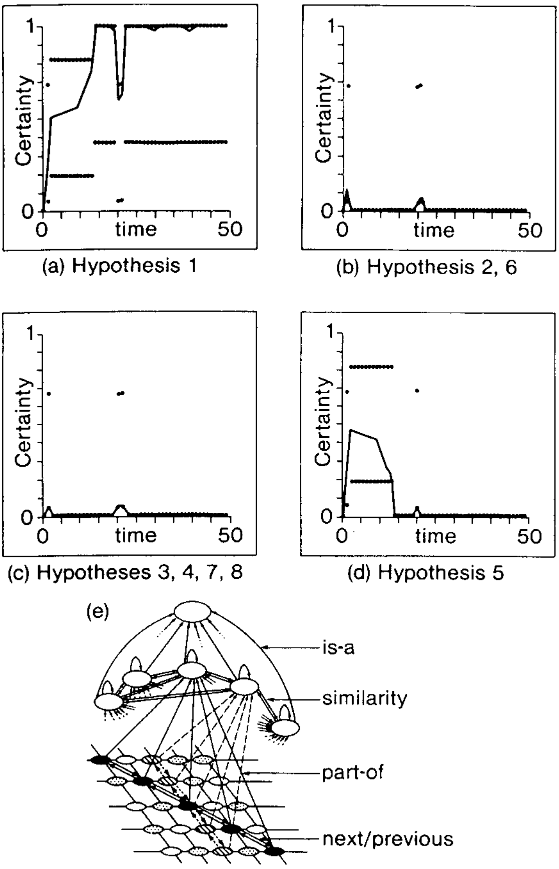
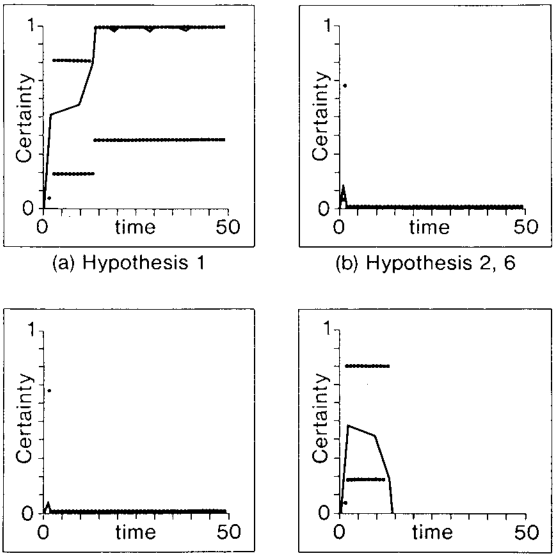


Figure 17

A direction-selection hypothesis structure, with stimuli on a grid, of varying strengths (coded by shading), and the resulting certainty-vs.-time profiles for each of the eight direction hypotheses. As the hypothesis number increases, the direction it represents rotates 45° clockwise, with the directions for hypotheses 1 and 2 shown. For example, hypothesis 5 represents the exact opposite direction to hypothesis 1. The temporal precedence relations among the grid stimuli are shown. Each stimulus is permitted to (re-)activate its PART-OF parent hypothesis. An IS-A relationship as top-down bias is added to this structure.



(a) Hypothesis 1

(b) Hypothesis 2, 6

(c) Hypotheses 3, 4, 7, 8

(c) Hypothesis 5

Figure 18
A direction-selection hypothesis structure, with stimuli on a grid, of varying strengths (coded by shading), and the resulting certainty-vs.-time profiles for each of the eight direction hypotheses. As the hypothesis number increases, the direction it represents rotates 45° clockwise, with the directions for hypotheses 1 and 2 shown. For example, hypothesis 5 represents the exact opposite direction to hypothesis 1. The temporal precedence relations among the grid stimuli are shown. The IS-A relationship is also present. Only the expected first stimulus of a hypothesized sequence is permitted to activate a hypothesis.

generated for the false hypothesis. The figures from Tsotsos 1981b that summarize the experimental findings are reproduced here as figure 19 and 20. The main results can be summarized as follows:

- When the number of competitors increases, the time to reach a decision also increases, roughly linearly.

- When varying amounts of noise are added, the slope of the curve increases in a smooth manner. The more noise, the longer it takes to reach a decision.

- When there is 50 percent noise (no information), no decisions can be reached.

These results are all very intuitive, yet it is satisfying to know that the temporal cooperative process possesses these characteristics.

