64 Relational Complexity and the Functions of Prefrontal Cortex

NINA ROBIN AND KEITH J. HOLYOAK

ABSTRACT We propose a theoretical framework for understanding the core cognitive functions of prefrontal cortex. Our general claim is that the prefrontal cortex is responsible for the creation and maintenance of explicit relational representations that guide thought and action. The framework provides a formal characterization of the types of representations that pose the greatest difficulty for animals with prefrontal damage, an explanation of the close tie between phylogenetic and ontogenetic development of the frontal lobes and cognitive capabilities, and an explanation of the deficits that can result from the use of impoverished representations.

Theoretical perspectives on prefrontal functions

The human prefrontal cortex presents us with a striking paradox. On the one hand it is widely believed to control the highest and most distinctively human forms of thinking; yet at the same time, extensive damage to this area of the brain is apt to yield only modest decrements in traditional measures of intelligence. This paradox offers a clear challenge to cognitive science. An adequate theory of mental representation and processing should be able to shed light on the relationship between high-level thinking and what we call intelligence. Resolution of the paradox may be achieved through the research strategy of cognitive neuroscience, which seeks closer links between models of brain and of cognition. Such links may provide needed constraints on the basic concepts used in theories of high-level cognition.

Despite the fact that patients with prefrontal damage perform in the normal range on standard intelligence tests, they are clearly impaired on a wide variety of cognitive functions, including planning, monitoring and modifying behavior, learning complex tasks, and temporal sequencing. Clinical reports reveal that these frontal patients fail to plan for future events, including social interactions, and have difficulty implementing formerly routine plans such as food shopping. Deficits in formulating strategies and plans have also been revealed in experimental studies of tasks involving hypothesis testing, spatial working memory, construction of categorization schemes, and problem solving. The diversity of deficits after frontal lobe damage has fostered the development of an equally diverse set of theories of frontal lobe functions. Recent theories have emphasized the function of the frontal lobes in executive control, attentional control based on internal representations of context, creation and use of abstract event knowledge, and sequencing of actions. (See Fuster, 1989, for a review.) Each of these perspectives appears to capture important insights into the patterns of deficits that have been observed. It has not been entirely clear, however, how the various approaches relate either to each other or to general theories of cognition. Work on frontal functions highlights a number of basic psychological concepts that have not been clearly defined. For example, what makes a task novel? At one extreme, any two experiences differ at least in terms of their spatiotemporal coordinates, so we could claim that everything is novel. At the other extreme, any two experiences have some similarity, such as "both happened to me" or "both happened in this century," so we could argue that nothing is entirely novel. Similar definitional problems cloud concepts such as complexity and context. How should complexity be measured? How is context different from any other feature of a task situation? And how are these concepts related to one another and to mechanisms for working memory and attentional control?

We believe it is possible to clarify, and perhaps even to answer, some of these questions by introducing a general framework for understanding *relational complexity*. We will argue that the central commonality linking

NINA ROBIN and KEITH J. HOLYOAK Department of Psychology, University of California, Los Angeles, Calif.

the various aspects of the frontal syndrome in humans and other animals involves the explicit representation and processing of relational information. The complexity of relations is closely related both to the abstraction of concepts and to distinct types of similarities between events. The prefrontal cortex appears to be responsible for the creation and maintenance of relational representations that guide thought and action. This basic function depends on a working-memory system that serves to bind elements into relational structures, which can then constrain behavior in a goal-directed manner across time. The framework provides a formal characterization of the types of representations that pose the greatest difficulty for animals with prefrontal damage; an explanation of the close tie between phylogenetic and ontogenetic development of the frontal lobes and cognitive capabilities; and an explanation of the deficits that can result from the use of impoverished representations. We will first introduce the framework, and then illustrate its application in interpreting empirical evidence obtained with humans and other primates.

A framework for defining relational complexity

Our thesis is that the functions of prefrontal cortex are linked to the acquisition and use of explicit relational knowledge in the service of a goal. Explicit knowledge, as we use the term here, is based on slow, effortful, and conscious processing that is directly dependent on working memory, and contrasts with implicit knowledge based on relatively rapid, effortless, and unconscious processing. By explicit *relational* knowledge we mean knowledge that differentiates roles from their fillers, and hence relates the latter to the former. The minimal requirement for inferring a distinction between role and filler is the ability to respond on the basis of a specific dimension of variation. At an earlier level, both phylogenetically and ontologically, an animal will react to individual objects without any explicit dimensional analysis. Work on children's classification reveals that even infants can match one apple to another on the basis of global similarity, but only later do children become capable of matching a red apple to a red block on the basis of their common color and despite differences on other dimensions (see Smith, 1989, for a review). Success at the latter task is indicative of differentiation between a role (e.g., "red thing") and the filler of the role (e.g., a particular red apple),

where the filler can be freely varied without altering the child's basic response to the dimension that defines the role. We will refer to the more primitive ability to react to global similarity between situations or objects as *holistic* processing, adopting a term often used in the developmental literature on classification.

