The ability to conceive of highly abstract concepts is a fundamental feature of human cognition. Using abstract mental representations, we can organize perceptions garnered from disparate experiences, develop novel solutions to problems we encounter, and predict future outcomes based on past experiences. All of these faculties have been shown to rely on the integrity of the lateral prefrontal cortex (e.g., Milner, 1964; Luria, 1966; Shallice, 1982; Duncan et al., 1995). Understanding abstract thought and the nature of abstract mental representations, therefore, may provide a useful framework for understanding the functions and organization of lateral prefrontal cortex.

In this chapter, we focus on the evidence and implications for an organization of prefrontal cortex according to different levels of abstraction in representational content. In the first section, we offer a historical review of some of the central concepts found in philosophical theories of abstraction and discuss how they motivate contemporary investigation in the field of cognitive neuroscience.

In the second section, we describe two recent functional neuroimaging experiments that examine the role of abstract representations in terms of lateral prefrontal cortex organization. The results of these experiments suggest a topographical organization of lateral prefrontal regions according to the level of abstraction in representational content (Fig. 6-1). This topography appears to follow an arcuate posterior-to-anterior direction, with concrete working memory representations corresponding to posterior prefrontal regions and representations at increasing levels of abstraction corresponding to progressively anterior regions.

HISTORICAL DEVELOPMENT IN THEORIES OF ABstraction

Discussions of abstraction are ubiquitous in cognitive neuroscience literature, with terms such as "abstract cognitive abilities," "abstract thought," and
Figure 6–1 Proposed arcuate topography in the human lateral prefrontal cortex. The arrow depicts the direction of the proposed gradient of abstraction, with increasing levels of abstraction in working memory representation located toward the anterior prefrontal cortex. Numbers indicate Brodmann areas in lateral prefrontal cortex. DLPFC, dorsolateral prefrontal cortex; RLPFC, rostrolateral prefrontal cortex; VLPFC, ventrolateral prefrontal cortex.

"abstract rules" almost invariably employed when lateral prefrontal functions are discussed (e.g., Milner, 1963; Luria, 1966; Baker et al., 1996; Christoff and Gabrieli, 2000; O’Reilly et al., 2002; Bunge et al., 2003; Miller et al., 2003; Sakai and Passingham, 2003). Despite this widespread use, there is considerable ambiguity concerning the meaning of this term. Sometimes “abstractness” is equated with difficulty of comprehension or lack of intrinsic form; oftentimes, the term is simply left undefined. For our purposes, we hearken back to its roots in the Latin abstrahere, meaning “to drag away,” and emphasize that abstract concepts are removed from specific instances.

The manner in which abstract concepts are “dragged away” from the specific has been the subject of much debate among thinkers. As George Berkeley wrote, “He who is not a perfect stranger to the writings and disputes of philosophers, must needs acknowledge that no small part of them are spent about abstract ideas” (Berkeley, 1734/1998). Here we review the development of some of the central philosophical concepts concerning abstraction.

**Philosophical Theories of Abstraction**

The idea for a distinction between abstract and concrete entities emerged as early as the time of Plato (c. 427–347 BCE), who contributed an early notion of
Abstraction of Mental Representations

abstraction with his theory of forms (Plato, 360 BCE/2003). Plato believed that separate from the flawed, imperfect realm of sensation is a perfect realm of forms, such as "beauty," "goodness," "equality," "likeness," "sameness," and "difference," which give structure to our world. In Plato's view, "sensibles," or objects of sensation, draw their characteristics from forms. Many beautiful objects exist, but all draw their common characteristic from a single form, namely, beauty.

Plato's view that forms and sensibles occupy different realms has been opposed by those who believe that forms and sensibles must inform each other, and as such, cannot be so completely separated. Aristotle (384–322 BCE), Plato's student, was one thinker who believed that sensation and form are inseparable. In Aristotle's view, form is embedded in sensation, and separating form from the sensory world, therefore, is a dubious matter (Aristotle, 350 BCE/2002). This idea of dynamic and interactive abstract and concrete concepts is important to cognitive theories of abstraction today.

