The Annals of the New York Academy of Sciences: Structure and Function of the Human Prefrontal Cortex, Vol. 769, 1995

A Computer Model of Prefrontal Cortex Function

Richard Levinson

Recom Technologies Inc.
NASA Ames Research Center
Computational Sciences Division
Mail Stop: 269-2, Moffett Field, CA 94035
E-mail: `levinson@ptolemy.arc.nasa.gov`

This manuscript describes a computer model of prefrontal cortex function that was designed by integrating the perspectives of neuropsychology and artificial intelligence. The model shows how several neuropsychological theories of frontal lobe function can be combined into a single computer model. The model also extends those component theories by focusing on information processing details that glue the pieces together.

This work is motivated by the following points: Neuropsychology and Artificial Intelligence (AI) describe complementary parts of a model for autonomous action. Neuropsychological models provide descriptions of how planning and reaction must be integrated for human autonomy, but the planning component is poorly understood and many information processing details have not been flushed out. In contrast, AI provides implementations of independent planning and reaction modules, but their integration is poorly understood. A computer model forces information processing issues to be addressed in detail. A computer model can also be tested more easily than models based on verbal descriptions because it produces behavior that can be compared directly with clinically observed behavior. In contrast, verbal models must be interpreted subjectively to predict and test their behavior.

NASA is interested in applying this model towards the development of autonomous instruments and spacecraft. Today's autonomous control technology is limited by executive function deficits that are similar to those found in frontal lobe patients. The control technology works well in pre-programmed situations, but it cannot reprogram itself to handle novel events. Our goal is to extend AI planning methods to simulate human executive functions in real-time closed-loop control applications. We also intend to use the model for a cognitive rehabilitation application. Earlier descriptions of the model can be found in [**?**] and [**?**].

An Information Processing Model

This model is a synthesis of several neuropsychological theories which will be reviewed in the next section. We designed our model by disassembling the functional components of the neuropsychological theories and then recombining them so that the functions are grouped together based on the type of knowledge representation and information processing involved. Thus the model components are organized from an information processing perspective. This organizing principle shows how the information processing requirements for computer implementation have guided our model definition.

An overview of our model is shown in Figure 1. The prefrontal cortex stores symbolic (linguistic) memories of conditioned action sequences called programs. In humans, these programs correspond to functional neural circuits (also called pathways). The programs encode memories of long duration action sequences that remain active to direct behavior in the absence of sensory stimuli. Routine sensory conditions activate *default* program pathways, and novel sensory conditions activate *deliberate* program variations.

