

# Exploring the use of Cognitive Models in AI Applications using the Stroop Effect

Charles Hannon and Diane J. Cook

University of Texas at Arlington  
Box 19015  
Arlington, Texas 75019  
{hannon, cook}@cse.uta.edu

## Abstract

Using a generalized adaptive framework for unified cognitive modeling, we replicate human performance on a standard Stroop task within an explanatory computational model of vision, language and higher order processing. Having shown the ability to generate similar results to its human counterpart on a Stroop Test evaluation, we discuss how a succession of similar tests on well-understood phenomena like the Stroop Effect can be used to refine a broader model of unified cognition which can then be used to improve general AI applications.

## Introduction

With the shift to intelligent agent-based systems, AI applications are relying on more detailed models of cognition. Direct cognitive modeling has most often focused on either a detailed model of an individual cognitive task (e.g., color or pattern recognition, etc.) or a higher level model of the general processing involved in a cognition ability (e.g., memory, learning, task planning and scheduling, etc.). Traditional symbolic systems have for the most part focused on the sequential processing of cognitive tasks with any parallelization being done in the generalized search and inference used in these processes (Anderson 1998). Connectionist approach's have explored the explanatory nature of massive parallelism but suffer from the complexity of building and understanding neural-based models of unified cognition (Levine 1991).

To attempt to gain from the strengths of both the symbolicist and connectionist approaches, we have constructed the Adaptive Modeling environment for Explanatory Based Agents (AMEBA) architecture by which the explanatory nature of massive parallelism can be explored using a symbolic framework (Hannon and Cook 2001). Using AMEBA's ability to construct large cognitive models from reusable components that capture aspects of more detailed cognitive functions, we propose a process by which an overlapping set of cognitive models are used to

refine our understanding of general cognitive processes. Since the implementation of these cognitive models in AMEBA relies on powerful computing concepts like component and agent mobility, multi-threaded process designs and rigid inter-process communication standards, AMEBA also allows us to study how AI applications can be constructed using these explanatory mechanisms.

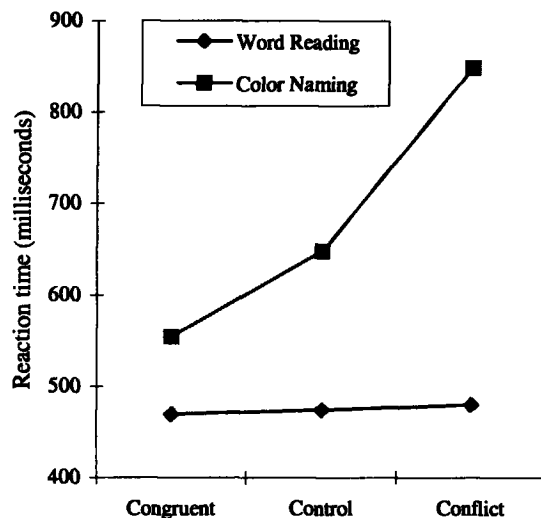
To date, we have examined two different computational models using AMEBA, one which models general language understanding and learning (Hannon and Cook 2001) and the one addressed here which explores the cognitive mechanism responsible for the Stroop Effect. The general language model addressed the complexity of combined utterance and discourse processing, symbol grounding, and the intentionality of the speech act, but proved difficult to quantifiably evaluate since it covered such a large set of overlapping cognitive skills. The Stroop Effect evaluation system was built to address a domain that still requires a moderately complex combination of vision and language processing skills but has a well understood and quantifiable expected result. By using the same components and structure in both models, we are able to provide support for our hypothesis that as the number of related computational cognitive models increases, so will our understanding of unified cognition and how to exploit this understanding in general AI applications.

## The Stroop Effect

The Stroop Effect was discovered by James R. Stroop (Stroop 1935). Simply stated it is:

The effect that the tendency to name a word will interfere with the ability to say the color in which the word is printed (Anderson 1995).

