

# Neuronal Models of Prefrontal Cortical Functions

STANISLAS DEHAENE<sup>a,b</sup> AND JEAN-PIERRE CHANGEUX<sup>c</sup>

<sup>a</sup>*Institut National de la Santé et de la Recherche Médicale  
Paris, France*

<sup>c</sup>*Institut Pasteur  
Paris, France*

In the last 10 years, considerable advances have been made toward understanding the functions of prefrontal cortex and their anatomical and neurophysiological counterparts.<sup>1,2</sup> Relating these behavioral and neurobiological data into a coherent picture, however, remains a challenging enterprise. The development of explicit models of prefrontal cortex architecture and functions can potentially help to bridge this gap by identifying the most relevant features of cellular and behavioral data, by testing the plausibility of hypotheses put forward to relate these data in a causal manner, by drawing attention to specific experimental predictions, and by pointing to unsolved questions.

We have tried to achieve some of these goals in the limited context of simple behavioral tests of prefrontal functions. Over the years, our modeling efforts focused successively on delayed-response tasks,<sup>3</sup> on the Wisconsin Card Sorting Test,<sup>4</sup> and on the possible contribution of prefrontal cortex to the development of numerical competence.<sup>5</sup> In each case, we speculated on the implementation of the task in neuronal networks, and we described computer simulations of formal neuronal networks whose properties reproduce, to some extent, the available neurophysiological and behavioral data. Although the details of the models varied, common principles of neural architecture were used for all tasks. In this review, three such principles will be discussed: (1) the distinction of levels of organization, with prefrontal circuits modulating lower-level networks, (2) the role of long-lasting neuronal activity in maintaining representations of task events, and (3) the interconnection of these representations with reward systems that compute the value associated with actions or events.

The general framework in which our work has developed is that of *neural or mental selectionist mechanisms*.<sup>6,7</sup> It stresses that organisms are not passively responsive to environmental inputs and do not absorb knowledge by instruction from an external teacher. Rather, higher organisms function in a projective mode in which hypotheses or prerepresentations are internally generated and are maintained or rejected depending on their adequacy to the situation at hand. Learning proceeds by selective elimination of spontaneously generated alternatives. Thus, behavior is not driven by quasireflex responses to stimulation, but by internal goals and by representations of past events and of future actions that may be

<sup>b</sup> Address correspondence to Dr. Stanislas Dehaene, Laboratoire de Sciences Cognitives et Psycholinguistique, Centre National de la Recherche Scientifique and Ecole des Hautes Etudes en Sciences Sociales, 54 Boulevard Raspail, 75270 Paris cedex 06, France.



relevant to these goals. In mammals, prefrontal cortical circuits seem instrumental in generating goals, in maintaining representations of goal-relevant information, and in selecting these representations as a function of their expected value for the organism. Our simulations illustrate how specialized prefrontal circuits may implement these functions.

### LEVELS OF ORGANIZATION

A first key feature of our models is their organization in multiple hierarchical and parallel levels. The notion that prefrontal cortex intervenes at a level of representation higher than that of other cortical circuits is common to almost all accounts of prefrontal cortical functions. Luria<sup>8</sup> considered the frontal lobes as crucial for the programming, regulation, and verification of activity. According to Grafman,<sup>9</sup> prefrontal cortex maintains and controls the execution of complex scripts or hierarchical plans for actions. Fuster<sup>1</sup> and Goldman-Rakic<sup>2</sup> have also emphasized the role of prefrontal cortex in modulating lower-level sensory-motor contingencies using working-memory representations of the organism's intentions and past knowledge. Finally in the Norman-Shallice theory,<sup>10</sup> prefrontal cortical areas together form a supervisory attentional system that can inhibit or select lower-level automatized thought or action schemata.

Our models<sup>3-5</sup> have implemented these ideas by introducing a layered architecture with multiple parallel mappings between hierarchical representations of sensory inputs and of intended motor outputs. At least three levels of processing have been distinguished: a direct mapping between sensory data and corresponding motor actions, an indirect mapping mediated by a working memory of past events, and another indirect mapping holding a representation of the rules of the task at hand. The last two indirect mappings, which are assumed to rely on prefrontal areas, modulate and select actions triggered at the lowest level (FIG. 1).

The functional role of such modulation is well illustrated by our model of delayed-response tasks, including the A-not-B task.<sup>11,12</sup> In these tasks, an object is hidden in one of two possible locations (A and B). The locations are then covered and the subject's attention is distracted for a short delay. At the end of this delay, one then measures the subject's ability to reach towards the appropriate location. When the task is made simpler by reducing the delay to zero, or by repeatedly hiding the object at the same location A, rhesus monkeys with prefrontal cortex lesions, as well as infants between 7 and 9 months of age whose prefrontal cortex is immature, succeed in reaching to the correct location. Hence, reaching to spatial locations and learning to always reach to a certain location are not dependent on prefrontal cortex, but rely on lower-level action schemata. In our simulations, these abilities were embodied in a direct sensorimotor pathway with slowly modifiable connections that linked a representation of the features of the objects to a representation of the available motor responses. This lower level of the model easily learned the reaching part of the task.