Explicit processing of dimensions, in which the frontal lobes appear to play a major role, can involve role-filler relationships at different levels of complexity. Here we will adapt a general taxonomy developed by Halford and his colleagues (Halford and Wilson, 1980; Halford, 1993), which has been related to working-memory requirements of cognitive tasks, cross-species cognitive comparisons, and levels of human cognitive development. Halford (1993) defines relational complexity in terms of the number of independently varying dimensions that must be considered together to generate an appropriate response. Each such dimension can be viewed as a variable argument or "slot" linked to a relational concept (i.e., a predicate, function, or operator); in a specific instantiation of the relational structure, the filler of each such slot is logically bound to it. For example, the structure

$$\text{red (apple-1)}$$

relates the predicate *red* to a certain red apple, apple-1. Apple-1 serves as the filler of the single argument slot associated with the predicate *red* (equivalent to the role of "red thing"), and thus is bound to the slot in this instantiation of the predicate. More generally, the filler of a slot is a logical constant bound to a variable.

In symbolic knowledge representations of this sort as they are typically used in artificial intelligence, the arguments of a predicate are viewed as a list. It is assumed that the arguments in the list can be processed sequentially, so that complexity would be linear with the number of arguments. However, Halford argues that complexity increases worse than linearly with the number of arguments. In his view, arguments psychologically correspond to dimensions of potential simultaneous variation that jointly determine a response. For example, the one-place predicate *red* picks out a single dimension, corresponding to objects that are red (e.g., apples, fire trucks, sunsets), where the appropriate response to each would be "true"; similarly, the two-place predicate *larger than* ranges over two dimensions corresponding to pairs of objects in which one is larger than the other (e.g., {dog, cat}, {truck, car}); and the three-place predicate *between* ranges over three

dimensions corresponding to appropriate triples of elements (e.g., {United States, Canada, Mexico}, as in *The United States is between Canada and Mexico*). The psychological benefit of an *n*-place predicate (i.e., a predicate with *n* distinct arguments or roles) is that it codes information about *n*-way interactions between the dimensions, which can then guide appropriate responses. Each dimension and each interaction between dimensions represents a potential source of variation relevant to determining the appropriate response. Psychologically, higher-level interactions provide a basis for responding to complex combinations of elements by virtue of their roles in the overall relational structure, where the response could not be determined by the individual elements considered in isolation from each other. The psychological cost associated with an *n*-place predicate is that it requires a representation of *n* dimensions of variation (i.e., *n* elements filling roles) that have to be considered simultaneously. Moreover, this representation must represent not only the individual elements, but also the binding of each element into its appropriate role.

To characterize the levels of relational complexity more formally, we will generally speak of functions that map a set of argument slots, each based on a dimension that varies independently of the others, to a response. Thus $f(D_1, D_2, \ldots D_n) \to R$ signifies that a function f applied to the fillers of n slots based on n dimensions, D, generates a response R, where the arrow represents response generation.[1] A taxonomy of levels of complexity can then be defined as follows:

Level 1. A function of 1 dimension provides an *attribute* mapping.

Level 2. A function of 2 dimensions provides a *relational* mapping.

Level 3. A function of 3 dimensions provides a *system* mapping.

Level 4. A function of 4 dimensions provides a *multiple system* mapping..

This taxonomy generates a hierarchy of abstraction for concepts. Holistic processing, which we assume does not depend on frontal functions, may be considered level 0 in the taxonomy, and does not involve abstraction of specific dimensions or representations of variables. Attribute mappings require abstraction of a single dimension of variation, the minimal requirement for differentiating a role from its filler. A relational mapping depends on the ability to relate two dimen-

sions of variation to each other. A system mapping makes it possible to define a relation between relations (i.e., a higher-order relation). A transitive ordering, for example, requires three elements for which the ordering of two pairs constrains the ordering of the third pair (e.g., *A before B* and *B before C* implies *A before C*, where the implication is based on the relation between two *before* relations).

Because of the cognitive burdens imposed by concepts of high dimensionality, reasoners will often shift representations and strategies to reduce the effective complexity of tasks. Halford (1993) suggests two basic mechanisms for complexity reduction: conceptual chunking and segmentation. Conceptual chunking involves collapsing a multidimensional concept into one based on fewer dimensions. When complex concepts are chunked, immediate access to higher-order interactions between dimensions is lost. A chunk is nonetheless informationally richer than a simple holistic representation, in that a chunk is constructed from a complex relational structure defined over symbols; furthermore, its internal structure is potentially recoverable by "unchunking." We assume that a variety of factors will affect the ease with which chunks can be formed. In particular, forming chunks from individual elements will be difficult when the elements are separated from each other in time or space, or when they do not consistently co-occur (i.e., when the same elements are often rearranged to form different relational combinations, yielding what is sometimes termed a *varied mapping* from inputs to response). As we will see, conditions that impede chunking, and hence increase relational complexity, often appear to increase the dependence of a task on prefrontal cortex.

Segmentation is a processing strategy that reduces task complexity by dividing a task into smaller components that can be processed independently and therefore serially. Often, chunking can be used to facilitate task segmentation. For example, suppose a task requires learning the order of *n* actions that must be performed sequentially to achieve a goal. If *n* is large, the complexity would be overwhelming if all the elements were considered in parallel. Suppose, however, that one first learns to perform action A before B, and then chunks this unit of activity, which we can refer to as AB. One can then learn to perform AB before C, chunking this new unit as ABC, and so on until the complete ordering has been mastered. Using such a segmentation strategy (which a benevolent teacher

might encourage), the task complexity of learning an ordering (of any number of items) can be reduced to a series of relational mappings.