Plato and Aristotle saw abstractness in objective terms, viewing forms and sensibles as entities that existed in the absolute. Later philosophers, such as John Locke and George Berkeley, viewed abstraction as a mental concept, and accordingly framed the discussion of the abstract-concrete distinction within the mind. In his Essay Concerning Human Understanding, John Locke distinguishes between particular and general ideas. Particular ideas are constrained to specific contexts in space and time. General ideas are free from such restraints and thus can be applied to many different situations. In Locke's view, abstraction is the process in which "ideas taken from particular beings become general representatives of all of the same kind" by dint of the mind's removing particular circumstances from an idea (Locke, 1690/1979).

Locke's idea of abstraction faced criticism from later thinkers, such as George Berkeley (1685–1753). In the introduction to his Treatise Concerning the Principles of Human Knowledge, Berkeley attacks Locke's notion that one individual quality of an object can be isolated from other qualities (Berkeley, 1734/1998). For example, he argued, it is impossible to imagine the motion of an object without also imagining its shape, color, and direction.

As an alternative to Locke's account, Berkeley argues that abstraction occurs through a shift in attention. In this view, a particular object can represent a group of objects when we focus attention on one of its qualities. For example, the image of a particular triangle, regardless of whether it is equilateral, isosceles, scalene, right, obtuse, or acute, can be used to represent all possible triangles when attention is focused on its feature of having three connecting line segments.

Berkeley thus introduced the notion that attention plays an important role in the process of extracting abstract ideas. The idea that abstraction occurs through focusing attention on a particular feature—a process today referred to as "selective attention"—has inspired many subsequent empirical developments and is central to a number of present-day theories of abstraction.
Cognitive Theories of Abstraction

The study of abstract ideas has more recently become a subject of investigations in the empirical disciplines of cognitive science, psychology, linguistics, and cognitive neuroscience. Contemporary thinkers in these fields have extended philosophical theories of abstraction by incorporating ideas of attention, perception, and neural connectivity.

One cognitive theory, developed in a series of works by George Lakoff and Mark Johnson, argues for a dichotomy of abstract and concrete concepts based on metaphorical understanding (Lakoff and Johnson, 1980a, b, 1999). More concrete concepts, such as “up,” “down,” “front,” “back,” “substance,” “container,” “motion,” and “food,” are understood directly from bodily experience. More abstract concepts, such as “time,” “emotions,” “communication,” “mind,” and “ideas,” are understood and structured in terms of more concrete and embodied concepts. For example, the abstract concept of “idea” can be understood metaphorically through the more concrete concepts of “commodity” (“How you package your ideas is important”) or “food” (“This idea is hard to digest”).

Notably, even the relatively concrete concepts described by Lakoff and Johnson are abstracted from more specific instances. For example, “food” is more abstract than the specific instances of “burrito” or “cheesecake.” We can thus interpret Lakoff and Johnson’s theory as part of a three-tier hierarchy of abstraction. Abstract concepts are understood by metaphorically mapping them onto relatively less abstract concepts, which in turn are learned by abstracting specific concrete instances encountered through sensation. Evidence for a similar three-level system of abstraction in representation is suggested by the neuroimaging findings presented later in this chapter.

In Lakoff and Johnson’s theory, the way in which highly abstract concepts are specifically mapped onto less abstract ones is determined by which metaphorical features are emphasized and which are de-emphasized, a process the authors refer to as “highlighting” and “hiding” (Lakoff and Johnson, 1980b). When we think of an argument in terms of war (“She shot down my argument,” “He attacked the weak points of my argument”), we highlight the combative aspects and hide the cooperative aspects of the situation. We have a different understanding of an argument if we think of it in terms of a process of achieving mutual understanding (“He responded to each point in my argument”), thus highlighting its cooperative aspects and hiding its combative aspects. Whereas Berkeley posited an idea of mapping concrete to abstract concepts with selective attention, Lakoff and Johnson emphasize the role of selective attention in mapping highly abstract, metaphorical concepts to relatively less abstract, nonmetaphorical concepts.

