Proceedings of the Annual Meeting of the Cognitive Science Society

**Title**

NOVA, Covert Attention Explored Through Unified Theories of Cognition

**Permalink**

https://escholarship.org/uc/item/8216b7kh

**Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

**Authors**

Wiesmeyer, Mark
Laird, John

**Publication Date**

1993

# NOVA, Covert Attention Explored Through Unified Theories of Cognition*

**Mark Wiesmeyer and John Laird**

Artificial Intelligence Laboratory
The University of Michigan
Ann Arbor, MI 48109

wiesmeye@caen.engin.umich.edu

### Abstract

Covert visual attention is a subtle part of human vision that has been widely researched in the psychology community. Most often visual attention is thought to involve movements of the eyes or head; however, covert visual attention does not involve overt movements of any sort. It has often been described in an homuncular sense as the "mind's eye."

This paper introduces both a new model of covert visual attention and a new approach in which to investigate attention. The approach is based on five assertions: *(1)* Development of models of attentional processes should occur in the context of a fixed, explicit model of nonattentional processes. *(2)* Evaluation of attentional models should occur in the context of complete tasks. *(3)* Judgment of the quality of an attentional model should be with respect to its ability to cover many tasks while maintaining constant parameters. *(4)* Computer implementation and simulation of an attentional model and the tasks it claims to cover should be used for demonstrating its sufficiency. *(5)* A process model (a model that seeks to correspond at some level of analysis to actual mechanisms of behavior) should be able to account for both the timing and the functions of behavior.

NOVA (Not Overt Visual Attention), the first operator-based model of covert visual attention, is based on the Model Human Processor [Card, Moran, and Newell, 1983], a model of nonattentional processes that has been applied successfully in Human-Computer Interaction (HCI) research. In this paper we review the results of using NOVA to model seven qualitatively different immediate-response tasks from the psychological literature. As a test of the sufficiency of NOVA, we implemented NOVA and each of the task models in the Soar cognitive architecture, a computer model of human behavior that has been proposed as the basis of Newell's "Unified Theories of Cognition" (UTC) [Newell, 1990]. NOVA is both a new theory of attention and a framework in which existing theories of attention have been unified.

## 1.0 Introduction

The major part of the motivation for our work, has come from Allen Newell. In particular, the paper "You can't play 20 questions with nature and win" [Newell, 1973], which set the stage for his work on UTC, has been an ever-present reminder of the need to develop theories of cognition like NOVA. Earlier psychologists also seemed to feel the need for overarching psychological theories that could unify large bodies of subtheory.

> *It is true that the discovery of attention did not result in any immediate triumph of the experimental method. It was something like the discovery of a hornet's nest: the first touch brought out a whole swarm of insistent problems.*
>
> ---Edward Titchener (1908)

The problems to which Titchener is referring arise because attention is only one component of a behaving intelligent organism. In order to understand what attention is, how it functions, and what its limitations and capabilities are, an explicit nonattentional model of the organism (including perception, cognition, motor) must be available in which to test attentional hypotheses.

NOVA has been able to explain covert attention phenomena by starting with the Model Human Processor (MHP) and extending and unifying it with well-known data that have been published in the behavioral and attentional literature. The models for each of the seven attentional experiments that are later described were developed in the order in which they are presented. During the process of modeling, NOVA was enlarged or altered as experimental data demanded, and the effect of each change on NOVA's ability to account for previously modeled experiments was checked. The power in the approach is that each isolated experiment, when combined with the other isolated experiments, acts as a constraint on the model. The result is a model that covers several experimental tasks, but has few parameters.

---

* This work was supported by grant NCC2-517 from NASA Ames.