When the task is made more difficult, however, by imposing a longer delay between cuing a location and letting the subject reach to it, and by changing the location of the hidden object from trial to trial, young infants and monkeys with



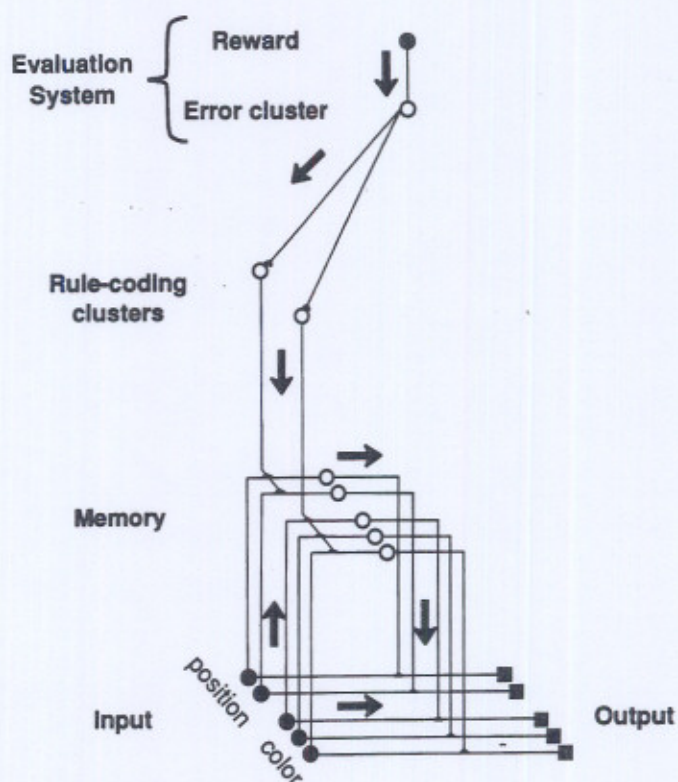


FIGURE 1. Schematic diagram of a neuronal model for delayed-response tasks,<sup>3</sup> illustrating a hierarchical and parallel organization with multiple levels of processing.

prefrontal lesions fail systematically. In the A-not-B task, they continue to reach towards the previously cued location A, even after the object has been shifted to location B. In other words, an immature or lesioned prefrontal cortex yields an impairment in inhibiting a previously learned response. Our simulation with the sensorimotor pathway only behaved quite similarly, making a systematic error of perseveration in reaching to location A (FIG. 2). In order to pass the test, just as neurologically intact monkeys and 12-month-old human infants do, our network had to be supplemented with a second pathway, parallel to the first, but which held a short-term memory of the cued location throughout the delay. Units at this higher level were not allowed to influence directly the motor output units, but only to modulate the connections of the lower level sensorimotor reaching pathway. They held a representation of the past location of the object and used it to bias reaching when two possible locations could be reached at after the end of the delay. These units therefore implemented a form of "working memory" which has been related to dorsolateral sectors of prefrontal cortex.<sup>1,13</sup>

At an even higher level of representation in our models, units that we called



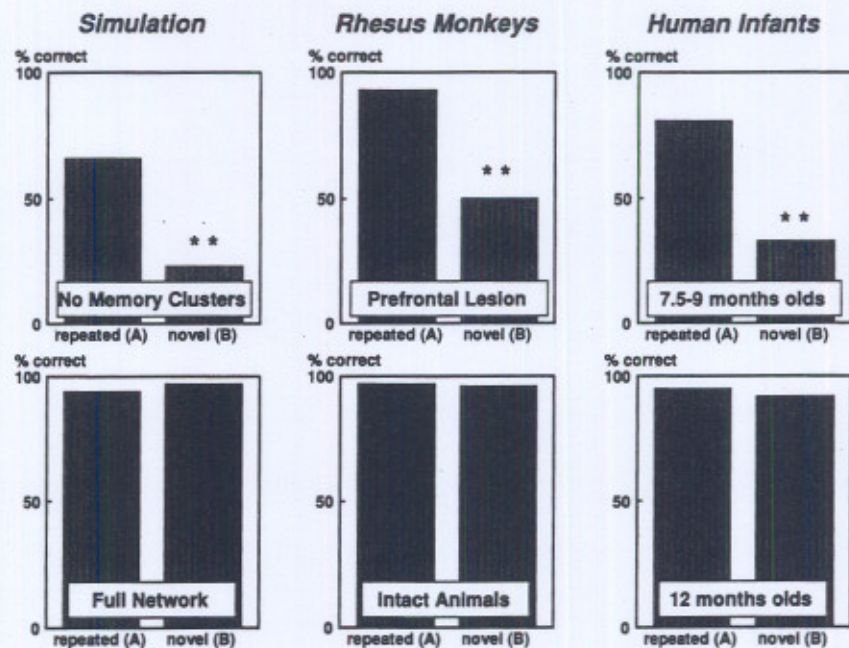


FIGURE 2. Performance of the simulated delayed-response model (*left*) as compared to actual data from rhesus monkeys (*middle*) and human infants (*right*) (data from ref. 12). The model with lesioned memory clusters, like monkeys with prefrontal lesions and 7.5–9-month-old infants, is able to reach to a hidden object after a delay when the object is repeatedly hidden at the same location, but not when the object is switched to a novel hiding location (*top graphs*). The full network model, like intact animals and older infants, reaches correctly to the hidden object in both cases (*bottom graphs*).