Other contemporary cognitive theories of abstraction do not rely on metaphor. For example, Lawrence Barsalou (1999) criticizes Lakoff and Johnson’s theory on the grounds that metaphorical mappings alone cannot produce adequate conceptual understanding. Knowing that an idea is like a commodity
and like food, for example, is hardly sufficient for understanding the concept of “idea.” Furthermore, Barsalou notes that Lakoff and Johnson ignore the possibility that much of what they denote “conceptual metaphor” may actually just function as polysemy. For instance, in the sentence “He attacked the weak points in my argument,” the word “attack” may function in two distinct ways—as both a physical action intended to cause harm and an attempted logical criticism.

Instead, Barsalou advocates for a theory wherein connections between concrete and abstract concepts are direct and nonmetaphorical. In this theory, concepts take the form of simulators, which are semantic clusters that can generate infinite further examples of a concept. As we encounter examples of objects, we encode their perceptual features and store them in our memories. These features will form into clusters, which eventually become simulators, with a frame of previously encountered common features and a set of infinite possible simulations that the frame can generate. For instance, our perceptual experience with various chairs has helped us form a concept of “chair.” We can now use this concept to simulate infinite further examples of chairs.

An abstract concept is understood in this system through three sequential processes. First, the concept is put into context by simulating an event sequence, or a system of projected actions associated with the concept. When we represent the abstract concept of “magic,” for example, we may simulate the actions of wand-waving, disappearance, and spontaneous transformation. Next, selective attention will extract features of this event sequence that are relevant to an understanding of the concept and to the internal state of the conceptual thinker. Thus, a skeptic may emphasize more the agents of artifice in his event simulations of the concept, whereas a believer may emphasize the miraculous features. Finally, introspective perceptual systems are employed in the interpretation of these selective concept attributes. These systems include emotional states; cognitive operations, such as search and comparison; and idiosyncratic systems of perceptual construal.

Barsalou’s theory provides a viable mechanism for concept dynamism and unconscious abstract representation. In addition, it builds on findings from neuroscience. Research showing the existence of neurons selective for certain object properties, such as color, orientation, and velocity, supports Barsalou’s proposed mechanism of conceptual clustering (Hubel and Wiesel, 1968; Wandell, 1995; Simoncelli and Heeger, 1998). This framework thus provides a promising preliminary link between theories of abstract thought and findings in neuroscience.

NEUROSCIENTIFIC EVIDENCE FOR ABSTRACTION OF REPRESENTATIONS IN LATERAL PREFRONTAL CORTEX

In the previous section, we reviewed some of the major theories concerning abstract notions and how these notions are represented in the mind. In this section, we will discuss how these abstract concepts are represented in the
brain. As mentioned earlier, our focus in this part will be on lateral prefrontal cortex and its functional organization. In particular, we argue that lateral prefrontal cortex is organized according to working memory representations at different levels of abstraction, with the most anterior part corresponding to the highest level of abstraction. In the following paragraphs, we will present some experimental results providing support for this conceptualization of prefrontal function.

Several studies on nonhuman primates have shown that prefrontal cortex plays a key role in abstract rule-guided behaviors (Wallis et al., 2001; Nieder et al., 2002; Bunge et al., 2003; Miller et al., 2003; see also Chapters 2 and 18). Single neurons in the monkey prefrontal cortex encode abstract rules and not simply the physical properties of the stimuli (Wallis et al., 2001). It has also been suggested that different prefrontal subregions deal distinctively with abstract and concrete information (Dias et al., 1996, 1997; see also Chapter 13). Dias and colleagues reported such distinction by using the intradimensional-extradimensional dynamic categorization task in marmoset monkeys. Deficits in intradimensional reversals, which require feature-level representation, were associated with orbitofrontal lesions, whereas deficits in extradimensional shifts, which require dimension-level representation, corresponded to dorsolateral prefrontal cortex (DLPFC) lesions. Building on the results from this study, O’Reilly and colleagues designed a computational model using a combination of activation-based working memory and frontal representations organized according to two different levels of abstraction (O’Reilly et al., 2002). Together, these studies introduced a distinction between concrete and abstract working memory representations in the brains of nonhuman primates.