Average human performance on a standard Stroop Test as reported in a 1984 study by Dunbar and MacLeod is given in Figure 1. When a subject is asked to read a word, her performance varies only slightly regardless of the relationship between the word pattern and the ink color. If



**Figure 1. Stroop Task Performance (Anderson 1995)**

the subject is asked to name the ink color of a word, the meaning of the word will have a major effect on the response time. Congruent combinations, where the word pattern is the name of the color of the ink, will improve performance over control cases where the word pattern is not a color name. Conflicting combinations, where the word pattern is the name of a color other than the color of the ink, will significantly interfere with the color-naming task.

Researchers have attempted to both clarify the scope of the phenomena and to create a neuropsychological model of an explanatory mechanism. While most research supports the theory that the automaticity of reading interferes with the procedural task of color naming, some argue that the interference on the color naming task is a result of the word's pattern and color forming a mental set (Besner, Stolz, and Boutilier 1997). There is some debate in both theories as to where this interference occurs.

Studies on the effect of frontal lobe lesions indicate an increased rate of interference on the Stroop test (Shimamura, 1995). This would seem to indicate that patients with frontal lobe damage have more difficulty in inhibiting the learned task of reading, although this damage could also be directly effecting attention control during the visual information encoding process. Studies of the effect of prefrontal lobe lesions which demonstrate that patients exhibit attention difficulty on similar task to the Stroop task would tend to support this later view (Knight and Grabowecky 1995).

In building a neuropsychological model to explain the Stroop effect, Virzi and Egeth used differences in performance based on the way the subject receives the stimulus and reports the results to suggest that the primary interference occurs during the translation mechanism

between spatial and semantic information (Virzi and Egeth 1985). MacLeod and Dunbar have been able to show a Stroop Effect on a learned task other than reading which has little or no language processing component (MacLeod and Dunbar 1988). These studies would seem to indicate that any observed translation interference is occurring at a conceptual level below the language-processing layer. Brega and Healy provide evidence that a sentence containing color descriptions appears to also generate Stroop interference (Brega and Healy 1999), but it is not clear whether this type of interference results from the same mechanism as the word pattern interference.

## AMEBA Architecture

The AMEBA architecture represents the refinement of generalized parallel tools to produce an environment for testing cognitive theories and building AI applications based on these models. The architecture runs on a SMP cluster using a Beowulf-like connection scheme of multiple high-speed networks.

AMEBA attempts to capture the explanatory force of a connectionist neural model while allowing the use of the better-understood representation and reasoning methods of symbolic AI. From a system perspective, it attempts to provide processor transparency within a parallel system and a flexible method of process and knowledge management. The key element that supports these requirements is the etheron process template. An etheron provides a container for an instance of any inference or routing mechanism needed by the system. Once contained, the etheron supports the mechanism with, 1) a standard way to load and store knowledge, 2) interfaces to AMEBA's management tools and 3) a generalized set of communication channels for talking with other etherons.

AMEBA models draw their explanatory depth from the environment's ability to support hierarchical cognitive processing. Using adaptive distributed processing and generalized inter-process communication, cognitive functions can be modeled at different levels of abstractions without changing the logical relationship between these functions. Thus, a function like the conceptual reasoning about the world and self can be simulated with a reasoning and knowledge storage system which has far less capacity than that of a real human. This allows us to preserve the overall model's explanatory depth, as long as we preserve explanatory relationships between cognitive components. To ensure that we preserve these relationships, our modeling research is driven by both the evidence from experimental psychology regarding the architecture of the mind and the evidence from neuro-physiology regarding the architecture of the brain.

## The Stroop Test System Design

In developing a computational model of the Stroop effect, our goal was to embed the resulting model in as realistic a

model of unified cognition as possible. Therefore, we tried to preserve as much as possible of the overall organization and operation of the components used to study the language understanding and learning since these components had already been shown to generate explanatory results. The resulting Stroop Test Response Evaluation Sub-System (STRESS) is presented in Figure 2. STRESS provides the flexibility to adaptively model a cognitive task by allowing the inference or knowledge used in one etheron to be changed without impacting the rest of the system. This ability was used during the modeling of the Stroop Effect to refine both the interactions between the Task and Concept reasoners and the internal actions of these two critical components.