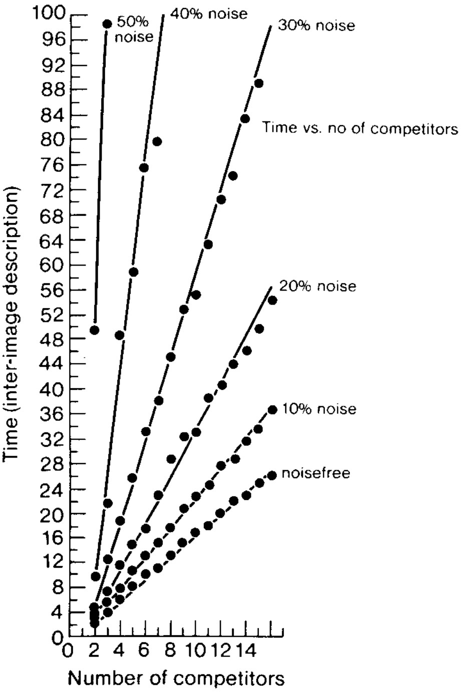


Figure 19
The results of experiments on time to decision with respect to size of discriminatory set for a particular setting of parameter values. In addition, varying amounts of random noise were included and the experiments repeated. The plots show a roughly linear relationship between time and number of competitors (as would be expected). Also, the addition of noise causes a graceful degradation of performance.

The effect on the exact same set of experiments of an added IS-A constraint are also described in Tsotsos 1981b. The slope of each curve, except the 50 percent noise curve, drops significantly. The added constraint can actually compensate for noise, to some degree, through feedback.

Temporal Sampling

Temporal sampling is an important issue. The guidelines presented earlier for the setting of compatibility values have presented some interesting possibilities for the determination of sampling rate. Intuitively, one might believe that the following play a role in the calculation of the sampling rate: the size of the discriminatory set, the expected noise level of the data, the

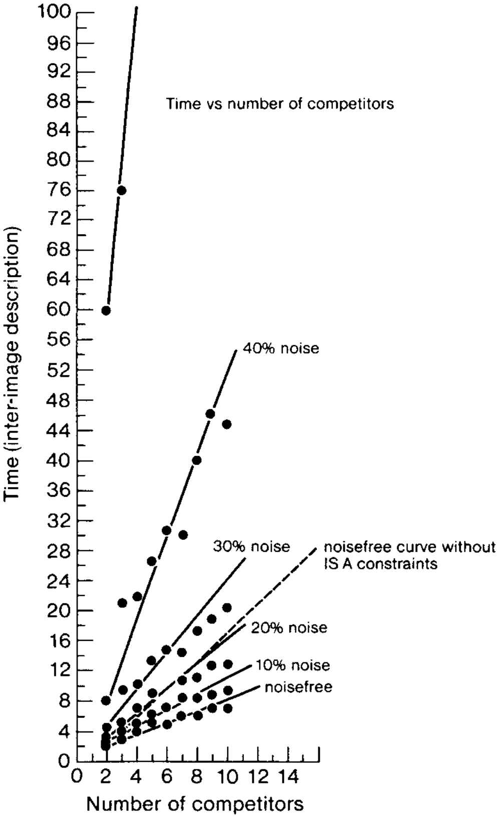


Figure 20
The same experimental setup as for figure 19 was used, with the addition of an IS-A constraint on the discriminatory set. It is clear that the time to decision is significantly decreased, even in the presence of moderate amounts of noise.

shortest duration for events in each discriminatory set, and, of course, the system response characteristics. With the size of the discriminatory set known, the time in iterations to make a useful decision was designed into the compatibilities; it is given by k_{self} for discriminatory set N_{SIM}. Proper performance is guaranteed for $k_{\text{self}} \geq |N_{\text{SIM}}| + 3$, and thus this compatibility can be set independently. Since k_{self} has units of "temporal measurements" where measurements are made "between" time samples, $k_{\text{self}} + 1$ gives the number of time samples required for k_{self} temporal measurements. The minimum event duration in seconds for a given discriminatory set is dur_{min}. If we assume the possibility of individualized compatibility values, where the subscript j ranges over all the individual groups, the minimum sampling rate can be given by

$$SR_{\text{min}} = \max_{j} \frac{(|N_{\text{SIM}}| + 4)_j}{(\text{dur}_{\text{min}})_j}.$$

Appropriate settings of k_{SIM} and $k_{\text{IS-A}}$ (that is, increasingly larger values) will compensate for reasonable amounts of noise. The resulting sampling rate must clearly be within the physical capabilities of the sampling system. If this is not true, it may be that changing k_{self} (that is, increasing this parameter) may shift it appropriately. If this is not possible, then the framework presented here is not applicable to the domain under consideration.