Humans, however, appear capable of higher forms of abstraction in mental representation than those exhibited by nonhuman primates. Examples of such higher-level abstraction in mental processing include integrating multiple relations simultaneously (Halfford, 1984; Thompson et al., 1997) and solving analogies (Holyoak and Kroger, 1995; Gentner et al., 2001). Results from neuroimaging studies show that behaviors that require the use of abstract rules, such as reasoning and problem-solving, specifically activate the most anterior part of lateral prefrontal cortex, also known as rostrolateral prefrontal cortex (RLPFC), or lateral Brodmann area (BA) 10 (Baker et al., 1996; Christoff and Gabrieli, 2000; Christoff et al., 2001, 2003; Bunge et al., 2003; Bunge et al., 2005). This region is activated during the processing of highly abstract mental representations, such as internally generated information (Christoff and Gabrieli, 2000; Christoff et al., 2003), future task operations (Sakai and Passingham, 2003), abstract hierarchies of goals (Koechlin et al., 1999; Braver and Bongiolatti, 2002) and even meta-awareness during mind-wandering (Christoff et al., 2004; Smith et al., 2006). Comparative studies of BA 10 in human and nonhuman primates have revealed that this area occupies a proportionally larger volume of brain in humans than in other primates (Semendeferi et al., 2001), although the extent of this difference continues to be debated (Holloway, 2002).
Abstraction of Mental Representations

When combined, the evidence from studies in humans and nonhuman primates suggests a possible organization of lateral prefrontal cortex based on the level of abstraction in working memory representation, a theory that we have recently proposed (Christoff, 2003). Here we describe two recent experiments designed to test this hypothesis directly.

Following Rules at Different Levels of Abstraction: The Role of Lateral Prefrontal Cortex

To examine the possibility that rule-guided behavior at different levels of abstraction activates different prefrontal subregions, we constructed a task that involved rule-guided behavior and rule reversal at three levels of abstraction (Christoff et al., manuscript under review-a). Thirteen healthy volunteers were recruited to undergo functional magnetic resonance imaging (fMRI) scanning while performing the task. The experiment comprised three sessions, each containing 140 trials grouped into three different conditions: concrete rule, first-order abstract rule, and second-order abstract rule (Fig. 6-2). In the concrete condition, the target was a single circle whose black side was oriented in one of four directions—right, left, up, or down—as determined by a cue at the beginning of each block. In the first-order abstract rule, the cue identified the target as a pair of circles whose orientation could be either “same” or “different.” Finally, in the second-order abstract condition, the target consisted of two pairs of circles that could be either “related” (two pairs that were either “same”- “same” or “different”- “different”) or “unrelated” (“same”- “different” or “different”- “same”).

Each condition started with a cue that determined both the level of abstraction (condition) and the applicable rule at each level, followed by a series of visual stimuli. In all conditions, the stimuli were two pairs of circles separated by a vertical line. Each circle consisted of one white and one black semicircle. In each condition, after a number of successive trials, the rule reversed (Fig. 6-3). There were five reversals per session. In all conditions, participants pressed the left button to indicate that the target was present on the screen, or the right button to indicate that the target was not present. Their response was followed by immediate feedback on every trial.

Statistical parametric mapping analysis was performed to determine which areas of the brain are activated during reversals at each level of abstraction (i.e., concrete, first order of abstraction, and second order of abstraction). The effects of reversal were modeled using parametric modulation within an event-related model (Buchel et al., 1998). Each reversal was modeled as an event, and a parametric regressor was then constructed using “time prior to reversal” (i.e., the time elapsed after the last reversal). This allowed us to model the effects of stronger engagement of executive control at reversals occurring after larger number of trials spent following the same rule (i.e., looking for the same target).