Judicious use of chunking and segmentation will be essential for performing any task that would otherwise exceed the reasoner's maximum limit on parallel processing of dimensions. Halford argues that in humans this maximum limit increases over the course of cognitive development to a maximum of four. Tasks used to test frontal functions in monkeys can, we believe, be plausibly performed using only relational mappings (coupled with the ability to form chunks and to learn tasks in segments), suggesting that the maximum dimensionality of concepts for nonhuman primates may be two.

Although Halford's taxonomy of relational complexity is purely formal, it suggests important constraints on the cognitive mechanisms required to learn and reason explicitly about relational contingencies. Selective attention is needed in order to focus on specific dimensions of the representation for an object or event, as must be done to move beyond primitive holistic processing. The system requires a working memory capable of processing n dimensions in parallel, where the size of n determines the maximum limit on complexity of individual concepts. This working memory, which can be distinguished from a short-term, declarative store used simply to maintain items over time, is required for relational processing. The working memory must code the bindings of elements into specific roles with respect to relational concepts. To maintain appropriate bindings, some form of attentional control must prevent "cross talk" that could blur the identities and role assignments of the individual elements. The bindings must be dynamic so that intermediate computations can be performed without permanently altering the representations of elements. In addition, task-irrelevant elements of the situation may need to be inhibited in order to prevent interference with the representation of goal-relevant elements. In order to segment a task, it will be necessary to swap dynamically between subsets of goal-relevant elements. Attentional control of the sort associated with executive functions will be required to adjudicate between conflicting responses generated by functions at different complexity levels. That is, a certain element considered in isolation may tend to trigger action A, while that same element in combination with one or more additional elements may trigger the incompatible action B. Often, a function based on an isolated element will trigger a response that is a useful default, but which is contextually inappropriate and hence must be inhibited in favor of a response that is a function of the relations between multiple elements considered together. Attentional control and dynamic binding is thus the key to reducing interference with adaptive problem solving.

Another crucial component of the overall system will be a mechanism for forming conceptual chunks. A conceptual chunk—a unitized representation of a complex relational structure—is closely related to the psychological concept of a schema. Formation of a chunk, or schema, requires a mechanism that can take the transient bindings of elements to roles in working memory and create a structure in long-term memory from which the bindings can be recovered by a mechanism for unchunking.

We propose that the major functions of the prefrontal cortex can be understood as aspects of an overall system for reasoning with and learning about explicit relational concepts. (This is not to deny that this region of the brain performs other functions as well.) Within any mammalian species, we conjecture that tasks requiring the highest complexity level attainable by the species will be maximally dependent on frontal functions. Our framework helps to integrate the various frontal functions that have been the focus of previous theories. Executive control of attention will be required in complex planning and problem solving because these operations intrinsically involve high levels of relational complexity. Contextual elements that need to be considered in combination with other elements will increase relational complexity, and also will require inhibition of responses based on individual elements. Abstract concepts will have relatively high dimensionality, and hence will be more dependent on prefrontal functions. Conditional contingencies based on multiple elements, especially when these are difficult to chunk due to their temporal or spatial separation, will also generate high levels of relational complexity. Once complex relations are chunked and stored in long-term memory, and appropriate task segmentations and specific role bindings have been learned, frontal involvement will be sharply reduced. Because conventional tests of intelligence largely tap previously acquired knowledge rather than reasoning with novel relational concepts, or even reasoning that

requires novel bindings of elements to roles in familiar schemas, such tests may often fail to detect the deficits that result from frontal damage.

Evidence of relational processing in prefrontal cortex

In this section we will review evidence that indicates how damage to the prefrontal cortex can impair the ability to create relational and system-level mappings in primates. While our analysis is based on data from both human and nonhuman primates, the application of animal results to humans must be done with several caveats (Fuster, 1989). First, neurological insult seldom respects anatomical boundaries. Tumors and closed head injury rarely produce focal deficits specific to a single functional region. Second, much of the lesion work on animals has involved bilateral ablation, whereas many of the human studies involve unilateral lesions. Thus, deficits may be more severe in animal subjects than in human frontal patients because these patients have the capacity to compensate with their remaining prefrontal cortex. An additional complication hinges on the fact that the prefrontal cortex appears evolutionarily more complex in humans than in nonhuman primates and other mammals. There may thus be important species differences in the complexity of the tasks governed by the prefrontal cortex in humans, as our analysis of relational complexity in fact suggests. Humans have more strategies available to compensate in complex situations.

SKETCH OF FUNCTIONAL NEUROANATOMY   Figure 64.1 provides a lateral view of a rhesus monkey brain. The frontal lobes correspond to the region anterior to the central sulcus. The frontal cortex can be divided into the prefrontal, premotor, and the cingulate or limbic cortex. The prefrontal cortex can be circumscribed on the basis of its connectivity to the thalamus: It receives afferent projections from the dorsomedial nucleus of the thalamus (Fuster, 1989). Figure 64.1 shows three major regions of the prefrontal cortex. The *dorsolateral* region, for which the principal sulcus is the most prominent morphological landmark in the rhesus monkey, has many reciprocal connections with the posterior cortex, receiving visual, auditory, and somatosensory information. There are also corticocortical connections with the orbital regions, as well as subcortical connections, particularly to the caudate nucleus. The *peri-*
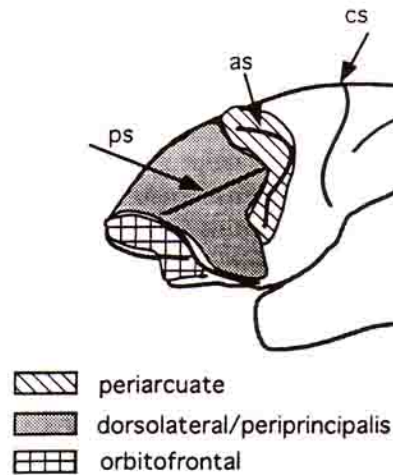