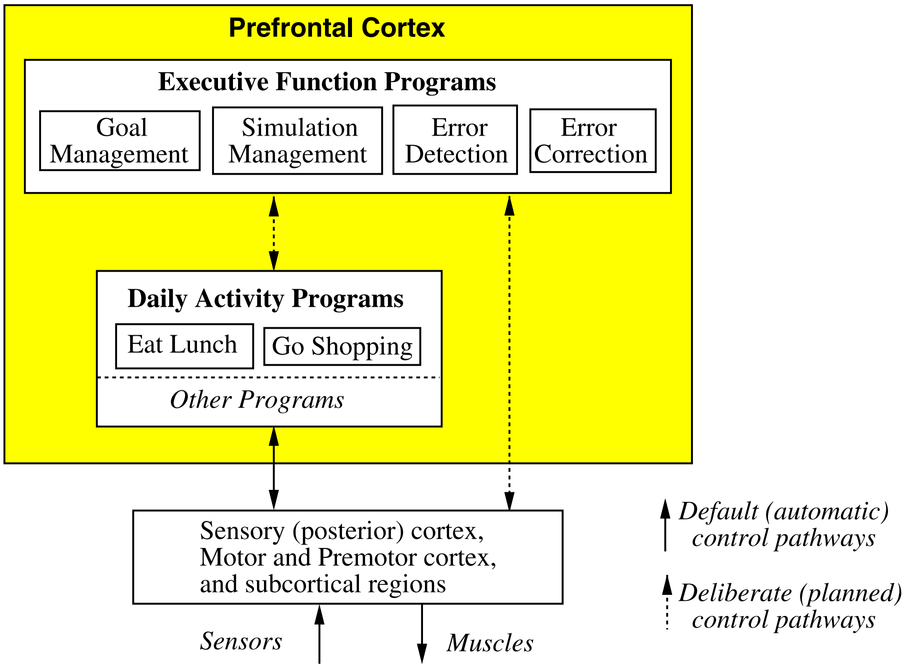

Figure 1: Programs stored in the prefrontal cortex

Some prefrontal programs encode behaviors for daily activities such as cooking, shopping, and bathing. Other programs encode executive function behavior for self-monitoring and self-programming. The *Executive Functions* determine when and how to replace the automatic, default programs with deliberately planned variations. These executive functions are summarized in Table 1.

Function Name	**Function Definition**	**Deficit**
Goal Management:	*Maintain* GOAL *structures that map* CONDITIONS *into reward values*	
Add	Add a new GOAL	Inactivity
Remove	Remove an old GOAL	Perseveration
Change	Change reward value associated with GOAL	Perseveration
Simulation Management:	*Simulate* PROGRAMS *to predict their effects*	
Goal	Triggers (re)planning when GOALS are modified	Unawareness
Condition	Triggers replanning when conditions change	Unawareness
Execution	Triggers replanning when execution failures occur	Unawareness
Deadline	Triggers replanning as deadline approaches	Unawareness
Error Detection:	*Analyze* SIMULATIONS *to detect program errors*	
Ineffective	Routine preconditions fail due to sensory conditions	Unawareness
Interfering	Routine succeeds, but causes another program to fail	Distractable
Irrelevant	Routine is triggered by sensory conditions and does not fail, but it is unrelated to *active* GOALS	Distractable, Stimulus-bound
Error Correction:	*Replace default* PROGRAMS *with deliberate variations*	
Inhibit	Override a default *start-condition* that is currently TRUE	Disinhibition
Start	Override a default *start-condition* that is currently FALSE	Poor initiation
Continue	Override a default *stop-condition* that is currently TRUE	Poor persistence
Stop	Override a default *stop-condition* that is currently FALSE	Perseveration
Choose	Override a default *choice point selection* in program body	Perseveration
Sequence	Override a default subroutine *order* in program body	Poor sequencing

Table 1: **The Executive Functions**. *Deficit* indicates the effect if function is impaired.

The main focus of our model is on the definitions of the *Executive Function* programs, and the four knowledge representations they process. These four representations are:
1) CONDITION structures that represent symbolic descriptions of environmental conditions.
2) GOAL structures that map sensory conditions into positive and negative reward values.
3) PROGRAM structures that link conditioned action sequences together.
4) SIMULATION structures that record the effects of simulating different program variations.

The information processing performed by the Executive Functions can be summarized as follows: First, the *Goal Management* program generates GOAL structures that identify a person's preferred conditions based on experience. Second, a set of PROGRAMS are learned from experience that represent automatic conditioned action sequences that achieve GOALS in routine situations. These *default* PROGRAMS map CONDITIONS automatically into effector commands in real-time without any deliberation. The main function of the Executive Functions is to monitor the automatic conditioned responses, and to anticipate, detect and correct any errors that occur due to novel CONDITIONS. The *Simulation Management* program produces SIMULATION structures that describe the predicted effects of executing any given PROGRAM in the context of current CONDITIONS. The *Error Detection* program then compares the SIMULATION predictions with the GOALS in order to detect potential errors. If errors are detected, then the *Error Correction* program overrides the default PROGRAM in order to maximize GOAL achievement.

An important feature of our model is that it is being implemented on top of an AI planning system, called PROPEL[?], that provides implemented methods for representing, executing, and simulating PROGRAMS. We have previously used PROPEL to demonstrate how a single PROGRAM representation can be used to produce both default reactions and deliberate plans[?]. This system combined both default and deliberate modes of action, but it was lacking a theory of how to coordinate the two modes. The executive function programs were designed to provide that capability.