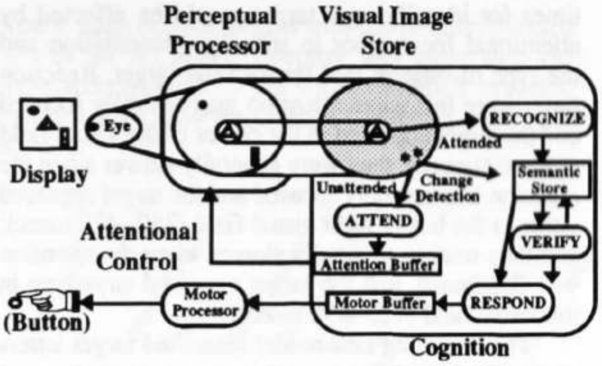

Figure 1: Nova

Figure 1 shows NOVA, which is similar functionally and structurally to the MHP. As in the MHP, each subsystem operates semi-autonomously at a particular rate called its "cycle time." The cycle time of Perception, 100 msec, is defined to be the time required for a stimulus to travel from the retina to the Visual Image Store, a part of Working Memory. Cognition is composed of Working Memory, Long-term Memory, and the Cognitive Processor. Cognition has a cycle time of 50 msec, which is the time required for an operator to be created or applied. Motor, which executes commands that Cognition deposits in Working Memory, has a cycle time of 70 msec. That is the time required for a Working Memory change to cause an overt motor response.

NOVA's major additions to the MHP include a perceptual dependency for the shift of covert visual attention (150 msec), the specification of three semantic operators (VERIFY, COUNT, RESPOND) and two visual operators (ATTEND and RECOGNIZE), and a more developed theory of visual features.

A "perceptual dependency" occurs when the creation of an operator that is critical to a task is dependent upon visual stimuli that may not yet be converted into Working Memory inputs. In the MHP, perceptual dependencies occur strictly due to stimulus presentation latencies, perceptual cycle times, and eye movements.

NOVA's 150 msec latency for a shift of attention is the sum of the time to create an ATTEND operator (50 msec), the time to apply that operator (50 msec), and the time required for attended visual information to arrive in the Visual Image Store from V4 (50 msec). Visual information has been shown to be attentionally gated at V4, a part of the visual cortex [Moran and Desimone, 1985].

The core concept of NOVA is that visual attention must precede identification, and both attention and identification are deliberate actions that are mediated by operators (ATTEND and RECOGNIZE). The attentional zoom lens is a single oval region of attentional focus that ranges across the visual field, separating it into attended and unattended regions. Perception delivers both attended features, which derive from the attended region, and unattended features, which derive from the unattended region, into the Visual Image Store, which is part of Working Memory. Both attended and unattended features are iconic (i.e., presemantic) representations. Unattended features are used to cue new shifts of attention. Attended features are used in the process of identification. Only after attended features are transformed into symbols through the application of the RECOGNIZE operator can the identities of visual events be reported or affect behavior. However, unidentified *changes* in the visual field can be reported or affect behavior without attention. The VERIFY operator monitors the symbolic contents of Working Memory allowing the task model to accomplish task goals, while the RESPOND operator produces overt actions. The COUNT operator counts attended items.

Differences in the availability of information for features from the attended and unattended regions is the result of visual attention and creates the need for the region of attention to move. There are four combinations of feature identity and location information in NOVA: *(1)* Known feature identity information and exact location information. Cognition "knows" that there is a specific event in the visual field and where it is. *(2)* Known feature identity information and inexact location information. Cognition knows that there is a specific event in the visual field in a general area of the visual field. *(3)* Unknown feature identity information and inexact location information. Cognition knows that there is an event in the visual field in a general area of the visual field, but does not know what it is. *(4)* Unknown feature identity information and no location information. Cognition knows that there is an event in the visual field, but it does not know what it is or where it is. These various combinations of feature information produce a discrete approximation of a gradient-based, zoom-lens attentional apparatus.

A discussion of the exact nature of features in NOVA is involved and out of the scope of this paper. Generally, however, in NOVA, shape and color stimuli are represented separately as two different classes of features. Perception delivers both simple shape features at the grain size of primitive features (e.g., lines, curves, orientations, angles) and complex shape features at the level of linguistic objects to the Visual Image Store. For a full explanation of features and other parts and results of the NOVA theory, please refer to the first author's thesis [Wiesmeyer, 1992].