The areas of prefrontal cortex that were activated during reversals at each level of abstraction are shown in Figure 6–4 (see color insert). In the concrete
a) Concrete rule condition

<table>
<thead>
<tr>
<th>Right</th>
<th>Left</th>
<th>Up</th>
<th>Down</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Concrete rule condition" /></td>
<td><img src="image" alt="Concrete rule condition" /></td>
<td><img src="image" alt="Concrete rule condition" /></td>
<td><img src="image" alt="Concrete rule condition" /></td>
</tr>
</tbody>
</table>

b) First-order abstract rule condition

<table>
<thead>
<tr>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="First-order abstract rule condition" /></td>
<td><img src="image" alt="First-order abstract rule condition" /></td>
</tr>
</tbody>
</table>

c) Second-order abstract rule condition

<table>
<thead>
<tr>
<th>Related</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Second-order abstract rule condition" /></td>
<td><img src="image" alt="Second-order abstract rule condition" /></td>
</tr>
</tbody>
</table>

Figure 6-2 Following rules at different levels of abstraction: stimuli and targets for the three conditions. Rules are enclosed in quotation marks, and correct responses are shown in parentheses. Dashed arrows indicate the targets.

condition, significant activation was observed in left insula, posterior ventrolateral prefrontal cortex (VLPFC), and right supplementary motor cortex (BA 6 and 8). In the first-order abstraction condition, activation was limited to anterior ventrolateral prefrontal cortex (BA 11/47). Finally, in the second-order abstraction condition, right VLPFC (BA 47), left VLPFC (BA 45), and left RLPFC (BA 10) showed significant activation. This RLPFC activation, however, was specific to the second-order abstraction condition, and was not observed during rule reversal at the concrete or first-order abstraction rules (Fig. 6–5).

These results provide additional evidence that RLPFC is recruited when rules at the highest order of abstraction are used to guide behavior, in agreement with previous research (Christoff et al., 2001, 2003; Bunge et al., 2003; Sakai and Passingham, 2003). On the other hand, more posterior prefrontal regions were engaged during executive control at lower orders of abstraction and during concrete rules—a finding consistent with previous results (Dias et al., 1996,
Abstraction of Mental Representations

Figure 6-3 An example of rule reversal during the second-order abstract condition. Rule reversals during the experiment included the following possibilities: “right” to “left,” “left” to “right,” “up” to “down,” and “down” to “up” (concrete condition); “same” to “different” and “different” to “same” (first-order abstract condition); “related” to “unrelated” and “unrelated” to “related” (second-order abstract condition).

One advantage of the study described here, however, is that it varied the level of abstraction in rule reversal at all three levels within each subject, thus allowing the regions recruited at different levels to be examined within the same data set. Within this study, the observed recruitment of different prefrontal subregions is strongly suggestive of organization topography based on levels of abstraction in mental representation.

A question that often arises when different levels of abstraction in executive processing are considered is whether such variation in abstraction is not the same as variation in difficulty of processing. A link between increasing level of abstraction and increasing complexity in processing, as indexed by reaction times and accuracy, is evident in the bulk of previous research (Christoff et al., 2001, 2003; Bunge et al., 2003; Sakai and Passingham, 2003). In fact, we have recently suggested that cognitive complexity may be one of the crucial factors in understanding the functions of higher-order regions, such as the anterior parts of prefrontal cortex (Christoff and Owen, 2006).

Even though this association between difficulty and level of abstraction has been addressed at the level of statistical analysis (Christoff et al., 2001, 2003),
Rule Representation

Figure 6–4 Group-averaged brain activations during rule reversal at different levels of abstraction. Ventrolateral (x, y, z = −44, 22, 0); orbitolateral (x, y, z = 48, 38, −12); and rostrolateral (x, y, z = −26, 48, 12). Prefrontal cortices showed significantly increased fMRI signal during the concrete, first-order abstract, and second-order abstract conditions, respectively (p < 0.001, uncorrected).

difficulty at the behavioral level almost invariably increases together with increasing level of abstraction. A similar simultaneous increase was also seen in this experiment, allowing for the argument that anterior prefrontal cortex recruitment was due to task difficulty, rather than the degree of abstractness. In the next section, we describe another study that directly aimed to dissociate the confounding effect of task difficulty and abstractness, while further testing the hypothesis for lateral prefrontal cortex organization according to abstraction in representation.