STRESS consists of three agents, 1) the Researcher, 2) the Evaluation Tool, and 3) the Subject. Each labeled ellipse in the figure represents a process component (an etheron) in the sub-system. A total of 26 etherons were used to model the sub-system, the major portion of these making up the Subject agent. Most of the Subject's components are the same as the ones used by a child agent in the language understanding and learning system. The basic differences between the Subject and a child agent of that system are; 1) the child agent had more Higher Order Process (HOP) reasoners to support tasks like discourse analysis, socialization and intentionality and 2) the child agent's vision components simulate simple object recognition while the Subject agent's vision system is based on a very simple emulation of a human eye and visual cortex.

Both the child agent of the general language model and the Subject agent of STRESS use a Concept Reasoner based on a semantic network that supports temporal-modal relationships between concept nodes. At a knowledge level, the Concept Reasoner of the child agent and the Subject are quite different. While the knowledge representation of the base reasoner has been used with very large networks, the STRESS conceptual network is quite small, containing less than a hundred concepts. This is because the child's Concept Reasoner is primarily used to store and retrieve all of the world knowledge needed to conduct a child-like discourse while the Subject's reasoner is primarily used to simulate both the constructive and destructive interference of spreading activation within the reasoner's semantic network. While the knowledge stored in the utterance processing etherons of the Subject agent is the same as the child agent, the STRESS system only uses the natural language interface between the Researcher and Subject to allow the Researcher to pass simple instructions to the Subject regarding the testing process.

Inter-agent communication in STRESS passes three types of information (natural language, vision information, and time synchronization) to and from different etherons within the same agent. This information is channelized in the inter-agent communication scheme to allow heterogeneous message passing between different interfaces in each agent. Intra-agent communication in all three STRESS agents relies on the Stimuli Routing Network (SRN) provided by AMEBA. Any etheron labeled as either a SRN or an Agent in Figure 2 is a SRN.