periarcuate
dorsolateral/periprincipalis
orbitofrontal

FIGURE 64.1   Lateral view of the brain of the rhesus monkey: cs, central sulcus; as, arcuate sulcus; ps, principal sulcus.

arcuate region (the arcuate sulcus and its surrounding area) is the caudal-most region of the prefrontal cortex. The *orbitofrontal* region is located ventral to the dorsolateral area, and in the rhesus monkey extends from the ventral lip of the principal sulcus (the only portion visible from the lateral view of figure 64.1) around the orbital convexity to the ventral surface. This area has cortical communication with the lateral areas of the prefrontal cortex as well as the limbic system. The posteromedial areas of the orbitofrontal region include the basal forebrain, which connects to the amygdala and hippocampus.

It should be clear from this description that the prefrontal cortex has reciprocal connections with a wide range of cortical and subcortical structures, including almost all areas of the central nervous system, enabling it to play an integrative role in the control of behavior. It is also clear that the prefrontal cortex is a heterogeneous region, both architecturally and functionally. Damage to different areas of the prefrontal cortex is associated with separable syndromes. In particular, damage to the dorsolateral area, including the principal sulcus and its surrounding area (i.e., the periprincipalis region), is associated with impaired planning, distractibility, and working-memory deficits that interfere with learning contingencies between stimuli and responses separated in time (Fuster, 1989; Goldman-Rakic, 1990; Knight, 1991). The periarcuate region is important for learning conditional contingencies between stimuli and responses, especially in situations

that do not involve significant delays (Petrides, 1985; Stamm, 1987). Insult to the orbitofrontal region is associated with loss of inhibitory control over environmentally cued responses, with associated affective disturbances, including personality changes, mood swings, and socially inappropriate behavior resulting from decreased inhibitions (Stuss and Benson, 1984; Fuster, 1989; Knight, 1991).

PHYLOGENETIC AND ONTOGENETIC TRENDS IN FRONTAL DEVELOPMENT   Increases in capability to handle relational complexity, both across species and within individuals over the course of development, are correlated with changes in the frontal lobes. This cortical region has been the latest to evolve, and its size and complexity increases more with phylogenetic development than does any other region. Ontogenetically, the prefrontal cortex is one of the last cortical areas to become fully myelinated, a process that may not be complete in humans until the beginning of the second decade; synaptic and neuronal density also increases significantly through childhood (Thatcher, 1991). This late development strongly suggests that the role of the prefrontal cortex in guiding behavior increases throughout development. Moreover, close parallels can be found between patterns of developmental changes in the coordination of electrical activity in the frontal cortex with that in posterior cortical systems, on the one hand, and developmental changes in attentional control and relational processing, on the other (Case, 1992). It is also the case that performance on a variety of tests that reveal frontal deficits in human adults shows pronounced developmental changes over middle to late childhood (Levin et al., 1991).

Although few in number, case studies of the consequences of childhood frontal damage provide additional support for the postulated link between frontal functions and relational learning. These studies reveal a pattern of delayed onset of behavioral deficits, followed by a period of modest progress, which ends with arrested development at adolescence (Ackerly and Benton, 1948, cited in Damasio, 1985; Grattan and Eslinger, 1991). The deficits that arise over the course of childhood include a variety of cognitive deficiencies related to organization and planning, coupled with lack of self-regulation and abnormal social behavior. Damasio (1985, 351) has concluded, "It seems probable that bilateral damage to the frontal lobes in infancy or childhood produces a more devastating effect on personality and cognitive ability than the same amount of damage sustained elsewhere in the brain at any time in the course of development." He particularly notes the striking contrast with cases of early hemispherectomy on either side, which demonstrate near-normal cognitive and social development. If the central function of the frontal lobes is to support learning of complex relational concepts, then it follows that frontal damage will be much more detrimental if it precedes the period when major concepts based on system and multiple-system mappings are normally acquired, which is middle childhood through adolescence. It is also possible that social cognition is especially dependent on the ability to respond to subtle relations between multiple cues, including contextual cues accumulated over time, and hence is particularly vulnerable to early frontal insults.