Note that $1/\text{dur}_{\text{min}}$ may be considered as the temporal frequency of the shortest event. If all discriminatory sets are considered, the value used in the sampling-rate determination is the maximum temporal frequency that is represented in the knowledge of the system, and thus the maximum temporal frequency that must be recognized. The standard Nyquist sampling rate for signal reproduction is $2F_{\text{max}}$, where F_{max} is the maximum frequency present in the signal. Therefore, we know that sampling at a higher rate will not yield new information. We can conclude that $[(|N_{\text{SIM}}| + 4)/2]_l$ must be performed per time sample. This satisfies our original goal, namely, that we are interested in a cooperative process such that the only decisions that are made are those that can be made within a small fixed number of iterations, and that this small number be sufficient for all events of interest. From an efficiency point of view, the smaller $|N_{\text{SIM}}|$ is, the fewer the iterations. This value can be kept small by appropriate use of concept organization along the IS-A dimension.

Finally, the PART-OF hierarchy of temporal concepts implicitly represents increasingly coarser levels of temporal as well as spatial resolution for concepts such as sequences. Using the relationship for sampling rate presented earlier, it is clear that, as temporal resolution becomes coarser, the sampling rate and thus the number of applications of the cooperative process iterations becomes smaller. This poses interesting possibilities for increasing the efficiency of the scheme, and it requires further exploration.

Discussion

Representational dimensions such as those described in the third major section of this paper have been prominent for some time in the literature on the representation of knowledge. The arguments for their use have been mostly qualitative, that is, they seem to have nice formal properties and lend themselves naturally to the construction of knowledge bases. In this work, it has been shown not only that these aspects are present but also that each representational dimension has a distinct role to play in an interpretation scheme. In fact, each has two important roles. One role is that of enabling multiple, interacting search mechanisms. This function should not be underestimated. Rule-based recognition paradigms, for example, only offer a single dimension of search. As was pointed out by Aiello (1983), such systems suffer from serious problems due to the one-dimensionality of the inference procedure. The conclusion is that goal-directed, event-directed, and model-directed inference mechanisms can most effectively compensate for one another's deficiencies if used in concert. For example, a data-directed scheme considers all the data and tries to follow through on every event generated. It can be nonconvergent, can produce only conclusions that are derivable directly or indirectly from the input data, and cannot focus or direct the search toward a desired solution. The goal-directed strategy is easy to understand and implement, and at each step of the execution the next step is predetermined. Rules are evaluated in the same order regardless of the input data. Thus, this strategy is inefficient and cannot exhibit a focus with respect to the problem being solved, since there is no mechanism that determines what is important and what is not. Finally, the model-directed approach, although the most efficient and the one that exhibits correct foci of problem-solving activity, has the disadvantage that its conclusions depend heavily on the availability of the correct model and initial focus. An incorrect initial focus will lead it to the examination of useless and incorrect analyses and will cause some perhaps relevant data to be ignored.

In our scheme, each dimension of search compensates for the failings of another, and thus as a whole the scheme offers a rich and robust framework. The several dimensions of search are tied to the knowledge-organization principles. Also, each organization dimension offers distinct and necessary contributions to the updating of hypothesis certainty, to the definition of neighborhoods and compatibilities, and to the maintenance of consistency within an interpretation:

- IS-A, besides offering a definition of global consistency of hypothesis certainty, plays the role of speeding up the convergence of results. This allows smaller temporal sampling rates. Owing to inheritance, the problem posed by the propagation of results disappears. IS-A also has an important part in the graceful recovery from poor predictions. Finally, feedback imposed by the IS-A hierarchy increases the stability of the cooperative process and partially compensates for the effects of noise disturbances.
- SIMILARITY plays the discrimination role, and is the only mechanism that allows for competition between hypotheses, enabling "best choice" selection. In conjunction with the exceptions that drive SIMILARITY activations, this is a strong feedback mechanism, enhancing the stability of the cooperative process. Moreover, it is central to the definition of temporal sampling rate and of compatibility values.
- PART-OF is the mechanism that permits the selection of the stronger of two equally consistent hypotheses on the basis of the strength of their components.
- Temporal Precedence assists in the discrimination of proper temporal order, which is important for temporal grouping and for temporal "gluing" of events into higher-order ones.