rule-coding clusters were allowed to modulate entire sets of connections at the lower levels (see also ref. 14). Their activation therefore drastically affected information processing. For instance, when the rule-coding cluster coding for "color" was activated, color information, rather than spatial information, was paid attention to and stored in the circuit memory. Changing the activity pattern over rule-coding clusters allowed for a very fast change in the rules used by the system, a performance analogous to that of normal human adults in the Wisconsin Card Sorting Test.<sup>15,16</sup> This part of the system therefore effectively performed functions attributed to the "central executive,"<sup>17</sup> "supervisory attentional system,"<sup>10</sup> or "attention for action" system<sup>18</sup> postulated by psychologists to underlie flexible task switching and controlling. Anatomically, this level may rely on orbitofrontal cortex,<sup>19</sup> the anterior cingulate,<sup>18</sup> and other areas forming a prefrontolimbic network.<sup>20</sup>

Our emphasis on levels of complexity should not imply that the connectivity in our models is purely hierarchical and pyramidal. The classical view that prefrontal cortex is an end-point in cortical processing is not tenable.<sup>20</sup> In our models, multi-



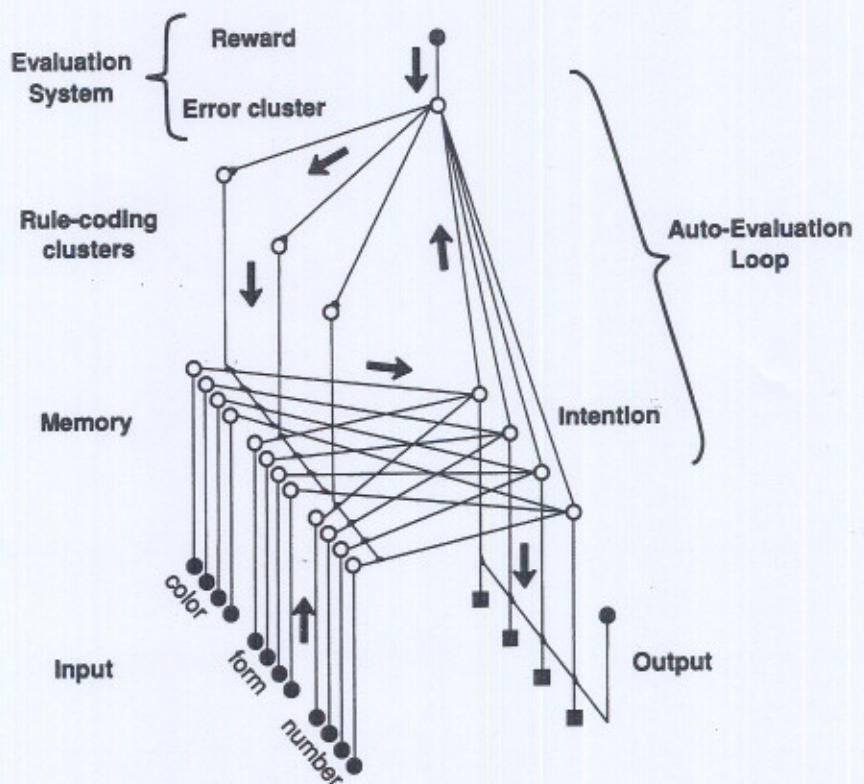


FIGURE 3. Schematic diagram of a neuronal model capable of solving the Wisconsin Card Sorting Test.<sup>4</sup> Note the central role of the evaluation system (*top*), which can selectively inactivate inappropriate rules at the lower levels (*left*), and can itself be internally activated by upcoming intentions via an auto-evaluation loop (*right*).

ple anatomical loops integrate prefrontal representations, together with sensory and motor representations, into parallel circuits for processing of color, form, location, or number information (Figs. 1 and 3). This bears some similarity to the known parallel cortico-cortical circuits linking prefrontal cortex with multiple areas such as posterior parietal, anterior and posterior cingulate, and occipital and superior temporal areas.<sup>20</sup> Recent observations have revealed parallel functional circuits for the representation of object identity and location information in distinct prefrontal areas.<sup>21</sup> Working memory for object identity, on the one hand, rests on the prefrontal area of the inferior convexity, in relation with the occipitotemporal "what" pathway. Working memory for object location, on the other hand, rests on dorsolateral prefrontal cortex, in close anatomical relation to the occipitoparietal "where" pathway. This anatomical organization is similar to the maintained segregation between color and location information at all levels of processing in our models (Figs. 1 and 3).



## LONG-LASTING NEURONAL ACTIVITY

It has been known for at least 20 years that prefrontal neurons can maintain a sustained level of firing for extended periods of time.<sup>22</sup> In our models, as in some others,<sup>23</sup> this is seen as a critical and specific property of prefrontal areas that enables them to hold on-line representations of past events, future intentions, and rules of behavior. We have simulated long-lasting firing using local recurrent excitatory connections within clusters of neurons coding for a given feature (e.g., the color red). These clusters coarsely model the known columnar organization of cortical areas. Theoretical analyses and simulations show that such clusters may possess two levels of activity that are stable in time. Either most neurons in the cluster are inactive, or most neurons fire at a sustained rate. In the latter case, neurons keep a sustained or "remanent" activity because they reactivate each other through fast recurrent connections. Simulations<sup>4</sup> indicate that the temporal activity profiles of neurons within bistable clusters resemble those seen in actual recordings (e.g., ref. 13).