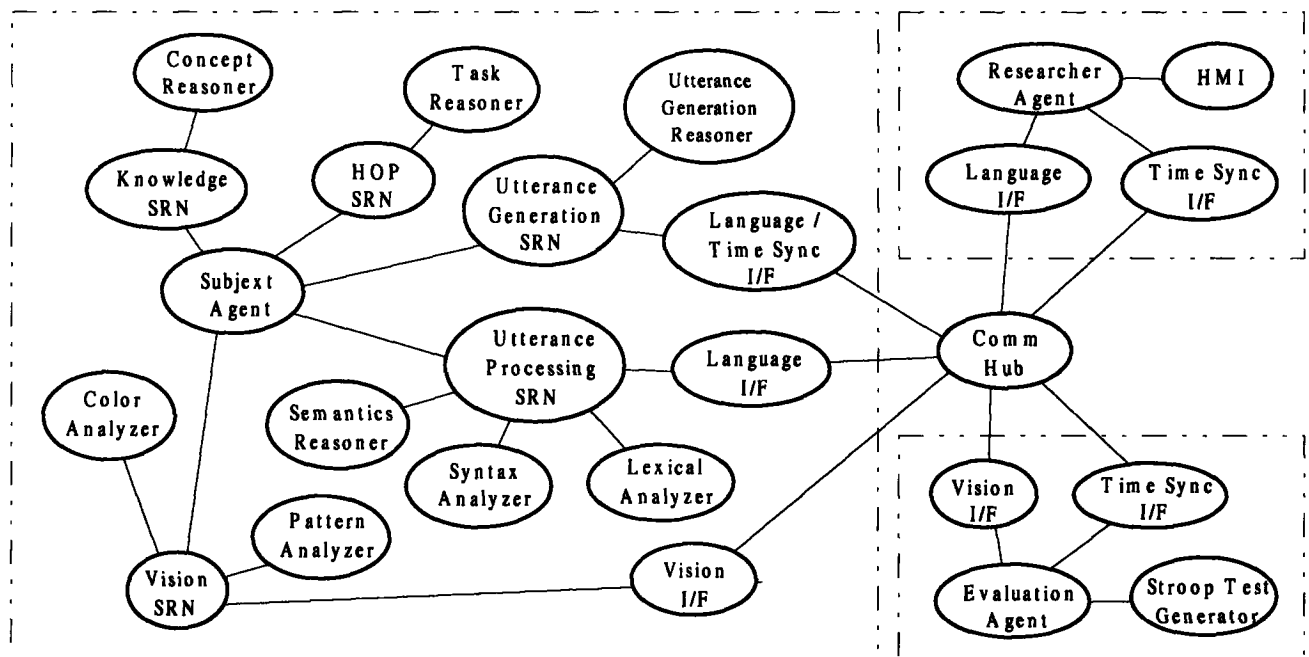


Figure 2. The STRESS Model

An SRN is best viewed as a special purpose etheron designed to support dynamic name-based routing of messages. (In reality, they are built from the same library of functions as any other etheron but just do not contain a very complex inference method and have all of their communication channels turned on.) SRNs can be used to route point-to-point, localcast, and broadcast messages. They are also designed to support message filtering and redirection; however, in STRESS we only used their localcast and broadcast capability.

All etherons labeled as I/F in Figure 2 serve the basic purpose of converting between inter-agent messages on one channel to stimuli messages used internally within an agent. They are best viewed as simulations of sense sensors and motor actuators. Any etheron labeled as an Analyzer is a coded response mechanism (or non-generic reasoner) based on AMEBA's generic database lookup tool. The Task Reasoner uses AMEBA's temporal logic based production system. The Subject's Utterance Generation and Semantics Reasoners are based on the same semantic network reasoner that is used by the Concept Reasoner.

### The Stroop Test System Operation

The heart of the Stroop Effect cognitive mechanism in STRESS is the Concept and Task Reasoners. The other etherons provide both cognitive task and implementation support for these two components. Table 1 provides a summary of the relevant stimuli types used in STRESS.

The STRESS system's two special purpose user interfaces are then attached to the Researcher and Evaluation Tool agents using the same user interface. At the Evaluation Tool interface a word pattern can be selected and the ink color of the word defined. This information is graphically presented to the user. At the Researcher interface, the Subject agent can be told (using natural language) to attend to either the word's pattern or its ink color. The user can then send the visual stimulus to the Subject by pressing a Send button.

When the Send button is pressed, the word pattern and its color are sent in a single message to the Subject as four values, the Red, Green, and Blue (RGB) values of the color and a pattern ID. At the same time, a Time Sync message is sent to the Researcher agent.

When the message containing the RGB values and pattern ID is received by the Vision I/F etheron of the Subject agent, it is turned into a SEE stimulus which is localcasted to the Color and Pattern analyzers. The Color Analyzer mixes the RGB values to calculate a spatial concept of the resulting color and broadcast this result in a COL stimulus. The Pattern Analyzer uses the pattern ID to generate a semantic concept of the word and broadcast this result as a PAT stimulus.

While every etheron in the Subject agent sees both the COL and PAT stimuli, only the Task Reasoner does anything with them. What it does depends on its current mental state which can be toggled between a word reading

**Table 1. Relevant Stimuli Types used in STRESS**

Stimuli Label	Information Contained
COL	A spatial color concept
PAT	A semantic word pattern concept
REF	A conceptual relationship
?REF	Question about a conceptual relationship
RSP	A deep structure verbal response
SEE	Raw vision data
SEM	An utterance semantic relationship
STM	A time synchronization pulse

and color matching task by the SEM stimuli generated as a result of the processing of an imperative utterance from the Researcher. In the word reading task mode, the Task Reasoner generates and broadcasts a RSP stimulus containing the semantic concept of the word pattern. This RSP is converted by the Utterance Generation etheron cluster into an utterance sent to the Researcher agent.

In the color matching task mode, the Task Reasoner generates and broadcasts a ?REF stimuli containing the question:

*?word-pattern color-name spatial-color*

When the Concept Reasoner gets the ?REF stimulus, it activates both the *word-pattern* and *spatial-color* concepts, then looks to see if there is a *color-name* intersection between the two concepts (i.e., you can follow a relation chain between the two concepts that contains any number of *is-a* relationships but one and only one *color-name* relationship). This intersection will only exist in the congruent case. It then broadcasts a REF (i.e., the answer to a ?REF) stimulus that says that the *word-pattern* is the color name.