Maintaining a Cognitive Mindset at Different Levels of Abstraction during Problem-Solving

To keep difficulty constant while varying the level of abstraction during executive processing, we used a verbal problem-solving task (Fig. 6–6) (Christoff et al., manuscript under review b). In this task, participants unscrambled anagrams to form words that were either highly abstract (e.g., “appeal,” “hope”), highly concrete (e.g., “desk,” “bottle”), or in the medium range of concrete-
Abstraction of Mental Representations

Figure 6-5 Condition-specific parameter estimates at the peak of activation in left rostrolateral prefrontal cortex (RLPFC) during rule reversal for concrete, first-order abstract, and second-order abstract conditions. The figure demonstrates significant recruitment of RLPFC specific to reversals of rules at second order of abstraction.

ness (e.g., “hero,” “path”). All words (nouns) were selected from the MRC psycholinguistic database (Wilson, 1988), with abstraction ratings according to Paivio et al. (1968). The words were clustered into groups of abstract, medium, and concrete, and matched for frequency, number of letters, and number of syllables.

Subjects were instructed to press a button once they had identified each word and to say the unscrambled word aloud, which provided the measure of reaction time for each trial. Their voices were recorded to obtain an estimate of accuracy. To help the subjects enter a specific mindset (abstract, medium, or concrete), a 2-second instruction period was displayed at the beginning of each block, during which the word “Abstract,” “Medium,” or “Concrete” was presented at the top part of the screen. This instruction remained during the entire block, indicating the category for the solutions to the following anagrams. After the instruction period, a set of anagrams was displayed at the bottom of the screen, one at a time. A pilot study was carried out before the actual experiment, testing different scrambled versions for accuracy and response time. The results of this pilot study were used to select those combinations of scrambled letters that yielded comparable difficulty, as measured by accuracy and response time, across the different levels of abstraction.
Rule Representation

Figure 6-6 Verbal problem-solving task (anagram) design. Subjects were shown instructions about the condition (concrete, medium, or abstract) for 2 seconds, followed by eight anagram-solving trials lasting 4 seconds each. The instructions remained on the screen while participants attempted to solve the anagrams. Solutions to the concrete examples are “Desk,” “Motor,” and “Bottle”; to the medium examples are “Trip,” “Dance,” and “Symbol”; and to the abstract examples are “Myth,” “Appeal,” and “Grace.”

Behavioral results from the fMRI experiment confirmed that reaction time and accuracy did not differ across different levels of abstraction (Fig. 6-7), allowing us to examine the regions of activation in the absence of difficulty variation. Each condition was compared with the other two conditions, resulting in three comparisons of interest. The observed activations (Fig. 6-8; see color insert) revealed a striking topography within lateral prefrontal cortex. RLPFC (BA 10/146) was the strongest region of activation in the prefrontal cortex when the abstract anagram solution was compared with the other two conditions. Right DLPFC (BA 46), on the other hand, was the strongest area of activation when the medium condition was compared with the other two. Finally, right VLPFC (BA 47/11) emerged as the most significant area of activation for concrete versus other conditions. These results provide the most conclusive evidence to date for lateral prefrontal cortex topography at different levels of abstractness.
SUMMARY AND DISCUSSION

In the first half of this chapter, we briefly reviewed a number of theoretical considerations regarding the ontology of abstract entities as well as the issues involved in the mental representation of these nonphysical entities. In the second half, we described some experimental findings concerning the cortical representation of information at different levels of abstraction in nonhuman and human primates. We presented the results from two recent neuroimaging studies that support the hypothesis that the human lateral prefrontal cortex is organized according to working memory representations at different levels of abstraction.