It should nevertheless be stressed that our attribution of long-lasting neuronal activity to local recurrent connections is a strictly theoretical hypothesis that awaits experimental confirmation. Alternative approaches attribute long-lasting firing to single-cell membrane properties, to distant recurrent loops with other brain areas, either cortical or subcortical,<sup>24,25</sup> or altogether disregard long-lasting firing as an important and specific property for the simulation of prefrontal functions.<sup>26,27</sup> Indeed, cells in many areas such as posterior parietal cortex or basal ganglia have also been recorded to keep a sustained level of firing during extended periods of time. In our interpretation, however, this could be due to a propagation of activation originating from local circuits forming neuronal clusters within prefrontal areas. Clearly, further research will be needed to decide between the single-cell, local circuit, and distant loop interpretations of sustained firing. Until then, our working hypothesis is that the ability to keep a sustained level of firing in the absence of sustained inputs from other areas is a specific property of prefrontal circuits that stems from strong local recurrent connections and that is not seen in other cortical regions.

What *function* does sustained firing serve? Goldman-Rakic,<sup>2</sup> Fuster,<sup>1,22</sup> and others have provided convincing evidence that prefrontal cell activity encodes a representation of past or future events. In our models, these events can be of several types. Consider for instance the model capable of solving the Wisconsin Card Sorting Test and depicted in FIG. 3. The cards that have to be sorted are encoded along three dimensions, according to the color, number, and form of the symbols on the card. Different assemblies of memory clusters are allocated to each of these parameters. Throughout a trial and the succeeding intertrial interval, these clusters hold in their activation pattern a memory of the parameters of the input card, long after it has been removed from sight. Activation from the memory clusters is then transmitted to an intention network that codes for the stack in which the network will place the input card. Hence, units at this level code for an anticipation of subsequent motor outputs ("intention") rather than for a memory of past inputs. Neurophysiologically, both types of units have been recorded in prefrontal cortex:<sup>28</sup> during the delay period of delayed-response task, some



neurons fire only after certain types of cues were presented (e.g., red cues), whereas other units fire only before certain types of movement are made (e.g., rightward movements).

Importantly, the memory and intention clusters in our models can be completely isolated from actual inputs and outputs. Memory units can maintain a sustained level of firing in the absence of their original inputs. Likewise, intention unit activity is not propagated to motor activators unless a "go" signal is received. The effect of this "go" signal is to potentiate the intention-to-motor connections and therefore to release a preprogrammed motor command. Before the "go" signal is received, the representations held in memory and intention units function as a "mental model" or "working memory space" that can be used to freely manipulate hypotheses independently of current input-output contingencies.

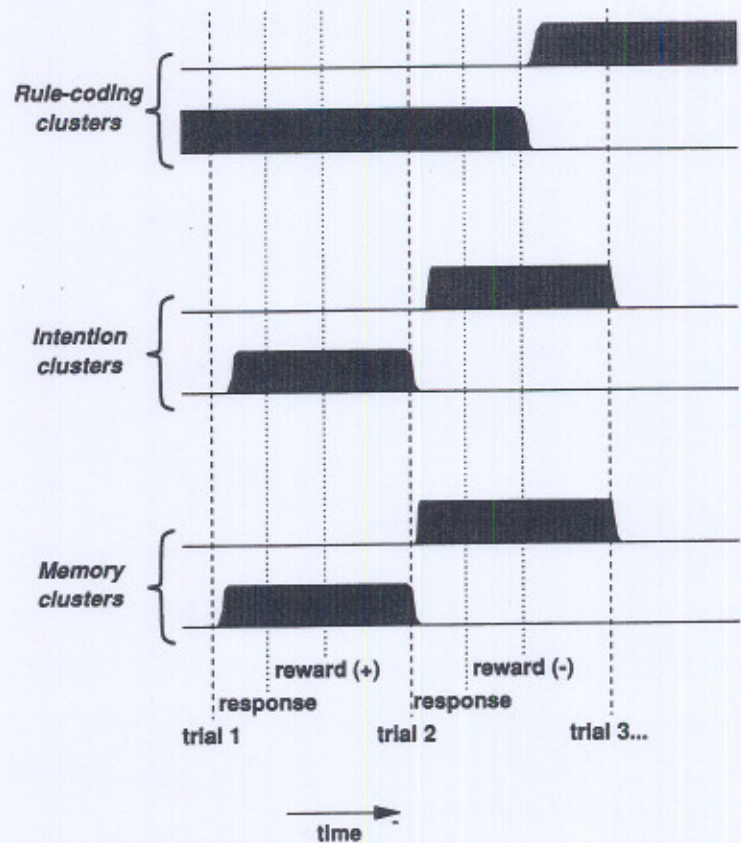
In the card sorting model, a third type of unit also shows long-lasting activation: the rule-coding clusters. Here, long-lasting firing is used to maintain on-line the behavioral rule which is currently being applied and which specifies how the lower-level network will sort the input cards. The Wisconsin Card Sorting Test imposes two contradictory requirements on the rule-coding system. First, a rule must be discarded as soon as it is found not to apply well to the present situation. Second, a successful sorting rule, once found, must be maintained and applied systematically across several trials. According to our model, prefrontal cortex is ideally suited for meeting these demands because rules can be represented as stable activity patterns over rule-coding clusters rather than as slowly modified connection weights. The bistable property of neuronal clusters makes it possible to keep a rule active as long as it is useful, and yet to immediately turn the corresponding circuit off if the sorting rule must be changed.

FIGURE 4 shows how the different kinds of units with long-lasting firing in our model are activated on two consecutive trials of the Wisconsin Card Sorting Test. Even before the first trial starts, a rule-coding cluster is already active. It codes for the sorting rule that the system is going to try first. Upon presentation of a card to be sorted, input units (not shown) are activated and the corresponding memory clusters therefore switch to an active state. The sorting rule is applied, leading to the activation of an intention cluster coding for the upcoming response of the network. When the "go" signal is received, activity is allowed to propagate from intention to output units which execute the intended response. Feedback is then received from the experimenter. On the first trial depicted in FIG. 4, feedback is positive and therefore the active rule-coding cluster is maintained throughout the duration of the trial. On the next trial, feedback is negative. The active rule-coding cluster is therefore turned off and another one is activated, thus implementing the switch from one sorting rule to another.

