S. Funahashi (Ed.)

Representation and Brain
S. Funahashi (Ed.)

Representation and Brain

With 101 Figures, Including 28 in Color
Preface

In our daily life, we perceive a variety of stimuli from the environment. However, among these, only a few selected stimuli are further processed in our brain. These selected stimuli are processed and integrated together with the information stored in long-term memory to generate an appropriate behavior. To achieve these processes, the brain needs to transform a variety of information into specific codes to be able to process it in the nervous system. For example, a digital computer uses a binary code to process information. A variety of information is transformed into a combination of 1’s and 0’s, or “on” and “off” states. When we push the “A” key on the keyboard, the character “a” appears on the monitor. In this case, the information that an “A” key was pushed is transformed into the binary code 01100001 in the computer, and this binary code is used to present the character “a” on the monitor. When we push the “A” key together with the “shift” key, however, this information is transformed into the binary code 01000001, and this code is used to present the character “A” on the monitor. Different characters and symbols are assigned to different binary codes in the computer using a specific common rule (e.g., ASCII format). A variety of information processed in the computer is represented using the specific rules of the binary code. This enables convenient and efficient use of any computer for a variety of purposes. By understanding the rules of information processing and the codes to represent information, we can understand how the computer works and how the computer processes information.

Similarly, specific common rules or common methods of coding can also be used in the brain for processing and representing information. How information is represented in the nervous system and how information that is represented in the nervous system is processed are the most important factors for understanding brain functions, especially for understanding neural mechanisms of higher cognitive functions such as thinking, reasoning, judging, and decision making. These subjects are the main themes of this book, in which we have tried to show the current state of our knowledge regarding information representation in the nervous system. We approach these subjects from four directions: studies on visual functions (Part I), motor functions (Part II), memory functions (Part III), and prefrontal cortical functions (Part IV).
How stimuli in our environment are represented in the nervous system is best known in the visual system. Part I focuses on this subject, particularly in visual information processing and visual image production. Perception of a visual target and neural responses to the target are strongly influenced by the context surrounding the target. The specific layout of the features surrounding the target can produce either a suppressive or a facilitatory effect on neural responses to the target. Ejima et al. investigated contextual effects of meta-contrast masking, chromatic induction, and the contour processing in the perception of an occluded object in human visual areas using fMRI. Brain activity in the higher-order visual areas showed more prominent specificity to stimulus attributes and contexts. Ejima et al. discuss this contextual modulation together with the possibility that the modulation plays different roles in different processing stages. By contrast, most theories of visual perception employ the concept of "top-down processing" (TDP). However, this concept has not been articulated in detail, and usually the term is used to indicate any use of stored knowledge in vision. Using data from a variety of research fields, Ganis and Kosslyn propose four distinct types of TDP ( reflexive processing, strategic processing, threshold lowering, and completion) that are engaged in different circumstances. They showed that reflexive processing occurs when bottom-up information automatically triggers TDP, which in turn sends feedback to an earlier process. Strategic processing occurs when executive control processes are used to direct a sequence of operations. Threshold lowering occurs when TDP increases the sensitivity of a neural population. Finally, completion occurs when TDP fills in missing information.

The neural mechanism of invariant object and face recognition is a major computational problem that is solved by the end of the ventral stream of the cortical visual pathway in the inferior temporal cortex. The formation of invariant representation is crucial for the areas that receive information from the inferior temporal cortex, so that when they implement learning based on one view, size, and position on the retina of the object, the learning is generalized later when the same object is seen in a different view, size, and position. Rolls presents evidence that invariant representations are provided by some neurons in the inferior temporal cortex and hypothesizes about how these representations are formed.

To efficiently interact with the environment, not only perception of objects and scenes but also maintenance of these visual representations for a brief period is crucial. Saiki focuses on some of the important issues regarding visual working memory for objects, scenes, and capacity; representational format (in particular feature binding); and manipulation of memory representations. Based on findings from both behavioral and functional brain imaging experiments, Saiki describes working memory for objects and scenes as a network of specialized brain regions.

Motor imagery and body schema are also good examples for understanding how information is represented in the nervous system. The “mirror neuron system” is an especially good model for understanding how we construct motor imagery and body schema and how we perceive and understand actions made by
others. Part II focuses on this subject. Recent neuroanatomical and functional data indicate that action itself and perception of action are linked, and that this link occurs through reciprocal parietofrontal circuits. These circuits not only are involved in different types of sensory-motor transformations but also allow the emergence of cognitive functions such as space perception and understanding an action. The representation of the goal of the action at the single neuron level is fundamental for these functions, because it constitutes internal knowledge with which the external word is matched. The mirror neuron system is an example of a neural mechanism through which we can recognize and interpret actions made by others by using our internal motor repertoire. Fogassi examined several types of motor functions, with a particular emphasis on those functions based on the mirror system, such as action understanding, understanding of intention, and imitation. Similarly, Murata describes the mirror system in the parietal cortex as possibly having a role in monitoring self-generated action by integrating sensory feedback and an efference copy. This idea is based on data showing that the parietal cortex receives somatosensory and visual inputs during body action, and that matching the efference copy of the action with the visual information of the target object is performed in the parietal area AIP (anterior intraparietal area).

The brain creates the neural representation of the motor imagery and the body schema by integrating motor, somatosensory, and visual information. Naito examined the neural representation of the motor imagery for hand movements using motor illusions, and showed that the execution, mental simulation, and somatic perception related to a particular action may share a common neural representation that uniquely maps its action in action-related areas in the human brain.

In addition to hand or arm movements, eye movements also play significant roles in our daily life. One of the most fundamental decisions in human behavior is the decision to look at a particular point in space. Recent advances in recording neural activity in the brain of epileptic patients have made it possible to identify the role of the brain areas related to the control of eye movements. Freyermuth et al. summarize some of the brain imaging studies on cortical control of saccades and describe recent findings obtained with intracranial recordings of brain activity in epileptic patients. They particularly focus on the subject of decision making in saccadic control by describing experiments during which subjects decide to make a saccade toward a particular direction in space.

Understanding how information is stored as memory in the brain is also a way to understand how a variety of information is represented internally in the nervous system. In order to discuss memory, we need to solve some important questions regarding representation, such as how the information is stored in the brain, what mechanism controls memory representation, and how flexible memory representation is. Part III focuses on this subject.

Examining the characteristics of the activity of place cells is one way to understand how flexible memory representation is. Brown and Taube examined whether place fields of hippocampal CA1 place cells are altered in animals receiving lesions in brain areas containing head direction (HD) cells. Although place
cells from lesioned animals did not differ from controls on many place-field characteristics, the signals (outfield firing rate, spatial coherence, and information content) were significantly degraded. Surprisingly, place cells from lesioned animals were more likely modulated by the directional heading of the animal. These results suggest that an intact HD system is not necessary for the maintenance of place fields, but lesions in brain areas that convey the HD signal can degrade this signal.

Sakurai explains why ensemble coding by cell assembly is a plausible view of the brain's neural code of memory and how we can detect ensemble coding in the brain. To consider ensemble coding, partial overlap of neurons among cell assemblies and connection dynamics within and among the assemblies are important factors. The actual dynamics include dependence on types of information-processing in different structures of the brain, sparse coding by distributed overlapped assemblies, and coincidence detection as a role of individual neurons to bind distributed neurons into cell assemblies. Sakurai describes functions of these parameters in memory functions.

Numbers are indispensable components of our daily life. We use them to quantify, rank, and identify objects. The verbal number concept allows humans to develop superior mathematical and logic skills, and the basic numerical capability is rooted in biological bases that can be explored in animals. Nieder describes its anatomical basis and neural mechanisms on many levels, down to the single-neuron level. He shows that neural representations of numerical information can engage extensive cerebral networks, but that the posterior parietal cortex and the prefrontal cortex are the key structures for this function in primates.

Finally, thinking is one of the most important subjects in the field of cognitive neuroscience. Thinking can be considered as neural processes in which a variety of internally representing information is processed simultaneously to produce a particular product in a particular context. Therefore, understanding how internally representing information is processed and how new information is created from representations being maintained in the nervous system must be the way to understand the neural mechanism of thinking. The prefrontal cortex is known to participate in this mechanism. Therefore, in Part IV, prefrontal functions are examined in relation to the manipulation and processing of internal representation.

An important function of the prefrontal cortex is the control and organization of goal-directed behavior. Wallis examines neural mechanisms underlying this function. This mechanism can be grouped in several levels according to the level of abstraction. At the simplest level, prefrontal neurons represent the expected outcomes of actions directed toward basic goals of homeostatic maintenance (e.g., maximizing energy intake or minimizing energy expenditure). At a more complex level, prefrontal neurons encode representations of arbitrary relationships between specific sensory stimuli and specific actions. Finally, at the most abstract level, prefrontal neurons represent rules and concepts. Wallis describes how these representations ensure optimal action selection for organisms.
The dorsolateral prefrontal cortex is known to participate in working memory. Working memory is a fundamental mechanism for many cognitive processes including thinking, reasoning, decision making, and language comprehension. Therefore, understanding neural mechanisms of working memory is crucial for understanding neural mechanisms of these cognitive processes. Funahashi describes experimental results showing the temporal change of information represented by a population of prefrontal activities during performances of spatial working memory tasks. He also points out widespread functional interactions among neighboring neurons and the dynamic modulation of functional interactions depending on the context of the trial, and concludes that functional interactions among neurons and their dynamic modulation depending on the context could be a neural basis of information processing in the prefrontal cortex.

Cognitive functions could be an outcome of the complex dynamic interaction among distributed brain systems. McIntosh describes two emerging principles which may help guide our understanding of the link between neural dynamics and cognition. The first principle is neural context, where the functional relevance of a brain area depends on the status of other connected areas. A region can participate in several cognitive functions through variations in its interactions with other areas. The second principle states that key nodes in distributed systems may serve as “behavioral catalysts,” enabling the transition between behavioral states. McIntosh suggests that an area either facilitates or catalyzes a shift in a dominant pattern of interactions between one set of regions to another, resulting in a change in the mental operation by virtue of its anatomical connections.

Thus, in this book, readers will find the current state of our knowledge regarding information representation in the nervous system. An analysis of single-neuron activity recorded from behaving animals has been used to reveal rules and codes that are used in the nervous system. We now know how light with a particular wave length is represented at the single-neuron level in the retina, the primary and secondary visual areas, and the inferior temporal cortex. We also know how the preparation of the movement toward a particular direction is represented in the activity of motor and premotor neurons. Advancement in neuroimaging techniques and the usage of a variety of behavioral paradigms in combination with advanced neurophysiological techniques will make it possible to approach these issues more closely.

I would like to express many thanks to the contributors of this book. I also would like to acknowledge that this publication is supported by the 21st Century COE (Center of Excellence) program of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) (Program number D-10, Kyoto University).

Shintaro Funahashi
Kyoto University
Kyoto, Japan
# Contents

Preface ........................................................................................................ V

Part I: Visual Information Processing and Visual Image Production

1  Visual Perception of Contextual Effect and Its Neural Correlates  
   Y. Ejima, S. Takahashi, H. Yamamoto, and N. Goda  ...................... 3

2  Multiple Mechanisms of Top-Down Processing in Vision  
   G. Ganis and S.M. Kosslyn ............................................................... 21

3  Invariant Representations of Objects in Natural Scenes in the Temporal Cortex Visual Areas  
   E.T. Rolls ......................................................................................... 47

4  Representation of Objects and Scenes in Visual Working Memory in Human Brain  
   J. Saiki ........................................................................................... 103

Part II: Motor Image and Body Schema

5  Action Representation in the Cerebral Cortex and the Cognitive Functions of the Motor System  
   L. Fogassi ......................................................................................... 123

6  Representation of Bodily Self in the Multimodal Parieto-Premotor Network  
   A. Murata and H. Ishida ................................................................. 151

7  Neuronal Correlates of the Simulation, Execution, and Perception of Limb Movements  
   E. Naito .......................................................................................... 177
8 Neural Basis of Saccadic Decision Making in the Human Cortex
   S. FREYERMUTH, J.-P. LACHAUX, P. KAHANE, and A. BERTHOZ . . . . . . 199

Part III: Memory as an Internal Representation
9 Neural Representations Supporting Spatial Navigation and Memory
   J.E. BROWN and J.S. TAUBE ...................................... 219

10 How Can We Detect Ensemble Coding by Cell Assembly?
    Y. SAKURAI .............................................................. 249

11 Representation of Numerical Information in the Brain
    A. NIEDER ............................................................... 271

Part IV: Manipulation of Internal Representation
12 Prefrontal Representations Underlying Goal-Directed Behavior
    J.D. WALLIS ............................................................. 287

13 The Prefrontal Cortex as a Model System to Understand Representation and Processing of Information
    S. FUNAHASHI .......................................................... 311

14 Large-Scale Network Dynamics in Neurocognitive Function
    A.R. McINTOSH ......................................................... 337

Subject Index ............................................................. 359
Part I
Visual Information Processing and Visual Image Production
1 Visual Perception of Contextual Effect and Its Neural Correlates

Yoshimichi Ejima¹, Shigeko Takahashi², Hiroki Yamamoto³, and Naokazu Goda⁴

1. Introduction

Gestalt principles of grouping, that is, how local signals are integrated across space to generate global percepts, have been central to the study of vision. Traditionally, visual information has been seen as ascending through a hierarchy of cortical areas, with cells at each successive stage processing inputs from increasingly larger regions of space. However, recent research using neurophysiological recordings in animals and functional magnetic resonance imaging (fMRI) in humans makes it increasingly clear that this traditional view is overly simplistic. The long-distance integration of visual signals can occur in the very early stages of processing. The response of cells in the primary visual cortex (V1) to stimulation of their receptive field can be modulated in a selective way by contextual stimuli lying far outside the receptive field in the receptive field surround. Although there is unquestionably some degree of serial processing from V1 to higher cortical areas, in many situations, activity in the early visual cortex is better correlated with, and may be more intimately involved in, perception than activity in later areas. Identifying the neural circuitry underlying these long-distance computations is crucial, because they may represent the neural substrates for feature grouping and figure-ground segregation. Here, in the first section, we have addressed this question by studying the cortical processing of simultaneous contrast effect of gratings patterns by fMRI measurement.

Whether perceptual grouping operates early or late in visual processing is an ongoing problem. One hypothesis is that elements in perceptual layouts are grouped early in vision, by properties of the retinal image, before meanings have

¹Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
²Department of Fine Arts, Kyoto City University of Arts, 13-6 Ohe-Kutsukake-cho, Nishikyo-ku, Kyoto 610-1197, Japan
³Graduate School of Human and Environmental Studies, Kyoto University, Yoshiden-nomatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan
⁴National Institute for Physiological Sciences, 38 Aza-saigou-naka, Myoudaiji-cho, Okazaki, Aichi 444-8585, Japan
been determined. Another hypothesis is that perceptual grouping operates on representation in higher cortical areas. How can early grouping be operationally defined and distinguished from higher-level grouping? To examine this question, we investigated grouping relative to meaning (see third section). By examining grouping relative to ambiguous achromatic display, we have demonstrated that perceptual grouping is based on the activation of early visual areas as well as on the activity of prefrontal cortical areas. This result permits the integration of cortical activity in early and higher cortical areas with regard to grouping by meaning.

2. Spatiotemporal Cortical Map of Contextual Modulation in the Human Early Visual Area

The appearance of a stimulus varies depending on the context in which the stimulus is presented. For instance, perceptual contrast of a stimulus is, in a complicated fashion, affected by surrounding stimuli depending on the relative contrast, orientation, and direction between the stimulus and the surround (Cannon and Fullenkamp 1991; Eijima and Takahashi 1985; Takeuchi and DeValois 2000; Xing and Heeger 2000). These lateral interaction effects, recently often called contextual effects, raised a fundamental question: how does local information interact to generate a percept across retinotopic space? It has been proposed that lateral interaction beyond “classical” receptive field in early visual areas is involved in complex functions such as contrast normalization and pop-out and figure-ground segregation (Kastner et al. 1997; Knierim and van Essen 1992; Lamme 1995; Sillito et al. 1995; Zipser et al. 1996).

Recent neuroimaging studies have demonstrated analogous, orientation-tuned lateral interaction in humans; responses to a local stimulus in early human visual areas were suppressed by adding surround stimuli (Kastner et al. 1998, 2001; Ohtani et al. 2002; Zenger-Landolt and Heeger 2003). However, the detailed spatial properties of the lateral interaction remain unclear. The spatial properties are important characteristics because they help to constrain the mechanism generating the modulation. The modulation may result from multiple mechanisms: long-range interaction, a short-range mechanism, and a mechanism sensitive to the local feature contrast at the boundary (Kastner et al. 1997; Knierim and van Esses 1992; Reinolds 1999; Sillito et al. 1995).

Here, we have analyzed the spatial distribution of response modulation in the human visual area during contextual change in perceptual contrast using a new fMRI analysis and visualization technique. Visual stimuli were composed of two test gratings or a circular surround grating or both (Fig. 1a). Stimuli were composed of test gratings in the two quarter (mainly upper right and lower left) visual fields at the middle eccentricities (2°–10°) and/or a large surround grating (0°–15° eccentricity), presented on a uniform gray background with a small fixation point. The test and surround gratings were spatially separated by small gaps. The orientations of the gratings were changed pseudo-randomly at 1 Hz (22.5°, 77.5°,
1. Neural Correlates for Contextual Effect

−22.5°, or −77.5°). Only oblique orientations were employed to avoid an “oblique effect” in fMRI response (Furmanski and Engel 2000). When the test and surround gratings were presented simultaneously, their relative orientation was kept constant at 0° (iso-orientation) or 90° (cross-orientation). Luminance contrast (0% [absence], 40%, or 80%), spatial frequency (2.0 cpd [cycle per degree] or 0.66 cpd), and relative spatial phase (0° or 180°) of the test and surround gratings were also manipulated. In the reference scan condition, a stimulus with the test component only (80% contrast, 2.0 cpd) and that with a surround component only (80% contrast, 2.0 cpd) were used. Retinotopic mapping experiments (Engel et al. 1994; Sereno et al. 1995) were also employed to locate visual areas for each subject. A checkered ring (2° width), expanding from the central (0.75°) to peripheral (16°), was used to map eccentricity in the visual field. A checkered 45° wedge, rotating counterclockwise, was used to map angular positions in the visual field. Subjects were instructed to stare continuously at the fixation point and to press a photo switch when the ring appeared at the center or at horizontal and vertical meridians to retain visual awareness.
Anatomical and functional MR measurements were conducted with a standard clinical 1.5-tesla scanner (General Electric Signa, Milwaukee, WI, USA) with echoplanar capability (Advanced NMR, Wilmington, MA, USA). Functional volumes were realigned (Woods et al. 1992). The functional volumes were then coregistered to the standard structural volume and thus to the reconstructed cortical surface. From the standard structural volume, the cortical surface was reconstructed for each hemisphere (Ejima et al. 2003). The retinotopic areas were determined by analyzing data from fMRI scans using phase-encoded stimuli. The boundaries between the areas V1d, V1v, V2d, V2v, V3, and VP, and V4v were determined based on the retinotopic criteria reported in previous studies (Ejima et al. 2003; Engel et al. 1994, 1997; Sereno et al. 1995).

The fMRI time series were sampled from each retinotopic area (V1d, V1v, V2d, V2v, V3, and VP). First, the surface regions of retinotopic areas defined by retinotopic mapping were inflated to 3-mm thickness, taking into account the thickness of gray matter, giving the gray matter regions of interest (GOIs). Next, the fMRI time series of the voxels overlapping with each GOI were sampled separately. The sampled voxel time series were then high-pass filtered to remove linear and nonlinear trends and converted to units of percent signal.

The fMRI time series sampled for each retinotopic area were then analyzed on the basis of the representation of eccentricity of the visual field. First, a set of “bands” was defined on the reconstructed surface based on the geodesic distance (the distance of the shortest path along the cortical surface) from the peripheral limit of the eccentricity map (16°) to the foveal region in each area. The bands were delineated so that the center of the bands differed in 1.5-mm steps in cortical distance and the width of each band was 3 mm. Thus, each band overlapped adjacent bands. We called these bands “iso-eccentricity bands.” Next, the voxel time series within each band were averaged, giving the mean time series of each eccentricity. Hereafter, the mean time series from repeated scans were averaged. The time series of each eccentricity were further averaged across stimulus cycles of the stimuli (24 or 48 cycles of two 16-s blocks). The event-related responses of different eccentricities were displayed as a spatiotemporal cortical map using an iso-contour plot with an interpolated pseudo-color format.

2.1. Surround Suppression

2.1.1. Retinotopic Localization of Stimulus Components

We first identified retinotopic subregions responding to a test stimulus component (located in the lower left and upper right quadrants of the visual field) and its surround stimulus components in each retinotopic area. Because the test and surround components were located at different eccentricities in the quarter visual fields, the regions responding to these components could be separated effectively in terms of retinotopy of the visual field eccentricity in each area. Figure 2a shows the results for right V1d, V2d, and V3, which represent the lower left quadrant of the visual field (Fig. 2, left panel) for one subject. These plots represent the
Fig. 2. Spatiotemporal maps of right V1d, V2d, and V3 for one subject. In each panel, mean time course at different eccentric locations for a stimulus cycle (two 16-s blocks) is plotted in pseudo-color format. The eccentric location was represented by cortical distance from a 16° eccentricity line. Increasing the distance (from lower to upper on vertical axis), the eccentricity varies from periphery to fovea. Schematic illustrations of the lower left visual field of stimuli in the two blocks are shown on the left of the pseudo-color plots. 

a Results for test-surround alternation (localizer condition). b Results for presenting an iso-oriented surround with a test component (iso-oriented surround condition). c Results for presenting a cross-oriented surround with a test component (cross-oriented surround condition). d Results for presenting an iso-oriented surround without test component (control condition). fMRI, functional magnetic resonance imaging.

Fig. 3. Spatiotemporal maps of V1, V2, and V3/VP, computed by averaging across hemispheres (right dorsal and left ventral). The dotted lines represent the test/surround boundaries, which were determined from the results of localizer scans.
spatiotemporal modulation of fMRI response concisely. Each horizontal trace represents the mean time course of an iso-eccentricity band, whose location is represented by the cortical distance (from peripheral to foveal region). Colors denote the signal values of the blood oxygen-action level-dependent (BOLD) contrast. These plots clearly show that the fMRI responses in middle eccentricities (about 15 mm wide) increased after presentation of the test with a hemodynamic delay, and those in more foveal and peripheral regions increased after the surround presentation. The spatial regions responding to the test were clearly separated from those of the surround.

We then averaged the spatiotemporal maps across subjects and hemispheres to improve signal-to-noise ratio further (Fig. 3a). As the extent of each area on the cortical surface was significantly different across subjects and hemispheres, we coregistered the map of each area before averaging so that the boundaries of the test and surround regions matched their average cortical distances using linear interpolation. To make averaged maps for each visual area V1, V2, and V3/VP, maps of the right dorsal and left ventral parts for two subjects were averaged.

2.1.2. Effect of Adding Iso- and Cross-Oriented Surround

Using similar stimulus setups and analysis, we examined how the response to the test component varied when the surround component was presented simultaneously. The test (40% contrast) was continuously presented during alternating two blocks. The high contrast surround (80% contrast, the same spatial phase as the test phase) was presented in only one of the two 16-s blocks. Because the test was always presented, any variation in the response was attributed to the effects by surround stimulation. We used two types of surround stimuli: an “iso-oriented” and a “cross-oriented” surround. In these stimuli, the relative orientation of the test and surround gratings was kept constant at either 0° (iso-orientation) or 90° (cross-orientation), while each orientation of the test and surround was changed at 1 Hz. Observers reported that the test was perceived as a lower contrast in the presence of the surrounds and the contrast suppressions were stronger for iso- than cross-orientation stimuli.

The pattern of the results indicated that the iso- and cross-oriented surround suppressed the test response in at least V1 and V2. In individual (Fig. 2b, c) and group-averaged maps (Fig. 3b, c), strong response modulations were observed in the foveal and peripheral surround regions. These reflected responses evoked by the surround component directly. In addition to these surround responses, response modulations in opposite phase to the surround responses were observed within regions centered on the test regions in V1 and V2. The responses at this region increased when the surround was absent and decreased when the surround was present, indicating that the surround suppressed the response to the test. The comparison of iso- and cross-orientation suggests that the magnitude of the suppression was stronger for iso-orientation than for cross-orientation.
The suppression of the test responses was therefore dependent on the relative orientation of the test and the surround. The orientation selectivity of the suppression was examined in detail in the next series of experiments.

To quantify the magnitude of the suppression and its spatial distribution for each stimulus condition, we computed the modulation amplitude of each iso-eccentricity band in a group-averaged map from the amplitude and phase of the stimulus frequency component (Fig. 4). The estimated modulation amplitude was plotted as a function of cortical distance (from peripheral to foveal region). Figure 4b,c shows that test responses were reduced in the presence of the iso- and cross-oriented surround in V1 and V2 but not in V3/VP. In V1 and V2, significant suppressive modulation was observed in narrow regions around the test center, which was about 7 mm distant from the test-surround boundary. These modulation patterns also indicated about 3-mm surround response spread into

![Diagram](image)

**Fig. 4.** Distribution of response modulations as a function of cortical distance from the 16° eccentricity line. Positive denotes that the response in the test-block was stronger than in the with-surround block, when the surround was presented. **Bars** show statistically significant modulations (Fourier $F$ test, $P < 1 \times 10^{-4}$). **a** Results for localizer condition. **b** Results for iso-oriented surround condition. **c** Results for cross-oriented surround condition. **d** Results for control condition.
the test region, which could cancel out the suppressive modulation of the test responses near the test-surround boundary.

2.1.3. Surround Modulation in the Absence of Test

In a separate control scan, we assessed whether surround suppression could be observed if the test was absent. If suppression was silent or much weaker in the absence of the test than in the presence of the test, it is likely that suppression is mediated by the interaction beyond Classified Receptive Field (CRF) (silent inhibition). We tested this by presenting a surround component in only one of the two 16-s blocks (test was not presented in both blocks).

The magnitude of the suppressive modulations without the test component in each area (Figs. 3d, 4d) was weaker than for the iso-orientation stimulus with the test (Figs. 3b, 4b), although it was similar to the cross-orientation stimulus (Figs. 3c and 4c). Thus, modulation without test, which might reflect the suppression of baseline activity in the test region or hemodynamic stealing uncorrelated with neural activity, was too weak to account for suppressive modulation, especially for the iso-orientation condition. Hence, we concluded that interaction beyond CRF predominantly contributes to suppression by the surround, at least for the iso-orientation condition.

Figures 3d and 4d show that surround responses in V3/VP spread into the center of the test region, probably because of relatively large CRFs (Smith et al. 2001). In V1 and V2, the spread of the surround response was limited to a smaller region near the test-surround boundary. The wide spread of response in V3/VP may explain why surround suppression of the test response was not evident in V3/VP, as shown in Figs. 3bc and 4bc. For our stimulus setting, suppressive modulations in the test region were likely to be largely canceled out by the spread surround responses. Cancellation by the spread surround response implies that the apparent magnitude of the suppression would depend on the size of CRF, the size of the stimulus, and spatial resolution of the measurement. If a larger stimulus was used, suppression might be observed in V3/VP. Conversely, if the modulations were measured in coarser spatial resolution, for instance, using large regions of interest (ROIs) such as the whole test region shown in Fig. 3a, modulation in the presence of the surround in V1 and V2 might be underestimated.

2.2. Orientation Selectivity of Lateral Interaction: Iso-Orientation Versus Cross-Orientation

In the second series of experiments, we compared responses to iso- and cross-orientation stimuli directly using a block-design paradigm to examine further the orientation selectivity of the surround suppression. The iso- and cross-orientation stimuli used in experiment 1 were presented in alternating 16-s blocks; hence, only the relative orientations of the test and surround gratings were changed between the blocks. We expected that the response in the test region would be stronger for the cross-orientation stimulus because iso-oriented surround sup-
presses the test response more strongly than cross-oriented surround. We also made measurements with mirror-symmetrical stimuli (test components in the lower right and upper left quadrants of the visual field) to test interhemisphere asymmetry.