## 2.0 Task Models

A "task model" is a collection of operators, control knowledge, and goals that when combined with task stimuli produces operator traces that may be used to explain task behavior. Each operator trace consists of a sequence of operator creations and applications that in conjunction with perceptual dependencies and motor cycle times account for functional behavior and yield timing estimates for the task. NOVA operator traces are similar to GOMS models as introduced with the MHP; however, GOMS uses much higher level operators than are employed in NOVA. NOVA may be thought of as GOMS modeling at the micro level.

Additional important characteristics of NOVA task models are that operators remain the same in all task models; task control knowledge and goals entirely specify the ordering of operators that are created and applied; and all operations take on the order of 50 msec.

The following subsections review the successes and shortcomings of the various task models. A task model for the Decay Experiments is presented in some detail, since it is representative of NOVA models and has not been published before---as have the precuing and search experiments [Wiesmeyer and Laird, 1990] and the counting experiments [Wiesmeyer, 1991].

### Precuing Experiments

The precuing task model was based on experimental results from Colegate, *et al.* (1973). The goal of these experiments was to vocally identify a cued letter from a group of 8 or 12 letters arrayed in a circle around a fixation point (only the 12-letter version of the experiment was modeled). Results showed that reaction times improved at a relatively constant rate as the cue preceded the letter target (maximum benefit occurred with a 250 msec precue). Mean response times were 574 msec (482--635 msec) for simultaneous presentation and 491 msec (434--569 msec) for 250 msec precued presentation.

The precuing task model identified cued letters. When the cue appeared with the letter wheel, the predicted reaction time was 570 msec. When the cue appeared 250 msec before the appearance of the letter wheel, the predicted reaction time was 470 msec. These predicted reaction times are close to the observed extreme reaction times. Additionally, the task model predicts three other reaction times (570 msec at 150 msec, 520 msec at 200 msec, and 470 msec at 350 msec). The relationship between precues and reaction times was not as linear as in the averaged experimental results.

### Crowding Experiments

The crowding task model was based on experimental results from LaBerge and Brown (1989). The goal of these experiments was to determine how reaction times for identifying a target would be affected by attentional focus prior to stimulus presentation and the type of objects that flanked the target. Reaction times were fast when attention was centrally focused and the target appeared in the center of the visual field (350--400 msec); they were generally slower when the attention was centrally focused and the target appeared either in the left or right visual field (350--470 msec); and they were consistently slower when the attention was distributed, and the target appeared anywhere in the visual field (420--490 msec).

The crowding task model identified target letters among flankers. When attention was centrally focused and the target appeared in the center of the visual field, the predicted reaction time was 320 msec. When attention was centrally focused and the target appeared either in the left or right visual field, the predicted reaction time was 470 msec. When attention was distributed, and the target appeared anywhere in the visual field, the predicted reaction time was also 470 msec.

### Decay Experiments

The decay task model was based on experimental results from Sperling (1960). The goal of these experiments was to identify as many letters as possible in a briefly presented display. Typical stimuli are shown in Figure 2. Reaction times were of no interest, although the time to respond after stimulus exposure was usually small. An average of 4.3 letters were identified.

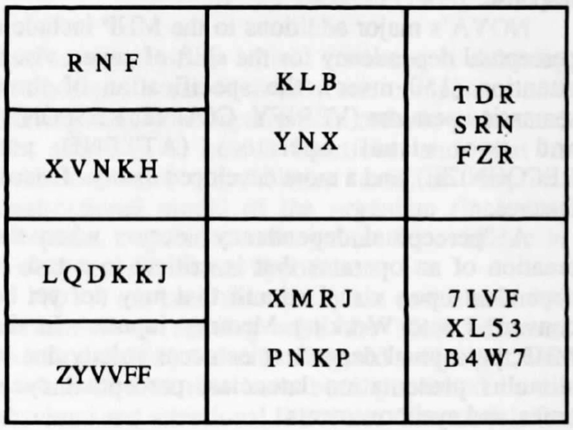

Figure 2: Decay Experiment Stimuli

In this task model, we assumed that identification and localization of letters proceeds serially in an item-by-item fashion until letter features supported by Perception completely decay. The operator trace shown in Figure 3 begins with letter stimulus presentation (top row, middle stimulus in Figure 2) and consists of a series of four shift/identification episodes corresponding to the four letters that can be identified and located before stimulus extinction occurs. Each episode begins with the creation (labeled "c") of an ATTEND operator and ends with the application (labeled "a") of a RECOGNIZE operator. Hatched areas show Perceptual, Cognitive, and Motor activity.