#### REWARD AND AUTO-EVALUATION SYSTEMS

The example of the Wisconsin Card Sorting Test illustrates the critical role that reward systems play in the architecture of our models. Most real organisms do not acquire information about the environment via a teacher that specifies desired levels of neuronal activity, as in the backpropagation algorithm. Rather,





**FIGURE 4.** Example of long-lasting activity in memory, intention, and rule-coding clusters during two simulated trials of the card sorting task. Negative reward during the second trial induces a change in the activity of rule-coding clusters.

they are exposed, often with some delay, to the positive or negative consequences of their actions, and they learn to anticipate future rewards and to adapt their representations and goals in order to optimize these rewards. Hence, the development of adequate *value systems*, which can evaluate internal representations and use this evaluation to direct behavior, is an essential part of learning and decision making.<sup>3-5,7,19,29</sup> Prefrontal cortex is richly interconnected with limbic areas such as the anterior cingulate as well as with subcortical nuclei, which have been postulated to provide information about the relevance and value of behavior. Furthermore, prefrontal cell activity is often modulated by the relevance, or value, of the situation to the organism. Hence, some sectors of prefrontal cortex obviously play an important role in evaluative functions.

In our models, the interaction between the representations held on-line in pre-



frontal cell activity and their evaluation in limbic and/or subcortical circuits has been modeled as a bidirectional pathway. On the one hand, external rewards such as food or punishment, which are received from the environment, can be transmitted to the appropriate representations and yield an immediate modification of behavior (external evaluation). On the other hand, the same circuitry can also be activated internally, with similar consequences on behavior, because the system has learned to anticipate that a given situation is likely to result in a negative or positive reward (internal evaluation or auto-evaluation).

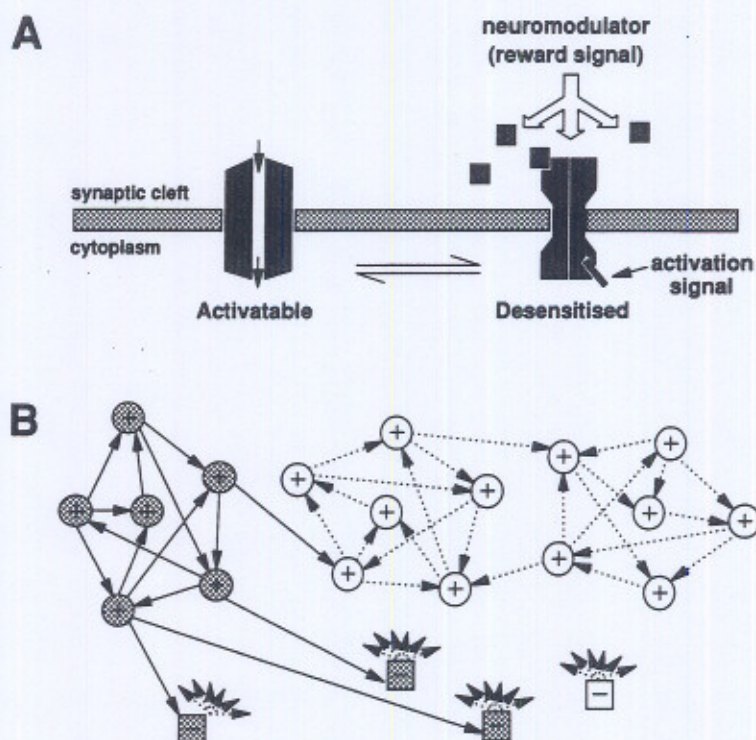
#### *External Evaluation*

In our model, external evaluation works as follows. The system, as a result of being in a certain activity state—for instance, with an active “color” rule cluster—performs a certain motor action (e.g., sorting a card by color). If the experimenter decides that this action was incorrect, negative reward is provided to the network. In turn, the reward input activates an error-coding cluster which signals that the network has performed an error. This error signal then has two effects on the rest of the network: a slow diffuse effect and a fast focal effect. First, the error signal is broadcast via diffuse neuromodulatory systems to all areas of the network, where it enables slow and diffuse modifications of connection strengths. The Hebbian rule that we use for synaptic modifications destabilizes recent neural activity if it led to negative reward, and stabilizes it if it led to positive reward, thus slowly increasing the chances of obtaining positive rewards in the future.

In parallel, the error signal is also sent to a targeted network, the rule-coding clusters, where it yields a fast desensitization of currently active synapses. A tentative molecular mechanism has been proposed for such desensitization.<sup>4</sup> A diffuse neuromodulator signaling recent negative reward (e.g., a catecholaminergic input), when occurring in conjunction with a molecular marker of recent postsynaptic activity such as an elevated intracellular concentration of calcium, would induce a reversible allosteric transition of postsynaptic receptor molecules toward a slow, desensitized state (FIG. 5). Whatever the exact molecular mechanism, the result of this fast focal effect of reward is to turn currently active rule-coding clusters back to an inactive state, thus letting other rule-coding clusters compete for the control of behavior. In the terminology of learning by selection, the rule-coding network functions as a *generator of diversity*. Spontaneous fluctuations in activity lead one rule-coding cluster to take control and inhibit the others. Subsequent negative rewards ensure the elimination of inadequate active clusters until, after several trials, one is found to yield only positive rewards.