We observed robust response modulations that were phase locked to the change in relative orientation of the test and surround. Figures 5 and 6 show group-averaged results. The responses in the test regions increased for the cross-orientation stimulus and decreased for the iso-orientation stimulus. The modulations were statistically significant in the test region for V1, V2, and V3/VP (Fig. 6a); thus, as expected, the test regions responded more strongly to the cross- than the iso-orientation stimulus. The modulations were also statistically significant for the mirror-symmetrical stimuli (Figs. 5b, 6b), indicating no interhemisphere asymmetry. These results supported that suppression of the test responses was stronger with the iso-oriented surround than the cross-oriented surround in V1 and V2, as suggested by the results of experiment 1. The results also suggested that similar orientation-selective modulation also occurred in V3/VP, which did not show the clear existence of surround suppression in experiment 1.

Note here that the spatial pattern of the response modulation in Figs. 5a,b and 6a,b reflects the configuration of the test and the surround. Interestingly, we

![Diagram of response modulation in V1, V2, and V3/VP](image-url)  
**Fig. 5.** Group-averaged spatiotemporal maps of V1, V2, and V3/VP to alternation of the cross- and the iso-orientation stimuli. **a** Results for the surround of the same spatial frequency and phase as the test (iso-phase surround condition). **b** Results for mirror-symmetrical iso-phase stimuli. **c** Results for the surround of the same spatial frequency but in antiphase (antiphase surround condition). **d** Results for the surround of low spatial frequency (low-SF surround condition).
observed modulation in the surround regions, which were in reverse relationship with those in the test regions (that is, decrease for the cross-orientation stimulus and increase for the iso-orientation stimulus). The magnitude of suppressive (positive) modulation tended to peak at the center of the test region, and the modulation reversed across the boundaries. This finding means that not only the response to the test component but also that to the high-contrast surround component could be modulated depending on the relative orientation; this suggests that lateral interaction could be antagonistic between the test and surround regions.

Although an antagonistic pattern of modulations was observed in many cases, in some cases the modulations peaked at the test-surround boundary and reduced as the distance from the boundary increased. This spatial pattern of response could be well explained by the local responses to feature contrast at the boundary (Kastner et al. 1997; Knierim and Van Essen 1992). It can be argued that the local responses could spread into the center of the test region, resulting in an apparent peak, especially for areas with larger CRF. According to our result
shown in Fig. 3d and the estimation of cortical point spread function (Engel et al. 1997), V3/VP is likely to be such a case for our stimulus, but V1 and V2 are not, as the spread in V1 and V2 was likely to be much smaller than the distance between the test-surround boundary and the center of the test region. Hence, although the responses to local feature contrast at the boundary are potential sources of response modulation around the test region, the contribution of these local responses to V1 and V2 was not dominant for our stimulus.

2.3. Effect of Relative Spatial Frequency and Phase

Finally, we manipulated the spatial frequency and phase of the surround grating and examined how these affected the response modulations for the iso- and cross-orientation. We used two conditions, an antiphase surround condition, in which the test and surround differed in the spatial phase, and a low spatial frequency (low-SF) surround condition, in which the test and surround differed by more than a octave in spatial frequency (0.66cpd for the surround, 2.0cpd for the test). According to psychophysical literature, these manipulations could affect subjects' perception. Our subjects reported that the apparent contrast of the test was slightly lower for iso-orientation than cross-orientation even when the test and surround differed in the spatial phase (antiphase surround condition), but not when they differed in spatial frequency (low-SF surround condition). The results showed that the antiphase surround produced suppressive response modulations in the test region for V2 and V3/VP (Figs. 5c, 6c) but was not statistically significant for V1. Suppressive response modulations were not observed in any areas in the low-SF surround condition (Figs. 5d, 6d). These tendencies were in accordance with the effects of these parameters on perceptual contrast.

2.4. Origin of Surround Suppression

The suppressive interaction observed in this study agrees with recent fMRI and Magnetoencephalography (MEG) studies that showed suppressive interaction among nearby stimuli in human early visual areas (Kastner et al. 1998, 2001; Ohtani et al. 2002; Zenger-Landolt and Heeger 2003). Kastner et al. (1998, 2001) have reported a suppressive effect that was analogous to “sensory suppression” within CRF at the level of a single neuron. Note that the surround suppression observed in our study was selective to the relative orientation of the test and surround, whereas sensory suppression is not; thus the observed surround suppression is not simply sensory suppression. The relative-orientation-selective surround suppression observed in our study is analogous to the response modulation beyond CRF that was found at the level of a single neuron in the cat and monkey primary visual cortex. For contrast parameters similar to ours, the iso-oriented surround presented outside CRF markedly suppressed neuronal response to a stimulus in CRF, while cross-oriented surround had a weak effect or none (Levitt and Lund 1997). Similar relative orientation-selective surround
suppression has also been found by optical imaging of cat area 17 (Toth et al. 1996), indicating that surround suppression beyond CRF could be observed in a large population of neurons. Our finding that surround suppression was weakened in the absence of the test also supported the contribution of the interaction beyond CRF. Hence, the surround suppression observed in our study reflects, in part, context-dependent suppression mediated by the interaction beyond CRFs.

3. Generating Meanings from Ambiguous Visual Inputs

Psychological projective tests, such as the Rorschach inkblot test, have a long history of use as mental assessments. This method requires respondents to construct meanings based on ambiguous patterns. Although there is controversy surrounding the projective tests, the Rorschach inkblot test is considered to be valid for identifying schizophrenia and other disturbances marked by disordered thoughts. A neuroimaging study on the generation of fluent speech during the Rorschach inkblot test focused on the brain activity in the superior temporal cortex, and showed that patients with formal thought disorders showed a reversed laterality of activation in the superior temporal cortex (Kircher et al. 2000). There are few data on the specific brain areas that are involved in semantic processing of visual inputs during the Rorschach test because the temporal cortex may be involved in the word generation process. Generation of meanings from ambiguous visual patterns may result from interactions between memory and perception. An electroencephalographic (EEG) investigation (Gill and O’Boyle 2003) showed that, when naming inkblot stimuli, bilateral activation of the parietal and occipital lobes was initiated and complemented by activation of the right frontal lobe, suggesting anterior regions of the right cerebral hemisphere contribute to the generation of meanings from visually ambiguous patterns. Identifying the brain areas that are involved in generation of meaning independent of word generation would further delineate meaning-related functions within the brain.

To assess meaning-related brain activity, we employed a naming task in which subjects were asked to name each of the visual stimuli covertly, as many as possible, while regional blood oxygen-action level-dependent (BOLD) contrast was measured using fMRI.

Figure 7a shows the visual materials employed. Three types of stimuli were used: five Rorschach inkblots (Cards I, IV, V, VI, and VII), five geometric shapes, and five face-like patterns. Rorschach inkblots and geometric figures were ambiguous, but their shapes provoked semantic associations with various categories, such as faces, particular animals, and so on. All were black- and white-figures (8.4° × 8.4°; 240 cd/m² for white). We used the event-related paradigm: each of the visual stimuli was presented with a duration of 15s followed by a 21-s uniform field with a fixation point; five stimuli of each stimulus type were presented in the order shown in Fig. 7a within one session, and two sessions were run for each stimulus type. In the naming task, subjects were asked to think “what this might be” and to name covertly, for each stimulus, as many items as possible.
No overt response, such as a verbal response or key press, was required, but subjects were required to assign as many names as possible. As a control experiment, we carried out fMRI measurements during passive viewing of the stimuli: subjects were instructed to concentrate on fixating on the central part of the stimulus and not to think about the visual stimuli.

All imaging used a General Electric 1.5-tesla scanner. The data were analyzed and visualized using our own in-house software (Ejima et al. 2003). The three-dimensional cortical surface was reconstructed for each. The reconstructed cortical surface was transformed into stereotaxic space (Talairach and Tournoux 1988). The fMRI data were processed using a temporal correlation analysis (Bandettinni et al. 1993). The linear trend and baseline offset for each voxel time series were first removed. A reference waveform was modeled with stimulus-related activation as a delayed “boxcar” function, taking into account the hemodynamic response lag convolved by a gamma function (phase delay = 3, pure delay = 2.5 s, time constant = 1.25 s) (Boynton et al. 1996). We employed an empirical procedure proposed by Baudewig et al. (2003), which introduced thresholds as $P$ values and defined thresholds with respect to the physiological noise distribution of individual statistical parametric maps. In the first step, a histogram of correlation coefficients, calculated using the reference functions, was used to estimate the underlying noise distribution of individual correlation images. In the second step, a Gaussian curve was fitted to the central portion of the histogram. In the third step, the distribution of correlation coefficients was rescaled into percentile ranks of the individual noise distribution. The percentile rank of 99.99% ($P = 0.0001$) served as the threshold for the identification of activation voxels. Voxels were deemed activated if they had a correlation coefficient greater than the threshold correlation coefficient defined by the $P$ value of 0.0001. We conducted an analysis in anatomically defined ROIs for the prefrontal, parietal, and occipital cortex. Ten ROIs were defined as regions within each gyrus, taking account of the anatomical structure and Brodmann areas (BAs) for each of the reconstructed cortical surfaces (Damasio 1990). Ten ROIs for the right hemisphere of one subject are shown in Fig. 8a. We sampled the fMRI data set from the full area of each ROI. Voxels within each region reaching a voxel threshold of $P < 0.0001$ were counted.

Behaviorally, in the naming task, the subjects reported that they assigned a much larger number of different names to the Rorschach inkblots and the geometric shapes than to the face-like stimuli, to which almost all subjects assigned one or two names. The regions of stimulus-related activation at the $P = 0.0001$ significance threshold were localized on the cortices of six subjects. Generation of meanings from ambiguous patterns resulted in extensive activation in regions including the prefrontal, parietal, and occipital cortex. Figure 7b shows brain activity for the Rorschach inkblots mapped on the reconstructed surface from the left and right hemispheres of one representative subject. Orange-yellow denotes the voxel locations activated during a naming task, and the colors represent $P$ values above 0.0001. Dark-blue-blue denotes the voxel locations activated during passive viewing, and the colors represent the $P$ values above 0.0001.
Fig. 7. a Visual stimuli employed in the experiment. b Activation maps for the naming task and the baseline (passive viewing). Areas selectively activated ($P < 0.0001$) are mapped on the reconstructed surface from the left and right hemispheres of one representative subject. Three views (lateral, posterior, and ventral) are presented. The visual stimuli were Rorschach inkblots. Orange-yellow colors denote the cortical regions showing significant activation during the naming task. The colors represent $P$ values above 0.0001. Dark-blue-blue colors denote cortical regions showing significant activation during passive viewing. The colors represent $P$ values above 0.0001. Red denotes the cortical regions showing significant activation common during the naming task and passive viewing.

a: Anatomical ROIs

b: Number of activated voxels
Red denotes the voxel locations activated commonly during the naming task and passive viewing. A similar activation pattern was observed for the geometric shapes. Multiple, bilateral regions in the prefrontal cortex were activated; these included bilateral prefrontal activation in the cortical areas lining the inferior frontal sulcus (IFS), middle frontal gyrus (MFG, BA46/8/9), and inferior frontal gyrus (IFG, BA44/45). The clustering in the lateral prefrontal activations was similar among subjects. There were foci of activity within the orbitofrontal cortex (OrbFC, BA11). The foci lay along the rostral and caudal parts of the lateral orbital sulcus. On the medial surface, we observed no significant consistent activation among subjects, and we did not observe significant activation in the anterior cingulated cortex.

Table 1 shows the means of the Talairach coordinates of the activated regions in the prefrontal cortex across six subjects. These regions showed consistent

<table>
<thead>
<tr>
<th>Table 1. Activation loci in the prefrontal cortex during naming task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left hemisphere coordinates</strong></td>
</tr>
<tr>
<td><strong>x</strong></td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Anterior MFG (BA46)</td>
</tr>
<tr>
<td>Mid-MFG (BA9)</td>
</tr>
<tr>
<td>Posterior MFG (BA8)</td>
</tr>
<tr>
<td>Anterior IFG (BA45)</td>
</tr>
<tr>
<td>Posterior IFG (BA44)</td>
</tr>
<tr>
<td>Orbitofrontal gyrus (BA11)</td>
</tr>
<tr>
<td>Anterior</td>
</tr>
<tr>
<td>Posterior</td>
</tr>
</tbody>
</table>

The numbers in parentheses denote ± standard error of mean; n = 6

MFG, middle frontal gyrus; IFG, inferior frontal gyrus

Red denotes the voxel locations activated commonly during the naming task and passive viewing. A similar activation pattern was observed for the geometric shapes. Multiple, bilateral regions in the prefrontal cortex were activated; these included bilateral prefrontal activation in the cortical areas lining the inferior frontal sulcus (IFS), middle frontal gyrus (MFG, BA46/8/9), and inferior frontal gyrus (IFG, BA44/45). The clustering in the lateral prefrontal activations was similar among subjects. There were foci of activity within the orbitofrontal cortex (OrbFC, BA11). The foci lay along the rostral and caudal parts of the lateral orbital sulcus. On the medial surface, we observed no significant consistent activation among subjects, and we did not observe significant activation in the anterior cingulated cortex.

Table 1 shows the means of the Talairach coordinates of the activated regions in the prefrontal cortex across six subjects. These regions showed consistent

**Fig. 8.** a Ten anatomical regions of interest (ROIs) defined on the reconstructed surface (right hemisphere) of one subject. MFG, middle frontal gyrus; IFG, inferior frontal gyrus; OrbG, orbitofrontal gyrus; SPL, superior parietal lobule; LOGs, superior part of lateral occipital gyr; LOGi, inferior part of lateral occipital gyr; MTG, middle temporal gyrus; ITG, inferior temporal gyrus; FG, fusiform gyrus; VI, area surrounding the calcarine sulcus including the lingual gyrus. b The number of suprathreshold voxels within each anatomical ROI, averaged across six subjects. The graph represents the number (± standard error of mean) of suprathreshold voxels (P < 0.0001) within ten ROIs. Black columns = left; white columns = right. Data are for the naming task: top for Rorschach inkblots, middle for geometric shapes, and bottom for face-like patterns.
activation across all subjects. During passive viewing, we did not observe activation in the regions of the prefrontal cortex. During the naming task, besides the prefrontal cortex, extensive activation was observed in the parietal (SPL), temporal (MTG, ITG), ventral (FG), and occipital (superior LOG, inferior LOG, V1) cortices. These regions were also activated during passive viewing, denoted by red and blue colors in Fig. 7b. Note that the activation in these regions during the naming task extended around the loci of activation observed during passive viewing, denoted by commonly activated voxels, and denoted by red in Fig. 7b. We also found that the amplitude of the fMRI signals in commonly activated regions was larger for the naming task than for the passive viewing.

Brain activation strongly depended on the ambiguity and/or stimulus characteristics of visual inputs. Figure 8b shows the number of suprathreshold voxels ($P < 0.0001$) within each of ten anatomical ROIs, averaged across six subjects for three stimulus types for the naming task. For face-like patterns, the number of activated voxels was markedly reduced for all the anatomical ROIs, comparing the data for the ambiguous patterns. There were some asymmetries between hemispheres in the activation, as measured by the number of suprathreshold voxels. All six subjects had more suprathreshold voxels in the inferior temporal gyrus (ITG) on the right than the left. There was significant right-sided asymmetry for the whole group ($t(6) = 3.15$, $P < 0.025$ for the geometric shapes; $t(6) = 2.65$, $P < 0.045$ for the Rorschach inkblots). There were more suprathreshold voxels within the right IFG, LOG, FG, and MTG than the left. On the other hand, there were more suprathreshold voxels in the left MFG and SPL than the right. The laterality effects, however, were not statistically significant ($P > 0.05$). The asymmetries in activation were not absolute: the naming task resulted in suprathreshold voxels bilaterally, but the relative ratio differed according to the cortical regions.

We adopted a covert procedure in which subjects generated names silently and found the activation pattern was very similar to that obtained by fMRI study of perceptual categorization (De Beek et al. 2000; Vogels et al. 2002) using a variant of the well-known dot pattern classification. The activated brain regions shown in the present study show the perceptual-conceptual memory system for visual inputs in the brain proposed in the literature. There are converging lines of evidence suggesting that conceptual and perceptual memory processes are subserved by different parts of the brain; regions in the inferior frontal gyrus, middle temporal gyrus, and inferior temporal gyrus are involved in conceptual memory processes, whereas the occipital cortex is involved in the perceptual memory process for visual stimuli and the superior temporal gyrus is involved in perceptual memory for auditory stimuli (Blaxton 1999; Buckner et al., 2000). We also found that the activation regions involved SPL and a part of the frontal eye field situated just behind the dorsolateral frontal region. These regions are known to be involved in processing spatial information in the visually guided saccade (Merriam et al. 2001). Eye tracking dysfunction is one of the established markers of risk for schizophrenia (Clementz and Sweeney 1990; Sweeney et al. 1998). The present study suggests that eye movement abnormalities observed for schizo-
phrenia during the Rorschach inkblot test (Fukuzako et al. 2002) are likely caused by the abnormalities in these brain regions.

**Acknowledgment.** We appreciate the assistance of C. Tanaka, K.M. Fukunaga, T. Ebisu, and M. Umeda with fMRI. This study was supported by the 21st Century COE Program (D-2, Kyoto University) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

**References**


2
Multiple Mechanisms of Top-Down Processing in Vision

GIORGIO GANIS$^{1,2,3}$ and STEPHEN M. KOSSLYN$^{3,4}$

1. Introduction

No animal could survive for long without perception. We must perceive the world, not only to find food, shelter, and mates, but also to avoid predators. Perception will fail if an animal does not register what is actually in the world. However, this simple observation does not imply that all processing during perception is “bottom up”—driven purely by the sensory input. Rather, bottom-up processing can be usefully supplemented by using stored information, engaging in processing that is “top down”—driven by stored knowledge, goals, or expectations. In this chapter we explore the nature of top-down processing and its intimate dance with bottom-up processing. We begin by considering basic facts about the primate visual system, and then consider a theory of its functional organization, followed by novel proposals regarding the nature of different sorts of top-down processing.

2. The Structure of Visual Processing in the Brain

An enormous amount has been learned about visual processing by studying animal models. In particular, the macaque monkey has very similar visual abilities to those of humans, and the anatomy of its visual system appears very similar to ours. Studies of the monkey brain have revealed key aspects of the organization of the visual system, namely, its hierarchical structure and the reciprocal nature of most connections between different visual areas of the brain. We briefly review key aspects of both characteristics of the brain next.

$^1$Department of Radiology, Harvard Medical School, Boston, MA 02115, USA
$^2$Massachusetts General Hospital, Martinos Center, Charlestown, MA 02129, USA
$^3$Department of Psychology, Harvard University, Cambridge, MA 02138, USA
$^4$Department of Neurology, Massachusetts General Hospital, Boston, MA 02142, USA
2.1. Hierarchical Organization

Over the past several decades, researchers have provided much evidence that the primate visual system is organized hierarchically. In the early 1960s and 1970s, Hubel and Wiesel’s electrophysiological findings, first in cats and then in nonhuman primates, strongly suggested a hierarchical relationship among early areas in the visual system; this inference was based on the increasing size and complexity of the receptive fields as one goes from striate cortex to areas farther along in the processing stream (Hubel and Wiesel 1962, 1965, 1968, 1974). The earliest areas of the visual system are organized topographically; space on the cortex represents space in the world, much as space on the retina represents space in the world (Felleman and Van Essen 1991; Fox et al. 1986; Heeger 1999; Sereno et al. 1995; Tootell et al. 1998; Van Essen et al. 2001). The higher-level areas are not organized topographically, but often represent information using population codes (Fujita et al. 1992; Miyashita and Chang 1988; Tanaka et al. 1991). In such codes, different neurons respond to complex visual properties, and shape is coded by the specific combination of neurons that are activated.

The work by Felleman and Van Essen (1991) charted the hierarchical organization of the entire visual system. They compiled a matrix of known anatomical connections among areas and showed that the pattern of connectivity could best be accounted for by a hierarchical structure with multiple parallel streams. The striate cortex (V1) was at the bottom of the entire hierarchy, and the inferotemporal (area TE) and parahippocampal (areas TH and TF) cortex were at the top of the ventral stream (which is specialized for object vision, registering properties such as shape and color) (Desimone and Ungerleider 1989).

The big picture of cortical organization provided by Felleman and Van Essen has been generally confirmed by computational analyses of the same dataset (Hilgetag et al. 1996), as well as by additional empirical approaches, such as those based on measuring the proportion of projecting supragranular layer neurons labeled by a suitable retrograde tracer (Vezoli et al. 2004). This neuroanatomical picture of a hierarchically organized visual system has also been confirmed by data from single unit recording studies of higher-level visual areas. For instance, area TE in the inferotemporal cortex has been shown to contain neurons with extremely large receptive fields (often encompassing the entire visual field), which are tuned to complex combinations of visual features (such as combination of shape fragments and textures); in contrast, neurons in lower-level areas, such as V4 (fourth visual area), have smaller receptive fields and are tuned to simpler feature combinations (Tanaka 1996).

2.2. Connections Among Areas

A considerable amount is now known about the connections among visual brain areas, and the evidence suggests that different connections are used in bottom-up and top-down processing.
2.2.1. Feed-Forward Connections and Bottom-Up Processing

Contiguous neurons in area V1 have contiguous receptive fields (i.e., regions of space in which they will respond to stimuli). A set of contiguous neurons in area V1 in turn projects to a single neuron in area V2 (secondary visual area), and this neuron has a larger receptive field than any of those neurons that feed into it. This “many-to-one” mapping continues up the hierarchy until the receptive fields become so large that the areas are no longer topographically organized.

The neuroanatomical findings and the properties of the receptive fields have given rise to numerous models that emphasize the feed forward nature of the ventral stream (Fukushima 1988; Riesenhuber and Poggio 1999; VanRullen et al. 2001; Wallis and Rolls 1997; Serre et al. 2007). Electrophysiological findings that document the fast onset of neural responses to visual stimuli at all levels in the ventral stream (i.e., the mean latency of neurons in area TE at the highest level of the hierarchy is just over 100 ms after stimulus onset) provided additional impetus for these models (Lamme and Roelfsema 2000). In these models, objects are identified during a feed forward pass throughout the ventral stream hierarchy, with increasingly complex information’s being extracted at higher levels in the system. For instance, in Riesenhuber and Poggio’s model of the ventral stream, the units farther upstream (from area V1 to TE) are tuned to increasingly complex features, all the way to units that are tuned to specific views of objects.

2.2.2. Feedback Connections and Top-Down Processing

Crucially for the topic at hand, and consistent with the connectivity pattern reported in earlier work by Rockland and Pandya (1979) and others, Felleman and Van Essen described not only feed-forward connections among areas in the primate visual system but also widespread feedback connections. They found a striking regularity in the pattern of laminar origin of feed-forward and feedback connections: although feed-forward connections originate in the supragranular layers (often layer III) and terminate in layer IV in the target area, feedback connections originate from neurons in layer VI and IIIA of the projecting area and end in layer I of the target area. Indeed, numerous other anatomical studies in nonhuman primates have confirmed that there are massive feedback connections at many levels in the visual system, including pathways from areas that are not traditionally considered visual areas (Barone et al. 2000; Budd 1998; Clavagnier et al. 2004; Rockland and Pandya 1979; Salin and Bullier 1995). For example, area V1 has been shown to receive direct feedback connections from many extrastriate regions including V2, V3, V4, TEO (temporo–occipital), TE, as well as from nonvisual areas, including the frontal eye fields, area 36, areas TH/TF, STP (superior temporal polysensory), and even the auditory cortex.

The feedback connections are not simply the inverse of feed-forward connections. The feed-forward connections display a lovely many-to-one mapping as they ascend the hierarchy, but there is nothing of the sort for the feedback connections. Instead, the feedback connections do not appear to be precisely
targeted, but rather often appear to meander (Budd 1998). Evidently, the feedback connections are not simply “replaying” information sent downstream.

2.2.3. Neuroanatomically Inspired Models of Top-Down Processing

A class of models of object vision has incorporated the finding of feedback connections in the visual system (Grossberg and Mingolla 1985; Li 1998; Mumford 1992; Ullman 1989, 1995). Generally, these models assume that feedback connections provide a mechanism by which top-down processing can occur, allowing relatively abstract information stored in higher-level visual areas to influence and constrain processing in lower-level visual areas. To illustrate the basic idea of why top-down processing is needed, researchers have created binarized photographs. In such photographs, gray-scale pixels are replaced with white if their brightness value is above a chosen threshold, or replaced with black if it is below this value. Because binarized images are highly degraded, pure bottom-up processes typically cannot organize them correctly into their constituent parts, and often one needs to use previously acquired knowledge about objects to identify the objects in them (Fig. 1).

The models just mentioned rest on algorithms that allow an interplay between stored information and online input. For instance, Mumford posits that higher-level visual areas try to find the best fit with the information they receive from lower-level visual areas by using the more abstract knowledge they store (e.g., a representation of a shape). The feedback connections allow higher-level visual areas to reconstruct the visual input in lower-level visual areas, based on such a best fit. The mismatch between the reconstructed visual input and the original input in lower-level visual areas (i.e., information not explained by the current

Fig. 1. This binarized picture illustrates the problems encountered by purely bottom-up approaches to vision. It is very difficult to parse correctly the fox at the center of the picture using purely bottom-up processing. Using top-down processing to exploit constraints imposed by knowledge of the shape of foxes makes the task much easier
fit in higher-level areas) is then sent forward, which can trigger another top-down processing cycle.

This class of models of top-down processing in the ventral stream has typically ignored the role of areas outside the ventral stream or has only postulated unspecified extravisual inputs. However, many nonvisual areas in the frontal and parietal lobe are connected to areas in the ventral stream (Petrides 2005).

Another class of models, in contrast, focuses on top-down influences that these nonvisual areas exert on areas in the ventral stream. For instance, the model of prefrontal function by Miller and Cohen (2001) has focused on the role of the prefrontal cortex in biasing processing in areas in the ventral stream.

Traditionally, the different classes of models have been pursued independently, although some of the terminology has overlapped. Unfortunately, the term “top-down processing” in vision has been used loosely in the neuroscientific literature to refer to a disparate range of phenomena. For instance, it has been used in the context of the neural effects of visual attention (Hopfinger et al. 2000), memory retrieval (Tomita et al. 1999), and phenomena such as illusory contours (Halgren et al. 2003).

In the remainder of the present chapter, we develop explicit distinctions between different types of visual top-down processes; these distinctions are cast within the context of a broad theory of the visual system that incorporates both bottom-up and top-down processes (Kosslyn 1994) as well as the role of nonvisual areas. Our aim is to make explicit some of the assumptions regarding top-down processes that are implicit in the literature and propose a first-order taxonomy, rather than to provide an exhaustive review of the top-down processing literature. In the following section we briefly summarize our theory of visual processing during visual object identification, relying on the background already provided, and then we proceed to describe how different types of top-down processing may operate within this system.

3. A Theory of the Functional Organization of Late Visual Processing in the Primate Brain

We propose that there are two general kinds of visual processes, “early” and “late.” Early visual processes rely entirely on information coming from the eyes whereas late processes rely on information stored in memory to direct processing. We must distinguish between early and late processes and the specific brain areas involved in vision: late processes can occur even in areas that are involved in the first stages of bottom-up processing (Lamme and Roelfsema 2000). Low-level visual areas are involved in both early (bottom-up) and late (top-down processing).

Vision, and more specifically object identification, is not a unitary and undifferentiated process. Indeed, similarly to memory operations such as encoding and recall, which are carried out by many subprocesses (Schacter 1996; Squire 1987), object identification is carried out by numerous subprocesses (for example,
those involved in figure-ground segregation, in shifting attention, in matching input to stored information). Our theory posits a specific set of component visual processes, with an emphasis on those involved in late visual processing; we call these components processing subsystems. A processing subsystem receives input, transforms it in a specific way, and produces a specific type of output; this output in turn serves as input to other subsystems.

Figure 2 illustrates the most recent version of our theory of processing subsystems. Although this diagram appears to imply sequentiality, the theory does not in fact assume that each processing subsystem finishes before sending output to the next. Rather, the theory posits that all processes are running simultaneously and asynchronously, and that partial results are continually being propagated through the system. Moreover, we assume that what shifts over time is how intensively a given process is engaged. Thus, the theory posits processing subsystems that operate in cascade, often operate on partial input, and send new outputs to other subsystems before they have completed processing (Kosslyn 1994).