Moreover, each of these representational dimensions is integrated into the certainty updating scheme in an intuitive manner through the use of separate compatibility values and the use of neighborhoods or "conceptual receptive fields" for hypotheses.

It should be clear from the discussions in this section that there is no optimal setting of compatibility values for an entire knowledge base. Setting the values to accommodate the worst case may cause too quick a decision for other cases. Thus, it is natural to consider individualized settings of compatibilities, with a particular set of values holding over each hypothesis in a discriminatory set. This would allow for the best performance characteristics. Each conceptual receptive field thus has an individualized compatibility profile that is computed automatically and dynamically depending on the hypothesis structure that it is involved with and on the only three domain-specific constants required for each hypothesis (namely, minimum duration, number of active competitors, and—if it is a sequence—maximum temporal separation of its parts).

We have achieved one of our original goals: that of discovering the conditions under which this cooperative process can reach decisions about time-varying events within a small, fixed number of iterations. This result has led to the first stages of a sampling theory. Sampling considerations, woefully missing from much of computer-vision research, are clearly necessary within the spatio-temporal context that is required for visual perception.

Conclusion

A framework for the integration of time into high level or attentive vision was described. The key elements are an organization of knowledge along several axes, including time; several search modes facilitated by the knowledge organization; a hypothesize-and-test reasoning framework; and a temporal cooperative process driven by the knowledge organization. The goal of the research was to tie together work in knowledge-representation theory with cooperative processes, which are important for vision. Knowledge or concept organization is seen as the key and not the internal form of knowledge packages. The analysis and the examples presented demonstrate that each of the common representational axes—IS-A, PART-OF, SIMILARITY, and Temporal Precedence—has a natural place within the temporal cooperative process, and, moreover, make important contributions to it. Temporal considerations have played a key role, and have led to a useful version of a relaxation rule for time-varying interpretation under the severe constraints that a changing environment places on the number of iterations that can be performed. Although this scheme has been successfully implemented within the ALVEN expert system, much work remains, particularly with its mathematical foundations.

Acknowledgments

I am indebted to Allan Jepson for his useful suggestions. David Fleet and Niels da Vitoria Lobo read drafts and provided many useful comments. Niels also assisted with the implementation and experimentation. This paper also benefited from discussions with Allen Hanson, Ed Riseman, and Roger Browse. Financial support was received from the Canadian Heart Foundation and from the Natural Sciences and Engineering Research Council of Canada. The author is a Fellow of the Canadian Institute for Advanced Research.

References

Aiello, N. 1983. A comparative study of control strategies for expert systems: AGE implementation of three variations of PUFF. In Proceedings of AAAI, Washington.

Allen, J. 1981. Maintaining Knowledge about Temporal Intervals. Report TR-86, Department of Computer Science, University of Rochester.

Anstis, S. 1978. Apparent motion. In *Handbook of Sensory Physiology,* ed. Held, Leibowitz, and Teuber (Springer).

Ballard, D., C. Brown, and J. Feldman. 1978. An approach to knowledge-directed image analysis. In *Computer Vision Systems,* ed. Hanson and Riseman (Academic).

Brachman, R. 1979. On the epistemological status of semantic networks. In *Associative Networks,* ed. Findler (Academic).

Brachman, R. 1982. What IS-A and isn't. In Proceedings of CSCSI-82, Saskatoon.

Braddick, O. 1974. A short range process in apparent motion. *Vision Research* 14: 519–528.

Brooks, R. 1981. Symbolic reasoning among 3D models and 2D images. *Artificial Intelligence* 17: 285–348.

Bugelski, B., and D. Alampay. 1962. The role of frequency in developing perceptual sets. *Canadian Journal of Psychology* 15: 205–211.

Cooper, L., and R. Shepard. 1973. Chronometric studies of the rotation of mental images. In *Visual Information Processing,* ed. Chase (Academic).

Down, B. 1983. Using Feedback in Understanding Motion. M. Sc. thesis, University of Toronto.

Dretske, F. 1981. *Knowledge and the Flow of Information.* MIT Press.

Gibson, J. J. 1957. Optical motions and transformations as stimuli for visual perception. *Psychological Review* 64, no. 5: 288–295.

Glazer, F. 1982. Multilevel Relaxation in Low Level Computer Vision. Report TR 82-30, Department of Computer and Information Science, University of Massachusetts, Amherst.