In simulations of the rule-coding cluster network, we found ranges of parameters for which the occurrence of one or two consecutive erroneous trials was sufficient to trigger an internal change in the sorting rule used. This is comparable to the performance of normal adult subjects in the Wisconsin Card Sorting Test, who rapidly switch to a new sorting rule when the previous one is found incorrect. When we lesioned either the reward network or the rule-coding network, however, perseverations were observed: the system continued to use the same sorting rule





**FIGURE 5.** Tentative mechanism for the shift in rule-coding cluster activity when negative reward is received (the "generator of diversity"):- (A) The simultaneous occurrence, on a given postsynaptic site, of a diffuse neuromodulator signaling recent negative reward and of a postsynaptic marker of recent cell activity, such as intracellular calcium, triggers an allosteric transition of postsynaptic receptor molecules towards a desensitized state (ion channel closed). (B) This molecular mechanism has the effect of depriving currently active neurons from supporting inputs from neighboring neurons belonging to the same cluster. The cluster therefore shifts from a stable active state to a stable inactive state, releasing other clusters from lateral inhibition and allowing them to compete for the control of behavior.

for several trials in a row, even after it was negatively rewarded several times. Hence, the lesioned network mimicked the perseverative behavior observed in patients with frontal lesions.<sup>15,16</sup>

Perseverations were observed in our model after many sorts of simulated lesions, including weakening the reward input, weakening the influence of the error signal on the rule-coding network, weakening or destroying the connections originating from rule-coding clusters, or eliminating the rule-coding clusters altogether. Hence, we would expect many different types of frontal and/or subcortical lesions to affect card sorting performance. We also found that our simulation performed poorly (but without perseverating) when the rule-coding layer was extended to include rules other than the three basic color, form, and number rules. This may,



to some extent, account for the failure of some normal subjects in the task.<sup>16</sup> Our contention is that the two sorts of failure have very different origins: normal subjects fail because they tend to try out complex sorting rules before having exhausted the simplest ones, whereas frontal patients fail because of an impairment in changing the current sorting rule in the face of negative reward.

#### *Internal Evaluation or Auto-Evaluation*

The ability to evaluate behavior internally and to anticipate future errors was also provided to the card sorting network by introducing modifiable connections from the intention units to the error cluster (FIG. 3). These connections, which formed an *auto-evaluation loop*, were able to learn when a pattern of activity over intention clusters had been associated with negative reward. If, later on the same trial, the same intention recurred, the error-coding cluster was spontaneously activated via the auto-evaluation loop, with the consequence that the sorting rule was immediately changed without actually having to try it on the environment.

Auto-evaluation, in conjunction with the previously described ability of memory and intention clusters to remain isolated from external inputs and outputs, provides our simulation with an internal work space in which representations can be manipulated and rules can be tried out until a satisfactory one is found. A precise sequence of neuronal activity is predicted (FIG. 6). When negative reward is received, it triggers a change in rule-coding cluster activity and a new sorting rule becomes active. This new rule, when applied to the memorized features of the input card, yields a new pattern of activity over intention clusters. If this pattern is again recognized by the auto-evaluation loop as likely to be negatively rewarded, the error cluster is internally activated, and the whole cycle starts over until a more satisfactory intention is found.

The notion of auto-evaluation and its relation to prefrontal cortex has now begun to receive experimental support. Damasio and colleagues<sup>19,30,31</sup> have studied patients who experience severe difficulties in decision making in everyday life and whose deficit is traceable to an impairment in evaluating whether a given image, situation, or plan has a positive or negative outcome. Because their auto-evaluation of ideas and plans is impaired, these patients do not know how to select a course of action other than by chance or by attempting to list all the events that could happen. Anatomically, these patients have lesions of the orbitofrontal cortex, which is a good candidate for a component of the auto-evaluation loop because of its strong connectivity with the limbic system.

In our model, the auto-evaluation loop is used exclusively for the internal detection of erroneous intentions. Recently, an electrophysiological correlate of such error-detection has been observed in normal humans.<sup>32-34</sup> In several reaction time tasks, a sharp focal negativity was recorded from medial frontal scalp electrodes only on trials in which the subject made an error. This error negativity occurred with a very short latency (about 70 ms) after the response was made, ruling out sensory feedback and suggesting that errors were internally monitored in parallel to the execution of the response. Anatomically, dipole modeling suggested that the generator of the error effect was located in the anterior cingulate cortex. Single-