Fig. 2. A theory of the functional architecture of the visual system (subcortical structures are not shown for simplicity). Note that each box represents a structure or process (e.g., the visual buffer) that is implemented in multiple areas (and often can itself be decomposed into more specialized processes, not discussed here). The associative memory subsystem is divided into long term (LT) and short term (ST). Some subsystems (associative memory and attention shifting) are implemented by spatially distant brain regions. These subsystems are connected by *arrowless lines*; for simplicity, inputs and outputs from these subsystems are only indicated for one of them. Connections that implement reflexive top-down processing are indicated by *red arrows* whereas those that implement strategic top-down processing are indicated in *green*. Note that the reflexive connections can also be engaged by strategic top-down processing (via inputs from the information shunting subsystem). Visual input from the lateral geniculate enters the visual buffer via the *black arrow at the bottom*.
3.1. Processing Subsystems Used in Visual Object Identification

The process of object identification can be conceptualized as the search for a satisfactory match between the input and stored memories. The specific set of subsystems involved, as well as their time-course of engagement, depends in part on the properties of the incoming visual stimulus. In all cases, however, the same architecture is used: the same types of representations and the same processes are used, but sometimes in different orders or more or less intensively. The first step in describing our theory is to summarize the processing subsystems at the most general level, as follows.

3.1.1 Visual Buffer

In primates, vision is carried out by a multitude of cortical areas, at least 32 in monkeys (Felleman and Van Essen 1991) and probably even more in humans (Sereno and Tootell 2005). Many of these areas, including V1 and V2, are organized topographically. That is, these areas use space on the cortex to represent space in the world. The specific pattern of activation in these areas reflects the geometry of the planar projection of a stimulus; in addition, focal damage to these areas causes scotomas, that is, blind spots at the spatial location represented by the damaged cortex. In our theory, the subset of topographically organized areas in the occipital lobe implements what we refer to as the visual buffer.

However, even if we think of the set of these topographically organized areas as a single functional entity, these areas are hierarchically organized themselves, and have somewhat different functional properties. We stress that although these areas are at the bottom of the visual hierarchy, and thus carry out bottom-up processes necessary for vision, they are also affected by top-down processes originating both within other portions of the visual buffer itself and from areas outside this structure.

3.1.2. Attention Window

Not all the information in the visual buffer can be fully processed. A subset of the information in the visual buffer is selected by an attention window, based on location, feature, or object of interest, for further processing (Brefczynski and DeYoe 1999; Cave and Kosslyn 1989; Posner and Petersen 1990; Treisman and Gelade 1980). There is good evidence that this attention window can be covertly shifted (Beauchamp et al. 2001; Corbetta and Shulman 1998; Posner et al. 1980) and it can also be split, at least in some specific circumstances, to include non-adjacent regions in the visual space (McMains and Somers 2004).

3.1.3. Object-Properties-Processing

As discussed earlier, many connections run from the topographically organized areas of the occipital lobe to other areas of the brain, giving rise to parallel
processing streams. The ventral stream runs from the occipital lobe to the inferior temporal lobe (Desimone and Ungerleider 1989; Haxby et al. 1991; Kosslyn 1994; Mishkin et al. 1983; Ungerleider and Mishkin 1982), where visual memories are stored (e.g., Fujita et al. 1992; Tanaka et al. 1991). The visual memories are stored, at least in the monkey brain, using a population code by which nearby cortical columns do not store information about nearby points in two-dimensional space but information about nearby points in feature space (Fujita et al. 1992; Miyashita and Chang 1988; Tanaka et al. 1991). These areas, implementing what we will refer to as the object-properties-processing subsystem, not only store information about shape and shape-related properties of objects and scenes, such as color and texture, but also match input to such stored information. Visual recognition of an object occurs when the content of the attention window matches a stored visual representation in this system.

### 3.1.4. Spatial-Properties-Processing

In addition to being able to determine the identity of objects, we can also determine their location in space. Our visual system accomplishes this by allocating different resources to extract and process information about properties that are inherent to objects (such as their shape or color) versus information about their spatial properties (such as their size and location). The object-properties-processing subsystem essentially ignores spatial information and produces position-invariant object recognition (Gross and Mishkin 1977; Rueckl et al. 1989). In contrast, spatial processing captures the very information discarded during object processing. Spatial processing is accomplished by the dorsal stream, a pathway that runs from the occipital lobe to the posterior parietal lobe (Andersen et al. 1985; Haxby et al. 1991; Kosslyn et al. 1998; Ungerleider and Mishkin 1982). In our theory, these posterior parietal regions embody what we refer to as the spatial-properties-processing subsystem.

According to our theory, the spatial-properties-processing subsystem not only registers spatial properties, but also constructs object maps, which indicate the locations of objects or parts of objects in space (cf. Mesulam 1990). The cortex that implements at least part of the spatial-properties-processing subsystem is topographically organized, and thus at least some of the representations used in this subsystem depict the locations of objects in space (Sereno et al. 2001).

### 3.1.5. Associative Memories

Outputs from both the object-properties-processing and spatial-properties-processing subsystems converge on associative memories. Associative memories specify links among representations. Our theory posits two classes of associative memories. On the one hand, short-term associative memory structures maintain information on line about which objects are in specific locations (Rao et al. 1997; Wilson et al. 1993). These memory structures are implemented in the dorsolateral prefrontal regions. On the other hand, long-term associative memory structures store more enduring associations among stored categories,
characteristics, situations, and events. If the outputs from the object-properties-processing and spatial-properties-processing subsystems match a stored representation in long-term associative memory, the information associated with it is accessed, leading to object identification. For instance, if the shape matches that of a cat, one can access information that it is a mammal, likes to drink milk, and sometimes sleeps much of the day. If no good match is found, the best-matching representation is used as a hypothesis of what the viewed object might be (we discuss this process in detail shortly). Long-term associative memory is implemented in Wernicke’s area, the angular gyrus, classic “association cortex” (e.g., area 19; Kosslyn et al. 1995), and parts of the anterior temporal lobes (Chan et al. 2001).

3.1.6. Information Shunting

In our theory, when the match of the input to representations in long-term associative memory is poor, then representations of distinctive visual parts and attributes of the best-matching object are retrieved and used by an information shunting subsystem to guide top-down search (Gregory 1970; Neisser 1967, 1976). Thus, by means of this process, the visual system actively seeks information to test hypotheses about the visual input. The information shunting subsystem operates in two related ways: First, it sends information to other subsystems, enabling them to shift the focus of attention to the likely location of distinctive parts or attributes. Second, simultaneously, the information shunting subsystem primes representations of these parts and attributes in the object-properties-processing subsystem (Kosslyn 1994; McAuliffe and Knowlton 2000; McDermott and Roediger 1994), which facilitates the ease of encoding these representations. The information shunting subsystem is implemented by one or more parts of the lateral prefrontal cortex (see Damasio 1985; Koechlin et al. 1999; Luria 1980; Petrides 2005; Posner and Petersen 1990). However, the lateral prefrontal cortex is a very large region, and so it is unlikely that implementing the information shunting subsystem is its only function; moreover, different regions of lateral prefrontal cortex in principle may implement specialized components of the information shunting subsystem, with each operating only on specific types of information (e.g., location versus shape).

3.1.7. Attention Shifting

Shifting the focus of attention to a new location or to a new attribute involves a complex subsystem, which is implemented in many parts of the brain (including the superior parietal lobes, frontal eye fields, superior colliculus, thalamus, and anterior cingulate (see Corbetta 1993; Corbetta and Shulman 1998; LaBerge and Buchsbaum 1990; Mesulam 1981; Posner and Petersen 1990). The attention shifting subsystem can shift the location of the attention window both covertly and overtly, such as occurs when we move our eyes, head, or body to look for new information.
3.2. Operation of the Subsystems Working Together

If an object is seen under optimal viewing conditions and is familiar, recognition (i.e., the match to visual representations in the object-properties-encoding subsystem) and identification (i.e., the match to representations in long-term associative memory) proceed very quickly and may be carried out entirely via bottom-up processes. However, if an object cannot be recognized and identified relatively quickly via bottom-up processing because it is unfamiliar or is seen under impoverished viewing conditions, then there is time for top-down processing to unfold. In this case, information about attributes of the best-matching object is accessed from long-term associative memory and then is used to direct attention to the location where a distinctive part or attribute should be found (and to prime the object’s representations in the object-properties-processing subsystem), and a new part or attribute is encoded into the visual buffer, beginning a new processing cycle. If the new part or attribute matches the primed representation in the object-properties-processing subsystem, this part or attribute is recognized, and this may lead to a good match in long-term associative memory—and the object will have been identified. If not, either additional parts or attributes of that object are sought or another hypothesis is generated (e.g., by taking the next best match in long-term associative memory) to guide a new search for a distinctive part or attribute.

4. Varieties of Top-Down Processing

In describing our theory, at different points we have invoked distinct types of top-down processing. In most theories, these different types are conflated and simply referred to as different instances of the same kind of activity. However, we propose here that these processes are in fact distinct, and that each operates only in specific circumstances. In the following, we outline a first-order taxonomy and discuss some of the empirical evidence that supports it. Our theory makes an initial distinction between two broad classes of top-down mechanisms.

4.1. Strategic Top-Down Processing

Strategic top-down processing relies on “executive control mechanisms” (which provide input to the information-shunting subsystem) to direct a sequence of operations in other brain regions, such as is used to engage voluntary attention or to retrieve stored information voluntarily. In the case of visual object recognition, strategic top-down processing is recruited when the initial encoding is not sufficient to recognize an object. In such circumstances, the best-matching information is treated as a hypothesis, which then is used both to shift the focus of attention and to prime representations in the object-properties-processing subsystem to facilitate the encoding of a sought part or characteristic.

For example, if a picture of a degraded object (degraded perhaps because it is partially hidden by another object, or is in poor lighting) is presented in a familiar
visual context, then the best-matching representation in long-term associative memory is treated as a “hypothesis,” which is used by the information shunting subsystem to direct attention to possible parts or characteristics that would identify the degraded object (Ganis et al. 2007). To illustrate, if a chair is presented in the context of a kitchen, the associations stored in long-term associative memory can be used to identify the chair even if only small parts of it are visible above and behind the table.

We also note that strategic top-down processing can also take place in the absence of any visual input, such as in some cases of visual imagery. In these cases, the information-shunting subsystem directs a sequence of operations entirely driven by endogenously generated information, which leads to the retrieval of stored information in the absence of an external stimulus (Kosslyn et al. 2006).

Two classes of evidence for strategic top-down processing have been reported, from nonhuman primates and from humans. We summarize important examples of each class of research in what follows.

4.1.1. Strategic Top-Down Processing in Nonhuman Primates

Although it is very difficult to obtain direct neural evidence for top-down processing because, at minimum, it entails recording neural activity from multiple sites in awake animals, some of this sort of direct evidence is available in nonhuman primates.

One of the most compelling studies that demonstrates strategic top-down processing in action was conducted by Tomita and collaborators on the retrieval of visual paired associates in monkeys (Tomita et al. 1999). In this study, monkeys first learned a set of paired-associates, and then the posterior parts of the corpus callosum were severed; after this surgery, the only remaining communication path between hemispheres was via the anterior parts of the corpus callosum, which connect the prefrontal cortex in the two hemispheres. Thus, following surgery, a visual stimulus presented to the left hemisphere (right visual hemifield) could only affect neural activity in the right inferotemporal cortex by means of an indirect route, which drew on the left prefrontal cortex, the anterior corpus callosum, and the right prefrontal cortex (the corresponding sets of structures hold for right hemifield presentation). As expected, the presentation of a cue to one hemifield resulted in robust bottom-up activation of stimulus-selective inferotemporal neurons in the contralateral hemisphere (with an average latency of 73 ms). Crucially, presentation of the same cue to the other hemifield (ipsilateral) resulted in delayed activity in neurons that were selective for the probe or the paired associate (with an average latency of 178 ms).

After the main experiment, fully severing the callosum abolished responses to the ipsilateral stimuli but left the responses to the contralateral stimuli intact, which showed that the previous results were not due to subcortical influences. These results provide good evidence that the prefrontal cortex sends top-down signals to inferotemporal neurons during the retrieval of visual information.
Fuster and his collaborators reported another study that documented strategic top-down processing in nonhuman primates; this study relied on a reversible cooling technique that temporarily inactivates an area (Fuster et al. 1985). In this study, spiking activity was recorded from single neurons in monkey inferotemporal cortex during a delay match-to-sample task with colors, while the prefrontal cortex was inactivated bilaterally via cooling probes. Before inactivation, during the delay period, neurons in inferotemporal cortex showed a sustained response with a clear preference for the color to be remembered. However, inactivation of prefrontal cortex by cooling impaired this selectivity profile. The critical finding was that inactivation of prefrontal cortex impaired the monkey’s performance in this task. These data indicate that top-down signals from prefrontal cortex are necessary for the maintenance of delayed, stimulus-specific activity in the inferotemporal cortex when no external stimuli are present.

Another study that documented strategic top-down processing in nonhuman primates was reported by Moore and Armstrong (Moore and Armstrong 2003). This study was designed to investigate the effects of electrical stimulation of the frontal eye fields on neural activity in area V4. When electrical stimulation of the frontal eye fields is strong enough, it produces systematic saccades to specific locations; the specific target location of the saccade depends on which specific part of the frontal eye fields is stimulated. The researchers recorded from neurons in V4 that had receptive fields at the location where specific stimulation of the frontal eye field directed a saccade. The activity of the V4 neurons to preferred and nonpreferred visual stimuli was also recorded without stimulation or with subthreshold stimulation of the frontal eye fields (subthreshold stimulation is not sufficient to elicit a saccade). The results showed that the responses of V4 neurons were enhanced when a preferred visual stimulus was within the neuron’s receptive field and the frontal eye fields were stimulated subthreshold (i.e., without generating a saccade) compared to when there was no stimulation. Crucially, this effect was not present without a visual stimulus or with a nonpreferred visual stimulus in the neuron’s receptive field. Furthermore, the effect was not present if the receptive field of the neuron did not cover the end location of the saccade that would be elicited by suprathreshold stimulation of the frontal eye fields. Thus, electrical stimulation of the frontal eye fields simulated the operation of covert attentional shifts on neural activity in V4 (note that the eye movements were monitored carefully, and trials during which the monkey was not fixating were excluded from the analyses). This finding provides direct evidence that strategic top-down signals can originate in the frontal eye fields and then can affect activation in at least one area that we include in the visual buffer.

Another study, reported by Buffalo and collaborators (Buffalo et al. 2005), charted the time-course over which areas in the ventral stream are engaged during strategic top-down processing. In this study, the researchers recorded single-unit and multunit activity in areas V4, V2, and V1, during a visual attention task on drifting greetings presented peripherally. Firing rates were compared for conditions when the animal paid attention to a stimulus inside versus outside
the receptive field of the neuron (while the monkey maintained fixation). As found in other studies, paying attention to a stimulus inside the receptive field of neurons in areas V1, V2, and V4 increased firing rates. The onset of the attentional effects revealed a striking pattern: area V4 showed the earliest onset (240 ms poststimulus onset), area V2 was next (370 ms), and area V1 was last (490 ms). These results strongly suggest that strategic top-down signals trickle down from higher-level visual areas that receive these signals from prefrontal or parietal areas.

### 4.1.2. Strategic Top-Down Processing in Humans

In humans, the evidence for strategic top-down processes is largely indirect, because of limitations of the noninvasive neuroimaging techniques employed. Nonetheless, such evidence is consistent with that from nonhuman primates. We discuss briefly three functional magnetic resonance imaging (fMRI) studies that document how strategic top-down processes affect neural activation in the ventral stream in humans.

Kastner and colleagues (Kastner et al. 1999) investigated the mechanisms by which visual attention affects activation in occipital cortex in the presence of multiple stimuli. The rationale for this study was grounded in a finding from single-cell studies of monkeys; this finding showed that responses of neurons in areas V2 and V4, to an otherwise ineffective stimulus, decrease when a second stimulus is presented in the neuron’s receptive field. However, this decrease in response is eliminated if the animal pays attention to the first stimulus, ignoring the other stimuli in the receptive field (Reynolds et al. 1999). To test the hypothesis that these same effects can occur in the human visual system, Kastner and collaborators presented four images in the upper right visual quadrant, either simultaneously (SIM) or in sequence (SEQ), in independent trials. In the “attend” condition (ATT), the participants were asked to maintain fixation and to pay attention to the images presented peripherally at a given spatial location and ignore the others. In the control condition, participants were asked simply to maintain fixation and ignore all visual stimuli in the periphery (UNATT). During ATT trials, 11 s before the onset of the visual stimuli there was a small cue at fixation, telling participants to direct attention covertly to the appropriate location in the visual field, waiting for the visual stimuli to be presented. Using this methodology, the researchers could monitor brain activity during attention in the absence of visual stimulation.

The results revealed increased activation during the expectation period (before the onset of the visual stimuli) in several visual areas, including V1, V2, V4, and TEO. The increased baseline activation was retinotopically specific, and it was strongest in visual areas TEO and V4. This increase in baseline firing rates in the absence of visual stimuli is very similar to that found in nonhuman primates during attentional tasks (Luck et al. 1997). In addition, Kastner and colleagues found greater activation in extrastriate visual areas in the response to visual
stimuli during the ATT condition relative to the UNATT condition. Furthermore, as expected, the increase was larger in the SIM than in the SEQ condition, because the inhibitory effects of multiple stimuli were not as strong when the stimuli were presented sequentially. Finally, the researchers did not observe an effect of type of presentation in area V1, probably because the receptive fields in area V1 are too small to encompass more than one of the stimuli used in this experiment.

Kastner and collaborators also examined activation in areas outside the ventral stream to gather evidence about the sources of the modulation of visual areas. They found that areas in the frontal lobe (specifically, the frontal eye fields and supplementary eye fields) and in the parietal lobe (the inferior parietal lobule and superior parietal lobule) showed robust increases in activation during the expectation period. The frontal eye fields, supplementary eye fields, and the superior parietal lobe areas did not display increased activation following presentation of the visual stimuli, which suggests that they were not driven by bottom-up information but rather were involved in strategic top-down processing. This result is consistent with the fact that the frontal eye fields and the supplementary eye fields have rich efferent (i.e., feedback) connections to areas in the ventral stream and posterior parietal cortex (Felleman and Van Essen 1991). Furthermore, this result is consistent with findings that these areas are engaged during covert shifts in attention (Beauchamp et al. 2001; Corbetta 1993; Corbetta and Shulman 1998), as described earlier.

Another fMRI study, performed by Ress et al., confirmed and complemented the Kastner et al. findings. In this study, participants were asked to respond when an annulus (inner radius = 3°/outer radius = 6°) appeared in the center of the screen. In the main condition, the contrast of the stimuli was adjusted for each participant so that correct detection occurred only on 75% of the trials. Target trials were intermixed with catch trials during which no stimulus was presented. A slow event-related paradigm was used to allow the analysis of activation during individual trials. Results showed that the regions in areas V1, V2, and V3 representing the annulus exhibited a robust BOLD response both during annulus-present and annulus-absent trials (note that the annulus-absent trials resemble the UNATT condition in the experiment by Kastner et al.). In fact, when the stimulus was presented at the lowest contrast levels (near-threshold), activation during the annulus-present trials was only slightly stronger than during the annulus-absent trials. Crucially, BOLD activity predicted the participant’s performance on the detection task: the greater the activity, the more likely the participant would correctly detect the presence or absence of the annulus. Furthermore, the annulus-absent BOLD response was much smaller during an independent condition that used blocks of trials with higher contrast stimuli (which were easier to detect), which suggests that this response reflected neural activation required to perform the more difficult detection task. Strategic top-down processes might increase sensitivity in low-level visual areas to incoming visual stimuli by pushing the neurons into a higher gain region of their operating range, where smaller differences in the input produce relatively larger differences in
response (Ress et al. 2000). Because higher firing rates are costly metabolically, it makes sense to have a mechanism that can increase sensitivity by increasing baseline firing rates only during difficult detection situations.

Although the study by Kastner et al. (1999) found parietal and frontal activation during the expectation period, the results do not provide direct evidence that these regions are actually a source of strategic top-down influences. Furthermore, Ress and collaborators did not sample the entire brain and therefore were unable to provide direct evidence of the source of the modulation of activation they observed in striate and extrastriate cortex. A more recent fMRI study used dynamic causal modeling to investigate the direction of influences among brain areas during visual mental imagery and visual perception (Mechelli et al. 2004). The key to this study is the comparison between mental imagery and perception. According to our theory, during visual imagery the information shunting subsystem retrieves stored representations of the structure of an object in long-term associative memory and sends information to the object- and spatial-properties-processing subsystems to activate the corresponding modality-specific representations. According to the theory, this activation process is identical to the priming that occurs during top-down hypothesis testing in perception; however, now the priming is so strong that activation propagates backward, and an image representation is formed in the visual buffer. The visualized shapes and spatial relations are retained (which is equivalent to holding them in “working memory”), and they can be inspected and identified in an image by the same attentional mechanisms used to inspect objects and locations during perception (Kosslyn et al. 2006).

Mechelli et al. compared blocks of trials in which participants formed visual mental images of faces, houses, or chairs with blocks of trials in which participants actually viewed the same objects. The researchers first measured intrinsic connectivity during visual perception and visual imagery, that is, the influence brain regions have on each other as a result of being in visual perception or visual imagery modes (regardless of the visual category); such intrinsic connectivity was then used as a baseline to quantify functional connectivity changes brought about by the experimental manipulation (visual category) within those modes. The intrinsic connectivity during visual perception revealed paths from occipital cortex and superior parietal cortex to ventral temporal cortex, whereas during visual mental imagery it revealed paths from superior parietal cortex and the precuneus to ventral temporal cortex. Analyses on the category specificity of the changes in functional connectivity showed that during visual perception there was an increase in functional connectivity from low-level visual cortices to the regions of ventral temporal cortex selective for the corresponding stimuli. For instance, the functional connectivity between inferior occipital cortex and the ventrototemporal region that responded the most to faces increased the most during the presentation of blocks of trials containing faces (compared to blocks of trials containing houses or chairs). In contrast, during visual mental imagery, the researchers found a selective increase in functional connectivity from prefrontal cortex and parietal cortex to these regions in ventral temporal cortex. However, only the strength of the path from prefrontal cortex was modulated by
stimulus category (i.e., faces, houses, or chairs). For instance, the functional connectivity between prefrontal cortex and the ventrotemporal region that responded the most to faces increased the most during the presentation of blocks of trials containing faces (compared to blocks of trials containing houses or chairs). Thus, the analysis of functional connectivity changes during visual imagery (compared to visual perception) suggests the existence of two types of strategic top-down influences on the ventral stream. The first one is an influence from the parietal cortex that is not modulated by stimulus category (i.e., it is the same regardless of the category of the visualized stimulus) and may reflect the operation of selective attention, whereas the second one is a category-specific signal that may be involved in the reconstruction of visual information in category-specific areas in the ventral stream.

4.2. Reflexive Top-Down Processing

We have so far been focusing on strategic top-down processing, which is under voluntary control. We also propose a second major class of top-down processing, which is automatic. Such reflexive top-down processing occurs between areas that are bidirectionally connected in the visual buffer, the object-properties-processing subsystem, and in long-term associative memory. Crucially, reflexive top-down processing is triggered by bottom-up signals without the intervention of the information shunting subsystem in the prefrontal cortex.

4.2.1. Reflexive Top-Down Processing in Nonhuman Primates and in Humans

The best evidence for reflexive top-down processing comes from studies in nonhuman primates that investigated how stimulus-driven neural activity in area V1 is affected by activity in higher-level visual areas. Similar processes probably take place in higher levels of the visual hierarchy and in long-term associative memory as well.

Some of the more compelling experiments use reversible inactivation techniques in anesthetized animals; these techniques rely on cooling or application of the inhibitor GABA. One critical finding is that inactivation of area V2 changes the response properties of neurons in area V1, generally making them less selective (Payne et al. 1996; Sandell and Schiller 1982). However, it is difficult to know how the anesthesia affected these results. Thus, findings obtained in awake monkeys, such as from the study reported by Lee and Nguyen (Lee and Nguyen 2001), are probably more compelling. In Lee and Nguyen’s study, activity was recorded from neurons in area V1 and V2 while monkeys viewed illusory contours (such the ones produced by the four black disks in Fig. 3) as well as corresponding real contours. The presentation paradigm was slightly different from that of previous studies that had failed to observe responses to illusory contours in V1 neurons (von der Heydt et al. 1984). In this new paradigm, four black disks centered at the corners of an imaginary square were first presented
for 400 ms. Then they were suddenly replaced with four corner disks in the same position, giving the impression that a white square had appeared in front of them, generating partial occlusion. There were also numerous control conditions, including squares defined by real contours.

The results of Lee and Nguyen’s study showed that, as expected, neurons in V1 responded vigorously to real contours that had the appropriate orientation, with a response onset of about 50 ms relative to stimulus onset. These same neurons, at least those in the superficial layers of V1, also responded to illusory contours with the same preferred orientation. However, the neurons responded more weakly to illusory contours, and, crucially, they began to respond about 55 ms later than they did to the real contours. The neural responses in area V2 to illusory contours were generally stronger than those in area V1 and began around 65 ms poststimulus, which was about 40 ms before the V1 response. One plausible interpretation of this finding is that V2 neurons, with their larger receptive field size compared to area V1 neurons, integrate more global information and can aid the contour completion process in V1 by sending feedback. The advantage of sending information back to area V1 is that V1 maintains a high-resolution version of the visual input (because of its small receptive field sizes), whereas higher-level visual areas have access to more abstract and global information required to parse the visual input into meaningful parts (e.g., surfaces) (Lee and Mumford 2003; Lee and Nguyen 2001). In addition to feedback processes, these findings may reflect properties of recurrent circuits within V1 itself, which may take as long to carry out some contour completion operations as feedback from area V2 (Girard et al. 2001).

Although Lee and Nguyen (2001) did not record from neurons outside of areas V1 and V2, the completion processes documented by this study are probably also affected by feedback from higher-level visual areas, such as inferotemporal cortex, perhaps on a different timescale. Indeed, monkeys with inferotemporal lesions have been shown to be severely and permanently impaired at shape discriminations based on illusory contours (Huxlin et al. 2000).

Furthermore, this putative role of feedback from higher-level areas is consistent with results from neuroimaging studies of humans, although the limitations
of the noninvasive techniques make inferences more difficult to draw. For instance, Halgren and colleagues (2003) used magnetoencephalography (MEG) to record brain activity while participants viewed arrays of shapes defined by illusory contours versus arrays that contained similar stimuli without illusory contours. The MEG activation to illusory contours was localized to the cortical surface by using a linear estimation approach that included noise-sensitivity normalization (Dale et al. 2000; Liu et al. 1998). The results revealed multiple waves of activation in the occipital polar cortex that suggested the operation of feedback loops. Specifically, following an earlier activation in the occipital pole around 100 ms after stimulus onset, a second wave of activation between about 140 and 190 ms after stimulus presentation spread from object-sensitive regions (i.e., brain regions that are activated when one views objects compared to textures) in the anterior occipital lobe back to foveal parts of areas V3, V3a, V2, and V1.

Finally, when reviewing the cognitive neuroscience literature on illusory contours, Seghier and Vuilleumier (2006) suggested that there may be two distinct feedback processing stages unfolding during the first 200 ms poststimulus: the first involves interactions between areas V1 and V2, and the second involves feedback to areas V1 and V2 from higher-level visual areas, such as lateral occipital complex (probably homologous to some object-sensitive inferotemporal regions in monkeys, that correspond to our object-properties-processing subsystem). According to our definition of reflexive top-down processing, both processing stages would be examples of reflexive top-down processes, even though they take place among different sets of areas in the ventral stream.

4.3. Modulating Interpretation

We further propose that both strategic and reflexive top-down processing can operate by altering the way earlier activation is interpreted. We can distinguish two types of such processing, which are directly analogous to changing the parameters $d'$ and $\beta$ in classical signal detection theory.

4.3.1. Changing Sensitivity

On the one hand, higher-level areas can increase or decrease the sensitivity (corresponding to $d'$) of the neurons that implement subsystems earlier in the processing stream (e.g., by increasing the baseline firing rates), making them more likely to detect the information to which they are selective. For example, in our theory, the attention shifting subsystem has the effect of priming some regions of the visual buffer (i.e., focusing the “attention window”). In addition, the information shunting subsystem passes information from long-term associative memory to the object-properties-processing subsystem; this information has the effect of increasing the sensitivity of neural populations in this subsystem for the expected parts or characteristics.