WORKING MEMORY AND INTEGRATION OF RELATIONS ACROSS DELAYS   The relational complexity of a task is almost inevitably increased when multiple dimensions of information must be integrated over time. Separation in time will hinder chunking, so the information will need to be processed as distinct but related units, making relational mapping necessary. At the same time, the need to maintain representations of at least one dimension over time until the other relevant information is presented will impose a burden on working memory and attentional control. It is therefore not surprising that one of the most robust findings obtained with monkeys after prefrontal lesions is that they exhibit impaired performance on delay tasks (Jacobson, 1935). Delay tasks all involve responding to a stimulus when there is a delay imposed between the stimulus and the response. The classical delay task involves baiting one of two identical food wells; after a fixed delay, the monkey must choose the baited food well. After a prespecified intertrial interval, the sequence is repeated with a new, randomly chosen food well. Successful performance on this task requires learning a functional rule that might be represented as a function of two variables,

$$\text{same } (X, \text{baited}_n) \rightarrow \text{choose } X,$$

where X is a food well and baited_n is the location baited on the current trial. For example, if the left food well has been baited, then the animal must choose it. Learning such a rule requires the capability to perform a relational mapping.[2] Performance also depends on

the ability to remember the most recently baited location during the delay, and to suppress any interfering memories or response tendencies from prior trials. That is, the critical bindings must be transitory, as the same elements play different roles from trial to trial.

The delayed match-to-sample task is similar to the basic delay task, except the relevant relation is nonspatial. A sample object is first presented, and then, after a delay, two objects are presented, one of which is the same object as the sample. The animal must learn to approach the previously shown object. The rule that must be acquired is,

$$\text{same } (X, \text{sample}_n) \rightarrow \text{choose } X,$$

which has the same logical form as the relational mapping required in the delay task. The primary difference between these two tasks is that in the basic delay task the animal must remember a particular spatial location across a delay, whereas in the delayed match-to-sample task the animal must remember a particular object identity.

Intact monkeys can succeed at both of the above tasks, as well as at more complex variations of each. In a variation of the basic delay task, called a delayed alternation task, the animal must learn to go to the location that was *not* baited on the prior trial. On the first trial, both food wells contain food and the animal is rewarded for approaching either one. On the next trial, the food well that the animal did not choose on the prior trial is baited (without the animal seeing this being done). The rule in a delayed alternation task is roughly,

$$\text{different } (X, \text{baited}_{n-1}) \rightarrow \text{choose } X,$$

Monkeys with lesions to the dorsolateral area, particularly the periprincipalis region, exhibit deficits in all these relational delay tasks, both spatial and nonspatial (Mishkin and Manning, 1978; Stamm, 1987). The magnitude of the deficit is proportional to the length of the delay (Fuster, 1989). Furthermore, lesions of the dorsolateral region do not impair performance on discrimination learning for visual stimuli (Petrides, 1985; Fuster, 1989), spatial stimuli (Passingham, 1985), or stimuli that differ in temporal duration (Goldman-Rakic, 1990).

A study by Rosenkilde, Rosvald, and Mishkin (1981) provides an especially informative contrast. They assessed performance on a temporal discrimination task using monkeys with dorsolateral, orbitofrontal, or periarcuate lesions. The temporal discrimination task involved learning that a 10-second delay signaled that the left food well had been baited, whereas a 30-second delay signaled that the right food well was baited. Monkeys were trained on the task preoperatively, and relearning was assessed approximately two weeks after surgery. Only the orbitofrontal group was impaired on relearning, and these animals exhibited significant perseverative tendencies. The results thus indicated that damage to the principal sulcus does not impair time discrimination per se. In terms of our complexity analysis, the temporal discrimination task used by Rosenkilde, Rosvald, and Mishkin can be represented as two attribute mappings,

$$\text{short (delay interval)} \rightarrow \text{choose left}$$

$$\text{long (delay interval)} \rightarrow \text{choose right,}$$

and hence is representationally simpler than the more complex conditional contingency tasks that require relational mappings. It thus appears that the deficits caused by periprincipalis damage are not tied simply to temporal coding per se; rather, the impairments are specific to tasks that require responses based on representations with at least the complexity of relational mappings that must be maintained in working memory.

Electrophysiological studies complement lesion studies in elucidating the functions of specific cortical regions in delay tasks. Studies based on single-cell recordings indicate that cells in the prefrontal cortex have increased rates of firing to specific task-related features, including the identity of the cue, the response, and the delay (Fuster, 1989, 1993; Goldman-Rakic, 1990). Neurons selectively responsive to stimuli presented in delay tasks have been found throughout the prefrontal cortex, although primarily in the region of the principal sulcus (Yamatani et al., 1990). The wide distribution of these cells may account for the fact that size of lesion is correlated with degree of deficit (Stamm, 1987). Furthermore, in the delayed match-to-sample task, many neurons respond to any object that is the same as the sample just presented, independently of the visual features of the object (Yamatani et al., 1990). Such cells are candidates for the neural basis for the coding of variables and bindings, in that they respond to the current instantiation of a relationship. In general, learning delay tasks based on relations between two dimensions of variation appears to involve changes in neuronal firing patterns in prefrontal cor-

tex, primarily in the dorsolateral but also in the peri-arcuate and orbitofrontal regions.

Although clear evidence of localization is difficult to obtain with humans, the broad pattern of findings with frontal patients is consistent with involvement of pre-frontal cortex in coding contingent relationships be-tween elements separated in time. Patients with frontal lobe damage usually are not significantly more im-paired than those with posterior damage on tests of short-term memory for information from visual, audi-tory, and kinesthetic modalities (Ghent, Mishkin, and Teuber, 1962). These negative findings involve tasks that do not require relational mappings (e.g., recalling the specific orientation of a line after a delay). In con-trast, frontal patients are selectively impaired on tasks that involve formulating contingencies based on spa-tiotemporally distinct events, such as delayed-response and delayed-alternation tasks (Freedman and Oscar-Berman, 1986).