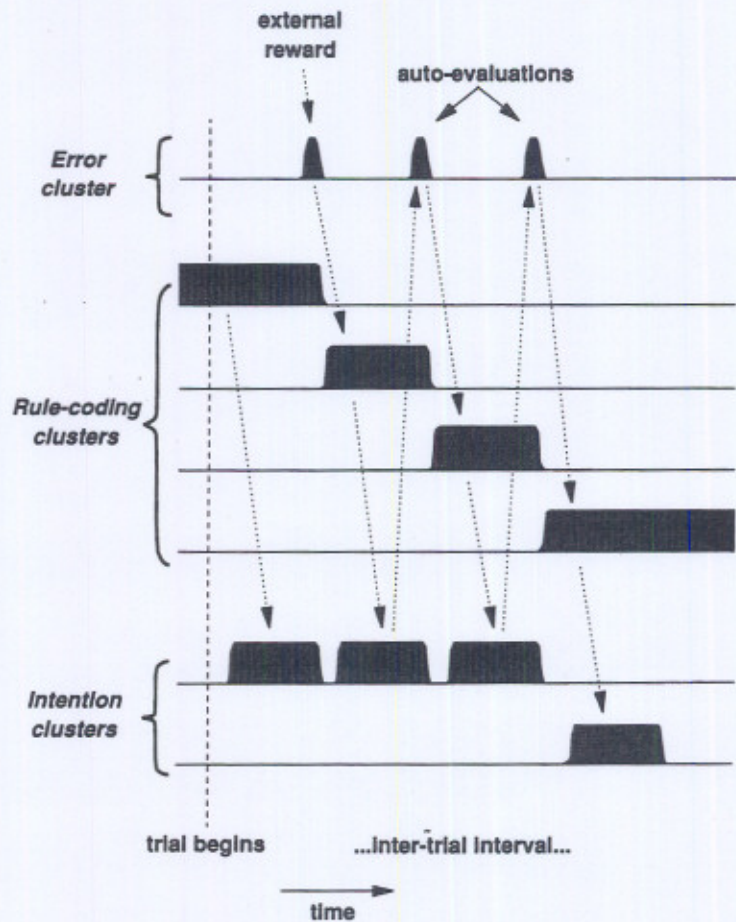


FIGURE 6. Simulated sequence of neuronal activity corresponding to an internal test of several rules. At the beginning of a trial, an intention is tried out and external negative reward is received. This sets up a cyclical sequence of activity in which (1) a new rule-coding cluster becomes active; (2) this leads to the activation of an intention cluster; (3) the current intention is recognized as being the same one that previously led to negative reward; and (4) the error cluster is activated via the auto-evaluation loop and the process repeats. The testing cycle stops when a rule is found which leads to a different intention, one that is not known to yield a negative reward.

cell recordings in monkeys have also revealed anterior cingulate cells that fire when the animal makes an error or when an expected reward fails to be delivered.<sup>35</sup> It is not known yet whether the anterior cingulate contributes to error detection, error correction, or both. However, its strong connectivity with multiple areas of prefrontal cortex fit well with its involvement in an anterior auto-evaluation circuit.



## CONCLUSION AND FUTURE PROSPECTS

According to our models, three aspects of frontal lobe circuitry are critical to an understanding of its contribution to cognitive functions. First, prefrontal areas are involved in hierarchically organized nested circuits that modulate lower-level sensory and motor circuits. Second, prefrontal circuits can sustain a long-lasting neuronal activity, which enables them to maintain over time representations of past events, anticipations of future events, and putative goals or rules for behavior, and to manipulate these representations in a purely internal manner. Third, prefrontal representations can be rapidly selected or eliminated based on an evaluation of performance by reward systems. Frontolimbic circuits forming an auto-evaluation loop endow the organism with a capacity to evaluate behavior internally instead of having to wait for external reinforcement.

The three properties of modulation of lower levels, representation detached from input-output contingencies, and auto-evaluation fit well with views of the frontal lobe as a "central executive" or "supervisory system"<sup>10,17</sup> that evaluates and controls cognitive processing. Nevertheless, our models of this supervisory system remain highly simplified. From the anatomical point of view, an important direction for future modeling will be to incorporate more realistic data on the anatomy and connectivity of known brain areas. From the functional point of view, tasks that are known to depend on the supervisory functions of prefrontal cortex, but cannot be handled by present network architectures, should be addressed in future simulations.

For instance, some prefrontal lesions are known to affect the ability to retrieve temporal-order information and use it to guide behavior.<sup>1</sup> Our models cannot address this issue yet because although they incorporate a mechanism for maintaining on-line representations of past events, the order in which these events occurred is not represented explicitly. Indeed, we know of no neuronal model of temporal-order judgments that has attempted to account for prefrontal impairments in the time domain. However, models for the role of frontal and subcortical areas in the production of temporal sequences of actions<sup>23-25</sup> might probably be extended to account for relative order judgments.

Another function not properly addressed by current models is the planning of a future sequence of actions. A classical test of planning is Shallice's Tower of London test,<sup>36</sup> a puzzle with pegs and movable disks in which the patient must find the shortest sequence of moves for achieving a given configuration of the disks. Solving the test requires the exploration of a tree of possible moves by trial and error. Normal subjects decompose the problem into a hierarchical sequence of subgoals; patients with prefrontal lesions may have a specific impairment in managing this subgoal hierarchy.<sup>37</sup> Auto-evaluation probably plays an important role in the process of subgoal selection by permitting the elimination of moves that diverge from the main goal. However, other components are also needed which have not yet been incorporated in models. Most notably, how could a hierarchy of goals and subgoals be represented within a neural network? And how could neural activity progress through a sequence of goals with automatic backtracking when errors or impasses are found? As our knowledge of prefrontal



functions increases, these questions stand out as important unsolved problems that will have to be addressed by theorists and neurobiologists alike.