This sort of “anticipatory” priming is strategic; a comparable kind of reflexive priming can occur in the presence of a highly constraining context. In such a case,
associations in long-term associative memory would be activated by the input, and may have the effect of reflexively providing feedback to increase the sensitivity to objects that are associated with the context (such as a nose in the context of a face). By the same token, sensitivity can also be reduced, for instance, to filter out unwanted stimuli. Some of the studies discussed earlier (Kastner et al. 1999; Ress et al. 2000) that found baseline increases caused by expectation in the absence of visual information illustrate this type of process taking place in multiple areas in the ventral stream.

4.3.2. Changing Decision Criterion

On the other hand, feedback can alter how much information is necessary to make a decision. For example, to the farmer gathering his cows at dusk, a passing shadow may be sufficient to recognize a cow. Such changes in criterion (β in classical signal detection theory) could affect two kinds of processing:

First, changes in criterion could affect simple detection thresholds. For example, they could alter how much activation of specific neurons in area V4 is necessary to register that one is viewing a particular color. Similarly, if one is expecting to see a handle on a cup, the threshold for that part could be lowered to the extent that only a small portion of the handle would be required to trigger the corresponding representations in the object-properties-processing and long-term associative memory subsystems.

Second, top-down processing could alter the threshold difference in activation required to decide between two or more alternatives. For example, a “winner take all” process probably takes place in the object-properties-processing subsystem, where representations of objects are mutually inhibitory so that only one representation of an object is activated at a time. Top-down processing could affect how much relative difference in activation is required for one representation to “win” over its competitors (Kosslyn 1994; Miller and Cohen 2001). For example, when viewing a kitchen scene, reflexive top-down processing from long-term associative memory may bias representations of “chair” so that they are not only activated by less input (i.e., their thresholds are lower), but they need not be activated much more strongly than representations of other objects.

4.4. Supplementing Input

In addition to altering how input is interpreted (via changing sensitivity or decision criterion), either strategic or reflexive top-down processing could have its effects by actually filling in missing information. In neural network models, such processing is called “vector completion” (Hopfield 1982). According to our theory, if representations in the object-properties-processing subsystem are primed strongly enough, feedback connections from the areas that implement this subsystem to the areas that implement the visual buffer can force activation in these early visual cortical areas. This activation in turn corresponds to a high-resolution visual mental image; all such imagery relies on such completion operations. Such images can be strategic, as when one intentionally tries to visualize,
or reflexive, as occurs when a partially degraded object is seen and one “automatically” fills in missing contours. As an illustrative example, say that you see just the very tip of the nipple on a baby bottle sticking out from under a cloth. The portion of the tip (its shape, size, texture, and color) is sufficiently distinctive that the “baby bottle” representation in the object-properties-processing subsystem is activated. However, the input is too degraded to be sufficient for recognition. In this case, top-down processes may “complete” the image, allowing one to “see” the remainder of the bottle (as a specific shape under the cloth), at the proper size and orientation to fit the visible features. If that image cannot be “fit” to the input, then the object must be something else. If the input image can be so “completed,” that would be evidence that the activated representation is appropriate.

We stress that the process of supplementing input is distinct from even an extreme case of modulating interpretation, either via altering sensitivity or criterion: no matter how much we increase sensitivity for a certain visual attribute or lower our decision threshold, missing information is not filled in. Completion involves actually adding information to a representation earlier in the processing sequence. Some of the studies discussed earlier (Mechelli et al. 2004; Tomita et al. 1999) are examples of strategic top-down processes that perform some form of pattern completion.

The distinction between modulating interpretation and supplementing input corresponds to the distinction between “inspecting” a pattern in a visual mental image and “generating” the image in the first place (Kosslyn 1994). The former process relies on mapping the input to a specific output, which is the interpretation; whereas the latter relies on using one representation to create another, which need not be fully interpreted in advance. For example, when asked what shape are Mickey Mouse’s ears, most people report that they visualize the cartoon character to answer. The process of visualizing relies on strategic top-down processing, where information in the visual buffer and object map is supplemented, creating a pattern that depicts the object. Once formed, this representation can then be interpreted; one can classify the ears as “round.” At the outset, however, the pattern was not necessarily interpreted in this way; such prior interpretation is not a prerequisite for generating an image (Kosslyn et al. 2006).

5. Summary and Conclusions

In this chapter we developed the beginnings of a taxonomy of top-down processes in vision. The first distinction we drew is between strategic versus reflexive top-down processes. Strategic top-down processes engage the information shunting subsystem in the frontal lobes and result in the modulation of processing in subsystems implemented in the ventral and dorsal streams. Examples of activities that engage this type of top-down processes are selective visual attention, working memory, and retrieval of visual information from long-term memory. In contrast, reflexive top-down processing is automatically engaged by bottom-up processing.
and does not recruit the information shunting subsystem. An example of this type of top-down process is the completion of illusory contours.

In addition, we also proposed that each of these two general types of top-down processing has different modes of operation. One mode consists of modulating the interpretation of outputs from processes; such modulation affects the sensitivity of processing (which corresponds to changing d', in signal detection parlance) or affects the decision criterion (which corresponds to β, in signal detection theory). Another mode consists of supplementing information that is present in a subsystem, such as via vector completion, which thereby can complete fragmentary patterns with stored information.

Although these distinctions are consistent with the available findings, not much extant evidence directly bears on them. One reason for the dearth of evidence lies in many technical limitations, such as the difficulty of establishing the precise time-course of neural information processing in humans. Another, perhaps more interesting reason, is that researchers heretofore have not been thinking about top-down processing from the present perspective. Only after researchers begin to consider distinctions of the sorts we have proposed are they likely to turn their attention to collecting relevant data.

References


Neisser U (1976) Cognition and reality. Freeman, San Francisco


3
Invariant Representations of Objects in Natural Scenes in the Temporal Cortex Visual Areas

Edmund T. Rolls

1. Introduction

Evidence on how information about visual stimuli is represented in the temporal cortical visual areas and on how these representations are formed is described. The neurophysiological recordings are made mainly in nonhuman primates, macaques, first because the temporal lobe, in which this processing occurs, is much more developed than in nonprimates, and second because the findings are relevant to understanding the effects of brain damage in patients, as will be shown. In this chapter, attention is paid to neural systems involved in processing information about faces, because with the large number of neurons devoted to this class of stimuli, this system has proved amenable to experimental analysis; because of the importance of face recognition and expression identification in primate, including human, social and emotional behavior; and because of the application of understanding this neural system to understanding the effects of damage to this system in humans. It is also shown that the temporal cortical visual areas have neuronal populations that provide invariant representations of objects. Although there is some segregation of face identity and object identity represetations in different cytoarchitectonic regions, the proportion of face-selective neurons in any one architectonically defined region reaches only 20%, (see Section 2).

In Section 2, I show that there are two main populations of face-selective neurons in the temporal cortical visual areas. The first population is tuned to the identity of faces and has representations that are invariant with respect to, for example, retinal position, size, and even view. These invariant representations are ideally suited to provide the inputs to brain regions such as the orbitofrontal cortex and amygdala that learn the reinforcement associations of an individual’s face, for then the learning, and the appropriate social and emotional responses, generalize to other views of the same face. Moreover, these inferior temporal cortex neurons have sparse distributed representations of faces, which are shown
to be well suited as inputs to the stimulus–reinforcer association learning mechanisms in the orbitofrontal cortex and amygdala that allow different emotional and social responses to be made to the faces of different individuals, depending on the reinforcers received. The properties of these neurons tuned to face identity or object identity are described in Sections 3–11. Section 12 describes a second main population of neurons that are in the cortex in the superior temporal sulcus, which encode other aspects of faces such as face expression, eye gaze, face view, and whether the head is moving. This second population of neurons thus provides important additional inputs to parts of the brain such as the orbitofrontal cortex and amygdala that are involved in social communication and emotional behavior. This second population of neurons may in some cases encode reinforcement value (e.g., face expression neurons), or provide social information that is very relevant to whether reinforcers will be received, such as neurons that signal eye gaze, or whether the head is turning toward or away from the receiver. Sections 13 and 14 show how the brain may learn these invariant representations of objects and faces. Section 15 shows how attention operates computationally in natural visual scenes, and Section 16 describes the biased competition approach to how attention can modulate representations in the brain. In Sections 17 and 18, I describe the representations of faces in two areas, the amygdala and orbitofrontal cortex, to which the temporal cortical areas have direct projections. I also review evidence (Section 18) that damage to the human orbitofrontal cortex can impair face (and voice) expression identification.

The orbitofrontal cortex is also shown to be involved in the rapid reversal of behavior to stimuli (which could be the face of an individual) when the reinforcement contingencies change, and therefore to have an important role in social and emotional behavior. Moreover, the human orbitofrontal cortex is shown to be activated in a simple model of human social interaction when a face expression change indicates that the face of a particular individual is no longer reinforcing. The representations in the orbitofrontal cortex are thus of the reward or affective value of the visual stimuli that are useful in emotional behavior, in contrast to the representations in the temporal cortical visual areas, where the representations that are built are primarily of the identity of the visual stimulus.

2. Neural Responses Found in Different Temporal Lobe Cortex Visual Areas

Visual pathways project by a number of cortico-cortical stages from the primary visual cortex until they reach the temporal lobe visual cortical areas (Baizer et al. 1991; Maunsell and Newsome 1987; Seltzer and Pandya 1978), in which some neurons that respond selectively to faces are found (Bruce et al. 1981; Desimone 1991; Desimone and Gross 1979; Desimone et al. 1984; Gross et al. 1985; Perrett et al. 1982; Rolls 1981, 1984, 1991, 1992a, 2000a, 2005, 2006; Rolls and Deco 2002). The inferior temporal visual cortex, area TE, is divided on the basis of cytoarchitecture, myeloarchitecture, and afferent input into areas TEa, TEM, TE3, TE2, and TE1. In addition, there is a set of different areas in the
cortex in the superior temporal sulcus (Baylis et al. 1987; Seltzer and Pandya 1978) (Fig. 1). Of these latter areas, TPO receives inputs from temporal, parietal, and occipital cortex; PGa and IPa from parietal and temporal cortex; and TS and TAA primarily from auditory areas (Seltzer and Pandya 1978).

Considerable specialization of function was found in recordings made from more than 2600 neurons in these architectonically defined areas (Baylis et al. 1987). Areas TPO, PGa, and IPa are multimodal, with neurons that respond to visual, auditory, and/or somatosensory inputs; the inferior temporal gyrus and adjacent areas (TE3, TE2, TE1, TEa, and TEM) are primarily unimodal visual areas; areas in the cortex in the anterior and dorsal part of the superior temporal sulcus (e.g., TPO, IPa, and IPg) have neurons specialized for the analysis of moving visual stimuli; and neurons responsive primarily to faces are found more frequently in areas TPO, TEa, and TEM, where they comprise approximately 20% of the visual neurons responsive to stationary stimuli, in contrast to the other temporal cortical areas, in which they comprise 4% to 10%. The stimuli that activate other cells in these TE regions include simple visual patterns such as gratings and combinations of simple stimulus features (Gross et al. 1985; Tanaka et al. 1990). Because face-selective neurons have a wide distribution, it might be expected that only large lesions, or lesions that interrupt outputs of these visual areas, would produce readily apparent face-processing deficits. Moreover, neurons with responses related to facial expression, movement, and gesture are more likely to be found in the cortex in the superior temporal sulcus, whereas neurons with activity related to facial identity are more likely to be found in the TE areas (Hasselmo et al. 1989a).

In human functional magnetic resonance imaging (fMRI) studies, evidence for specialization of function is described (Grill-Spector and Malach 2004; Haxby et al. 2002; Spiridon and Kanwisher 2002) related to face processing (in the
fusiform face area, which may correspond to parts of the macaque inferior temporal visual cortex in which face neurons are common; to face expression and gesture (i.e., moving faces) (in the cortex in the superior temporal sulcus, which corresponds to the macaque cortex in the superior temporal sulcus); to objects (in an area that may correspond to the macaque inferior temporal cortex in which object but not face representations are common, as already described); and to spatial scenes (in a parahippocampal area, which probably corresponds to the macaque parahippocampal gyrus areas in which neurons are tuned to spatial view and to combinations of objects and the places in which they are located (Georges-François et al. 1999; Robertson et al. 1998; Rolls 1999c; Rolls and Kesner 2006; Rolls and Xiang 2005, 2006; Rolls et al. 1997b, 1998, 2005). However, there is much debate arising from these human fMRI studies about how specific each region is for a different type of function, in that such studies do not provide clear evidence on whether individual neurons can be very selective for face identity versus face expression versus objects and thereby convey specific information about these different classes of object; whether each area contains a mixture of different populations of neurons each tuned to different specific classes of visual stimuli, or neurons with relatively broad tuning that respond at least partly to different classes of stimuli; and about the fine-grain topology within a cortical area. The single-neuron studies in macaques described above and below do provide clear answers to these questions. The neuronal recording studies show that individual neurons can be highly tuned in that they convey information about face identity, or about face expression, or about objects, or about spatial view. The recording studies show that within these different classes, individual neurons by responding differently to different members of the class convey information about whose face it is, what the face expression is, etc., using a sparse distributed code with an approximately exponential firing rate probability distribution. The neuronal recording studies also show that each cytoarchitectonically defined area contains different proportions of face identity versus object neurons, but that the proportion of face-selective neurons in any one area is not higher than 20% of the visually responsive neurons in an area, so that considerable intermixing of specifically tuned neurons is the rule (Baylis et al. 1987). The neuronal recording studies also show that at the fine spatial scale, clusters of neurons extending for approximately 0.5–1 mm with tuning to one aspect of stimuli are common (e.g., face identity, or the visual texture of stimuli, or a particular class of head motion), and this can be understood as resulting from self-organizing mapping based on local cortical connectivity when a high dimensional space of objects, faces, etc., must be represented on a two-dimensional cortical sheet (Rolls and Deco 2002; Rolls 2008).

3. The Selectivity of One Population of Neurons for Faces

The neurons described in our studies as having responses selective for faces are selective in that they respond 2 to 20 times more (and statistically significantly more) to faces than to a wide range of gratings, simple geometric stimuli, or
complex three-dimensional (3-D) objects (Baylis et al. 1985, 1987; Rolls and Deco 2002; Rolls 1984, 1992a, 1997, 2000a, 2007). The recordings are made while the monkeys perform a visual fixation task in which, after the fixation spot has disappeared, a stimulus subtending typically 8° is presented on a video monitor (or, in some earlier studies, while monkeys perform a visual discrimination task). The responses to faces are excitatory, with firing rates often reaching 100 spikes/s, sustained, and have typical latencies of 80–100 ms. The neurons are typically unresponsive to auditory or tactile stimuli and to the sight of arousing or aversive stimuli. These findings indicate that explanations in terms of arousal, emotional or motor reactions, and simple visual feature sensitivity are insufficient to account for the selective responses to faces and face features observed in this population of neurons (Baylis et al. 1985; Perrett et al. 1982; Rolls and Baylis 1986). Observations consistent with these findings have been published by Desimone et al. (1984), who described a similar population of neurons located primarily in the cortex in the superior temporal sulcus that responded to faces but not to simpler stimuli such as edges and bars or to complex non-face stimuli (see also Gross et al. 1985).

These neurons are specialized to provide information about faces in that they provide much more information (on average, 0.4 bits) about which (of 20) face stimuli is being seen than about which (of 20) non-face stimuli is being seen (on average, 0.07 bits) (Rolls and Tovee 1995a; Rolls et al. 1997a). These information theoretical procedures provide an objective and quantitative way to show what is “represented” by a particular population of neurons, and indicate that different categories of visual stimulus are represented by different populations of inferior temporal cortex neurons (see also Hasselmo et al. 1989a).

4. The Selectivity of These Neurons for Individual Face Features or for Combinations of Face Features

Masking out or presenting parts of the face (e.g., eyes, mouth, or hair) in isolation reveal that different cells respond to different features or subsets of features. For some cells, responses to the normal organization of cut-out or line-drawn facial features are significantly larger than to images in which the same facial features are jumbled (Perrett et al. 1982; Rolls et al. 1994). These findings are consistent with the hypotheses developed below that by competitive self-organization some neurons in these regions respond to parts of faces by responding to combinations of simpler visual properties received from earlier stages of visual processing, and that other neurons respond to combinations of parts of faces and thus respond only to whole faces. Moreover, the finding that for some of these latter neurons the parts must be in the correct spatial configuration shows that the combinations formed can reflect not just the features present, but also their spatial arrangement; this provides a way in which binding can be implemented in neural networks (Elliffe et al. 2002; Rolls and Deco 2002). Further evidence that neurons in these regions respond to combinations of features in
the correct spatial configuration was found by Tanaka et al. (1990) using combinations of features that are used by comparable neurons to define objects.

5. Distributed Encoding of Face and Object Identity

An important question for understanding brain function is whether a particular object (or face) is represented in the brain by the firing of one or a few gnostic (or “grandmother”) cells (Barlow 1972), or whether instead the firing of a population or ensemble of cells each with different profiles of responsiveness to the stimuli provides the representation. It has been shown that the representation of which particular object (face) is present is rather distributed. Baylis, Rolls, and Leonard (1985) showed this with the responses of temporal cortical neurons that typically responded to several members of a set of 5 faces, with each neuron having a different profile of responses to each face. In a further study using 23 faces and 45 non-face natural images, a distributed representation was again found (Rolls and Tovee 1995a), with the average sparseness being 0.65. The sparseness of the representation provided by a neuron can be defined as

\[ a = \left( \frac{\sum_{i=1}^{S} r_i/S}{\sum_{i=1}^{S} (r_i/S)} \right)^2 \]

where \( r_s \) is the mean firing rate of the neuron to stimulus \( s \) in the set of \( S \) stimuli [see Rolls and Treves (1998) and Franco et al. (2007)]. If the neurons are binary (either firing or not to a given stimulus), then \( a \) would be 0.5 if the neuron responded to 50% of the stimuli and 0.1 if a neuron responded to 10% of the stimuli. If the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the changes of firing rate, that is, the active responses of the neurons, were used in the sparseness calculation, then the “response sparseness” had a lower value, with a mean of 0.33 for the population of neurons.

The distributed nature of the representation can be further understood by the finding that the firing rate distribution of single neurons when a wide range of natural visual stimuli are being viewed is approximately exponentially distributed, with rather few stimuli producing high firing rates, and increasingly large numbers of stimuli producing lower and lower firing rates (Baddeley et al. 1997; Franco et al. 2006; Rolls and Tovee 1995a; Treves et al. 1999) (Fig. 2). The sparseness of such an exponential distribution of firing rates is 0.5. It has been shown that the distribution may arise from the threshold nonlinearity of neurons combined with short-term variability in the responses of neurons (Treves et al. 1999).

Complementary evidence comes from applying information theory to analyze how information is represented by a population of these neurons. The information required to identify which of \( S \) equiprobable events occurred (or stimuli were shown) is \( \log_2 S \) bits. (Thus, 1 bit is required to specify which of 2 stimuli was shown, 2 bits to specify which of 4 stimuli was shown, 3 bits to specify which of 8 stimuli was shown, etc.) The important point for the present purposes is that
if the encoding was local (or grandmother cell-like), then the number of stimuli encoded by a population of neurons would be expected to rise approximately linearly with the number of neurons in the population. In contrast, with distributed encoding, provided that the neuronal responses are sufficiently independent, and are sufficiently reliable (not too noisy), the number of stimuli encodable by the population of neurons might be expected to rise exponentially as the number of neurons in the sample of the population was increased. The information available about which of 20 equiprobable faces had been shown that was available from the responses of different numbers of these neurons is shown in Fig. 3. First, it is clear that some information is available from the responses of just one neuron, on average, approximately 0.34 bits. Thus, knowing the activity of just one neuron in the population does provide some evidence about which stimulus was present. This evidence that information is available in the responses of individual neurons in this way, without having to know the state of all the

Fig. 2. Firing rate distribution of a single neuron in the temporal visual cortex to a set of 23 face ($F$) and 45 non-face images of natural scenes. The firing rate to each of the 68 stimuli is shown. $P$, a face profile stimulus; $B$, a body part stimulus such as a hand. (After Rolls and Tovee 1995a)
E.T. Rolls

other neurons in the population, indicates that information is made explicit in the firing of individual neurons in a way that will allow neurally plausible decoding, involving computing a sum of input activities each weighted by synaptic strength, to work (see following). Second, it is clear (see Fig. 3) that the information rises approximately linearly, and the number of stimuli encoded thus rises approximately exponentially, as the number of cells in the sample increases (Abbott et al. 1996; Rolls and Treves 1998; Rolls et al. 1997a).

This direct neurophysiological evidence thus demonstrates that the encoding is distributed, and the responses are sufficiently independent and reliable, that the representational capacity increases exponentially with the number of neurons in the ensemble (Fig. 4). The consequence of this is that large numbers of stimuli, and fine discriminations between them, can be represented without having to measure the activity of an enormous number of neurons. [It has been shown that the main reason why the information tends to asymptote, as shown in Fig. 3, as the number of neurons in the sample increases is just that the ceiling is being approached of how much information is required to discriminate between the set of stimuli, which with 20 stimuli is \( \log_2 20 = 4.32 \) bits (Abbott et al. 1996; Rolls et al. 1997a)].
Fig. 4. The number of stimuli (in this case from a set of 20 faces) that are encoded in the responses of different numbers of neurons in the temporal lobe visual cortex, based on the results shown in Fig. 3. The decoding method was dot product (DP, open circles) or probability estimation (PE, filled circles). (After Rolls, Treves, and Tovee 1997; Abbott, Rolls, and Tovee 1996)

It has in addition been shown that there are neurons in the inferior temporal visual cortex that encode view invariant representations of objects, and for these neurons the same type of representation is found, namely distributed encoding with independent information conveyed by different neurons (Booth and Rolls 1998).

The analyses just described were obtained with neurons that were not simultaneously recorded, but we have more recently shown that with simultaneously recorded neurons similar results are obtained, that is, the information about which stimulus was shown increases approximately linearly with the number of neurons, showing that the neurons convey information that is nearly independent (Panzeri et al. 1999b; Rolls et al. 2004). [Consistently, Gawne and Richmond (1993) showed that even adjacent pairs of neurons recorded simultaneously from the same electrode carried information that was approximately 80% independent.] In the research described by Panzeri et al. (1999b), Rolls et al. (2003b), and Franco et al. (2004), we developed methods for measuring the information in the relative time of firing of simultaneously recorded neurons, which might be significant if the neurons became synchronized to some but not other stimuli in a set, as postulated by Singer (1999). We found that for the set of cells currently available, almost all the information was available in the firing rates of the cells, and very little (not more than approximately 5% of the total information) was available about these static images in the relative time of firing of different simultaneously recorded neurons (Franco et al. 2004; Panzeri et al. 1999b; Rolls et al. 2003b, 2004). Thus, the evidence is that for representations of faces and objects in the inferior temporal visual cortex (and of space in the primate hippocampus and of odors in the orbitofrontal cortex; see Rolls et al. 1996, 1998), most of the information is available in the firing rates of the neurons.
To obtain direct evidence on whether stimulus-dependent synchrony is important in encoding information in natural and normal visual processing, we (Aggelopoulos et al. 2005) analyzed the activity of simultaneously recorded neurons using an object-based attention task in which macaques searched for a target object to touch in a complex natural scene. In the task, object-based attention was required as the macaque knew which of the two objects he was searching for. Feature binding was required in that two objects (each requiring correct binding of the features of that object but not the other object) were present, and segmentation was required to segment the objects from their background. This is a real-world task with natural visual scenes, in which, if temporal synchrony was important in neuronal encoding, it should be present. Information theoretical techniques were used to assess how much information was provided by the firing rates of the neurons about the stimuli and how much by the stimulus-dependent cross-correlations between the firing of different neurons that were sometimes present. The use of information theoretic procedures was important, for it allowed the relative contributions of rates and stimulus-dependent synchrony to be quantified (Franco et al. 2004). It was found that between 99% and 94% of the information was present in the firing rates of inferior temporal cortex neurons, and less that 5% in any stimulus-dependent synchrony that was present, as illustrated in Fig. 5 (Aggelopoulos et al. 2005). The implication of these results is that any stimulus-dependent synchrony that is present is not quantitatively important, as measured by information theoretical analyses under natural scene conditions; this has been found for the inferior temporal cortex, a brain region where features are put together to form representations of objects (Rolls and Deco 2002), and where attention has strong effects, at least in scenes with blank backgrounds (Rolls et al. 2003a). The finding as assessed by information theoretical methods of the importance of firing rates and not stimulus-dependent synchrony is consistent with previous information theoretic approaches (Franco et al. 2004; Rolls et al. 2003b, 2004). It would of course also be of interest to test the same hypothesis in earlier visual areas, such as V4, with quantitative, information theoretical, techniques. In connection with rate codes, it should be noted that a rate code implies using the number of spikes that arrive in a given time, and that this time can be very short, as little as 20 to 50ms, for very useful amounts of information to be made available from a population of neurons (Rolls 2003; Rolls and Tovee 1994; Rolls et al. 1994, 1999, 2006b; Tovee and Rolls 1995; Tovee et al. 1993).

The implications of these findings for the computational bases of attention are important. First, the findings indicate that top-down attentional biasing inputs could, by providing biasing inputs to the appropriate object-selective neurons, facilitate bottom-up information about objects without any need to alter the time relations between the firing of different neurons. The neurons to which the top-down biases should be applied could in principle be learned by simple Hebbian associativity between the source of the biasing signals, in for example the prefrontal cortex, and the inferior temporal cortex neurons (Rolls and Deco 2002). Thus, rate encoding would be sufficient for the whole system to implement attention, a conclusion supported by the spiking network model of attention of Deco
Two stimuli were presented at the same time on the screen and neuronal activity was measured when the monkey was looking at stimulus 1 or stimulus 2, and the information was calculated.

Fig. 5. Right. The information available from the firing rates (Rate Inf) or from stimulus-dependent synchrony (Cross-Corr Inf) from populations of simultaneously recorded inferior temporal cortex neurons about which stimulus had been presented in a complex natural scene. The total information (Total Inf) is that available from both the rate and the stimulus-dependent synchrony, which do not necessarily contribute independently. Left. Eye position recordings and spiking activity from two neurons on a single trial of the task. (Neuron 31 tended to fire more when the macaque looked at one of the stimuli, $S_-$, and neuron 21 tended to fire more when the macaque looked at the other stimulus, $S_+$. Both stimuli were within the receptive field of the neuron.) (After Aggelopoulos et al. 2005)

and Rolls (2005c), in which nonlinear interactions between top-down and bottom-up signals without specific temporal encoding can implement the details of the interactions found neurophysiologically in V4 and V2. Second, the findings are consistent with the hypothesis that feature binding is implemented by feature combination neurons which respond to features in the correct relative spatial locations (Elliffe et al. 2002; Rolls and Deco 2002), and not by temporal synchrony and attention (Singer 1999; Singer and Gray 1995; von der Malsburg 1990).

With respect to the synchrony model, von der Malsburg (1990) suggested that features that should be bound together would be linked by temporal binding. There has been considerable neurophysiological investigation of this possibility (Singer 1999; Singer and Gray 1995). A problem with this approach is that temporal binding might enable features 1, 2, and 3 (which might define one stimulus) to be bound together and kept separate from, for example, another stimulus
consisting of features 2, 3, and 4, but would require a further temporal binding (leading in the end potentially to a combinatorial explosion) to indicate the relative spatial positions of the 1, 2, and 3 in the 123 stimulus, so that it can be discriminated from 312, for example. Thus, temporal synchrony could it seems at best be useful for grouping features (e.g., features 1, 2, and 3 are part of object 1, and features 4 and 6 are part of object 2), but would not, without a great deal more in the way of implementation, be useful to encode the relative spatial positions of features within an object or of objects in a scene.