In the first study, a rule-reversal task with three conditions at three levels of abstraction was used. Different areas in prefrontal cortex showed significant activation during executive control in each condition. The most anterior prefrontal region, RLPFC, was only activated during executive control at the highest level of abstraction. Posterior prefrontal cortices, on the other hand, were recruited for executive control at lower levels of abstraction. In the second study, we varied the levels of abstraction in mental mindset during a verbal problem-solving task in which anagram solutions were either highly concrete, highly abstract, or at a medium level of abstraction. To our knowledge, this is the first study that dissociates task difficulty from abstraction in mental representation. When the abstract condition was compared with the other two conditions, RLPFC was the only prefrontal subregion that showed
significant activation. Mental mindset at a medium level of abstraction during this task was associated with the strongest DLPFC recruitment, and concrete mindset was associated with VLPFC recruitment. These findings suggest an arcuate topography in the human lateral prefrontal cortex (Fig. 6–1), based on working memory representations at different gradients of abstraction.

The idea that lateral prefrontal cortex in primates is organized according to the content of working memory representations was first developed by Goldman-Rakic (1996). According to this notion, VLPFC is specialized for object representations in working memory, whereas DLPFC is specialized for spatial representations. The evidence for such modality-specific conceptualization of prefrontal function, however, is mixed (Owen, 1997), and on a number of occasions, it has been openly challenged by the more consistently observed process-specific separation between the ventral and dorsal parts of

Figure 6–8 Group-average brain activations for anagram-solving at different levels of abstraction, each compared with the average activation for the other two conditions. Concrete blocks were associated with activation in ventrolateral prefrontal cortex (x, y, z = 32, 36, -16). Medium blocks were associated with activation in right dorsolateral prefrontal cortex (x, y, z = 46, 42, 24). Abstract blocks were associated with activation in right rostrolateral prefrontal cortex (x, y, z = 36, 48, 4) [p < 0.05 corrected]. Abs, abstract; med, medium; con, concrete.
Abstraction of Mental Representations

lateral prefrontal cortex (Petrides, 1996; D'Esposito et al., 1998; Owen et al., 1999). The hereby proposed organization of prefrontal function, while focusing on the content of working memory representation, regards as important not its modality, but the level of abstractness of informational content. One of the main advantages of this account is that it includes all lateral prefrontal subregions—anterior prefrontal cortex (RLPFC) as well as ventral (VLPFC) and dorsal (DLPFC) regions—within the same conceptual framework.

Another account of lateral prefrontal cortex organization that encompasses all of these regions, however, was proposed a number of years ago (Christoff and Gabrieli, 2000), incorporating the ventral-dorsal process-based distinction (Petrides, 1996; D'Esposito et al., 1998; Owen et al., 1999) in conjunction with the mental operations known to recruit anterior prefrontal cortex. This was an exclusively process-based account and proved extremely successful in conceptualizing the results of subsequent studies. The question might arise then: What is a better way of understanding prefrontal organization—using a process-based or a representation-based account? It is our view that both are necessary and indeed complementary ways of conceptualizing prefrontal functions. Process-based accounts appear to be most useful in the context of incremental extraction of information from the environment, whereby the products of one mental process are used by another mental process. A good example of such a paradigm was developed and used by D'Esposito et al. (2000), in which the manipulation of working memory information is performed on the products of encoding. Other examples include paradigms from the reasoning (Christoff et al., 2001; Kroger et al., 2002) and problem-solving (Baker et al., 1996) literature. In most of these cases, incremental recruitment of prefrontal cortex regions in a posterior-to-anterior direction is observed as additional mental processes are added, possibly reflecting the interactions between adjacent prefrontal subregions (Fig. 6–9A).

Figure 6–9 Examples of the hypothesized intrinsic (A) and extrinsic (B) interactions involving prefrontal subregions and their suggested topography based on patterns of connectivity found in studies in nonhuman primates (Pandya and Barnes, 1987). The direction of arrows indicates the hypothesized direction of prefrontal attentional control.
Rule Representation

Representation-based accounts, on the other hand, appear to be more useful in the context of attentional and cognitive control processes that require the maintenance of particular information in working memory for the purposes of biasing representations in regions outside of prefrontal cortex—as in the selective attention account proposed by Desimone and Duncan (1995). In these cases, recruitment of different prefrontal subregions may be observed in the absence of activation in other prefrontal regions, as in the results presented here from the anagram solution task. Thus, representation-based accounts of prefrontal function may be more useful in describing attentional control processes, presumably reflecting the interactions between prefrontal regions and other regions of the brain (Fig. 6-9B).