## REFERENCES

1. FUSTER, J. M. 1989. *The Prefrontal Cortex*. 2nd edit. Raven. New York.
2. GOLDMAN-RAKIC, P. S. 1987. Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. In *Handbook of Physiology*. F. Plum & V. Mountcastle, Eds. Vol. 5: 373-417. American Physiological Society. Bethesda, MD.
3. DEHAENE, S. & J. P. CHANGEUX. 1989. A simple model of prefrontal cortex function in delayed-response tasks. *J. Cognit. Neurosci.* 1: 244-261.
4. DEHAENE, S. & J. P. CHANGEUX. 1991. The Wisconsin Card Sorting Test: Theoretical analysis and modelling in a neuronal network. *Cereb. Cortex* 1: 62-79.
5. DEHAENE, S. & J. P. CHANGEUX. 1993. Development of elementary numerical abilities: A neuronal model. *J. Cognit. Neurosci.* 5: 390-407.
6. CHANGEUX, J. P. & S. DEHAENE. 1989. Neuronal models of cognitive functions. *Cognition* 33: 63-109.
7. EDELMAN, G. 1987. *Neural Darwinism*. Basic Books. New York.
8. LURIA, A. R. 1966. *The Higher Cortical Functions in Man*. Basic Books. New York.
9. GRAFMAN, J. 1989. Plans, actions, and mental sets: Managerial knowledge units in the frontal lobes. In *Integrating Theory and Practice in Clinical Neuropsychology*. E. Perecman, Ed.: 93-138. Erlbaum. Hillsdale, NJ.
10. SHALLICE, T. 1988. *From Neuropsychology to Mental Structure*. Cambridge University Press. Cambridge, MA.
11. PIAGET, J. 1954. *The Construction of Reality in the Child*. Basic Books. New York.
12. DIAMOND, A. & P. S. GOLDMAN-RAKIC. 1989. Comparison of human infants and rhesus monkeys on Piaget's A-not-B task: Evidence for dependence on dorsolateral prefrontal cortex. *Exp. Brain Res.* 74: 24-40.
13. FUNAHASHI, S., C. J. BRUCE & P. S. GOLDMAN-RAKIC. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61: 331-349.
14. COHEN, J. D., K. DUNBAR & J. MCCLELLAND. 1990. On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychol. Rev.* 97: 332-361.
15. MILNER, B. 1963. Effects of brain lesions on card sorting. *Arch. Neurol.* 9: 90-100.
16. NELSON, H. E. 1976. A modified card sorting test sensitive to frontal lobe defects. *Cortex* 12: 313-324.
17. BADDELEY, A. 1986. *Working Memory*. Oxford University Press. New York.
18. POSNER, M. I. & S. DEHAENE. 1994. Attentional networks. *Trends Neurosci.* 17: 75-79.
19. DAMASIO, A. R. 1994. *Descartes' error: Emotion, reason, and the human brain*. G. P. Putnam. New York.
20. GOLDMAN-RAKIC, P. S. 1988. Topography of cognition: Parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* 11: 137-156.
21. WILSON, F. A. W., S. P. O'SCALAIDHE & P. S. GOLDMAN-RAKIC. 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260: 1955-1958.
22. FUSTER, J. M. 1973. Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *J. Neurophysiol.* 36: 61-78.
23. GUIGON, E., B. DORIZZI, Y. BURNOD & W. SCHULTZ. 1995. Neural correlates of learning in the prefrontal cortex of the monkey: A predictive model. *Cereb. Cortex* 5: 135-147.
24. BERNS, G. S. & T. J. SEJNOWSKI. How the basal ganglia make decisions. In *Neurobiology of Decision-Making*. A. R. Damasio, Ed. Springer Verlag. Berlin. In press.
25. DOMINEY, P., M. ARBIB & J. P. JOSEPH. A model of cortico-striatal plasticity for learning oculomotor associations and sequences. *J. Cognit. Neurosci.* In press.



26. COHEN, J. D. & D. SERVAN-SCHREIBER. 1992. Context, cortex, and dopamine: A connectionist approach to behavior in schizophrenia. *Psychol. Rev.* 99: 45-77.
27. LEVINE, D. S. & P. S. PRUEITT. 1989. Modelling some effects of frontal lobe damage—novelty and perseveration. *Neural Networks* 2: 103-116.
28. NIKI, H. & M. WATANABE. 1976. Prefrontal unit activity and delayed response: Relation to cue location versus direction of response. *Brain Res.* 105: 79-88.
29. FRISTON, K. J., G. TONONI, G. N. REEKE, O. SPORNS & G. M. EDELMAN. 1994. Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neurosci.* 59: 229-243.
30. DAMASIO, A. R., D. TRANEL & H. DAMASIO. 1990. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* 41: 81-94.
31. ESLINGER, P. J. & A. R. DAMASIO. 1985. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* 35: 1731-1741.
32. DEHAENE, S., M. I. POSNER & D. M. TUCKER. 1994. Localization of a neural system for error detection and compensation. *Psychol. Sci.* 5: 303-305.
33. FALKENSTEIN, M., J. HOHNSBEIN, J. HOORMANN & L. BLANKE. 1990. Effects of errors in choice reaction time tasks on the ERP under focused and divided attention. In *Psychophysiological Brain Research*. C. H. M. Brunia, A. W. K. Gaillard & A. Kok, Eds.: 192-195. Tilburg University Press, Tilburg, the Netherlands.
34. GEHRING, W. J., B. GOSS, M. G. H. COLES, D. E. MEYER & E. DONCHIN. 1993. A neural system for error detection and compensation. *Psychol. Sci.* 4: 385-390.
35. NIKI, H. & M. WATANABE. 1979. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res.* 171: 213-224.
36. SHALLICE, T. 1982. Specific impairments of planning. *Phil. Trans. R. Soc. London B* 298: 199-209.
37. GOEL, V. & J. GRAFMAN. Are the frontal lobes implicated in "planning" functions? Interpreting data from the tower of Hanoi. *Neuropsychologia*. In press.