It is unlikely that there are further processing areas beyond those described where ensemble coding changes into grandmother cell (local) encoding. Anatomically, there does not appear to be a whole further set of visual processing areas present in the brain; and outputs from the temporal lobe visual areas such as those described, are taken to limbic and related regions such as the amygdala and orbitofrontal cortex, and via the entorhinal cortex to the hippocampus, where associations between the visual stimuli and other sensory representations are formed (Rolls and Deco 2002; Rolls 2005). Indeed, tracing this pathway onward, we have found a population of neurons with face-selective responses in the amygdala (Leonard et al. 1985; Rolls 2000b) and orbitofrontal cortex (Rolls et al. 2006a), and in the majority of these neurons, different responses occur to different faces, with ensemble (not local) coding still being present. The amygdala in turn projects to another structure that may be important in other behavioral responses to faces, the ventral striatum, and comparable neurons have also been found in the ventral striatum (Williams et al. 1993).

6. Advantages of the Distributed Representation of Objects and Faces for Brain Processing

The advantages of the distributed encoding found are now considered, and apply to both fully distributed and to sparse distributed (but not to local) encoding schemes, as explained elsewhere (Rolls 2008; Rolls 2005; Rolls and Deco 2002; Rolls and Treves 1998).

6.1. Exponentially High Coding Capacity

This property arises from a combination of the encoding being sufficiently close to independent by the different neurons (i.e., factorial), and sufficiently distributed. Part of the biological significance of the exponential encoding capacity found is that a receiving neuron or neurons can obtain information about which one of a very large number of stimuli is present by receiving the activity of relatively small numbers of inputs (of the order of hundreds) from each of the neuronal populations from which it receives. In particular, the characteristics of the actual visual cells described here indicate that the activity of 15 would be able to encode 192 face stimuli (at 50% accuracy); of 20 neurons, 768 stimuli; of 25 neurons; 3 072 stimuli; of 30 neurons, 12 288 stimuli; and of 35 neurons, 49 152 stimuli (the values are for the optimal decoding case) (Abbott et al. 1996). Given
that most neurons receive a limited number of synaptic contacts, of the order of several thousand, this type of encoding is ideal. (It should be noted that the capacity of the distributed representations was calculated from ensembles of neurons each already shown to provide information about faces. If inferior temporal cortex neurons were chosen at random, 20 times as many neurons would be needed in the sample if face-selective neurons comprised 5% of the population. This brings the number of inputs required from an ensemble up to reasonable numbers given brain connectivity, a number of the order of the thousands of synapses being received by each neuron.) This type of encoding would enable, for example, neurons in the amygdala and orbitofrontal cortex to form pattern associations of visual stimuli with reinforcers such as the taste of food when each neuron received a reasonable number, perhaps in the order of hundreds, of inputs from the visually responsive neurons in the temporal cortical visual areas which specify which visual stimulus or object is being seen (Rolls 1990, 1992a,b; Rolls and Deco 2002; Rolls and Treves 1998). It is useful to realize that although the sensory representation may have exponential encoding capacity, this does not mean that the associative networks that receive the information can store such large numbers of different patterns. Indeed, there are strict limitations on the number of memories that associative networks can store (Rolls and Treves 1990, 1998; Treves and Rolls 1991). The particular value of the exponential encoding capacity of sensory representations is that very fine discriminations can be made, as there is much information in the representation, and that the representation can be decoded if the activity of even a limited number of neurons in the representation is known.

One of the underlying themes here is the neural representation of faces and objects. How would one know that one had found a neuronal representation of faces or objects in the brain? The criterion suggested (Rolls and Treves 1998) is that when one can identify the face or object that is present (from a large set of stimuli, which might be thousands or more) with a realistic number of neurons, say of the order of 100, and with some invariance, then one has a useful representation of the object.

The properties of the representation of faces, of objects (Booth and Rolls 1998), and of olfactory and taste stimuli, have been evident when the readout of the information was by measuring the firing rate of the neurons, typically over a 20-, 50-, or 500-ms period. Thus, at least where objects are represented in the visual, olfactory, and taste systems (e.g., individual faces, odors, and tastes), information can be read out without taking into account any aspects of the possible temporal synchronization between neurons, or temporal encoding within a spike train (Aggelopoulos et al. 2005; Franco et al. 2004; Panzeri et al. 1999b; Rolls et al. 1997a, 2003b, 2004; Tovee et al. 1993).

6.2. Ease with Which the Code Can Be Read by Receiving Neurons

For brain plausibility, it is also a requirement that neurons should be able to read the code. This is why when we have estimated the information from populations
of neurons, we have used in addition to a probability estimating measure (PE, optimal, in the Bayesian sense), also a dot product measure, which is a way of specifying that all that is required of decoding neurons would be the property of adding up postsynaptic potentials produced through each synapse as a result of the activity of each incoming axon (Abbott et al. 1996; Rolls et al. 1997a). It was found that with such a neurally plausible algorithm (the dot product, DP, algorithm), which calculates which average response vector the neuronal response vector on a single test trial was closest to by performing a normalized dot product (equivalent to measuring the angle between the test and the average vector), the same generic results were obtained, with only a 40% reduction of information compared to the more efficient (PE) algorithm. This is an indication that the brain could utilize the exponentially increasing capacity for encoding stimuli as the number of neurons in the population increases. For example, by using the representation provided by the neurons described here as the input to an associative or autoassociative memory, which computes effectively the dot product on each neuron between the input vector and the synaptic weight vector, most of the information available would in fact be extracted (Franco et al. 2004; Rolls and Deco 2002; Rolls and Treves 1990, 1998; Treves and Rolls 1991).

6.3. Higher Resistance to Noise

This, like the next few properties, is an advantage of distributed over local representations, which applies to artificial systems as well, but is presumably of particular value in biological systems in which some of the elements have an intrinsic variability in their operation. Because the decoding of a distributed representation involves assessing the activity of a whole population of neurons, and computing a dot product or correlation, a distributed representation provides more resistance to variation in individual components than does a local encoding scheme (Panzeri et al. 1996; Rolls and Deco 2002).

6.4. Generalization

Generalization to similar stimuli is again a property that arises in neuronal networks if distributed but not if local encoding is used. The generalization arises as a result of the fact that a neuron can be thought of as computing the inner or dot product of the stimulus representation with its weight vector. If the weight vector leads to the neuron having a response to one visual stimulus, then the neuron will have a similar response to a similar visual stimulus. This computation of correlations between stimuli operates only with distributed representations. If an output is based on a single input or output pair, then if either is lost, the correlation drops to zero (Rolls and Treves 1998; Rolls and Deco 2002).

6.5. Completion

Completion occurs in associative memory networks by a similar process. Completion is the property of recall of the whole of a pattern in response to any part
of the pattern. Completion arises because any part of the stimulus representation, or pattern, is effectively correlated with the whole pattern during memory storage. Completion is thus a property of distributed representations, and not of local representations. It arises, for example, in autoassociation (attractor) neuronal networks, which are characterized by recurrent connectivity. It is thought that such networks are important in the cerebral cortex, where the association fibers between nearby pyramidal cells may help the cells to retrieve a representation that depends on many neurons in the network (Rolls and Deco 2002; Rolls and Treves 1998).

6.6. Graceful Degradation or Fault Tolerance

This also arises only if the input patterns have distributed representations, and not if they are local. Local encoding suffers sudden deterioration once the few neurons or synapses carrying the information about a particular stimulus are destroyed.

6.7. Speed of Readout of the Information

The information available in a distributed representation can be decoded by an analyzer more quickly than can the information from a local representation, given comparable firing rates. Within a fraction of an interspike interval, with a distributed representation, much information can be extracted (Panzeri et al. 1999a; Rolls et al. 1997a; Rolls et al. 2006b; Treves 1993; Treves et al. 1996, 1997). In effect, spikes from many different neurons can contribute to calculating the angle between a neuronal population and a synaptic weight vector within an interspike interval (Franco et al. 2004; Rolls and Deco 2002). With local encoding, the speed of information readout depends on the exact model considered, but if the rate of firing needs to be taken into account, this will necessarily take time, because of the time needed for several spikes to accumulate in order to estimate the firing rate.

7. Invariance in the Neuronal Representation of Stimuli

One of the major problems that must be solved by a visual system is the building of a representation of visual information that allows recognition to occur relatively independently of size, contrast, spatial frequency, position on the retina, angle of view, etc. This is required so that if the receiving associative networks (in, e.g., the amygdala, orbitofrontal cortex, and hippocampus) learn about one view, position, etc., of the object, the animal generalizes correctly to other positions or views of the object. It has been shown that the majority of face-selective inferior temporal cortex neurons have responses that are relatively invariant with respect to the size of the stimulus (Rolls and Baylis 1986). The median size change tolerated with a response of greater than half the maximal response was
12 times. Also, the neurons typically responded to a face when the information in it had been reduced from 3-D to a 2-D representation in gray on a monitor, with a response which was on average 0.5 of that to a real face. Another transform over which recognition is relatively invariant is spatial frequency. For example, a face can be identified when it is blurred (when it contains only low spatial frequencies), and when it is high-pass spatial frequency filtered (when it looks like a line drawing). It has been shown that if the face images to which these neurons respond are low-pass filtered in the spatial frequency domain (so that they are blurred), then many of the neurons still respond when the images contain frequencies only up to 8 cycles per face. Similarly, the neurons still respond to high-pass filtered images (with only high spatial frequency edge information) when frequencies down to only 8 cycles per face are included (Rolls et al. 1985). Face recognition shows similar invariance with respect to spatial frequency (Rolls et al. 1985). Further analysis of these neurons with narrow (octave) bandpass spatial frequency filtered face stimuli shows that the responses of these neurons to an unfiltered face can not be predicted from a linear combination of their responses to the narrow band stimuli (Rolls et al. 1987). This lack of linearity of these neurons, and their responsiveness to a wide range of spatial frequencies, indicate that in at least this part of the primate visual system recognition does not occur using Fourier analysis of the spatial frequency components of images.

Inferior temporal visual cortex neurons also often show considerable translation (shift) invariance, not only under anesthesia (see Gross et al. 1985), but also in the awake behaving primate (Tovee et al. 1994). It was found that in most cases the responses of the neurons were little affected by which part of the face was fixated, and that the neurons responded (with a greater than half-maximal response) even when the monkey fixated 2° to 5° beyond the edge of a face which subtended 8° to 17° at the retina. Moreover, the stimulus selectivity between faces was maintained this far eccentric within the receptive field.

Until recently, research on translation invariance considered the case in which there is only one object in the visual field. What happens in a cluttered, natural, environment? Do all objects that can activate an inferior temporal neuron do so whenever they are anywhere within the large receptive fields of inferior temporal cortex neurons (Sato 1989)? If so, the output of the visual system might be confusing for structures which receive inputs from the temporal cortical visual areas. In an investigation of this, it was found that the mean firing rate across all cells to a fixated effective face with a noneffective face in the parafovea (centered 8.5° from the fovea) was 34 spikes/s. On the other hand, the average response to a fixated non-effective face with an effective face in the periphery was 22 spikes/s (Rolls and Tovee 1995b). Thus these cells gave a reliable output about which stimulus is actually present at the fovea, in that their response was larger to a fixated effective face than to a fixated noneffective face, even when there are other parafoveal stimuli effective for the neuron.

It has now been shown that the receptive fields of inferior temporal cortex neurons, while large (typically 70° in diameter) when a test stimulus is presented
against a blank background, become much smaller, as little as several degrees in diameter, when objects are seen against a complex natural background (Rolls et al. 2003a). Object representation and selection in complex natural scenes is considered in Section 9.

8. A View-Independent Representation of Faces and Objects

It has also been shown that some temporal cortical neurons reliably responded differently to the faces of two different individuals independently of viewing angle (Hasselmo et al. 1989b), although in most cases (16/18 neurons) the response was not perfectly view independent. Mixed together in the same cortical regions are neurons with view-dependent responses (Hasselmo et al. 1989b). Such neurons might respond for example to a view of a profile of a monkey but not to a full-face view of the same monkey (Perrett et al. 1985a). These findings, of view-dependent, partially view-independent, and view-independent representations in the same cortical regions are consistent with the hypothesis discussed below that view-independent representations are being built in these regions by associating together neurons that respond to different views of the same individual.

Further evidence that some neurons in the temporal cortical visual areas have object-based rather than view-based responses comes from a study of a population of neurons that responds to moving faces (Hasselmo et al. 1989b). For example, four neurons responded vigorously to a head undergoing ventral flexion, irrespective of whether the view of the head was full face, of either profile, or even of the back of the head. These different views could only be specified as equivalent in object-based coordinates. Further, for all of the ten neurons that were tested in this way, the movement specificity was maintained across inversion, responding, for example, to ventral flexion of the head irrespective of whether the head was upright or inverted. In this procedure, retinally encoded or viewer-centered movement vectors are reversed, but the object-based description remains the same. It is an important property of these neurons that they can encode a description of an object that is based on relative motions of different parts of the object and which is not based on flow relative to the observer. The implication of this type of encoding is that the upper eyelids closing could be encoded as the same social signal that eye contact is being broken independently of the particular in-plane rotation (tilt, as far as being fully inverted) of the face being observed (or of the observer’s head).

Also consistent with object-based encoding is the finding of a small number of neurons that respond to images of faces of a given absolute size, irrespective of the retinal image size or distance (Rolls and Baylis 1986).

Neurons with view-invariant responses of objects seen naturally by macaques have also been found (Booth and Rolls 1998). The stimuli were presented for 0.5 s on a color video monitor while the monkey performed a visual fixation task.
The stimuli were images of ten real plastic objects that had been in the monkey’s cage for several weeks to enable him to build view-invariant representations of the objects. Control stimuli were views of objects that had never been seen as real objects. The neurons analyzed were in the TE cortex in and close to the ventral lip of the anterior part of the superior temporal sulcus. Many neurons were found that responded to some views of some objects. However, for a smaller number of neurons, the responses occurred only to a subset of the objects (using ensemble encoding), irrespective of the viewing angle. Further evidence consistent with these findings is that some studies have shown that the responses of some visual neurons in the inferior temporal cortex do not depend on the presence or absence of critical features for maximal activation (Perrett et al. 1982; Tanaka 1993, 1996). For example, Mikami et al (1994) have shown that some TE cells respond to partial views of the same laboratory instrument(s), even when these partial views contain different features. In a different approach, Logothetis et al. (1994) have reported that in monkeys extensively trained (over thousands of trials) to treat different views of computer-generated wire-frame “objects” as the same, a small population of neurons in the inferior temporal cortex did respond to different views of the same wire-frame object (Logothetis and Sheinberg 1996). The difference in the approach taken by Booth and Rolls (1998) was that no explicit training was given in invariant object recognition, as Rolls’ hypothesis (1992a) is that view-invariant representations can be learned by associating together the different views of objects as they are moved and inspected naturally in a period that may be in the order of a few seconds.

9. The Representation of Objects in Complex Natural Scenes

9.1. Object-Based Attention and Object Selection in Complex Natural Scenes

Object-based attention refers to attention to an object. For example, in a visual search task the object might be specified as what should be searched for, and its location must be found. In spatial attention, a particular location in a scene is pre-cued, and the object at that location may need to be identified.

Much of the neurophysiology, psychophysics, and modeling of attention has been with a small number, typically two, of objects in an otherwise blank scene. In this section, I consider how attention operates in complex natural scenes, and in particular describe how the inferior temporal visual cortex operates to enable the selection of an object in a complex natural scene.

To investigate how attention operates in complex natural scenes, and how information is passed from the inferior temporal cortex (IT) to other brain regions to enable stimuli to be selected from natural scenes for action, Rolls et al. (2003a) analyzed the responses of inferior temporal cortex neurons to stimuli presented in complex natural backgrounds. The monkey had to search for two
objects on a screen, and a touch of one object was rewarded with juice, and that of another object was punished with saline (Fig. 6). Neuronal responses to the effective stimuli for the neurons were compared when the objects were presented in the natural scene or on a plain background. It was found that the overall response of the neuron to objects was hardly reduced when they were presented in natural scenes, and the selectivity of the neurons remained. However, the main finding was that the magnitudes of the responses of the neurons typically became much less in the real scene the further the monkey fixated in the scene away from the object (Fig. 7). It is proposed that this reduced translation invariance in natural scenes helps an unambiguous representation of an object that may be the target for action to be passed to the brain regions which receive from the primate inferior temporal visual cortex. It helps with the binding problem, by reducing in natural scenes the effective receptive field of at least some inferior temporal cortex neurons to approximately the size of an object in the scene.

It is also found that, in natural scenes, the effect of object-based attention on the response properties of inferior temporal cortex neurons is relatively small, as illustrated in Fig. 8 (Rolls et al. 2003a). The results summarized in Fig. 8 for $5^\circ$ stimuli show that the receptive fields were large ($77.6^\circ$) with a single stimulus
in a blank background (top left) and were greatly reduced in size (to 22.0°) when presented in a complex natural scene (top right). The results also show that there was little difference in receptive field size or firing rate in the complex background when the effective stimulus was selected for action (bottom right, 19.2°), and when it was not (middle right, 15.6°) (Rolls et al. 2003a). (For comparison, the effects of attention against a blank background were much larger, with the receptive field increasing from 17.2° to 47.0° as a result of object-based attention, as shown in Fig. 8.) The computational basis for these relatively minor effects of object-based attention when objects are viewed in natural scenes is considered in Section 15.

These findings on how objects are represented in natural scenes make the interface to memory and to action systems simpler, in that what is at the fovea can be interpreted (e.g., by an associative memory in the orbitofrontal cortex or amygdala) partly independently of the surroundings, and choices and actions can be directed if appropriate to what is at the fovea (Ballard 1993; Rolls and Deco 2002).

9.2. The Representation of Information About the Relative Positions of Multiple Objects in a Scene

These experiments have been extended to address the issue of how several objects are represented in a complex scene. The issue arises because the relative spatial locations of objects in a scene must be encoded (and is possible even in
3. Visual Object and Face Representations

Fig. 8. Summary of the receptive field sizes of inferior temporal cortex neurons to a 5° effective stimulus presented in either a blank background (blank screen) or in a natural scene (complex background). The stimulus that was a target for action in the different experimental conditions is marked by T. When the target stimulus was touched, a reward was obtained. The mean receptive field diameter of the population of neurons analyzed, and the mean firing rate in spikes/s, is shown. The stimuli subtended 5° × 3.5° at the retina, and occurred on each trial in a random position in the 70° × 55° screen. The dashed circle is proportional to the receptive field size. Top row: responses with one visual stimulus in a blank (left) or complex (right) background. Middle row: responses with two stimuli, when the effective stimulus was not the target of the visual search. Bottom row: responses with two stimuli, when the effective stimulus was the target of the visual search. (After Rolls et al. 2003)
short presentation times without eye movements) (Biederman 1972) (and this has been held to involve some spotlight of attention); and because as shown above what is represented in complex natural scenes is primarily about what is at the fovea, yet we can locate more than one object in a scene even without eye movements. Aggelopoulos and Rolls (2005) showed that with five objects simultaneously present in the receptive field of inferior temporal cortex neurons, although all the neurons responded to their effective stimulus when it was at the fovea, some could also respond to their effective stimulus when it was in a parafoveal position 10° from the fovea. An example of such a neuron is shown in Fig. 9. The asymmetry is much more evident in a scene with five images present (Fig. 9A) than when only one image is shown on an otherwise blank screen (Fig. 9B). Competition between different stimuli in the receptive field thus reveals the asymmetry in the receptive field of inferior temporal visual cortex neurons. The asymmetry provides a way of encoding the position of multiple objects in a scene. Depending on which asymmetrical neurons are firing, the population of neurons provides information to the next processing stage, not only about which image is present at or close to the fovea, but where it is with respect to the fovea.

Fig. 9. A The responses (firing rate with the spontaneous rate subtracted, means ± SEM) of one neuron when tested with five stimuli simultaneously present in the close (10°) configuration with the parafoveal stimuli located 10° from the fovea. B The responses of the same neuron when only the effective stimulus was presented in each position. The firing rate for each position is that when the effective stimulus for the neuron was in that position. The $P$ value is that from the ANOVA calculated over the four parafoveal positions. (After Aggelopoulos and Rolls 2005)
This information is provided by neurons that have firing rates, which reflect the relevant information, and stimulus-dependent synchrony is not necessary. Top-down attentional biasing input could thus, by biasing the appropriate neurons, facilitate bottom-up information about objects without any need to alter the time relationships between the firing of different neurons. The exact position of the object with respect to the fovea, and effectively thus its spatial position relative to other objects in the scene, would then be made evident by the subset of asymmetrical neurons firing.

This is, thus, the solution that these experiments indicate is used to the representation of multiple objects in a scene (Aggelopoulos and Rolls 2005), an issue which has previously been difficult to account for in neural systems with distributed representations (Mozer 1991) and for which “attention” has been a proposed solution.

10. Learning of New Representations in the Temporal Cortical Visual Areas

To investigate the hypothesis that visual experience might guide the formation of the responsiveness of neurons so that they provide an economical and ensemble-encoded representation of items actually present in the environment, the responses of inferior temporal cortex face-selective neurons have been analyzed while a set of new faces were shown. It was found that some of the neurons studied in this way altered the relative degree to which they responded to the different members of the set of novel faces over the first few (1–2) presentations of the set (Rolls et al. 1989b). If in a different experiment a single novel face was introduced when the responses of a neuron to a set of familiar faces was being recorded, it was found that the responses to the set of familiar faces were not disrupted, while the responses to the novel face became stable within a few presentations. It is suggested that alteration of the tuning of individual neurons in this way results in a good discrimination over the population as a whole of the faces known to the monkey. This evidence is consistent with the categorization being performed by self-organizing competitive neuronal networks, as described below and elsewhere (Rolls 1989a; Rolls and Deco 2002; Rolls and Treves 1998; Rolls et al. 1989a).

Further evidence that these neurons can learn new representations very rapidly comes from an experiment in which binarized black-and-white images of faces that blended with the background were used. These images did not activate face-selective neurons. Full gray-scale images of the same photographs were then shown for ten 0.5-s presentations. It was found that in a number of cases, if the neuron happened to be responsive to that face, when the binarized version of the same face was shown next, the neurons responded to it (Tovee et al. 1996). This is a direct parallel to the same phenomenon that is observed psychophysically and provides dramatic evidence that these neurons are influenced by only a very few seconds (in this case 5 s) of experience with a visual stimulus. We have
shown a neural correlate of this effect using similar stimuli and a similar paradigm in a positron emission tomography (PET) neuroimaging study in humans, with a region showing an effect of the learning found for faces in the right temporal lobe and for objects in the left temporal lobe (Dolan et al. 1997).

Such rapid learning of representations of new objects appears to be a major type of learning in which the temporal cortical areas are involved. Ways in which this learning could occur are considered below. It is also the case that there is a much shorter term form of memory in which some of these neurons are involved, for whether a particular familiar visual stimulus (such as a face) has been seen recently, for some of these neurons respond differently to recently seen stimuli in short-term visual memory tasks (Baylis and Rolls 1987; Miller and Desimone 1994; Xiang and Brown 1998), and neurons in a more ventral cortical area respond during the delay in a short-term memory task (Miyashita 1993; Renart et al. 2000).

11. The Speed of Processing in the Temporal Cortical Visual Areas

Given that there is a whole sequence of visual cortical processing stages including V1, V2, V4, and the posterior inferior temporal cortex to reach the anterior temporal cortical areas, and that the response latencies of neurons in V1 are about 40 to 50 ms, and in the anterior inferior temporal cortical areas approximately 80 to 100 ms, each stage may need to perform processing for only 15 to 30 ms before it has performed sufficient processing to start influencing the next stage. Consistent with this, response latencies between V1 and the inferior temporal cortex increase from stage to stage (Thorpe and Imbert 1989). In a first approach to this issue, we measured the information available in short temporal epochs of the responses of temporal cortical face-selective neurons about which face had been seen. We found that if a period of the firing rate of 50 ms was taken, then this contained 84.4% of the information available in a much longer period of 400 ms about which of 4 faces had been seen. If the epoch was as little as 20 ms, the information was 65% of that available from the firing rate in the 400-ms period (Tovee et al. 1993). These high information yields were obtained with the short epochs taken near the start of the neuronal response, for example, in the poststimulus period of 100 to 120 ms. Moreover, we were able to show that the firing rate in short periods taken near the start of the neuronal response was highly correlated with the firing rate taken over the whole response period, so that the information available was stable over the whole response period of the neurons (Tovee et al. 1993). We were able to extend this finding to the case when a much larger stimulus set, of 20 faces, was used. Again, we found that the information available in short (e.g., 50-ms) epochs was a considerable proportion (e.g., 65%) of that available in a 400-ms-long firing rate analysis period (Tovee and Rolls 1995). These investigations thus showed that there was considerable information about which stimulus had been seen in short time epochs near the start.
of the response of temporal cortex neurons. Moreover, we have shown that the information is available in the number of action potentials from each neuron (which might be 1, 2, or 3) in these short time periods (a rate code), and not in the order in which the spikes arrive from different neurons (Rolls et al. 2006b).

The next approach has been to use a visual backward masking paradigm. In this paradigm, there is a brief presentation of a test stimulus that is rapidly followed (within 1–100ms) by the presentation of a second stimulus (the mask), which impairs or masks the perception of the test stimulus. It has been shown (Rolls and Tovee 1994) that when there is no mask inferior temporal cortex neurons respond to a 16-ms presentation of the test stimulus for 200 to 300ms, far longer than the presentation time. It is suggested that this reflects the operation of a short-term memory system implemented in cortical circuitry, the importance of which in learning invariant representations is considered below in Section 13. If the pattern mask followed the onset of the test face stimulus by 20ms (a stimulus onset asynchrony of 20ms), face-selective neurons in the inferior temporal cortex of macaques responded for a period of 20 to 30ms before their firing was interrupted by the mask (Rolls and Tovee 1994; Rolls et al. 1999). We went on to show that under these conditions (a test-mask stimulus onset asynchrony of 20ms), human observers looking at the same displays could just identify which of six faces was shown (Rolls et al. 1994).

These results provide evidence that a cortical area can perform the computation necessary for the recognition of a visual stimulus in 20 to 30ms (although it is true that for conscious perception, the firing needs to occur for 40–50ms; see Rolls 2003). This condition provides a fundamental constraint that must be accounted for in any theory of cortical computation. The results emphasize just how rapidly cortical circuitry can operate. Although this speed of operation does seem fast for a network with recurrent connections (mediated by, e.g., recurrent collateral or inhibitory interneurons), analyses of networks with analogue membranes that integrate inputs, and with spontaneously active neurons, do show that such networks can settle very rapidly (Rolls and Treves 1998; Treves 1993; Treves et al. 1996). This approach has been extended to multilayer networks such as those found in the visual system, and again very rapid propagation (in 40–50ms) of information through such a four-layer network with recurrent collaterals operating at each stage has been found (Panzeri et al. 2001). The computational approaches thus show that there is sufficient time for feedback processing using recurrent collaterals within each cortical stage during the fast cortical processing of visual inputs.

12. Different Neural Systems Are Specialized for Face Expression Decoding and for Face Recognition

It has been shown that some neurons respond to face identity and others to face expression (Hasselmo et al. 1989a). The neurons responsive to expression were found primarily in the cortex in the superior temporal sulcus, whereas the neurons
responsive to identity (described in the preceding sections) were found in the inferior temporal gyrus including areas TEa and TEM. Information about facial expression is of potential use in social interactions (Rolls 1984, 1986a,b, 1990, 1999b, 2005). Damage to this population may contribute to the deficits in social and emotional behavior that are part of the Kluver–Bucy syndrome produced by temporal lobe damage in monkeys (Leonard et al. 1985; Rolls 1981, 1984, 1986a,b, 1990, 1999b, 2005).

A further way in which some of these neurons in the cortex in the superior temporal sulcus may be involved in social interactions is that some of them respond to gestures, for example, to a face undergoing ventral flexion, as described above and by Perrett et al. (1985b). The interpretation of these neurons as being useful for social interactions is that in some cases these neurons respond not only to ventral head flexion, but also to the eyes lowering and the eyelids closing (Hasselmo et al. 1989a). These two movements (head lowering and eyelid lowering) often occur together when a monkey is breaking social contact with another. It is also important when decoding facial expression to retain some information about the head direction of the face stimulus being seen relative to the observer, for this is very important in determining whether a threat is being made in your direction. The presence of view-dependent, head and body gesture (Hasselmo et al. 1989b), and eye gaze (Perrett et al. 1985b), representations in some of these cortical regions where face expression is represented is consistent with this requirement. In contrast, the TE areas (more ventral, mainly in the macaque inferior temporal gyrus), in which neurons tuned to face identity (Hasselmo et al. 1989a) and with view-independent responses (Hasselmo et al. 1989b) are more likely to be found, may be more related to a view-invariant representation of identity. Of course, for appropriate social and emotional responses, both types of subsystem would be important, for it is necessary to know both the direction of a social gesture, and the identity of the individual, to make the correct social or emotional response.