If the *color-name* intersection fails, the Concept Reasoner broadcasts a REF stimulus indicating that no intersection exists. According to the theory of spreading activation, if the *word-pattern* is a conflicting color-name, its activation should inhibit all other color-name concepts for a given set of time. This is simulated in the Concept Reasoner by testing the relation, *word-pattern is-a color-name*, and stalling the inference thread if this question answers True.<sup>1</sup> If the Task Reasoner gets the REF stimulus

<sup>1</sup> While stalling the thread of execution which is responsible for color lookup in the concept reasoner might at first appear to be simply a trick to make the simulation work, at the model's current level of abstraction this is the best way to simulate the expected result of spreading activation on a color-lookup task. A more general component would need to control the spreading activation at an individual-concept level, but the resulting method would still be only a simulation of the actual brain process even if it was completely connectionist-based.

reporting failure, it sends a ?REF stimuli containing the question:

*spatial-color spatial-color-of ?Color-Name*

When the Concept Reasoner gets this ?REF stimulus, it tests for a *spatial-color-of* intersection between the *spatial-color* and the *Color-Name* (the root concept of all color names) concepts. This intersection cannot fail (since we know the name of all spatial colors generated by the Color Analyzer) but the time at which the test starts can be delayed if the *word-pattern* concept was a color name since the inference thread may still be stalled. Thus, the color name for both the control and conflict case will be found in this second inference, but in the case of a conflict, the inference has to wait until the simulated inhibiting field is removed from all other color names.

When the final REF stimulus containing the color name answer is broadcast by the Concept Reasoner, this stimulus is detected by the Task Reasoner which has been primed to expect it (when it sent either the first or second ?REF). Since this is now a semantic concept, the Task Reasoner generates and broadcasts a RSP stimulus containing the semantic concept of the color name. This RSP is again converted by the Utterance Generation etheron cluster into an utterance sent to the Researcher agent.

When the Language/Time Sync IF etheron of the Subject agent gets the word to be sent to the Researcher, it sends this over the language channel of the inter-agent messaging system. It also sends a Time Sync message over the time channel. When the Researcher agent's Language I/F etheron gets the response word, it is converted to an utterance stimulus which is displayed at the Researcher agent's user interface. At the Time Sync I/F, the Time Sync message is compared to the one sent by Evaluation Test agent and a duration of test stimulus is sent to the Researcher agent's user interface.

## Results

STRESS was tested on an SMP cluster of 16 total processors. The topology of this system is eight two-processor SMP machines connected via two switched and one hubbed 100baseT networks. Tests were run on 1, 3 and 4 SMP machines by first dividing the model across agent boundaries, and then, dividing the Subject agent across two SMP machines. In our language understanding and learning system, we have been able to see some speed up by breaking the system across both agent and etheron cluster boundaries; however, STRESS is not big enough to gain much direct speed up.

The results of a series of tests running STRESS on a single SMP machine are presented in Figure 3. These values represent the average trial time of the values that clustered around the median time of test. This method was used instead of a straight average since some evaluation tests took as much as three times as long as the median.

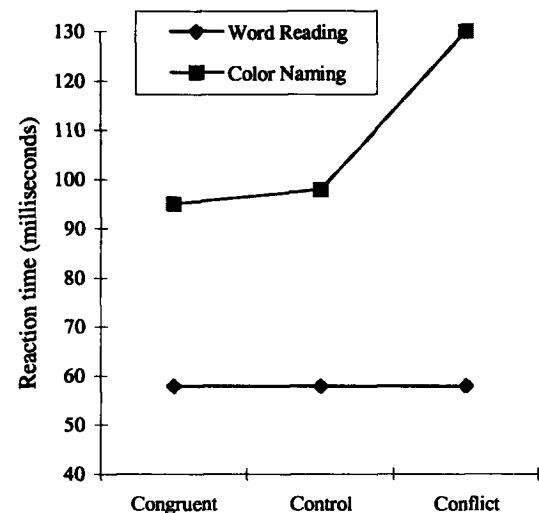


Figure 3. STRESS Performance

While access to the high-speed networks being used to link our SMP cluster is carefully controlled, it is suspected that these long test times were a result of network interference caused by utility messages within the network.