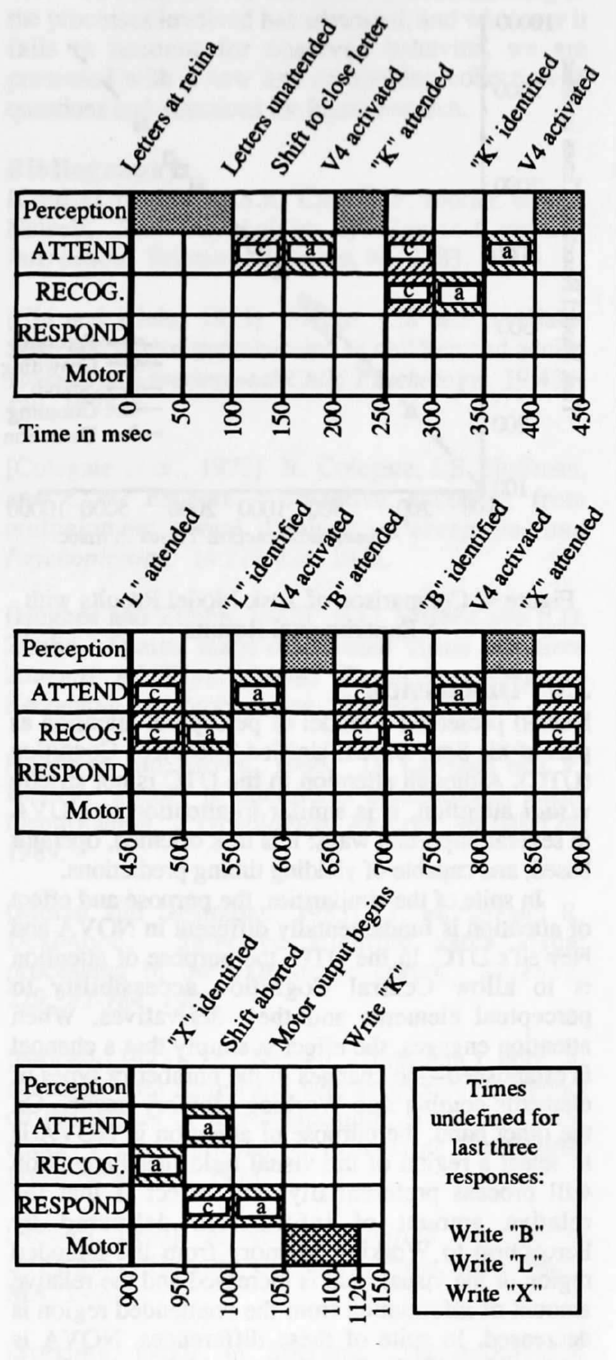

Figure 3: Operator Trace for Decay Experiments

Since stimulus presentation was so brief, an approximate latency of stimulus decay could be determined by simply reading the time off the trace when the last identification completed (950 msec). A final shift of attention is attempted at 950 msec but is aborted, because letter features in Working Memory are retracted due to decay of the stimulus in Perception. Following the shift/identification activity, responses are made. The task model does not commit to latencies for the responses, because response times were not a variable of interest in the experiment and were not recorded.

### Illusory Conjunction Experiments

The illusory conjunction task model was based on experimental results from Treisman and Schmidt (1982). For each trial of these experiments, the primary goal was to identify two numbers that flanked three colored letters. The secondary goal was to identify and locate the central colored letters. Displays were masked shortly after presentation. Results showed that subjects performed the primary task very accurately, reported about 50% of the shape and color features present in each trial, and mislocated a few of the identified shapes and colors. About 12% of the mislocated shapes and colors produced illusory conjunctions.