13. Possible Computational Mechanisms in the Visual Cortex for Face and Object Recognition

The neurophysiological findings described above, and wider considerations on the possible computational properties of the cerebral cortex (Rolls 1989a,b, 1992a; Rolls and Treves 1998), lead to the following outline working hypotheses on object (including face) recognition by visual cortical mechanisms (Rolls and Deco 2002).

Cortical visual processing for object recognition is considered to be organized as a set of hierarchically connected cortical regions consisting at least of V1, V2, V4, posterior inferior temporal cortex (TEO), inferior temporal cortex (e.g., TE3, TEa and TEM), and anterior temporal cortical areas (e.g., TE2 and TE1), as shown in Fig. 10. There is convergence from each small part of a region to the succeeding region (or layer in the hierarchy) in such a way that the receptive
field sizes of neurons (e.g., 1° near the fovea in V1) become larger by a factor of approximately 2.5 with each succeeding stage (and the typical parafoveal receptive field sizes found would not be inconsistent with the calculated approximations of, e.g., 8° in V4, 20° in TEO, and 50° in inferior temporal cortex) (Boussaoud et al. 1991) (see Fig. 10). Such zones of convergence would overlap continuously with each other. This connectivity would be part of the architecture by which translation invariant representations are computed. Each layer is considered to act partly as a set of local self-organizing competitive neuronal networks with overlapping inputs. The region within which competition would be implemented would depend on the spatial properties of inhibitory interneurons and might operate over distances of 1 to 2 mm in the cortex. These competitive nets operate by a single set of forward inputs leading to (typically nonlinear, e.g., sigmoid) activation of output neurons; of competition between the output neurons mediated by a set of feedback inhibitory interneurons that receive from many of the principal (in the cortex, pyramidal) cells in the net and project back (via inhibitory interneurons) to many of the principal cells, which serves to decrease the firing rates of the less active neurons relative to the rates of the more active neurons; and then of synaptic modification by a modified Hebb rule, such that synapses to strongly activated output neurons from active input axons strengthen and those from inactive input axons weaken (Rolls and Deco 2002; Rolls and Treves 1998).
Translation, size, and view invariance could be computed in such a system by utilizing competitive learning operating across short time scales to detect regularities in inputs when real objects are transforming in the physical world (Rolls 1992a, 2000a; Rolls and Deco 2002; Wallis and Rolls 1997). The hypothesis is that because objects have continuous properties in space and time in the world, an object at one place on the retina might activate feature analyzers at the next stage of cortical processing, and when the object was translated to a nearby position, because this would occur in a short period (e.g., 0.5 s), the membrane of the postsynaptic neuron would still be in its “Hebb-modifiable” state (caused, for example, by calcium entry as a result of the voltage-dependent activation of NMDA receptors, or by continuing firing of the neuron implemented by recurrent collateral connections forming a short term memory), and the presynaptic afferents activated with the object in its new position would thus become strengthened on the still-activated postsynaptic neuron. It is suggested that the short temporal window (e.g., 0.5 s) of Hebb modifiability helps neurons to learn the statistics of objects moving in the physical world, and at the same time to form different representations of different feature combinations or objects, as these are physically discontinuous and present less regular correlations to the visual system. Földiák (1991) has proposed computing an average activation of the postsynaptic neuron to assist with translation invariance. I also suggest that other invariances, for example, size, spatial frequency, rotation, and view invariance, could be learned by similar mechanisms to those just described (Rolls 1992a). It is suggested that the process takes place at each stage of the multiple layer cortical processing hierarchy, so that invariances are learned first over small regions of space, and then over successively larger regions; this limits the size of the connection space within which correlations must be sought.

Increasing complexity of representations could also be built in such a multiple layer hierarchy by similar competitive learning mechanisms. To avoid the combinatorial explosion, it is proposed that low-order combinations of inputs would be what is learned by each neuron. Evidence consistent with this suggestion that neurons are responding to combinations of a few variables represented at the preceding stage of cortical processing is that some neurons in V2 and V4 respond to end-stopped lines, to tongues flanked by inhibitory subregions, or to combinations of colors (see references cited by Rolls 1991); in posterior inferior temporal cortex to stimuli that may require two or more simple features to be present (Tanaka et al. 1990); and in the temporal cortical face processing areas to images which require the presence of several features in a face (such as eyes, hair, and mouth) to respond (Perrett et al. 1982; Yamane et al. 1988). It is an important part of this suggestion that some local spatial information would be inherent in the features which were being combined (Elliffse et al. 2002). For example, cells might not respond to the combination of an edge and a small circle unless they were in the correct spatial relationship to each other. [This is in fact consistent with the data of Tanaka et al. (1990) and with our data on face neurons (Rolls et al. 1994), in that some faces neurons require the face features to be in the correct spatial configuration, and not jumbled.] The local spatial information in
the features being combined would ensure that the representation at the next level would contain some information about the (local) arrangement of features. Further low-order combinations of such neurons at the next stage would include sufficient local spatial information so that an arbitrary spatial arrangement of the same features would not activate the same neuron, and this is the proposed, and limited, solution that this mechanism would provide for the feature-binding problem (Elliffe et al. 2002).

It is suggested that view-independent representations could be formed by the same type of computation, operating to combine a limited set of views of objects. The plausibility of providing view-independent recognition of objects by combining a set of different views of objects has been proposed by a number of investigators (Koenderink and Van Doorn 1979; Logothetis et al. 1994; Poggio and Edelman 1990; Ullman 1996). Consistent with the suggestion that the view-independent representations are formed by combining view-dependent representations in the primate visual system is the fact that in the temporal cortical areas, neurons with view-independent representations of faces are present in the same cortical areas as neurons with view-dependent representations (from which the view-independent neurons could receive inputs) (Booth and Rolls 1998; Hasselmo et al. 1989b; Perrett et al. 1987). This solution to “object-based” representations is very different from that traditionally proposed for artificial vision systems, in which the coordinates in 3-D space of objects are stored in a database, and general-purpose algorithms operate on these to perform transforms such as translation, rotation, and scale change in 3-D space (Ullman 1996), or a linked list of feature parts is used (Marr 1982). In the present, much more limited but more biologically plausible scheme, the representation would be suitable for recognition of an object, and for linking associative memories to objects, but would be less good for making actions in 3-D space to particular parts of, or inside, objects, as the 3-D coordinates of each part of the object would not be explicitly available. It is therefore proposed that visual fixation is used to locate in foveal vision part of an object to which movements must be made, and that local disparity and other measurements of depth then provide sufficient information for the motor system to make actions relative to the small part of space in which a local, view-dependent, representation of depth would be provided (Ballard 1990; Rolls and Deco 2002).


To test and clarify the hypotheses just described about how the visual system may operate to learn invariant object recognition, we have performed simulations that implement many of the ideas just described and which are consistent with and based on much of the neurophysiology summarized here. The network simulated (VisNet) can perform object, including face, recognition in a biologically plausible way, and after training shows, for example, translation and view
invariance (Rolls and Deco 2002; Rolls and Milward 2000; Wallis and Rolls 1997; Wallis et al. 1993; Rolls 2008; Rolls and Stringer 2006b).

In the four-layer network, the successive layers correspond approximately to V2, V4, the posterior temporal cortex, and the anterior temporal cortex. The forward connections to a cell in one layer are derived from a topologically corresponding region of the preceding layer, using a Gaussian distribution of connection probabilities to determine the exact neurons in the preceding layer to which connections are made. This schema is constrained to preclude the repeated connection of any cells. Each cell receives 100 connections from the 32 × 32 cells of the preceding layer, with a 67% probability that a connection comes from within 4 cells of the distribution center. Figure 11 shows the general convergent network architecture used, and may be compared with Fig. 10. Within each layer, lateral inhibition between neurons has a radius of effect just greater than the radius of feed-forward convergence just defined. The lateral inhibition is simulated via a linear local contrast-enhancing filter active on each neuron. (Note that this differs from the global “winner-take-all” paradigm implemented by Földiák 1991.) The cell activation is then passed through a nonlinear cell activation function, which also produces contrast enhancement of the firing rates.

So that the results of the simulation might be made particularly relevant to understanding processing in higher cortical visual areas, the inputs to layer 1 come from a separate input layer that provides an approximation to the encoding found in visual area 1 (V1) of the primate visual system.

The synaptic learning rule used can be summarized as follows:

\[
\delta w_{ij} = k \ m_i \ r'_j \ \text{and} \\
m'_i = (1 - \eta) r'_i + \eta m'_i \ 
\]

Fig. 11. Hierarchical network structure of VisNet
where \( r'_j \) is the \( j^{th} \) input to the neuron, \( r_i \) is the output of the \( i^{th} \) neuron, \( w_{ij} \) is the \( j^{th} \) weight on the \( i^{th} \) neuron, \( \eta \) governs the relative influence of the trace and the new input (typically 0.4–0.6), and \( m^t_i \) represents the value of the \( i^{th} \) cell’s memory trace at time \( t \). In the simulation, the neuronal learning was bounded by normalization of each cell’s dendritic weight vector, as in standard competitive learning (Rolls and Treves 1998; Rolls and Deco 2002).

To train the network to produce a translation-invariant representation, one stimulus was placed successively in a sequence of nine positions across the input, then the next stimulus was placed successively in the same sequence of nine positions across the input, and so on through the set of stimuli. The idea was to enable the network to learn whatever was common at each stage of the network about a stimulus shown in different positions. To train on view invariance, different views of the same object were shown in succession, then different views of the next object were shown in succession, and so on. It has been shown that the network can learn to form neurons in the last layer of the network that respond to one of a set of simple shapes (such as “T, L, and +”) with translation invariance, or to a set of five to eight faces with translation, view, or size invariance, provided that the trace learning rule (and not a simple Hebb rule) is used (Figs. 12, 13) (Rolls and Deco 2002; Wallis and Rolls 1997).

There have been a number of investigations to explore this type of learning further. Rolls and Milward (2000) explored the operation of the trace learning rule used in the VisNet architecture, and showed that the rule operated especially well if the trace incorporated activity from previous presentations of the same object, but no contribution from the current neuronal activity being produced by the current exemplar of the object. The explanation for this is that this temporally asymmetrical rule (the presynaptic term from the current exemplar, and the trace

![Discrimination Factor vs Cell Rank](image)

**Fig. 12.** Comparison of VisNet network discrimination when trained with the trace learning rule, with a \( HEBB \) rule (no trace), and when not trained (random, \( RAND \)) on three stimuli, +, T, and L, at nine different locations. (After Wallis and Rolls 1997)
from the preceding exemplars) encourages neurons to respond to the current exemplar in the same way as they did to previous exemplars. It is of interest to consider whether intracellular processes related to LTP (long term potentiation) might implement an approximation of this rule, given that it is somewhat more powerful than the standard trace learning rule described above. Rolls and Stringer (2001) went on to show that part of the power of this type of trace rule can be related to gradient descent and temporal difference learning (Sutton and Barto 1998). Elliffe et al. (2002) examined the issue of spatial binding in this general class of hierarchical architecture studied originally by Fukushima (1980, 1989, 1991), and showed how by forming high spatial precision feature combination neurons early in processing, it is possible for later layers to maintain high precision for the relative spatial position of features within an object, yet achieve invariance for the spatial position of the whole object.

These results show that the proposed learning mechanism and neural architecture can produce cells with responses selective for stimulus identity with considerable position or view invariance (Rolls and Deco 2002). This ability to form invariant representations is an important property of the temporal cortical visual areas, for if a reinforcement association leading to an emotional or social response is learned to one view of a face, that learning will automatically generalize to other views of the face. This is a fundamental aspect of the way in which the brain is organized to allow this type of capability for emotional and social behavior (Rolls 1999b, 2005). Further developments include operation of the system in a cluttered environment (Stringer and Rolls 2000), generalization from trained to untrained views of objects (Stringer and Rolls 2002), a new training algorithm named continuous transformation learning (Stringer et al. 2006), and a unifying theory of how invariant representations of optic flow produced by rotating or looming objects could be produced in the brain (Stringer 2007; Rolls and Stringer 2006).

Fig. 13. Response profiles for two fourth-layer neurons in VisNet, discrimination factors 4.07 and 3.62, in the L, T, and + invariance learning experiment. (After Wallis and Rolls 1997)
15. Object Representation and Attention in Natural Scenes: A Computational Account

The results described in Section 9 and summarized in Fig. 8 show that the receptive fields of inferior temporal cortex neurons were large (77.6°) with a single stimulus in a blank background (top left) and were greatly reduced in size (to 22°) when presented in a complex natural scene (top right). The results also show that there was little difference in receptive field size or firing rate in the complex background when the effective stimulus was selected for action (bottom right) and when it was not (middle right) (Rolls et al. 2003a).

Trappenberg et al. (2002) have suggested what underlying mechanisms could account for these findings and simulated a model to test the ideas. The model utilizes an attractor network representing the inferior temporal visual cortex (implemented by the recurrent excitatory connections between inferior temporal cortex neurons) and a neural input layer with several retinotopically organized modules representing the visual scene in an earlier visual cortical area such as V4 (Fig. 14). The attractor network aspect of the model produces the property that receptive fields of IT neurons can be large in blank scenes by enabling a weak input in the periphery of the visual field to act as a retrieval cue for the object attractor. On the other hand, when the object is shown in a complex background, the object closest to the fovea tends to act as the retrieval cue for the attractor, because the fovea is given increased weight in activating the IT module because the magnitude of the input activity from objects at the fovea is greatest because of the cortical higher magnification factor of the fovea incorporated into the model. [The cortical magnification factor can be expressed as the number of millimeters of cortex representing 1° of visual field. The cortical magnification factor decreases rapidly with increasing eccentricity from the fovea (Cowey and Rolls 1975; Rolls and Cowey 1970).] This difference results in smaller receptive fields of IT neurons in complex scenes because the object tends to need to be close to the fovea to trigger the attractor into the state representing that object. (In other words, if the object is far from the fovea in a cluttered scene, then the object will not trigger neurons in IT that represent it, because neurons in IT are preferentially being activated by another object at the fovea.) This may be described as an attractor model in which the competition for which attractor state is retrieved is weighted toward objects at the fovea.

Attentional top-down object-based inputs can bias the competition implemented in this attractor model, but have relatively minor effects (in for example increasing receptive field size) when they are applied in a complex natural scene, because then as usual the stronger forward inputs dominate the states reached. In this network, the recurrent collateral connections may be thought of as implementing constraints between the different inputs present to help arrive at firing in the network that best meets the constraints. In this scenario, the preferential weighting of objects close to the fovea because of the increased magnification factor at the fovea is a useful principle in enabling the system to provide useful
The attentional object biasing effect is much more marked in a blank scene, or a scene with only two objects present at similar distances from the fovea, which are conditions in which attentional effects have frequently been examined. The results of the investigation (Trappenberg et al. 2002) thus suggest that attention may be a much more limited phenomenon in complex, natural scenes than in reduced displays with one or two objects present. The results also suggest...
suggest that the alternative principle, of providing strong weight to whatever is close to the fovea, is an important principle governing the operation of the inferior temporal visual cortex, and in general of the output of the ventral visual system in natural environments. This principle of operation is very important in interfacing the visual system to action systems, because the effective stimulus in making inferior temporal cortex neurons fire is in natural scenes usually on or close to the fovea. This means that the spatial coordinates of where the object is in the scene do not have to be represented in the inferior temporal visual cortex, nor passed from it to the action selection system, as the latter can assume that the object making IT neurons fire is close to the fovea in natural scenes (Rolls and Deco 2002; Rolls et al. 2003a).

There may of course be in addition a mechanism for object selection that takes into account the locus of covert attention when actions are made to locations not being looked at. However, the simulations described in this Section suggest that in any case covert attention is likely to be a much less significant influence on visual processing in natural scenes than in reduced scenes with one or two objects present.

Given these points, one might question why inferior temporal cortex neurons can have such large receptive fields, which show translation invariance (Rolls 2000a; Rolls et al. 2003a). At least part of the answer to this may be that inferior temporal cortex neurons must have the capability to have large receptive fields if they are to handle large objects (Rolls and Deco 2002). A V1 neuron, with its small receptive field, simply could not receive input from all the features necessary to define an object. On the other hand, inferior temporal cortex neurons may be able to adjust their size to approximately the size of objects, using in part the interactive attentional effects of bottom-up and top-down effects described elsewhere in this chapter.

The implementation of the simulations is described by Trappenberg et al. (2002), and some of the results obtained with the architecture (Fig. 14) follow. In one simulation, only one object was present in the visual scene in a plain background at different eccentricities from the fovea. As shown in Fig. 15A by the line labeled “simple background,” the receptive fields of the neurons were very large. The value of the object bias $k^{IT}p_{TA}$ was set to 0 in these simulations. Good object retrieval (indicated by large correlations) was found even when the object was far from the fovea, indicating large IT receptive fields with a blank background. The reason that any drop is seen in performance as a function of eccentricity is because some noise was present in the recall process. This finding demonstrates that the attractor dynamics can support translation invariant object recognition even though the translation invariant weight vectors between V4 and IT are explicitly mapped by a modulation factor derived from the cortical magnification factor.

In a second simulation, individual objects were placed at all possible locations in a natural and cluttered visual scene. The resulting correlations between the target pattern and the asymptotic IT state are shown in Fig. 15A with the line labeled “natural background.” Many objects in the visual scene are now
Fig. 15. Correlations as measured by the normalized dot product between the object vector used to train IT and the state of the IT network after settling into a stable state with a single object in the visual scene (blank background) or with other trained objects at all possible locations in the visual scene (natural background). There is no object bias included in the results shown in A, whereas an object bias is included in the results shown in B, with $k^{ITBIAS} = 0.7$ in the experiments with a natural background and $k^{ITBIAS} = 0.1$ in the experiments with a blank background.

Competing for recognition by the attractor network, and the objects around the foveal position are enhanced through the modulation factor derived from the cortical magnification factor; this results in a much smaller size of the receptive field of IT neurons when measured with objects in natural backgrounds.

In addition to this major effect of the background on the size of the receptive field, which parallels and, we suggest, may account for the physiological findings outlined above, there is also a dependence of the size of the receptive fields on the level of object bias provided to the IT network. Examples are shown in Fig. 15B where an object bias was used. The object bias biases the IT network toward the expected object with a strength determined by the value of $k^{ITBIAS}$ and has the effect of increasing the size of the receptive fields in both blank and natural backgrounds (compare Fig. 15B to Fig. 15A). This models the effect found neurophysiologically (Rolls et al. 2003a).

Some of the conclusions are as follows. When single objects are shown in a scene with a blank background, the attractor network helps neurons to respond to an object with large eccentricities of this object relative to the fovea. When the object is presented in a natural scene, other neurons in the inferior temporal cortex become activated by the other effective stimuli present in the visual field, and these forward inputs decrease the response of the network to the target stimulus by a competitive process. The results found fit well with the neurophysiological data, in that IT operates with almost complete translation invariance when there is only one object in the scene, and reduces the receptive field size of its neurons when the object is presented in a cluttered environment. The model described here provides an explanation of the responses of real IT neurons in natural scenes.
In natural scenes, the model is able to account for the neurophysiological data that the IT neuronal responses are larger when the object is close to the fovea, by virtue of fact that objects close to the fovea are weighted by the cortical magnification factor. The model accounts for the larger receptive field sizes from the fovea of IT neurons in natural backgrounds if the target is the object being selected compared to when it is not selected (Rolls et al. 2003a). The model accounts for this by an effect of top-down bias, which simply biases the neurons toward particular objects, compensating for their decreasing inputs produced by the decreasing magnification factor modulation with increasing distance from the fovea. Such object-based attention signals could originate in the prefrontal cortex and could provide the object bias for the inferotemporal cortex (Renart et al. 2000, 2001; Rolls and Deco 2002). Important properties of the architecture for obtaining the results just described are the high magnification factor at the fovea and the competition between the effects of different inputs, implemented in the foregoing simulation by the competition inherent in an attractor network.

We have also been able to obtain similar results in a hierarchical feed-forward network where each layer operates as a competitive network (Deco and Rolls 2004). This network thus captures many of the properties of our hierarchical model of invariant object recognition (Elliffe et al. 2002; Rolls 1992a; Rolls and Deco 2002; Rolls and Milward 2000; Rolls and Stringer 2001, 2006; Stringer and Rolls 2000, 2002; Stringer et al. 2006; Wallis and Rolls 1997), but incorporates in addition a foveal magnification factor and top-down projections with a dorsal visual stream so that attentional effects can be studied (Fig. 16).

Deco and Rolls (2004) trained the network described shown in Fig. 16 with two objects, and used the trace learning rule (Rolls and Milward 2000; Wallis and Rolls 1997) to achieve translation invariance. In a first experiment, we placed only one object on the retina at different distances from the fovea (i.e., different eccentricities relative to the fovea); this corresponds to the blank background condition. In a second experiment, we also placed the object at different eccentricities relative to the fovea, but on a cluttered natural background.

Figure 17 shows the average firing activity of the inferior temporal cortex neuron specific for the test object as a function of the position of the object on the retina relative to the fovea (eccentricity). In both cases (solid line for blank background, dashed line for cluttered background) relatively large receptive fields are observed, because of the translation invariance obtained with the trace learning rule and the competition mechanisms implemented within each layer of the ventral stream. (The receptive field size is defined as the width of the receptive field at the point where there is a half-maximal response.) However, when the object was in a blank background (solid line in Fig. 17), larger receptive fields were observed. The decrease in neuronal response as a function of distance from the fovea is mainly the effect of the magnification factor implemented in V1. On the other hand, when the object was in a complex cluttered background, the effective size of the receptive field of the same inferior temporal cortex neuron shrinks because of competitive effects between the object features and the background features in each layer of the ventral stream. In particular, the global
Fig. 16. Cortical architecture for hierarchical and attention-based visual perception. The system is essentially composed of five modules structured such that they resemble the two known main visual paths of the mammalian visual cortex. Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream leads ventrally through V2–V4 and IT (inferior temporal visual cortex) and is mainly concerned with object recognition. The occipitoparietal stream leads dorsally into PP (posterior parietal complex) and is responsible for maintaining a spatial map of an object’s location. The solid lines with arrows between levels show the forward connections, and the dashed lines show the top-down back-projections. Short-term memory systems in the prefrontal cortex (PF46) apply top-down attentional bias to the object or spatial processing streams. (After Deco and Rolls 2004)
character of the competition expressed in the inferior temporal cortex module (caused by the large receptive fields and the local character of the inhibition, in our simulations, between the two object specific pools) is the main cause of the reduction of the receptive fields in the complex scene.

Deco and Rolls (2004) also studied the influence of object-based attentional top-down bias on the effective size of an inferior temporal cortex neuron for the case of an object in a blank or a cluttered background. To do this, we repeated the two simulations but now considered a non-zero top-down bias coming from prefrontal area 46v and impinging on the inferior temporal cortex neuron specific for the object tested (Fig. 18). We plot the average firing activity normalized to the maximum value to compare the neuronal activity as a function of the eccentricity. When no attentional object bias is introduced (a), shrinkage of the receptive field size is observed in the complex background (dashed line). When attentional object bias is introduced (b), the shrinkage of the receptive field because of the complex background is slightly reduced (dashed line). Rolls et al. (2003a) also found that in natural scenes that the effect of object-based attention on the response properties of inferior temporal cortex neurons was relatively

![Fig. 17.](image-url)
small. They found only a small difference in the receptive field size or firing rate in the complex background when the effective stimulus was selected for action versus when it was not. In the framework of the model (Deco and Rolls 2004), the reduction of the shrinkage of the receptive field is caused by the biasing of the competition in the inferior temporal cortex layer in favor of the specific IT neuron tested, so that it shows more translation invariance (i.e., a slightly larger receptive field). The increase of the receptive field of an IT neuron, although small, produced by the external top-down attentional bias offers a mechanism for facilitation of the search for specific objects in complex natural scenes.


Visual attention exerts top-down influences on the processing of sensory information in the visual cortex, and therefore is intrinsically associated with intercortical neural interactions. Thus, elucidating the neural basis of visual attention is an excellent paradigm for understanding the basic mechanisms of intercortical neurodynamics. Recent cognitive neuroscience developments allow a more direct
study of the neural mechanisms underlying attention in humans and primates. In particular, the work of Chelazzi et al. (1993) has led to a promising account of attention termed the biased competition hypothesis (Desimone and Duncan 1995; Reynolds and Desimone 1999). According to this hypothesis, attentional selection operates in parallel by biasing an underlying competitive interaction between multiple stimuli in the visual field toward one stimulus or another, so that behaviorally relevant stimuli are processed in the cortex while irrelevant stimuli are filtered out. Thus, attending to a stimulus at a particular location or with a particular feature biases the underlying neural competition in a certain brain area in favor of neurons that respond to the location, or the features, of the attended stimulus.

Neurodynamical models for biased competition have been proposed and successfully applied in the context of attention and working memory. In the context of attention, Usher and Niebur (1996) introduced an early model of biased competition. Deco and Zihl (2001) extended Usher and Niebur’s model to simulate the psychophysics of visual attention by visual search experiments in humans. Their neurodynamical formulation is a large-scale hierarchical model of the visual cortex whose global dynamics is based on biased competition mechanisms at the neural level. Attention then appears as an emergent effect related to the dynamical evolution of the whole network. This large-scale formulation has been able to simulate and explain in a unifying framework visual attention in a variety of tasks and at different cognitive neuroscience experimental measurement levels (Deco and Rolls 2005a), namely, single cells (Deco and Lee 2002; Rolls and Deco 2002), fMRI (Corchs and Deco 2002), psychophysics (Deco and Rolls 2005a; Rolls and Deco 2002), and neuropsychology (Deco and Rolls 2002). In the context of working memory, further developments (Deco and Rolls 2003) managed to model in a unifying form attentional and memory effects in the prefrontal cortex, integrating single-cell and fMRI data, and different paradigms in the framework of biased competition.

In particular, Deco and Rolls (2005c) extended previous concepts of the role of biased competition in attention by providing the first analysis at the integrate-and-fire neuronal level, which allows the neuronal nonlinearities in the system to be explicitly modeled, to investigate realistically the processes that underlie the apparent gain modulation effect of top-down attentional control. In the integrate-and-fire model, the competition is implemented realistically by the effects of the excitatory neurons on the inhibitory neurons and their return inhibitory synaptic connections; this was also the first integrate-and-fire analysis of top-down attentional influences in vision that explicitly models the interaction of several different brain areas. Part of the originality of the model is that in the form in which it can account for attentional effects in V2 and V4 in the paradigms of Reynolds et al. (1999) in the context of biased competition, the model with the same parameters effectively makes predictions which show that the “contrast gain” effects in MT (Martinez-Trujillo and Treue 2002) can be accounted for by the same model. These detailed and quantitative analyses of neuronal dynamical systems are an important step toward understanding the operation of complex
processes such as top-down attention, which necessarily involve the interaction of several brain areas. They are being extended to provide neurally plausible models of decision making (Deco and Rolls 2003, 2005b, 2006).

In relation to representation in the brain, the impact of these findings is that they show details of the mechanisms by which representations can be modulated by attention, and moreover can account for many phenomena in attention using models in which the firing rate of neurons is represented and in which stimulus-dependent synchrony is not involved.

17. A Representation of Faces in the Amygdala

Outputs from the temporal cortical visual areas reach the amygdala and the orbitofrontal cortex, and evidence is accumulating that these brain areas are involved in social and emotional responses to faces (Rolls 1990, 1999b, 2000b, 2005; Rolls and Deco 2002). For example, lesions of the amygdala in monkeys disrupt social and emotional responses to faces, and we have identified a population of neurons with face-selective responses in the primate amygdala (Leonard et al. 1985), some of which may respond to facial and body gestures (Brothers et al. 1990). In humans, bilateral dysfunction of the amygdala can impair face expression identification, although primarily of fear (Adolphs et al. 1995; Adolphs et al. 2002), so that the impairment seems much less severe than that produced by orbitofrontal cortex damage.