As Figure 3 demonstrates, we were able to get results which are in many ways similar to the average human performance on the Stroop test. We believe the differences in both the relative time between the reading and color naming task and the differential impact of constructive and destructive Stroop interference can both be explained by the crudeness of our current computational model of unified cognition. The current model provides far too much independent control over the time that the mismatched color inhibitor field remains active in the conceptual reasoner, but we feel that we have taken the necessary steps to justify that our choice of inhibit time is relative to the sum of the rest cycles of the expected number of neurons that would chain-fire during a similar activity in the brain. While the mechanism we used for generating the Stroop Effect cannot be fully proven by this current research, the current research does provide some interesting results which seem to support the following conclusions:

- the difference in word pattern reading and color naming can be explained by the amount of effort required to translate the spatial color concept to a semantic concept representing that color.
- the constructive and destructive Stroop Effect on the color-naming task can be partially explained by the concept priming and the generation of an inhibiting field on similar concepts suggested by the Spreading Activation Theory.
- symbolically based testing environments, like AMEBA, can be used to demonstrate these types of analytical results within the context of a larger unified model.

## Future Work

The current work with STRESS can be extended to gain a further understanding of both the Stroop Effect and its impact on a broader unified model of cognition. One clear extension would be to study the effect of different modality relationships (e.g., spatial-to-semantic, spatial-to-spatial, etc.) between the input stimuli and the output response. The effect of modality has been addressed in a number of psychology experiments so there already exist a large body of evidence on which to base such a model. The explanatory depth of the method used to simulate spreading activation in the conceptual reasoner also needs to be improved to support a more general simulation of this phenomenon and to test the assumptions we made in selecting an appropriate inhibitor field activation time.

While AMEBA is designed to allow models to be further and further refined, our current ability to model general vision, language and higher order processes can already be broadened and refined based on the current Stroop Effect model. As expected, STRESS pointed out several minor weaknesses in our original language learning and understanding model. It is this interplay of models which we hope will allow AMEBA to contribute to the overall field of cognitive science.

The next detailed modeling domain we would like to address is the Wisconsin Card Sorting Test (WCST). The WCST will provide an opportunity to study such things as mental sets and task planning. In addition, the further modeling of language generation needs to be addressed.

## Conclusion

STRESS shows that it is possible to build a cognitive model of the Stroop Effect that can be examined symbolically. While it was primarily designed to further study AMEBA as a modeling environment and to discover weaknesses in our general language model, the model does point out which formal theories of the Stroop Effect cognitive task seem most reasonable from an early computation model of its mechanism. As more computational models of well-known human performance test are studied within the common framework provided by AMEBA, our ability to improve general AI applications based on these models should continue to increase. We may never need to directly simulate the Stroop Effect in an AI application, but an understanding of how it can be modeled should aid in the construction of language and vision processing elements for future AI-based interfaces.

## References

- Anderson, J. R. 1995. *Cognitive Psychology and its Implications*. New York: W. H. Freeman and Company.
- Anderson, J. R. and Lebiere, C. 1998. *Atomic Components of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Besner, D., Stolz, J. A., and Boutilier, C. 1997. The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, 4, 221-225.

Brega, A.G. and Healy, A.F 1999. Sentence Interference in the Stroop Test. *Memory & Cognition* 27(5):768-778.

Hannon, C. and Cook, D. J. 2001. Developing a Tool for Unified Cognitive Modeling using a Model of Learning and Understanding in Young Children. To appear in *The International Journal of Artificial Intelligence Tools*.

Knight and Grabowecky 1995. Escape from Linear Time: Prefrontal Cortex and Conscious Experience. In Gazzaniga, M. S. (Editor-in-Chief) *The Cognitive Neurosciences*. MIT Press: Cambridge 1357-1371.

Levine, D. S. 1991. *Introduction to Neural and Cognitive Modeling*. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacCleod and Dunbar 1988. Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14:126-135.

Martindale, C. 1991. *Cognitive Psychology, A Neural-Network Approach*. Brooks/Cole Publishing Co., Pacific Grove, CA.

Shimamura, 1995. Memory and Frontal Lobe Function. In Gazzaniga, M. S. (Editor-in-Chief) *The Cognitive Neurosciences*. MIT Press: Cambridge 803-813.

Stroop, R. 1935. Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology* 18:643-662.

Virzi, R. and Egeth, H. 1985. Toward a translational model of Stroop interference. *Memory & Cognition* 13(4):304-319.