The ability retrospectively or prospectively to se-quence actions and events also involves formulating relational and higher-order mappings. When subjects are required to recall the order of recently presented items, patients with frontal lobe damage are impaired in both delayed and immediate recall conditions (Lewinsohn et al., 1972). For example, Lewinsohn and colleagues presented two items sequentially on each trial. Each item had three features that had to be re-called: a picture, a background pattern, and a back-ground color. The background patterns were highly similar across items; in addition, pictures and back-grounds were repeated in different combinations across trials, making chunking difficult. Recall was scored for the specific features of each item as well as for the order of the two items. Frontal patients showed deficits for both immediate and delayed recall.

Because the stimuli were constructed in a manner that would tend to prevent processing the items as chunks, subjects presumably were forced to treat the three features of each item as separate elements. Cor-rect recall of any feature of an item in relation to the corresponding feature in the other item would there-fore involve computing at least two relations: the part-whole relation between a feature and the item in which it occurred, and the relation between the order of the two items. For example, suppose I_A and I_B are the two items to be recalled, and f_1 is a feature of I_A. Correct recall of f_1 in its appropriate position would require processing the relations

$$\text{part-of } (f_1, I_A)$$

and

$$\text{before } (I_A, I_B).$$

Considering these two relations together involves three dimensions of variation (the feature and the two items), and therefore constitutes a system mapping. The level of relational complexity is thus far greater for this ordered recall task than for one that simply requires unordered recall of individual items. Thus for humans, as for monkeys, frontal involvement in temporal and spatial coding is linked to relational complexity.

The ability to sequence events is a crucial compo-nent of planning. Shallice (1982) and Owen et al. (1990) found that patients with frontal lobe damage were impaired in planning a sequence of moves that would rearrange an initial pattern of colored beads into a goal state. Efficient performance on this task depends upon the ability to break the task into sub-goals and then reach each of the subgoals. The repre-sentational complexity of the task depends on the number of subgoals and their relationship to one an-other. In general, planning a set of actions requires at least relational-level mappings, because the choice of action depends on the relationship between the goal and an available operator.

LEARNING CONDITIONAL CONTINGENCIES THAT DO NOT REQUIRE TEMPORAL BRIDGING On the basis of evi-dence that the deficits resulting from prefrontal lesions are not specific to spatial or visual discriminations, or to delay per se, it has been argued that dorsolateral lesions impair the more general ability to learn rela-tionships between distinct temporal and spatial events (Petrides, 1985, 1987; Fuster, 1989). Even when tem-poral gaps are not involved, however, it is possible to design tasks that require learning conditional contin-gencies with the complexity of relational mappings. Petrides (1985) compared monkeys with periarcuate lesions to those with lesions of the principal sulcus re-gion on learning a nontemporal conditional contin-gency task. On this task, monkeys were rewarded when they opened a lit box in the presence of a toy clown or when they opened an unlit box in the presence of a yellow disc. (Both boxes were always presented.) Mon-keys with periarcuate lesions were severely impaired, while monkeys with lesions in the principal sulcus were slower than normal controls but able to learn the task.

However, there were no differences between groups when monkeys were rewarded for opening the lit box and not otherwise.

There is a crucial difference between tasks involving conditional versus nonconditional contingencies in terms of our complexity analysis: The former task depends on forming two relational mappings, whereas the latter depends on only a single attribute mapping (or even on holistic processing). The above conditional task required learning two relational rules,

appear-together (toy clown, lit box) → open lit box,

appear-together (yellow disc, unlit box) → open unlit box.

In contrast, the nonconditional contingency can be learned using a much simpler representation, such as

appear (lit box) → open lit box.

In general, it seems that although deficits in delay tasks are greatest when the periprincipalis region is damaged, deficits in nontemporal conditional contingency learning are greatest after damage to the periarcuate area (Petrides, 1985, 1986; Stamm, 1987; see Petrides, 1987, for a review). It thus seems that both these prefrontal regions play roles in binding elements into relations. The periprincipalis region is especially relevant to the maintenance of relevant dimensions of variation in working memory as relations are dynamically formed, whereas the periarcuate region is central to the formation of relations between elements when temporal gaps are not a major factor.

Similar deficits have been found in human frontal patients when they perform nontemporal tasks that nonetheless involve high relational complexity. For instance, one of the most frequent findings with human patients who have suffered frontal lobe insult is poor ability to categorize stimuli (Milner, 1964; Stuss and Benson, 1984; Owen et al., 1990). Stuss et al. (1983), for example, found that patients with prefrontal leucotomies were impaired on the ability to identify similarities between three of four objects. On each trial, subjects were shown a card with pictures of four objects. Three of the four objects could be classified in two ways on the basis of their values on the dimensions of color, form, size, and orientation. Subjects were first required to identify three similar items, and then to verbally state the reason for the similarity. Next, subjects were required to identify three *other* similar items, and again provide a verbal explanation of the basis for the grouping. The patients were able to point to a correct grouping of the items on the first categorization task, but often gave inconsistent or concrete responses when explaining their decisions. Furthermore, they were markedly impaired in their ability both to create and to explain a second, alternative grouping of the items.