18. A Representation of Faces in the Orbitofrontal Cortex

Rolls et al. (2006a) have found a number of face-responsive neurons in the orbitofrontal cortex, and they are also present in adjacent prefrontal cortical areas (Wilson et al. 1993). The orbitofrontal cortex face-responsive neurons, first observed by Thorpe et al. (1983), then by Rolls et al. (2006a), tend to respond with longer latencies than temporal lobe neurons (140–200 ms typically, compared with 80–100 ms); they also convey information about which face is being seen, by having different responses to different faces (Fig. 19), and are typically rather harder to activate strongly than temporal cortical face-selective neurons, in that many of them respond much better to real faces than to 2-D images of faces on a video monitor (Rolls and Baylis 1986). Some of the orbitofrontal cortex face-selective neurons are responsive to face gesture or movement. The findings are consistent with the likelihood that these neurons are activated via the inputs from the temporal cortical visual areas in which face-selective neurons are found. The significance of the neurons is likely to be related to the fact that faces convey information that is important in social reinforcement, both by conveying face expression (Hasselmo et al. 1989a), which can indicate reinforcement, and by encoding information about which individual is present, also important in evaluating and utilizing reinforcing inputs in social situations (Rolls et al. 2006a).
We have also been able to obtain evidence that non-reward used as a signal to reverse behavioral choice is represented in the human orbitofrontal cortex (for background, see Rolls 2005). Kringelbach and Rolls (2003) used the faces of two different people, and if one face was selected then that face smiled, and if the other was selected, the face showed an angry expression. After good performance was acquired, there were repeated reversals of the visual discrimination task. Kringelbach and Rolls (2003) found that activation of a lateral part of the orbitofrontal cortex in the fMRI study was produced on the error trials, that is,
when the human chose a face and did not obtain the expected reward (Figs. 20, 21). Control tasks showed that the response was related to the error, and the mismatch between what was expected and what was obtained, in that just showing an angry face expression did not selectively activate this part of the lateral orbitofrontal cortex. An interesting aspect of this study that makes it relevant to human social behavior is that the conditioned stimuli were faces of particular individuals and the unconditioned stimuli were face expressions. Moreover, the study reveals that the human orbitofrontal cortex is very sensitive to social feedback when it must be used to change behavior (Kringelbach and Rolls 2003, 2004; Rolls 2005).

To investigate the possible significance of face-related inputs to the orbitofrontal cortex visual neurons described above, we also tested the responses to faces of patients with orbitofrontal cortex damage. We included tests of face (and also voice) expression decoding, because these are ways in which the reinforcing quality of individuals is often indicated. Impairments in the identification of facial
and vocal emotional expression were demonstrated in a group of patients with ventral frontal lobe damage who had socially inappropriate behavior (Hornak et al. 1996; Rolls 1999a). The expression identification impairments could occur independently of perceptual impairments in facial recognition, voice discrimination, or environmental sound recognition. The face and voice expression problems did not necessarily occur together in the same patients, providing an
indication of separate processing. Poor performance on both expression tests was correlated with the degree of alteration of emotional experience reported by the patients. There was also a strong positive correlation between the degree of altered emotional experience and the severity of the behavioural problems (e.g., disinhibition) found in these patients. A comparison group of patients with brain damage outside the ventral frontal lobe region, without these behavioral problems, was unimpaired on the face expression identification test, was significantly less impaired at vocal expression identification, and reported little subjective emotional change (Hornak et al. 1996; Rolls 1999a). To obtain clear evidence that the changes in face and voice expression identification, emotional behavior, and subjective emotional state were related to orbitofrontal cortex damage itself, and not to damage to surrounding areas, which is present in many closed head injury patients, we performed further assessments in patients with circumscribed lesions made surgically in the course of treatment (Hornak et al. 2003). This study also enabled us to determine whether there was functional specialization within the orbitofrontal cortex, and whether damage to nearby and connected areas (such as the anterior cingulate cortex) in which some of the patients had lesions could produce similar effects. We found that some patients with bilateral lesions of the orbitofrontal cortex had deficits in voice and face expression identification, and the group had impairments in social behavior and significant changes in their subjective emotional state (Hornak et al. 2003). The same group of patients had deficits on a probabilistic monetary reward reversal task, indicating that they have difficulty not only in representing reinforcers such as face expression, but also in using reinforcers (such as monetary reward) to influence behavior (Hornak et al. 2004). Some patients with unilateral damage restricted to the orbitofrontal cortex also had deficits in voice expression identification, and the group did not have significant changes in social behavior, or in their subjective emotional state. Patients with unilateral lesions of the anteroventral part of the anterior cingulate cortex and/or medial prefrontal cortex area BA9 were in some cases impaired on voice and face expression identification, had some change in social behavior, and had significant changes in their subjective emotional state. Patients with dorsolateral prefrontal cortex lesions or with medial lesions outside the anterior cingulate cortex and medial prefrontal BA9 areas were unimpaired on any of these measures of emotion. In all cases in which voice expression identification was impaired, there were no deficits in control tests of the discrimination of unfamiliar voices and the recognition of environmental sounds.

These results (Hornak et al. 2003) thus confirm that damage restricted to the orbitofrontal cortex can produce impairments in face and voice expression identification, which may be primary reinforcers. The system is sensitive, in that even patients with unilateral orbitofrontal cortex lesions may be impaired. The impairment is not a generic impairment of the ability to recognize any emotions in others, in that frequently voice but not face expression identification was impaired, and vice versa. This implies some functional specialization for visual versus auditory emotion-related processing in the human orbitofrontal cortex. The results
also show that the changes in social behavior can be produced by damage restricted to the orbitofrontal cortex. The patients were particularly likely to be impaired on emotion recognition (they were less likely to notice when others were sad, or happy, or disgusted); on emotional empathy (they were less likely to comfort those who are sad, or afraid, or to feel happy for others who are happy); on interpersonal relationships (not caring what others think, and not being close to his/her family); and were less likely to cooperate with others; were impatiente and impulsive; and had difficulty in making and keeping close relationships. The results also show that changes in subjective emotional state (including frequently sadness, anger, and happiness) can be produced by damage restricted to the orbitofrontal cortex (Hornak et al. 2003). In addition, the patients with bilateral orbitofrontal cortex lesions were impaired on the probabilistic reversal learning task (Hornak et al. 2004). The findings overall thus make clear the types of deficit found in humans with orbitofrontal cortex damage, and can be directly related to underlying fundamental processes in which the orbitofrontal cortex is involved (see Rolls 2005), including decoding and representing primary reinforcers (including face expression), being sensitive to changes in reinforcers, and rapidly readjusting behaviour to stimuli when the reinforcers available change.

The results (Hornak et al. 2003) also extend these investigations to the anterior cingulate cortex (including some of medial prefrontal cortex area BA9) by showing that lesions in these regions can produce voice and/or face expression identification deficits and marked changes in subjective emotional state.

It is of interest that the range of face expressions for which identification is impaired by orbitofrontal cortex damage (Hornak et al. 1996; Hornak et al. 2003; Rolls 1999a) is more extensive than the impairment in identifying primarily fear face expressions produced by amygdala damage in humans (Adolphs et al. 2002; Calder et al. 1996) (for review, see Rolls 2005). In addition, the deficits in emotional and social behavior described above that are produced by orbitofrontal cortex damage in humans seem to be more pronounced than changes in emotional behavior produced by amygdala damage in humans, although deficits in autonomic conditioning can be demonstrated (Phelps 2004). This result suggests that in humans and other primates the orbitofrontal cortex may become more important than the amygdala in emotion, and possible reasons for this, including the more powerful architecture for rapid learning and reversal that may be facilitated by the functional architecture of the neocortex with its highly developed recurrent collateral connections, which may help to support short-term memory attractor states, are considered by Rolls (2005; 2008).

19. Conclusions

Neurophysiological investigations of the inferior temporal cortex are revealing at least part of the way in which neuronal firing encodes information about faces and objects and are showing that one representation implements several types of invariance. The representation found has clear utility for the receiving
networks. These neurophysiological findings are stimulating the development of computational neuronal network models, which suggest that part of the process involves the operation of a modified Hebb learning rule with a short-term memory trace to help the system learn invariances from the statistical properties of the inputs it receives. Neurons in the inferior temporal cortex, which encode the identity of faces and have considerable invariance and a sparse distributed representation, are ideal as an input to stimulus–reinforcer association learning mechanisms in the orbitofrontal cortex and amygdala that enable appropriate emotional and social responses to be made to different individuals. The neurons in the cortex in the superior temporal sulcus, which respond to face expression, or for other neurons to eye gaze, or for others to head movement, encode reinforcement-related information that is important in making the correct emotional and social responses to a face. Neurons of both these main types are also found in the orbitofrontal cortex (Rolls et al. 2006a) and are important in human social and emotional behavior, which is changed after damage to the orbitofrontal cortex. A more comprehensive description of the reinforcement-related signals and processing in brain regions such as the orbitofrontal cortex that are important in emotional and social behavior, and how these depend on inputs from the temporal cortex visual areas, is provided in Emotion Explained (Rolls 2005).

Acknowledgments. The author has worked on some of the investigations described here with N. Aggelopoulos, P. Azzopardi, G.C. Baylis, H. Critchley, G. Deco, P. Földiák, L. Franco, M. Hasselmo, J. Hornak, M. Kringelbach, C.M. Leonard, T.J. Milward, D.I. Perrett, S.M. Stringer, M.J. Tovee, T. Trappenberg, A. Treves, and G. Wallis, and their collaboration is sincerely acknowledged. Different parts of the research described were supported by the Medical Research Council, PG8513790; by a Human Frontier Science Program grant; by an EC Human Capital and Mobility grant; by the MRC Oxford Interdisciplinary Research Centre in Cognitive Neuroscience; and by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience.

References
3. Visual Object and Face Representations


Fukushima K (1989) Analysis of the process of visual pattern recognition by the neocognitron. Neural Networks 2:413–420
Rolls ET (1992a) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond B 335:11–21
Rolls ET (1999c) Spatial view cells and the representation of place in the primate hippocampus. Hippocampus 9:467–480
Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 65:38–48
Rolls ET, Tovee MJ (1995b) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the visual field. Exp Brain Res 103:409–420


Rolls ET, Franco L, Aggelopoulos NC, Perez JM (2006b) Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. Vision Res 46:4193–4205


4 Representation of Objects and Scenes in Visual Working Memory in Human Brain

JUN SAIKI

1. Explicit Human Scene Understanding

We live in the world with multiple objects, many of which move independently. To appropriately interact with objects, it is insufficient to form and maintain a static representation of multiple objects. Instead, the ability to actively update object representations is indispensable. Moreover, given that the early-level vision system decomposes visual objects into different features such as color, motion, shape, etc., the formation of object representation must require integration of visual features. Therefore, to achieve the goal of interaction with objects, our cognitive system needs to integrate features, maintain objects, and actively update representations in a coordinated fashion. Visual working memory undoubtedly plays a crucial role in this complicated process, and great progress has been made, through various studies, toward understanding how visual memory operates.

This chapter first reviews studies regarding three major components of scene understanding: feature binding, object maintenance, and dynamic updating. This review reveals that although the neural correlates of component processes themselves are fairly well understood, understanding of coordination mechanisms remains nebulous. The second half of the chapter describes a few studies that probe those mechanisms of coordination, provides a hypothetical scheme for coordination, and discusses directions of future research.

1.1. Structure of Human Scene Understanding

Figure 1 shows a schematic illustration of functional architecture of human scene understanding. Scene understanding can be separated into two aspects: rapid perception of scene gist and layout and explicit scene understanding. These facets are analogous to “vision at a glance” and “vision with scrutiny,” respectively, as
referred to by reverse hierarchy theory (Hochstein and Ahissar 2002), and to the coherence theory terms “setting” and “object” (Rensink 2000). The global structure of a scene is rapidly perceived by bottom-up and parallel mechanisms, which in turn guide locus of focused attention necessary for explicit scene understanding. Although perceiving gist and layout is an important aspect of scene understanding, and some interesting empirical data and theoretical ideas addressing rapid perception have been developed (Thorpe et al. 1996; Torralba and Oliva 2003), there are still relatively few investigations regarding the brain mechanisms underlying these phenomena (Epstein and Kanwisher 1998). Given the relative paucity of data regarding this topic, the rapid understanding of scene gist is not discussed here in any further detail.

The term “explicit scene understanding” refers to understanding an event composed of several multidimensional objects. Each object contains multiple visual features and is dynamic; this corresponds to an everyday situation where one is interacting with one or several objects, in the context of additional distractors. Air traffic control, driving an automobile or flying an airplane, and monitoring a complex system such as a nuclear power plant are a few examples. Successful scene understanding requires selection of appropriate objects, formation of bound object representations, and maintenance and updating of multiple object representations.

There are, then, three major components of explicit scene understanding (see Fig. 1): feature binding, maintenance of objects, and updating of objects. The feature-binding mechanism integrates visual features of an object into a coherent
representation. Object representations are presumably maintained for at least a short period of time for future use, or further processing. Moreover, if objects in the external world are moving or changing their properties, object representations need to be updated. Although there are other components in scene understanding, such as response selection and inhibition of irrelevant objects, this chapter focuses on the aforementioned components, which are closely linked to visual working memory.

2. Components of Explicit Scene Understanding

2.1. Maintenance of Object Representations

Holding a mental representation of visual stimuli for a short period of time is the core mechanism of visual working memory, and a large number of studies exploring this mechanism have been carried out. This line of research originated from single-unit recording studies of macaque monkeys (Funahashi et al. 1989; Fuster and Alexander 1971), and expanded, in the 1990s, to human functional brain imaging studies.

The basic experimental paradigm in human functional brain imaging studies is the “delayed matching to sample” (DMS) task used in single-unit recording studies. In the DMS task, an observer is shown one stimulus for a brief period (encoding phase), followed by a prescribed blank period (maintenance phase). Afterward, the test stimulus is presented, and the observer judges whether the test and study stimuli are the same (retrieval phase). Various studies have observed sustained delay activities in parietal, temporal, and frontal regions. Recently, event-related functional magnetic resonance imaging (fMRI) design allows us to decompose brain activity into encoding, maintenance, and retrieval phases. A brief review of these recent studies follows.

First, human brain imaging studies revealed results analogous to single-unit recording data, showing that anterior and posterior activities are different regarding robustness to interference (Miller et al. 1996). Miller et al. found that sustained activity during the maintenance period in the prefrontal cortex endures the presentation of an interfering stimulus, whereas activity in the parietal and temporal cortices is greatly reduced by interference. Similar results were reported by Sakai et al. (2002) using fMRI. Furthermore, Courtney et al. (1997) showed that sustained activity in frontal areas shows different patterns, suggesting different functional roles played by various areas in prefrontal cortex. For example, the posterior midfrontal region showed the greatest transient activity and the smallest sustained activity. By contrast, the anterior midfrontal region showed the smallest transient and the greatest sustained activities, whereas the inferior frontal region showed an intermediate pattern of activity.

The finding of differential sustained activity in prefrontal cortex motivates investigations of functional roles of prefrontal cortex, in turn giving rise to dissenting views about how the prefrontal cortex is functionally organized. One
view claims that functional organization is based on maintenance of information content (Haxby et al. 2000); namely, that the superior frontal cortex (in particular, the superior frontal sulcus, or SFS), maintains spatial working memory (Courtney et al. 1998) while the ventrolateral frontal cortex maintains object working memory (Courtney et al. 1997). Alternatively, another view proposes that prefrontal functions are organized based on mental processing (D’Esposito et al. 1998; Owen et al. 1996; Petrides 1994). Proponents of this argument contend that the ventrolateral prefrontal cortex is chiefly responsible for maintenance of memoranda, whereas the dorsolateral and superior prefrontal areas are associated with manipulation of memory representations. This argument is far from settled, but it should be noted that functional organization by information content and by processing are not necessarily mutually exclusive. Curtis and D’Esposito (2003) proposed that whereas maintenance of different types of information occurs in specific frontal premotor areas and posterior regions, including the parietal and temporal cortices, the role of the dorsolateral prefrontal cortex (DLPFC) is processing based and domain independent. Recently, Mohr et al. (2006) reported evidence for this proposal, identifying areas that showed content-specific and processing-independent activation (SFS for spatial tasks; and inferior frontal sulcus, or IFS, for color tasks), and others showing processing-specific and content-independent activation (anterior middle frontal gyrus and inferior frontal junction).

Another important issue concerning visual working memory maintenance is the identification of those regions sensitive to memory load, which presumably play a crucial role in maintaining memory representations. In fMRI studies on visual working memory load, various load manipulations have been employed, such as duration of blank period (Jha and McCarthy 2000; Leung et al. 2002), the number of objects sequentially presented (Druzgal and D’Esposito 2003; Linden et al. 2003), the n-back task (Braver et al. 1997; Cohen et al. 1997), and the number of objects simultaneously presented (Jha and McCarthy 2000; Song and Jiang 2006; Todd and Marois 2004; Xu and Chun 2006). Pessoa et al. (2002), in addressing this issue, investigated the relationship between fMRI signals and task performance. Observed load-sensitive regions are not consistent among studies that use different manipulations, or even among those using similar manipulations. The focus of this chapter is those studies that investigate the effect of memory load using simultaneous presentation of multiple objects, given that this manipulation is the most relevant to the issue of explicit scene understanding.

Todd and Marois (2004) investigated the effect of memory load using a change-detection task used by Luck and Vogel (1997), whereby briefly presented multiple objects were compared with probe objects, which were presented following a blank period, to judge the presence of object change. Behavioral data suggest that the memory load effect is such that accurate performance holds up to about three objects, beyond which accuracy declined significantly. Todd and Marois showed that intraparietal and intracerebelum sulci (IPS/IOS) revealed brain activation proportional to memory load; that is, activation increased as the set size
increased up to three objects but did not increase beyond that. Beyond the IPS/IOS, significant brain activation was observed in the anterior cingulate and ventral-occipital regions, neither of which showed load-dependent activity. Using event-related potentials, Vogel and Machizawa (2004) obtained quite a similar pattern of results.

Subsequent studies using a similar experimental paradigm revealed additional properties of maintenance-related brain activity. Xu and Chun (2006) showed that whereas the inferior IPS maintains fixed number of objects at different spatial locations, the superior IPS and lateral occipital cortex maintain objects depending on their complexity. Song and Jiang (2006) manipulated both the number of objects and task-relevant features. They showed that prefrontal, occipitotemporal, and posterior parietal regions are sensitive to the number of objects, to featural difference, and to both object and feature, respectively.

2.2. Dynamic Updating of Object Representations

Generally speaking, processing and control are key aspects of working memory and have been extensively studied. This line of research is concerned primarily with response-related processing and control (Rowe et al. 2000), and not necessarily with visual scene understanding. One well-known example of this type of inquiry is the n-back task. Another type of updating, less related to response selection, is multiple object tracking (MOT).

In a MOT task, observers are shown several homogeneous objects (usually discs) randomly placed on the display. Before a trial begins, certain discs flash, indicating that they are targets. Then, all discs move in a random direction, similar to Brownian motion, for several seconds. At the end of the trial, observers are asked to identify the initially presented target discs. This task requires observers to continuously update target locations to successfully track them. Thus, task performance reflects dynamic updating of multiple object representations. Behavioral studies revealed that people can track up to four or five targets successfully (Pylyshin and Storm 1988). A few studies incorporating the MOT have shown that activities of frontoparietal network reflect the monitoring load (Culham et al. 1998; Culham et al. 2001; Jovicich et al. 2001).

It should be noted that activities reflecting manipulation of representations as observed by the n-back task and other similar tasks, and, as observed in the MOT, are quite different: the former is in the DLPFC, and the latter is in the frontoparietal network. This difference might be attributable specifically to type of manipulation. In the n-back task, the manipulation of memory representation involves response-related property (i.e., updating an item to be compared), whereas the multiple object tracking involves spatiotemporal manipulation of objects, which is not directly related to response selection. Alternatively, differences between the n-back task and MOT might reflect the type of representations to be manipulated. The n-back task requires updating of multiple memory representations, whereas MOT requires updating of memory representation with current perceptual representation.
Several studies on MOT have revealed more details about the functional characteristics of certain brain regions. The first study using MOT (Culham et al. 1998) found that various regions in the frontoparietal network are activated. Subsequently, Jovicich et al. (2001) searched for regions sensitive to attentional load by applying linear trend analysis, subsequently reporting primarily posterior parietal activations, including the intraparietal sulcus (IPS) and superior parietal lobule (SPL), thus suggesting that the posterior parietal area is important for dynamic tracking of object representations. Alternatively, Culham et al. (2001) used a series of regression analyses to distinguish between areas sensitive to attentional load and those sensitive only to task setting. According to Culham et al. (2001), the IPS and superior frontal sulcus (SFS) are load sensitive, whereas the SPL and frontal eye field (FEF) are only sensitive to task. Although there is some inconsistency between Jovicich et al. (2001) and Culham et al. (2001), particularly in terms of the load sensitivity in the SPL, Culham et al. appear to be more consistent with other studies in a broader context. For example, in studying parietal activation, Todd and Marois (2004) showed load-sensitive activation only in IPS in a visual working memory task; in examining frontal activation, Courtney et al. (1998) showed activation in the SFS with a spatial working memory task.

2.3. Feature Binding to Form Object Representation

In the literature, the problem of feature binding has been discussed largely in the context of visual object representation. The essence of this problem is the question of how the brain represents multiple objects simultaneously. Although this problem appears to be trivial at first glance, its importance becomes clear when one considers characteristics of visual information processing in the brain. Early visual areas mainly decompose visual input into separate features; thus, the higher-level vision system needs to integrate these features. At the same time, it is known that higher-level visual areas, such as inferior temporal (IT) cortex, have quite large receptive fields, which usually cover multiple objects. The question raised, then, is how the neurons in the IT can tell the correct combination of visual features of multiple objects. There are several hypotheses for the solution of the binding problem, including binding by selective focused attention (Treisman 1988), binding by temporal synchrony of neural activities (Singer and Gray 1995), and spatial coding of IT neurons (Rousselet et al. 2004). In-depth perusal of these hypotheses is beyond the scope of the present discussion, beyond noting that there is still no answer. However, several empirical studies, regarding feature binding in both perception and memory, merit attention.

Feature binding has been discussed mainly in the context of visual perception. In principle, however, feature binding is also associated with the memory system for multiple objects. Recently, binding in visual working memory has also been investigated. Both in perception and in visual working memory, research findings using functional brain imaging techniques are quite equivocal.

In studies of perceptual feature binding, the role of attention in binding has been investigated. Treisman (1988) postulated that spatial attention to an object...
location is essential to bind features of the object. According to this view, the parietal lobe, particularly the posterior parietal cortex, is expected to play an important role. In a neuropsychological study, Freedman-Hill et al. (1995) found that patients with parietal lesions showed severe impairment in tasks requiring feature binding, such as the conjunction search task. On the other hand, Ashbridge et al. (1999) failed to demonstrate the involvement of parietal cortex in feature binding, per se. Results of functional brain imaging studies are also equivocal. Elliot and Dolan (1997), and Rees et al. (1997) failed to show parietal lobe activity, using positron emission tomography (PET) and fMRI experiments, respectively. On the other hand, Shafritz et al. (2002), conducting fMRI experiments, reported activity in the parietal cortex associated with spatial attention, a finding consistent with Treisman’s hypothesis. Corbetta et al. (1995) and Woiciulik and Kanwisher (1999) also found activation in the parietal cortex.

In studies of feature binding in memory, Munk et al. (2002) used a delayed discrimination task to compare “what,” “where,” and conjunction conditions. They found sustained delay period activities in the parietal and frontal areas and a nonadditive relationship among spatial memory, nonspatial memory, and their conjunction. Areas known to be active for spatial memory, namely, the IPL, SPL, and superior frontal gyrus (SFG), showed highest activity in the “where” condition, as did areas for nonspatial memory, such as the inferior frontal gyrus (IFG), in the “what” condition, respectively. Both spatial and nonspatial memory areas showed the second highest activity in the conjunction condition. This pattern of results suggests that solution of the conjunction task is not carried out by the addition of signals but by a partial recruitment of component processes.

In the study reported by Munk et al. (2002), an area rather uniquely showing activity in the conjunction task was the medial superior frontal cortex, or supplementary motor area (SMA). The SMA is known to be related to task set control (Rushworth et al. 2004); thus, this SMA activity may reflect processing operations, rather than maintenance per se. Another prefrontal area known to be active in feature binding in memory is the anterior prefrontal cortex, Broadmann’s area (BA) 10 (Mitchell et al. 2000; Prabhakaran et al. 2000). Prabhakaran et al. (2000) compared bound and separate conditions in a working memory task of letter and location. In the bound condition, letters were presented in the locations to be remembered, whereas in the separate condition, letters were presented centrally, and locations were indicated by parentheses placeholders. The task is to judge whether a single probe is composed of a letter and a location presented in the sample display, regardless of their combination. They reported significantly higher anterior prefrontal activation (BA 9, −10, and −46) in the bound condition than in the separate condition, and found that various parietal and temporal activations were significantly lower in the bound condition. Mitchell et al. (2000) also reported BA 10 activity in the conjunction task.

Another study of visual working memory showing significant BA 10 activity is that described by Ranganath et al. (2004), who compared the delayed paired associate (DPA) task with delayed matching to sample (DMS) task, and showed that BA 10 and the posterior hippocampus had significantly higher activity in the DPA task than in the DMS task.
Taken together, many studies on visual working memory that show anterior prefrontal activities, including BA 10 and anterior SMA, share some common characteristics. Although the tasks just reviewed are rather different, they all concern combining two relatively independent pieces of information. Ramnani and Owen (2004) recently reviewed studies of working memory that have shown activity in BA 10, and proposed the hypothesis that BA 10 integrates the outcomes of two or more separate cognitive operations.

3. Coordination of Component Processes

Thus far, we have reviewed investigations of visual working memory that have examined the component processes of scene understanding: maintenance of memory representations, updating of object representation and feature binding to form object representations. Because each component itself poses so many important theoretical issues, most of the previous work has focused on a single mechanism. However, to fully understand cognitive mechanisms underlying explicit scene understanding, we need to investigate the coordination of these components.

3.1. Two Possible Mechanisms of Binding and Maintenance

With regard to the relationship between binding and memory maintenance, two possible general mechanisms of formation, maintenance, and manipulation of scene representations may be considered. In the first, feature binding precedes working memory maintenance, implying that working memory can maintain and update object representations with bound features. Alternatively, a working memory system may primarily maintain representations without bound features, and only selected subsets of memory representations receive binding operation, when necessary. The former and the latter views may be termed “binding memory” and “feature memory”, respectively. The binding memory view postulates that perceptual binding operations make object representation stable, and that multiple object representations are stored within visual working memory. The feature memory view assumes that perceptual feature binding is quite fragile, and that its products are not maintained by default.

The studies reviewed here are equivocal regarding this issue. A few of them appear to support the binding memory view. For example, Todd and Marois (2004) claim that feature-bound object representations are mainly stored in the IPS, showing that brain activity in this region is proportional to memory load. Similarly, Culham et al. (2001) show that the IPS and SFS reveal activity sensitive to the load in the multiple object tracking task. Combined with the finding by Shafritz et al. (2002), showing that perceptual feature binding is carried out in the SPL and the anterior IPS, multiple object representations may be formed, maintained, and updated in the parietal cortex, particularly in the IPS. Figure 2 shows the functional roles of various brain areas, according to the binding
memory view. In this model, the main functions of explicit scene understanding, such as binding, maintenance, and updating, are executed in the parietal area, and the functional role of the prefrontal cortex is mainly task-related control of memory processing. By contrast, some studies locate the critical role of feature binding in memory in the prefrontal cortex, particularly the anterior prefrontal area (Mitchell et al. 2000; Mohr et al. 2006; Prabhakaran et al. 2000; Ranganath et al., 2004). Taking into account of Ramnani and Owen’s hypothesis that the functional role of BA10 is the integration of two or more relatively independent cognitive operations, and of the fact that BA10 has connections predominantly with other prefrontal regions (Ramnani and Owen 2004), the feature binding in BA10 suggest that the formation of object representation by binding feature representations occurs in the prefrontal cortex, not in the parietal cortex, which is consistent with the feature memory view. Figure 3 is a schematic illustration of functional roles, according to the feature memory view.