The illusory conjunction task model produced shape and color mismatches. It showed that errors in feature conjunction could be modeled as the removal of location information from features in a special part of the Visual Image Store called Integrative Visual Memory (where attended features are accumulated).

### Search Experiments

The search task model was based on experimental results from Treisman and Gelade (1980), Treisman and Gormican (1988), and Wolfe *et al.* (1989). The goal of these experiments was to determine the presence or absence of a target among a field of distracters in as little time as possible. Results showed that reaction times were either fast and relatively independent of the number of distracters or were slower and linearly related to the number of distracters. Searches for targets defined by single attributes tended to be of the faster variety, while search for targets defined by both shapes and colors tended to be of the slower variety. Search slopes were found to range from 2 to 128 msec per item and intercepts were found to range from 397 to 682 msec.

The search task model found targets among distracters. It showed that a single model of search could account for both of the typical reaction time behaviors seen in experimental data. The central notion of the task model is that search rate per item is related to the number of items that the attentional mechanism can scan at a time. This number, the group size, is related to the contrast between the target and distracters. The task model did not account for the large range of intercepts, because the intercepts did not generally correlate with identifiable parameters of the experimental stimuli. (All task models for search had an intercept of 470 msec.)

**Counting Experiments**

The counting task model was based on results from Chi and Klahr (1975). The goal of the counting experiment that we modeled was to count from one to ten dots as quickly as possible. The results showed that for up to about three of four dots counting is very quick and that after that it becomes significantly slower per dot. Straight line fits for Chi's results show that there is a shallow slope of 46 msec with an intercept of 495 msec for the first three items, and a steeper slope of 307 msec with an intercept of -442 msec for subsequent items.

The counting task model counted items using two processes. In the first process, up to three or four items were counted in a single region of attention. In the second process, any remaining objects (more than three or four) were counted each in its own region of attention. The reaction times predicted by the task model for each number of items were accurate to about 10 msec.

**Detection Experiments**

The detection task model was based on results from Hughes and Zimba (1985). The goal of these experiments was to detect the onset of a luminance increment (brightening) either in a cued or noncued position in the visual field. Results showed that the effects of attention were only hemifield specific. When cues were in the expected hemifield, reaction times improved by about 10 msec with respect to a neutral cue condition, and when cues were in the unexpected hemifield, reaction times worsened by about 30 msec with respect to a neutral cue condition. Typical reaction times were 190--240 msec in the expected hemifield and 250--290 in the unexpected hemifield.

The detection task model reported (simulated) luminance increments. It predicted reaction times for only two classes of conditions: *(1)* neutral cues and targets in the expected hemifield (220 msec) and *(2)* targets in the unexpected hemifield (270 msec). This task model was significantly different from others in that the notion of a region of attention was not involved, since attentional effects were instead ascribed to the early creation and application of the RESPOND operator.

**Comparison of Task Models with Experimental Results**

Figure 4 compares the timing estimates of the task models to the experimental results that they sought to explain. Note that the task models for decay and illusory conjunctions are not included in the figure, because they did not yield timing estimates. The reaction times for search are based on a Treisman and Gelade, Experiment 2, the difficult condition. The figure shows that all experimental results fall close to the dashed 45 degree line. In this type of comparative figure, since the abscissa and ordinate are at the same scale, that is exactly where the best results should lie. The average absolute error for data points provided by NOVA task models is 17.2 msec.