This pattern of performance can be directly related to the complexity of each of the successive tasks. The initial categorization could be done on the basis of holistic processing based on the overall similarities of the items. However, to identify the basis for the similarity requires forming a relational mapping involving a dimension shared by three items but not the fourth. The second categorization task is yet more demanding in that success requires establishing a relational mapping that is different from that used for the first grouping. Representing the difference between two relational mappings has the complexity of a system mapping. Such results support our claim that the frontal lobes play an increasingly large role in cognitive tasks as the relational complexity of the tasks increases, even when temporal bridging is not required.

INHIBITION OF INTERFERENCE FROM RIVAL RESPONSES
In normal information processing, selection of items or actions is accompanied by inhibition of competing responses. When the contextually appropriate action is based on a high level of relational complexity, it will be necessary to inhibit alternative responses based on simpler representational levels, such as responses to single, isolated elements, as well as any previously established rival responses. Damage to the prefrontal cortex will lead to formation of impoverished representations of relations. This in turn will decrease inhibition of rival responses and hence lead to greater interference.

Such decreased inhibition appears to be selectively associated with damage to the orbitofrontal region. One of the classical paradigms used to measure response interference is the "go, no-go" paradigm, in which success requires learning to make a response to one stimulus but to inhibit that same response to a similar stimulus. Lesions of the orbitofrontal region lead to higher error rates on this task than do lesions of the dorsolateral cortex (Fuster, 1989). Damage to the orbitofrontal region has been shown to lead to increased interference from perseverative responses, an error pattern that has also been observed in delayed match-to-sample tasks with a single pair of stimuli, and

in object-alternation tasks (Mishkin and Manning, 1978).

Drewe (1975) found that humans with frontal lobe damage were impaired on learning a go, no-go task. This type of error is also commonly observed in patients' performance on the Wisconsin Card Sorting Test. On this task, the subject must match a set of cards that vary in color, shape, and number to a set of target cards on the basis of experimenter feedback. The task is to match the cards in the deck to the target cards according to one of the three dimensions. After the subject has achieved criterion on one dimension, the experimenter changes the concept, and the subject must learn to categorize along a different dimension based on the feedback. Frontal patients tend to perseverate in their initially learned response pattern after the experimenter has shifted the concept (Milner, 1964; Stuss and Benson, 1984).

Conclusion

Our review of the functions of prefrontal cortex leaves us in broad agreement with the conclusions of Fuster (1989), who argued that converging evidence from lesion studies, single-cell recordings, and electrophysiological measures have shown that the prefrontal cortex subserves three primary functions: maintaining representations of elements in working memory to process cross-temporal relationships; learning conditional contingencies; and providing resistance to interference. Experimental lesion studies with monkeys indicate that these functions have separate anatomical loci: The dorsolateral region acts as a working memory to code relationships that occur across temporal discontinuities; the periarcuate region is central to learning conditional contingencies that do not involve temporal gaps; and the orbitofrontal region is critical to minimizing interference. However, all three loci act together to constrain behavior in a goal-directed manner across time.

Because relational complexity of tasks is maximal when novel combinations of elements must be processed, damage to the prefrontal cortex is especially deleterious when it occurs at a young age, before the major relational schemas required for mature thinking and behavior have been acquired. Once relational schemas have been formed, their subsequent use will be less dependent on prefrontal functions. Thus both humans and other primates are able to perform previously acquired habits, even those that span temporal

intervals, despite suffering frontal injury as adults. The relative independence of well-learned responses from frontal control accounts in part for the fact that frontal patients often do not show major impairment on conventional measures of intelligence. In addition, the tester, who guides the patient's attentional focus, may to some degree play the part of the patient's frontal lobes, minimizing interference and distraction (Stuss and Benson, 1984). The testing environment itself can facilitate task segmentation and thus reduce a task's relational complexity. Nonetheless, damage to the prefrontal cortex will continue to produce deficits in tasks involving relations between novel combinations of elements, including tasks in which established schemas must be reinstantiated in new ways.

Our framework for analyzing relational complexity is in general agreement with previous theories of frontal lobe functions, but at the same time may prove useful in developing closer ties between these theories and computational models of thought. The diverse tasks that show impairment after frontal damage—including planning, sequencing of actions, using context to modulate social behavior, learning contingencies between spatiotemporally separate stimuli and responses, and forming flexible categories—all share a common requirement: Independently varying elements must be bound to specific roles with respect to relations. Future work should include task analyses to isolate the representational units required for successful task performance. In addition, theoretical work should be directed at the development of models of relational processing that capture the role of working-memory and attentional constraints.

NOTES

1. We assume that the generation of the response need not be explicitly represented by the animal; hence, only the arguments that serve as inputs to the function—and not the response produced as its output—are treated as dimensions of variation. This assumption is reflected in our notation: The dimensions are included within parentheses as arguments to the function, and "$\rightarrow R$," which is outside of the parentheses, signifies the mapping from the set of inputs to a response.

2. The required representation may actually be less complex than a full relational mapping in the sense of Halford

(1993), because the two arguments can be assigned separately on the basis of temporal cues. Similar caveats apply to all the two-slot representations suggested for tasks performed by nonhuman primates.

REFERENCES

CASE, R., 1992. The role of the frontal lobes in the regulation of cognitive development. *Brain Cogn.* 20:51–73.

DAMASIO, A. R., 1985. The frontal lobes. In *Clinical Neuropsychology*, ed. 2, K. M. Heilman and E. Valenstein, eds. New York: Oxford University Press, pp. 339–375.