Hanson, A., and E. Riseman. 1978. VISIONS: A computer system for interpreting scenes. In *Computer Vision Systems,* ed. Hanson and Riseman (Academic).

Hartley, D. 1749. Observations on man, his frame, his duty, and his expectations. Reprinted in *A Source Book in the History of Psychology,* ed. Herrnstein and Boring (Harvard University Press, 1968).

Hay, J. 1966. Optical motions and space perception: An extension of Gibson's analysis. *Psychological Review* 73, no. 6: 550–565.

Hinton, G., and T. Sejnowski. 1983. Optimal perceptual inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, D.C.

Hummel, R., and S. Zucker. 1980. On the Foundations of Relaxation Labelling Processes. Report TR-80-7, Department of Electrical Engineering, McGill University.

Julesz, B. 1980. Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Philosophical Transactions of the Royal Society of London* 290: 83–94.

Julesz, B., and R. Schumer. 1981. Early visual perception. *Annual Review of Psychology* 32: 575–627.

Kanade, T. 1980. Survey: Region segmentation: Signal vs. semantics. *Computer Graphics and Image Processing* 13: 279–297.

Kandel, E., and J. Schwartz, eds. 1981. *Principles of Neural Science.* Elsevier/North-Holland.

Levesque, H., and J. Mylopoulos. 1979. A procedural semantics for semantic networks. In *Associative Networks,* ed. N. Findler (Academic).

Levine, M. 1978. A knowledge-based computer vision system. In *Computer Vision Systems,* ed. Hanson and Riseman (Academic).

Mackworth, A. K. 1978. Vision research strategy: Black magic, metaphors, mechanisms, miniworlds, and maps. In *Computer Vision Systems,* ed. Hanson and Riseman (Academic).

Mackworth, A., and W. Havens. 1982. Representing visual knowledge. *IEEE Computer* 16, no. 10: 90–98.

Marr, D. 1982. *Vision.* Freeman.

Mill, J. 1829. Analysis of the phenomena of the human mind. Reprinted in *A Source Book in the History of Psychology,* ed. Herrnstein and Boring (Harvard University Press, 1968).

Minsky, M. 1975. A framework for representing knowledge. In *Phychology of Computer Vision,* ed. Winston (McGraw-Hill).

O'Rourke, J., and N. Badler. 1980. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 522–536.

Palmer, S. 1975. The effects of contextual scenes on the identification of objects. *Memory and Cognition* 3: 519–526.

Sabbah, D. 1981. Design of a highly parallel visual recognition system. In Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver.

Tenenbaum, J., and H. Barrow. 1977. Experiments in interpretation guided segmentation. *Artificial Intelligence* 8: 241–274.

Terzopoulos, D. 1982. Multi-Level Reconstruction of Visual Surfaces. AI Lab Memo 671, Massachusetts Institute of Technology.

Treisman, A. 1982. Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology* 8, no. 2: 194–214.

Treisman, A., and G. Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12: 97–136.

Treisman, A., and H. Schmidt. 1982. Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14: 107–141.

Tsotsos, J. 1981a. Temporal event recognition: An application to left ventricular performance assessment. In Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver.

Tsotsos, J. 1981b. On classifying time-varying events. In Proceedings of the Conference on Pattern Recognition and Image Processing, Dallas.

Tsotsos, J. 1983. Medical knowledge and its representation: Problems and perspectives. In Proceedings of IEEE MEDCOMP' 83.

Tsotsos, J. K., J. Mylopoulos, H. D. Covvey, and S. W. Zucker. 1980. A framework for visual motion understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 563–573.

Tsotsos, J. K., D. Covvey, J. Mylopoulos, and P. McLaughlin. 1984. The role of symbolic reasoning in left ventricular performance assessment: The ALVEN system. In *Ventricular Wall Motion*, ed. Sigwart and Heintzen (Georg Thieme Verlag).

Wertheimer, M. 1923. Untersuchung zur Lehre von der Gestalt. II. *Psychologische Forschung* 4: 301–350.

Zucker, S. W. 1978a. Vertical and horizontal processes in low level vision. In *Computer Vision Systems*, ed. Hanson and Riseman (Academic).

Zucker, S. W. 1978b. Production systems with feedback. In *Pattern-Directed Inference Systems*, ed. Waterman and Hayes-Roth (Academic).

Zucker, S. W., R. A. Hummel, and A. Rosenfeld. 1977. An application of relaxation labelling to line and curve enhancement. *IEEE Transactions on Computers* 26: 394–403.