We will now further describe how the executive functions process the above representations. To illustrate the concepts, we will use the following simplified PROGRAM that represents the routine for doing laundry every Thursday. The `Do-Laundry` program represents a conditional action sequence that will repeatedly wash the clothes, wait 20 minutes, and then dry the clothes, until the laundry basket is empty.

Program Name:	`Do-Laundry`
Start Condition:	`Today is Thursday and Detergent is not empty`
Stop Condition:	`Laundry basket is empty or Detergent is empty`

Program Steps:	`Put clothes and detergent in washer`
	`Start washer`
	`Wait 20 minutes`
	`Put clothes in dryer and start dryer`

The **Goal Management** program generates and updates the GOAL representations that map CONDITIONS into positive and negative reward values. GOALS can be added, removed, or changed. An initial set of top level GOALS for safety and health are predetermined. New (sub)GOALS are then generated when the Simulation Manager generates SIMULATIONS with hypothetical conditions that are associated with high rewards. Without a well defined and maintained set of GOALS and values, goal-driven behavior is impossible.

The **Simulation Management** program generates and updates the SIMULATION representations. When this program is triggered by one of the four events listed in Table 1, it uses PROPEL's program simulator to predict the effects of a given program in the context of current CONDITIONS. The simulation process is coupled with an AI search method that allows different program variations to be simulated and tested against the GOALS in order to find the most effective behavior. The resulting SIMULATIONS are used for Error Detection and to identify predicted conditions that serve as *plan assumptions* which need to be monitored.

Simulation is triggered by the following events: A GOAL event occurs when the Goal Manager modifies a GOAL structure. Adding new goals will trigger the initial simulation of the default response to the goal. Removing goals or changing their reward value will trigger replanning that involves backtracking through the space of program variations. The CONDITION event triggers replanning if an updated sensory condition conflicts with a predicted condition that was generated by the simulator (a plan assumption). For instance, it may be necessary to delay the start of `Do-Laundry` if the washing machine breaks on Wednesday, the day before laundry day. The EXECUTION event triggers replanning when a default program fails *during* execution, as when someone drops a coffee mug. Unlike with the CONDITION event, this is the case when program execution has already begun and the motor pathways have already been activated. The DEADLINE event occurs when a GOAL's deadline moves within some threshold temporal distance. This event could trigger replanning using updated sensory CONDITIONS five minutes before execution.

Simulation Management allows the brain to detach mental activity from real-time events in order to reason about SIMULATIONS of the past and the future. This enables the mental independence from sensory CONDITIONS that is required for abstract, non-concrete thought. Implementation of this component will require resolution of many open issues. One issue involves choosing the best order in which to try different program variations when replanning. Other issues involve deciding how much time to spend planning, and understanding how to simulate actions at an abstract level without simulating every little detail.

The **Error Detection** program analyzes SIMULATIONS to detect the three error patterns described in Table 1. INEFFECTIVE routines are detected if the simulation record indicates that a program's preconditions could not be satisfied. For example, if there is no detergent to wash the clothes. This is called a *program failure*. INTERFERING programs are detected when the SIMULATION indicates that multiple PROGRAMS conflict over a shared resource, resulting in at least one program failure. For example, doing laundry may interfere with taking a hot shower if the hot water is limited, or a phone call may distract someone from moving the clothes from the washer to the dryer. IRRELEVANT routines are detected if the SIMULATION does not show any program failures, but it doesn't show any goal achievement either. For instance, taking the garbage out on the regularly scheduled night is irrelevant if it won't be collected due to a holiday. A frontal patient who begins to bake cookies whenever she sees an oven also illustrates the type of deficit that can occur if this function is impaired. If one of these error conditions is detected, then the Error Correction program is activated.

The **Error Correction** program analyzes the SIMULATION and PROGRAM structures in order to generate PROGRAM variations that override inappropriate default programs. Routine PROGRAMS contain default *start* and *stop* conditions that automatically trigger the start and stop of program execution. The first four Error Correction methods in Table 1 are INHIBIT, START, CONTINUE and STOP. These methods correspond to the 2 by 2 matrix produced by overriding the *start* and *stop* conditions in the *true* and *false* cases. For example, the start of the above `Do-Laundry` program can be deliberately postponed from Thursday until Saturday by using INHIBIT on Thursday and START on Saturday. Also, the default stop conditions of `Do-Laundry` can be overridden by using CONTINUE when the detergent runs out or by using STOP when there is not enough time to wash all the clothes in the basket.