DREWE, E. A., 1975. Go-no go learning after frontal lobe lesions in humans. *Cortex* 11:8–16.

FREEDMAN, M., and M. OSCAR-BERMAN, 1986. Bilateral frontal lobe disease and selective delayed response deficits in humans. *Behav. Neurosci.* 100:337–342.

FUSTER, J. M., 1989. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe.* New York: Raven.

FUSTER, J. M., 1993. Frontal lobes. *Curr. Opin. Neurobiol.* 3:160–165.

GHENT, L., M. MISHKIN, and H. L. TEUBER, 1962. Short-term memory after frontal lobe lesion in man. *J. Comp. Physiol. Psychol.* 55:705–709.

GOLDMAN-RAKIC, P. S., 1990. Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. In *Progress in Brain Research 85: The Prefrontal Cortex: Its Structure, Function and Pathology*, H. B. M. Uylings, C. G. Van Eden, J. P. De Bruin, M. A. Corner, and M. G. Feenstra, eds. New York: Elsevier, pp. 325–336.

GRATTAN, L., and P. ESLINGER, 1991. Frontal lobe damage in children and adults: A comparative review. *Dev. Neuropsychol.* 7:283–326.

HALFORD, G. S., 1993. *Children's Understanding: The Development of Mental Models.* Hillsdale, N.J.: Erlbaum.

HALFORD, G. S., and W. H. WILSON, 1980. A category theory approach to cognitive development. *Cognitive Psychol.* 12:356–411.

JACOBSON, C. F., 1935. Functions of frontal association area in primates. *Arch. Neurol. Psychiatry* 33:558–569.

KNIGHT, R. T., 1991. Evoked potential studies of attention capacity in human frontal lobe lesions. In *Frontal Lobe Function and Dysfunction*, H. S. Levin, H. M. Eisenberg, and A. L. Benton, eds. New York: Oxford University Press, pp. 139–155.

LEVIN, H., K. CULHANE, J. HARTMANN, K. EVANKOVICH, A. MATTSON, H. HARWARD, G. RINGHOLZ, L. EWING-COBBS, and J. FLETCHER, 1991. Developmental changes in performance on tests of purported frontal lobe functioning. *Dev. Neuropsychol.* 7:377–395.

LEWINSOHN, P. M., R. E. ZIELER, J. LIBET, S. EYEBERG, and G. NIELSON, 1972. Short-term memory: A comparison between frontal and nonfrontal right- and left-hemisphere brain-damaged patients. *J. Comp. Physiol. Psychol.* 81:248–255.

MILNER, B., 1964. Some effects of frontal lobectomy in man. In *The Frontal Granular Cortex and Behavior*, W. J. Warren and K. Akert, eds. New York: McGraw-Hill, pp. 313–334.

MISHKIN, M., and F. J. MANNING, 1978. Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res.* 1443:313–323.

OWEN, A., J. DOWNES, B. SAHAKIAN, C. POLKEY, and T. ROBBINS, 1990. Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia* 28:1021–1034.

PASSINGHAM, R. E., 1985. Prefrontal cortex and the sequencing of movement in monkeys Macaca mulatta. *Neuropsychologia* 23:453–462.

PETRIDES, M., 1985. Deficits in non-spatial conditional associative learning after periarcuate lesions in the monkey. *Behav. Brain Res.* 16:95–101.

PETRIDES, M., 1986. The effects of periarcuate lesions in the monkey on the performance of symmetrically and asymmetrically reinforced visual and auditory go, no-go tasks. *J. Neurosci.* 6:2054–2063.

PETRIDES, M., 1987. Conditional learning and primate cortex. In *The Frontal Lobes Revisited*, E. Perecman, ed. N.J.: Erlbaum, pp. 91–108.

ROSENKILDE, C. E., H. E. ROSVOLD, and M. MISHKIN, 1981. Time discrimination with positional responses after selective prefrontal lesions in monkeys. *Brain Res.* 210:129–144.

SHALLICE, T., 1982. Specific impairments of planning. *Phil. Trans. R. Soc. Lond. [Biol.]* 298:199–209.

SMITH, L. B., 1989. From global similarities to kinds of similarities: The construction of dimensions in development. In *Similarity and Analogical Reasoning*, S. Vosniadou and A. Ortony, eds. Cambridge: Cambridge University Press, pp. 146–178.

STAMM, J. S., 1987. Monkey's delayed response deficit solved. In *The Frontal Lobes Revisited*, E. Perecman, ed. N.J.: Erlbaum, pp. 73–89.

STUSS, D. T., and D. F. BENSON, 1984. Neuropsychological studies of the frontal lobes. *Psychol. Bull.* 95:3–28.

STUSS, D. T., D. F. BENSON, E. F. KAPLAN, W. S. WEIR, M. A. NAESER, I. LIEBERMAN, and D. FERRILL, 1983. The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia* 21:235–248.

THATCHER, R. W., 1991. Maturation of the human frontal lobes: Physiological evidence for staging. *Dev. Neuropsychol.* 7:397–419.

YAMATANI, K., O. TAKETOSHI, H. NISHIJO, and A. TAKAKU, 1990. Activity and distribution of learning-related neurons in monkey Macaca fuscata prefrontal cortex. *Behav. Neurosci.* 104:503–531.