It is our belief that process and representation are flip sides of the same mental coin. Using fMRI forces us to treat relatively large brain regions as the basic unit of our analyses and theorizing. As such, “process” implies some interaction and information exchange between a region of interest and another brain region. On the other hand, “representation” implies a state of activation in a region of interest, without explicitly relating it to activation states in other brain regions. When a discussion is focused exclusively on a particular brain division, such as prefrontal cortex (as is the case in this chapter), the functions of particular prefrontal subregions may be best understood in terms of representational account (e.g., anterior prefrontal cortex supports working memory representations at high levels of abstraction). If the discussion is expanded to include areas outside prefrontal cortex, however, a process-based account may become relevant as well (e.g., anterior prefrontal cortex biases representations in anterior temporal cortex in support of the retrieval of semantic information at high levels of abstraction). To fully understand the organization of prefrontal function, it will be necessary to consider the functions of its subregions in terms of both process and representation—as well as the nature and systematicity of its intrinsic and extrinsic connections and interactions.

Another advantage of the hereby proposed account of prefrontal organization is that it is corroborated by findings from human neurodevelopment and patterns of brain connectivity in nonhuman primates. It has been suggested that, in human infants, maturation of prefrontal cortex starts from the most posterior part of prefrontal cortex and progresses toward the most anterior part (Diamond, 1991). These neurodevelopmental changes appear to be contemporaneous with the development of rule-learning abilities in preschoolers (Casey et al., 2000), who can first learn only concrete rules, and later can reason at increasingly higher levels of abstraction (Jacques and Zelazo, 2001; Bunge and Zelazo, 2006; see also Chapter 19). In addition, studies examining the connectivity patterns between prefrontal subregions and other brain areas have revealed a systematic topography whereby progressively more anterior prefrontal regions receive and send projections preferentially to regions of progressively higher levels of sensory integration (Pandya and Barnes, 1987). This pattern of connectivity supports the possibility that attentional
Abstraction of Mental Representations

control processes may occur in parallel streams from more anterior prefrontal regions to areas of higher-order sensory integration (Fig. 6–9B).

Earlier in this chapter, we emphasized the process of selective attention that occurs ubiquitously in the discussions by philosophers and modern cognitive scientists alike. Berkeley's process of focusing attention on a particular feature, Lakoff and Johnson's "highlighting and hiding" in understanding abstract concepts through metaphor, and Barsalou's idea for the application of selective attention in the extraction of particular features of a generalized event sequence all represent accounts of the process of abstraction, prominently featuring this aspect of attention. At the same time, neuroscientific accounts of prefrontal functions have also strongly emphasized the process of selective attention (Desimone and Duncan, 1995; Everling et al., 2002) and attentional control processes in general (Cohen et al., 1996; Braver and Cohen, 2001) in understanding prefrontal functions. In view of the strong association between prefrontal functions and abstraction, the processes of selective attention will likely turn out to be crucial for our neuroscientific theories of abstraction. Although no such theory currently emphasizes this process, our discussion here suggests that this would be necessary to advance understanding of the way prefrontal cortex represents abstract information and implements the process of abstraction. It is with such advancement that we will ultimately be able to better understand not only the neural but also the corresponding cognitive processes that enable this most uniquely human phenomenon.

ACKNOWLEDGMENTS Support for the preparation of this chapter came from the Tula Foundation, the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Michael Smith Foundation for Health Research. The authors thank Rachelle Smith for help with data collection as well as helpful comments and suggestions at the chapter's conception. We are also greatly indebted to Alan Gordon for his extensive and thoughtful feedback and his invaluable help at the later manuscript preparation stage.

NOTE
1. The term "arcuate" is used here in its sense of "curved" or "formed in the shape of an arc."

REFERENCES


Christoff K, Keramatian K, Dove A, Owen AM (manuscript under review a) Executive processing at different levels of abstraction.

Christoff K, Keramatian K, Smith R, Maedler B (manuscript under review b) Prefrontal organization and levels of abstraction in working memory representation.


Abstraction of Mental Representations