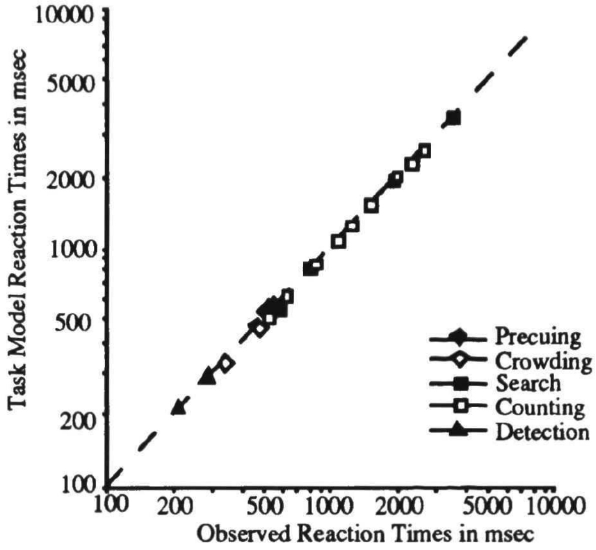

Figure 4: Comparison of Task Model Results with Experimental Results

## 3.0 Discussion

Newell presented a model of perceptual attention as part of his Soar-Based, Unified Theory of Cognition (UTC). Although attention in the UTC is not strictly *visual* attention, it is similar to attention in NOVA in several important ways: it is task oriented, operator based, and capable of yielding timing predictions.

In spite of the similarities, the purpose and effect of attention is fundamentally different in NOVA and Newell's UTC. In the UTC, the purpose of attention is to allow Central Cognition accessibility to perceptual elements and their derivatives. When attention engages, the effect is simply that a channel is established---no changes in the number or types of elements coming into Working Memory occur. On the other hand, the purpose of attention in NOVA is to select a region of the visual field that Perception will process preferentially. The effect is that the relative amount of information delivered by Perception to Working Memory from the attended region of the visual field is increased and the relative amount of information from the unattended region is decreased. In spite of these differences, NOVA is compatible with the Soar theory, which means that it can be viewed as an alternative to the UTC model of attention.

Much of inspiration for the NOVA theory and the supporting methodology for behavioral modeling presented in this paper is derived from an old idea [Newell, 1973]. The idea is that much gain may be made in psychological research from an approach that emphasizes the synthesis of observed data into a consistent integrated model. The beauty in Newell's approach to modeling is that it is in some sense always successful. Whenever a model accounts for observed behavior, we feel that our understanding of the processes involved has advanced, and whenever it fails to account for observed behavior, we are presented with a new and compelling collection of questions and directions for future research.

## Bibliography

[Card *et al.*, 1983] S.K. Card, T.P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Erlbaum, Hillsdale, NJ, 1983.

[Chi and Klahr, 1975] M.T.H. Chi and D. Klahr. Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, 19:434-439, 1975.

[Colegate *et al.*, 1973] R. Colegate, J.E. Hoffman, and C.W. Eriksen. Selective encoding from multielement visual displays. *Perception and Psychophysics*, 14:217-224, 1973.

[Hughes and Zimba, 1985] H.C. Hughes and L.D. Zimba. Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 11:409-430, 1985.

[LaBerge and Brown, 1989] D. LaBerge and V. Brown. Theory of attentional operations in shape identification. *Psychological Review*, 96:101-124, 1989.

[Moran and Desimone, 1985] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229:782-784, 1985.

[Newell, 1973] A. Newell. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W.G. Chase editor, *Visual Information Processing*. Academic Press, New York, 1973.

[Newell, 1990] A. Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA, 1990.

[Sperling, 1960] G. Sperling. The information available in brief visual presentations. *Psychological Monographs*, 74:1-29, 1960.

[Titchener, 1908] E.B. Titchener. Attention as sensory clearness. In *Lectures on the Elementary Psychology of Feeling and Attention*, pages 171-206. MacMillan, New York, 1908.

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

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

[Treisman and Gormican, 1988] A. Treisman and S. Gormican. Feature Analysis in Early Vision: Evidence from Search Asymmetries. *Psychological Review*, 95:15-48, 1988.

[Wiesmeyer and Laird, 1990] M.D. Wiesmeyer and J.E. Laird. A Computer Model of Visual Attention. *Twelfth Annual Conference of the Cognitive Science Society*, Boston, 1990.

[Wiesmeyer, 1991] M.D. Wiesmeyer. An Operator-Based Model of Rapid Visual Counting. *Thirteenth Annual Conference of the Cognitive Science Society*, Chicago, 1991.

[Wiesmeyer, 1992] M.D. Wiesmeyer. An Operator-Based Model of Covert Visual Attention. PhD Thesis, The University of Michigan, Ann Arbor, 1992

[Wolfe, *et al.* 1989] J.M. Wolfe, K.R. Cave, and S.L. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15:419-433, 1989.