Underlying Neuropsychological Theories

Our model combines the following neuropsychological theories, and it extends them by elaborating on the information processing details described above. Luria's early description of the frontal lobes as system for the *"programming, regulation, and verification of activity"*[**?**] is perhaps the best one sentence description of prefrontal cortex functionality. However, it is too general to use as a design for computer implementation. We therefore looked for the most detailed descriptions of executive function dimensions we could find, and we selected the models proposed by Sohlberg and Mateer[**?**], and by Lezak[**?**].

Sohlberg and **Mateer** developed the Executive Function Behavioral Ratings Scale (EFBRS)[**?**], to assess behavioral dysfunction after head injury. The EFBRS includes three main executive function components which are further divided into subcomponents. The first component, *Selection and Execution of Cognitive Plans*, involves the ability to describe goals and procedures, to determine appropriate action sequences, to initiate activity, to repair plans, and to maintain persistent effort until a task is completed. The second component of the EFBRS, *Time Management*, involves the ability to generate realistic schedules, and to perform the scheduled activities within given time constraints. The third component, *Self-regulation*, involves using feedback to control behavior and to inhibit inappropriate reactions.

Lezak describes four essential components of executive function[**?**]. First, *Goal Formulation* is the ability to generate and select descriptions of desirable future states. Second, *Planning* involves the selection of steps, elements, and sequences needed to achieve a goal. This requires the ability to recognize and evaluate choices. Third, *Carrying Out Activities* involves the ability to start, stop, maintain, and switch between planned actions. These subcomponents led directly to several of the Error Correction methods shown in Table 1. Fourth, *Effective Performance* involves the ability to monitor and repair activities.

We began the design of the executive function programs by making a list of the executive function components as described by the EFBRS and by Lezak. In order to facilitate computer implementation, we then reorganized those functions based on the type of information being processed. We have tried to capture all of the functionality of these two models in our own. However, due to our functional reorganization, there is not always a one-to-one mapping between their components and our own. Instead, some of their components can be found distributed *across* several of our model components.

The **Norman** and **Shallice** model of the frontal lobes as a Supervisory Attentional System (SAS)[**?**] corresponds strongly to our model's Executive Functions, and is also based (loosely) on an AI information processing model. This model, like our own, proposes two modes of action: automatic *default* responses for routine situations, and *deliberate* planned responses for novel situations. Shallice and Burgess propose that the frontal lobes serve as a Supervisory Attentional System that is required to: (1) inhibit undesirable automatic responses, and (2) generate and execute desirable new responses. The SAS model helped to define the role and function of our executive functions. It also supported our independently developed method for using both default and deliberate action modes. Our work extends the SAS model by providing an implemented PROGRAM representation and by elaborating on exactly how the SAS may detect and correct program errors in novel situations.

Grafman proposes that the prefrontal cortex contains representations called Structured Event Complexes (SECs)[?]. SECs encode memories of long duration action sequences that guide behavior through well-learned activities. SECs correspond strongly to PROGRAMs in our model. Grafman describes SEC hierarchies, where the highest SECs are called Managerial Knowledge Units (MKUs) because they manage other SECs. These MKUs correspond to our model's Executive Function PROGRAMS. This SEC-MKU model led us to adopt the idea that the daily activities and the executive functions in our model are all represented as PROGRAMS, rather than requiring a different representation for the executive functions.

Our work extends the SEC model by describing an implemented PROGRAM representation and a planning system that can be used to anticipate, detect and correct program errors. Also, our Executive Function programs define a specific set of MKUs that have not previously been discussed in detail. Our model predicts that two other representation types, GOALS and SIMULATIONS, are stored in the prefrontal cortex in addition to the SEC-like PROGRAMS. Grafman also elaborates on issues we have not addressed such as the effect of different neural activation patterns on SEC development and learning.

Stuss' view of self-awareness [?] is also an important part of our model. Frontal lobe patients often have intact awareness of their sensory environment despite impaired awareness of the interactions between themselves and the environment. Stuss proposes that frontal system damage can impair the awareness of one's *Self* as a continuity from the past into the future. This concept of self is represented in our system by the GOALs, which identify a value system that is learned from experience. In our model, *self-awareness* is the process of relating one's GOALS to external CONDITIONS and SIMULATIONS. This is achieved by the combined functions of our Simulation Management and Error Detection programs.

Evaluation Methodology

The model's definitions of the executive functions and the representations they manipulate identify hypothetical neurological functions and connections and mental representations that can be tested for validity in various ways. Clinical studies can be designed to test for the presence of these model elements. Our plan is to have frontal lobe patients perform daily living tasks in a computer simulated world. We will then connect our model to the same simulated world and compare the performance of our model with that of the patient. The model will be "lesioned" to produce behavior that is similar to that of the patient. This will allow us to compare the behaviors of the model and patient directly rather than subjectively interpreting the behavior predicted by a verbal model. The predicted deficits produced by these lesions are shown in Table 1.

Another benefit of our model is that we can test it on large scale tasks that last minutes, hours, or even days. PROPEL's program representation is expressive enough that we can represent complex daily living activities such as making dinner. In contrast, many frontal lobe models are based on analysis of simplistic tasks such as the Towers of London (or Hanoi), Block Design, or the Delayed Response Task. We propose that directly measuring the model's performance on large-scale daily activities such as shopping and cooking will be a better measure of the ecological validity of our model.

Conclusion

The PROPEL substrate for our system is currently implemented, providing our basic ability to represent, simulate and execute PROPEL programs. The Executive Function programs, however, are still under development. Many difficult technical design and implementation issues remain unresolved. Our current efforts are directed towards extending the system to support the model's full functionality.

We have presented a computer model of prefrontal cortex function that combines several frontal lobe and executive function models based on information processing principles. Before the model is complete, many open research questions will have to be answered from both the neuropsychological and the information processing perspectives. We hope our model definition provides motivation and a common language for future interdisciplinary efforts to improve upon this start.

Acknowledgements

Thanks to McKay Sohlberg, Tom Boyd, Jordan Grafman, Jeffrey Englander, Richard Delmonico, Peter Robinson, and Steve Farmer for providing many useful comments that have influenced the shape of this work.

References

[1] Levinson, R. 1994. Human Frontal Lobes and AI Planning Systems. *Proceedings of the Second International Conference on AI Planning Systems.* AAAI Press, Menlo Park, CA.

[2] Levinson, R. 1995. An interdisciplinary theory of autonomous action. 1995 AAAI Stanford Spring Symposium on Extending Theories of Action, Stanford, CA.

[3] Levinson, R. 1995. A General Programming Language for Unified Planning and Control. *Artificial Intelligence* journal special issue on Planning and Scheduling. *To appear 1995.*

[4] Luria, A. R. 1973. *The Working Brain.* Basic Books Inc. New York.

[5] Sohlberg, M. M. & C. Mateer. 1989. *Introduction to Cognitive Rehabilitation.* Guilford Press, New York.

[6] Lezak, M. 1983. *Neuropsychological Assessment.* Oxford University Press. New York, NY.

[7] Shallice, T., Burgess, P. 1991. Higher-Order Cognitive Impairments and Frontal Lobe Lesions in Man. In *Frontal Lobe Function and Dysfunction*, edited by H. Levin, H. Eisenberg, and A. Benton. Oxford University Press, New York.

[8] Grafman, J. 1989. Plans, Actions and Mental Sets: Managerial Knowledge Units in the Frontal Lobes. *Integrating Theory and Practice in Clinical Neuropsychology.* E. Perecmen, Eds. Erlbaum Press. Hilldale, NJ

[9] Stuss, D. 1991. Disturbance of Self-Awareness after Frontal System Damage. *Awareness of Deficit After Brain Injury.* Prigatano and Schacter, Eds. Oxford University Press. New York