To evaluate the plausibility of the binding and the feature memory views, one important, but largely unexplored, empirical issue is whether multiple object tracking and change-detection tasks really reflect feature-bound representations of objects. If so, based on the findings of neural correlates of these tasks, the binding memory view is the more plausible account. By contrast, if these tasks are not reflecting the function of feature-bound object representations, the feature memory view is more plausible. The following section discusses the author’s work using a new experimental paradigm to address this issue.
3.2. Multiple Object Permanence Tracking (MOPT)

To examine whether multiple object tracking and change-detection tasks reflect updating and maintenance of feature-bound object representations, we devised an experimental paradigm called “multiple object permanence tracking,” or MOPT (Saiki 2003a,b; Saiki and Miyatsuji 2005, 2007). In the MOPT task, four to six discs of different colors or shapes are placed at equal eccentricity, then rotated behind a windmill-shaped occluder (Fig. 4). In the middle of the rotation sequence, the colors of two discs may be switched during an occlusion. The task of the observer was to detect whether a color switch has occurred. The switch detection task requires memory for feature binding, because memory for colors or locations alone cannot detect a switch between two colors (see Wheeler and Treisman 2002 for discussion of a similar idea). The speed of disc rotation was manipulated by relative motion of discs and occluder to investigate the effect of motion in a parametric manner. The factor of motion speed manipulates memory load in general and spatiotemporal manipulation cost in particular (Mohr et al. 2006). Conceptually, the MOPT task is a mixture of MOT task and a type of change-detection task that investigates binding memory (Wheeler and Treisman 2002). Thus, if the MOT and change-detection tasks reflect the function of feature-bound object representations, MOPT task performance would be expected to be analogous to MOT and change-detection performance. In general, however, switch detection was markedly impaired as motion speed increased (Saiki 2003a,b).

Severe impairment of task performance with object motion suggests that feature binding is not maintained during tracking. The next question, then, is
whether feature binding can be maintained with stationary objects. Subsequent studies employed a multidimensional version of MOPT in combination with the type-identification paradigm (Fig. 4b). Each object was defined by the conjunction of color and shape, and observers were required to identify the type of feature switch (color only, shape only, color and shape, or none), instead of the presence of change alone (Saiki and Miyatsuji 2005, 2007). The multidimensional MOPT, with type identification, showed that estimated capacity is less than two objects, suggesting that memory for feature binding is severely limited even when objects are stationary. At the very least, one may reasonably conclude that feature binding information in memory, if any, is not accessible for explicit judgment about feature combination.

3.3. **Neural Substrates of Multiple Object Permanence Tracking**

Behavioral data derived from multiple object permanence tracking experiments suggest that change-detection and multiple object tracking tasks do not reflect
maintenance and manipulation of feature-bound object representations. If this is the case, brain activity during the MOPT task should not be confined to those regions activated by the change-detection and MOT tasks, such as the IPS, SPL, and SFS. To test this prediction, Imaruoka et al. (2005) conducted an fMRI experiment using the MOPT task. The experiment was an epoch-related design, with object movement (moving and stationary) and change-detection (test and control) as experimental factors. Each object (disc) was defined by color, and observers were required to detect a color switch by pressing a button in the test conditions. In the control condition, four discs with an identical color were used, and observers pressed a key when the color turned to gray. A whole-brain analysis revealed that the MOPT task induced brain activity largely extending from the posterior parietal area to the anterior frontal area in both object-moving and object-stationary conditions (Fig. 5). In general, both stationary and moving conditions showed activity in the same areas, with stronger activity in the moving condition. Clearly, the activated areas are quite different from those activated in both change-detection tasks (Todd and Marois 2004) and MOT tasks (Culham et al. 2001; Jovicich et al. 2001). Note that this difference cannot be attributed to task difficulty, because the behavioral performance in the MOPT experiment was comparable to the MOT (Jovicich et al. 2001) and the large set size conditions in the change-detection tasks (Todd and Marois, 2004). By contrast, the pattern of brain activities was quite consistent with studies showing BA10 activity, suggesting that feature binding and/or spatiotemporal manipulation of object information in the MOPT requires anterior prefrontal activity.

Regions of interest (ROI) analyses further revealed different functional roles of various areas, which are largely consistent with previous studies. The posterior IPS showed load-sensitive activity, whereas the anterior IPS showed task-dependent activity. This pattern is consistent with that described by Xu and Chun (2006), using a change-detection task, and Culham et al. (2001), using the MOT task. In regard to prefrontal activity, the effects of task (switch detection

![Fig. 5. Brain activation patterns in MOPT task. Panels a–c and d–f represent results from object-moving (MOVE) and object-stationary (STAY) conditions, respectively. (Reproduced from Imaruoka et al. 2005, with permission)
vs. control) and load (stationary vs. moving) are independent in the anterior PFC and the IFG. This pattern of activity was qualitatively different from that in posterior regions, suggesting that the frontoparietal network and the anterior/inferior PFC play different functional roles in performing the MOPT task.

However, the epoch-related design in Imaruoka et al. (2005) prevents us from further analysis, given that activities in the maintenance and change-detection periods were confounded. Recently, Takahama et al. (2005) conducted a follow-up to the Imaruoka study, using an event-related design and modifying the experimental paradigm in several points. First, with extensive practice trials before the fMRI sessions, the accuracy of behavior data was set to be quite high, and there was no substantial difference in task difficulty across conditions. Second, visual stimuli in the maintenance period of control conditions now were exactly the same as those in the experimental conditions, so that differences in brain activity could be said to reflect top-down control of memory maintenance. Third, activities in the maintenance and change-detection periods were now decomposed by event-related design. Although still preliminary, the results were largely consistent with those of Imaruoka et al., but with some new findings. Regarding activity in posterior areas, maintenance activity showed the pattern similar to that of Imaruoka et al. (2005). By contrast, the event-related design revealed further qualifications about anterior prefrontal activity. The effect of load (moving versus stationary) was observed during the maintenance period such that the moving condition showed stronger activation in both control and test conditions, without task effect. The effect of task (binding versus control conditions) was instead observed during the change detection. These data suggest that manipulation of memory representation during the maintenance period increases anterior prefrontal activity, whereas binding of color and location affects the memory retrieval and matching process. Manipulation-related activity in the anterior prefrontal area is consistent with Mohr et al. (2006), and binding-related activity at the time of change detection appears to imply that the anterior PFC is not the storage place of feature binding, but, rather, is involved in carrying out judgments based on a change in feature binding. Because all the reports of binding-related activity in the anterior PFC used epoch-related design (Imaruoka et al. 2005; Mitchell et al. 2000; Prabhakaran et al. 2000), this interpretation is consistent with those previous studies.

Taken together, studies using the MOPT paradigm support the feature memory view, in the sense that maintenance and updating of feature-bound object representations cannot be carried out autonomously within the frontoparietal network. Updating of color-location binding requires activity of the anterior PFC, suggesting that the conventional MOT task is unlikely to be actually investigating the tracking of feature-bound object representations. Although memory for color-location binding cannot fully function within the frontoparietal network, there are some alternative explanations regarding the functional architecture of binding memory. First, as suggested by Wheeler and Treisman (2002), visual working memory is inherently feature based, and memory judgment on feature conjunction, such as switch detection, is carried out by combining states of two
feature-based memory systems. Alternatively, memory representations in the inferior IPS may be feature-bound, but the frontoparietal network cannot detect a change in feature combination autonomously. In other words, representations in the inferior IPS may be implicitly feature-bound, but explicit detection of change requires prefrontal activity. This implicit binding idea is consistent with some behavioral data (Kahneman et al. 1993) but still fails to explain why the explicit detection of single feature change from bound representations can be carried out without prefrontal activity. Further investigations are necessary to resolve this issue.

4. Toward a Full Understanding of Human Scene Understanding

This chapter focused on explicit scene understanding, defined as maintaining dynamic states of multiple objects having multiple features, and has discussed the associated underlying brain mechanisms. Component processes of explicit scene understanding, namely, feature binding, maintenance, and updating, have been studied extensively. By contrast, there are few studies regarding the interaction of the component processes. We have discussed in detail one such interaction, the relationship between feature binding and memory process. Globally, explicit scene understanding involves a large network of brain regions, including the anterior PFC and the frontoparietal network. Currently, it is too early to exactly detail the functional roles of different regions in explicit scene understanding. Further studies, using various new tasks and data analytical techniques, such as functional connectivity analyses, are necessary.

Beyond the issue of explicit scene understanding, we need to investigate the functional relationships of explicit scene understanding with other mechanisms. Given that explicit scene understanding appears to be more capacity limited than previously thought, it is crucial to understand its relationship to selective attention and to rapid perception of layout and gist.

References

4. Explicit Scene Understanding

Part II
Motor Image and Body Schema
5
Action Representation in the Cerebral Cortex and the Cognitive Functions of the Motor System

LEONARDO FOGASSI$^{1,2}$

1. Introduction

How is sensory information processed to become perception? What is the relationship between sensory and motor areas? Has the motor cortex some role in the construction of perception? Until 20 years ago, the answers to these questions consisted of a kind of hierarchical model, according to which perception was considered as the final outcome of high-order sensory elaboration. The final percept was then fed to the motor cortex, which used it for building those motor programs suitable for executing movements in response to sensory stimuli. According to this model, the motor cortex comes after sensory cortices and does not have any influence on the elaboration of perception.

The neurophysiological, neuroanatomical, and brain imaging data of the past 20 years, together with clinical findings, point to another view of cortical functioning quite different from the previous one. First, neuroanatomical data demonstrated that parietal and frontal cortex are *reciprocally* connected, and that these connections are the substrate for several sensorimotor transformations involving different effectors, such as eyes, arm, and hand (Rizzolatti et al. 1998). Second, neurophysiological studies showed that the motor cortex is not simply involved in motor programming and execution but has a major role in coding goal-directed actions (Rizzolatti et al. 2004). Third, brain imaging experiments showed that sensory input and cognitive activities strongly activate motor areas. Fourth, clinical observation showed that there can be dissociation between perception for recognition and perception for action (Goodale and Westwood 2004).

In this chapter, I review data, in particular from neurophysiology and brain imaging, that enable us to build a new conceptual framework for cortical functioning and for the definition of the cortical motor system. Most of the chapter is dedicated to the mirror neurons system and its functional implications.

---

$^1$Department of Neuroscience, Section Physiology, Via Volturno 39, I-43100 Parma, Italy
$^2$Department of Psychology, B.go Carissimi 10, I-43100 Parma, Italy
2. A New View of the Motor System

The first systematic studies that tried to build a map of movements in the agranular frontal cortex, by using the technique of electrical stimulation applied to the cortical surface, were those of Woolsey and coworkers in monkeys (1951) and of Penfield and coworkers in humans (1937). The studies of both groups brought the discovery of two complete motor representations in the agranular frontal cortex, the primary motor cortex (MI) on the lateral brain convexity and the supplementary motor area (SMA, MII) on the mesial aspect of each hemisphere. MI contains a complete motor map involving axial, proximal, and distal joints. In MII, there is a larger representation of axial and proximal movements, whereas distal and facial movements are much less elicited. The data of these and subsequent more detailed stimulation experiments raised a debate on whether the motor cortex represents muscles or movements around joints and on whether simple or complex movements are elicited by electrical stimulation. In parallel, since the beginning of single-neuron recordings in awake animals, many researchers have looked for possible single-neuron or population coding, at the level of the motor cortex, of movement parameters such as force, direction, and amplitude (Evarts 1968; Georgopoulos et al. 1982; Kalaska et al. 1989).

More recent data allow us to look at the motor cortex from another perspective. Although until the end of 1970s the motor cortex had been subdivided, in particular from a functional point of view, into a few gross regions, architectonic and histochemical studies allowed recognizing a higher number of subdivisions of agranular frontal cortex. In this chapter, in describing these subdivisions, I adopt the classification of Matelli and coworkers (1985, 1991). Most importantly, these anatomical subdivisions appear to correspond to an at least equal number of functional areas. Furthermore, each anatomical subdivision of agranular frontal cortex has a main bidirectional connection with an area of parietal cortex (see below). The mosaic organization of the motor cortex and its rich connectivity with the parietal cortex are suggestive of a sensorimotor system more complex than previously thought. However, the main finding that challenges the previous views on the motor cortex is the discovery that motor cortical neurons encode the goal of motor acts instead of movement parameters. Although this aspect has been investigated more in depth in the ventral premotor cortex (areas F4 and F5), it appears to be a core function of the cortical motor system. Only neurons of area F1 (primary motor cortex) seem to code single movements. To have a better comprehension of what coding the goal means, the properties of areas F4 and F5 are described below, with more emphasis given to area F5, deeply investigated in the recent decade.

2.1. Coding Motor Acts in Ventral Premotor Area F4

Area F4 constitutes the caudal part of ventral premotor cortex (Matelli et al. 1985; Fig. 1). It is connected with posterior parietal areas VIP (ventral intraparietal area) and PFG (Luppino et al. 1999; Rozzi et al. 2005). Its electrical micro-
stimulation evokes neck, arm, and face movements, and very often the evoked movements consist of a combination of two or three body parts (Fogassi et al. 1996a; Gentilucci et al. 1988). Most F4 neurons are activated by somatosensory and visual stimuli (Fogassi et al. 1996b; Gentilucci et al. 1988).

The most interesting category of F4 neurons is represented by bimodal, somatosensory, and visual, neurons. These neurons have tactile receptive fields (RFs) on the face, arm, or trunk and respond to three-dimensional (3-D) objects moved toward the tactile RF (Fogassi et al. 1996a,b; Graziano et al. 1994, 1997). The visual RFs of F4 bimodal neurons have two important features: (a) they are three dimensional, that is, they begin to activate only when the stimulus is close...
to the monkey, within reaching distance (they have been called peripersonal RFs); and (b) the visual RF does not shift when the eyes move, that is, it is coded in somatocentric coordinates. What is the functional meaning of these visual responses? F4 motor neurons code axial and proximal motor acts toward 3-D objects in space such as orienting toward or avoiding an object, reaching an object, or bringing it to the mouth (Fogassi et al. 1996b; Gentilucci et al. 1988).

The motor and sensory properties of F4 neurons would allow us to conclude, according to the classic concept of a serial processing of sensory information, that the visual input reaching F4 is instrumental in providing spatial sensory information for the programming and execution of the different types of motor acts represented in this area. However, a different explanation can be proposed, namely, that the visual input evokes always a “motor representation” of the stimulus. In other words, every time an object is introduced in the monkey’s peripersonal space, a pragmatic representation of the associated motor act is immediately retrieved. This representation, depending on the context, can be either executed or remain at the level of representation. In both cases, there is the activation of an internal “motor knowledge.”

The “motor representation” interpretation of the multisensory responses within F4 posits that the function of the different sensory inputs would be that of retrieving different types of motor representations, according to the body part on which, or near which, the tactile, visual, or auditory stimuli are applied. A visual stimulus near the monkey could evoke an avoiding action, a mouth approaching action, or a reaching action directed toward the body. Similarly, a far peripersonal stimulus could elicit an orienting action or a reaching action away from the body. If the context is suitable, these actions will be executed; conversely, they remain as potential actions, enabling space perception. Thus, the motor representations of F4 neurons can accomplish two tasks. First, they can play a major role in the sensorimotor transformation for facial, axial, and proximal actions. Second, they can code space directly, in motor terms, using the same coordinate system of the effector acting in that portion of peripersonal space.

2.2. Coding Motor Acts in Ventral Premotor Area F5

Area F5 is located in the rostral part of the ventral premotor cortex, bordered caudally by area F4 and rostrally by the inferior limb of the arcuate sulcus, which separates it from the prefrontal cortex (see Fig. 1). Microstimulation of area F5 evokes hand and mouth movements (Ferrari et al. 2003a; Gentilucci et al. 1988). F5 motor neurons discharge when a monkey performs goal-related hand and mouth motor acts (see references in Rizzolatti et al. 2004) such as grasping, manipulating, holding, tearing, and digging out objects. Some of them discharge in relation to the abstract goal of the motor act, for example, when the monkey grasps food with the hand or with the mouth, thus coding the goal of grasping independently of the effector used for executing that motor act. There are also F5 motor neurons coding more specific motor acts such as precision grip, finger prehension, or power grip. Thus, F5 motor neurons constitute a “vocabulary” of
motor acts. Note that this vocabulary is not only used for overt motor act execution, but constitutes a sort of motor knowledge, because to each goal-related motor act is associated also the representation of the consequences of that act.

The presence of a vocabulary of motor acts in premotor cortex is very important because it can be accessed by several sensory inputs, in particular through parietofrontal connections. This access enables individuals to match sensory information with the motor knowledge, this processing giving rise to a first-person understanding of the external world. The neuronal examples reported here better explain this latter concept.

Beyond purely motor neurons, which constitute the majority of all F5 neurons, area F5 contains also two categories of visuomotor neurons, the motor properties of which are indistinguishable from those of motor neurons. However, they have peculiar visual responses. The first category of visuomotor neurons is formed by neurons responding to the observation of objects of particular size, shape, and orientation (“canonical” neurons). The size or the shape of the observed object for which they are selective correlates very well with the specific type of motor act they code (Rizzolatti et al. 2004). A recent investigation of the properties of these neurons (Murata et al. 1997; Raos et al. 2006) demonstrated that the visual response of canonical neurons is coded in motor terms, that is, it can be better interpreted as a kind of motor representation of the object. In other words, while the pictorial object description necessary for recognizing and discriminating objects is represented in the inferior temporal cortex, the motor vocabulary of the act contained in the premotor cortex enables coding a pragmatic object description.

3. Mirror Neurons

3.1. General Properties of Mirror Neurons

The second category of F5 visuomotor neurons constitutes “mirror” neurons. These neurons discharge both when a monkey performs a hand motor act and when it observes another individual (a human being or another monkey) performing a similar hand motor act in front of it (Fig. 2). Differently from canonical neurons, they do not discharge to the simple observation of food or other interesting objects. They also do not discharge, or discharge much less, when the observed action is mimicked without the target object. The response is generally weaker or absent when the effective action is executed by using a tool instead of the hand. Thus, the only visual stimulus effective in triggering a mirror neuron response is a hand–object interaction (Gallese et al. 1996; Rizzolatti et al. 1996a), suggesting that mirror neurons have an important role in recognizing the goal of motor acts performed by others.

The observed motor acts most effective in activating mirror neurons correspond to those contained in F5 motor vocabulary, for example, grasping, manipulating, tearing, and holding objects. Grasping is the most represented motor act.
Among neurons responding to the observation of grasping there are some that are very specific, coding the type of observed grip. Thus, mirror neurons can present different types of visual selectivity: selectivity for the observed motor act and selectivity for the way in which the observed act is performed.

### 3.2. Mouth Mirror Neurons

The first discovered category of mirror neurons included only neurons responding to observation of hand motor acts. Later we found another category of mirror neurons, activated by the observation and execution of mouth motor acts (“mouth mirror neurons”; Ferrari et al. 2003a). Most mouth mirror neurons respond to
observation of ingestive motor acts such as biting, tearing with the teeth, sucking, or licking. Their specific properties are very similar to those described above. A smaller but significant percent of mouth mirror neurons respond specifically to the observation of mouth communicative gestures belonging to the monkey repertoire, such as lips smacking, lips protrusion, or tongue protrusion. Mouth mirror neurons of this subcategory do not respond, or respond very weakly, to the observation of ingestive motor acts.

3.3. Motor Properties of F5 Mirror Neurons: The Matching System

Mirror neurons activate during execution of goal-related motor acts. The great majority of mirror neurons present a good congruence between visual and motor responses. On the basis of the congruence, mirror neurons were assigned to two major categories: “strictly congruent” neurons, constituting about 30% of all F5 mirror neurons, in which observed and executed motor acts coincide, and “broadly congruent” neurons (60% of all mirror neurons), in which the coded observed and the coded executed motor act are similar but not identical. Note that the congruence has been observed also for mouth mirror neurons.

The congruence between the visual and the motor response of mirror neurons is probably their most important characteristic, because it suggests that every time a motor act is observed, there is an automatic activation of the motor circuits of the observer encoding a similar motor act; this suggests the existence of an observation–execution matching system at the single neuron level. This matching system constitutes the basis for understanding motor acts done by others, starting from an internal motor knowledge.

What is the reason for the presence of strictly and broadly congruent mirror neurons? Strictly congruent mirror neurons are very likely involved in a detailed analysis of the observed motor act. These neurons could be suitable for imitation (see below). In contrast, broadly congruent neurons can generalize across different ways of achieving the same goal, thus probably enabling a more abstract type of coding. This coding could be used not only for understanding others’ motor acts but also for appropriately reacting within a social environment.

3.4. Mirror Neurons and Understanding of the Goal

The finding that mirror neurons respond only to observation of hand–object interaction and not to observation of the same motor act, when mimed, suggests that mirror neurons may play a crucial role in understanding the goal of another individual motor act. One could argue, however, that the visual response of these neurons may represent a simple visual description of a biological stimulus, and this could be not enough for attributing the function of goal understanding to mirror neurons.
This possible criticism was addressed in an experiment in which monkeys had to observe partially hidden motor acts (Umiltà et al. 2001). In a first condition of the experiment, the monkey observed a fully visible hand motor act (usually grasping) directed toward an object (“Full vision” condition). In a second condition it observed the same action, but its final part, in which the hand closed around the object and took possession of it, occurred behind an opaque screen (“Hidden” condition). In this latter condition the monkey knew that an object was present behind the screen. There were also two control conditions (“Mimicking in full vision” and “Hidden mimicking”) in which the same motor act was mimed without the object, in full vision and behind the screen, respectively. Note that the only difference between “Hidden” and “Hidden mimicking” conditions consisted of the fact that in the latter the monkey knew that there was no object behind the screen.

The results showed that the majority of mirror neurons tested in this study discharged not only in “Full vision” condition, but also during observation of hand motor acts in the “Hidden” condition (Fig. 3). However, when the hidden motor act was mimicked, there was no activation. These findings indicate that mirror neurons respond also when the monkey knows which will be the final outcome of the motor act, without seeing its crucial final part. Another suggestion in this line comes from a few mirror neurons recorded in this experiment. In these neurons, discharging during the “Hidden” condition, the discharge was anticipated in comparison to the “Full vision” condition, suggesting the ability of these neurons of predicting the final goal of the motor act, basing on its initial visual cues. Thus, the mirror neurons mechanism appears to have an important role in the understanding of the goal of a motor act (Gallese et al. 1996; Rizzolatti et al. 1996a, 2004).

Another demonstration of the involvement of mirror neurons in understanding the goal of a motor act was given by the discovery in area F5 of mirror neurons that become active when monkeys not only observe, but also hear the sound of, a motor act (“audio-visual” mirror neurons; Kohler et al. 2002). The responses of these neurons are specific for the type of seen and heard motor acts. For example, they respond to peanut breaking when the motor act is only observed, only heard, or both heard and observed, and do not respond to the vision and sound of another motor act, or to unspecific sounds. Most noisy biological motor acts are coded by these peculiar mirror neurons. Note that often the neuron discharge to the simultaneous presentation of both the visual and the acoustic inputs is higher than the response to either of the inputs when presented alone. These data show that the acoustic input has access to the vocabulary of motor acts contained in the motor cortex of a listener, thus accessing to the content of the motor act. Note that the capacity of representing the content of a motor act independently of the modality used to access this content is typical of language.

A characteristic of all the described categories of mirror neurons is that the sensory response is congruent with the motor act belonging to the internal motor repertoire (the “internal motor knowledge”) of the observing/listening individual. However, in daily life it is common observation that we can understand
Fig. 3. Example of a mirror neuron responding during observation of grasping in both “Full vision” and “Hidden” condition. A, C Observation of goal-directed or mimed grasping, respectively, in full vision. B, D Observation of goal-directed or mimed grasping, respectively, in the hidden condition. In every panel, from top to bottom, the rasters and histogram of the neuron response and the schematic drawing of the experimenter motor act are shown. The gray frame in conditions B and D indicates the screen interposed between the monkey and the experimenter’s hand in the two hidden conditions. The asterisk indicates the location of a stationary marker that was attached at the level of the crossing point where the experimenter’s hand disappeared behind the screen in the hidden conditions. The line above each raster represents the kinematics of the experimenter’s hand movement: the downward deflection of the line means that the hand is approaching the stationary marker (the minimum corresponding to the moment in which the hand is closest to the marker). Histograms bin width = 20 ms. (Modified from Umiltà et al. 2001)
motor acts also when they do not belong directly to our repertoire (for example, a bird flying) or when they are done by an artifact (for example, a robot arm). The original observations on mirror neurons indicated that observation of motor acts performed with tools did not produce any response. However, in a recent experiment (Ferrari et al. 2005a), we recorded from neurons of the most lateral part of area F5 and found a subcategory of mirror neurons that discharged when the monkey observed motor acts performed by an experimenter with a tool (a stick or a pair of pliers), named tool-responding mirror neurons. It is noteworthy that this response was stronger than that obtained when the monkey observed a similar motor act made with a biological effector (the hand or the mouth). For example, one neuron could respond to the observation of picking up a piece of food with a stick, but not when the same act was done with the hand. In some extreme cases the observation of the act made with the tool elicited an excitatory response, while that of the same act made with the hand elicited an inhibition of the neuron discharge. Motorically, these neurons responded when the monkey executed hand and mouth motor acts.

Although at first glance the visual and the motor responses to these neurons seem incongruent, the congruency can be found in the fact that they share the same general goal, that is, taking possession of an object and modifying its state. To explain the presence of this peculiar type of mirror neurons, we hypothesized that after a relatively long visual exposure to tool actions (as occurred during long-lasting experiments) a visual association between the hand and the tool is created, so that the tool becomes a kind of prolongation of the hand (for a similar interpretation, see Maravita and Iriki 2004). Independently of the mechanism used for their formation, the importance of tool-responding mirror neurons can reside in the fact that they enable extending the goal understanding capacity to motor acts that do not strictly correspond to internal motor representations.

4. Formation of the Mirror Neuron Matching System

One of the most interesting questions about mirror neurons is how they could have formed. This is not just an issue for the curiosity of neurophysiologists, but it is at the core of the matching system. Although to understand the ontogeny of this matching system we would need to study it in newborn monkeys, neuroanatomical and single-neuron data of adult monkeys help to at least propose a model of the circuit in which this matching might occur. First, which is the origin of the visual input to mirror neurons? Perrett and coworkers (Perrett et al. 1989, 1990) showed that in the anterior part of the superior temporal sulcus (STSa) there are neurons responding to the sight of biological motion, that is, to the observation of other individuals’ movements, performed with different body parts, such as head or legs. One subcategory of these neurons is specific for the observation of hand–object interactions, such as picking up, tearing, and manipulating. In contrast to mirror neurons, they do not discharge when the monkey executes the same hand motor acts. However, they could be the source of the visual information about biological actions to be used for matching motor act observation with
motor act execution. STSa is not directly connected with area F5, but is anatomically linked with the prefrontal cortex and the inferior parietal lobule (IPL) (Cavada and Goldman-Rakic 1989; Rozzi et al. 2005; Seltzer and Pandya 1994). Thus, indirect connections could provide the sought matching. The pathway passing through the prefrontal cortex seems different because there is only a weak connection between area F5 and prefrontal area 46 (Matelli et al. 1986). The “parietal” pathway appears to be a better candidate, because the IPL, and in particular areas PF and PFG, are very strongly connected with ventral premotor cortex, including the sector where mirror neurons have been found (Cavada and Goldman-Rakic 1989; Matelli et al. 1986; Matsumura and Kubota 1979; Muakkassa and Strick 1979; Petrides and Pandya 1984; Rozzi et al. 2005).

Generally speaking, the inferior parietal lobule has been classically considered as an associative cortex, in which somatosensory and visual input are combined to allow, for example, space perception. However, old and recent work assigned to this lobule also motor functions and the property of visuomotor integration (Fogassi and Luppino 2005).

Ferrari et al. (2003) have explored the functional properties of the IPL and found that a large percent of neurons respond during execution of motor acts. The representation of these acts in the lobule follows a gross somatotopy, with the mouth motor field located rostrally, the hand and arm motor fields in an intermediate position, and, finally, the eye field located caudally. Very interestingly, the responses of most IPL motor neurons, as in ventral premotor cortex, code the goal of the motor acts and not specific movement parameters.

Beyond purely motor neurons, in IPL there are also somatosensory, visual, and sensorimotor neurons. Among visuomotor neurons we found, very likely in cytoarchitectonic area PFG (Gregoriou et al. 2006), neurons responding to the sight of hand–object interactions and during the execution of the same motor acts (Fogassi et al. 1998, 2005; Gallese et al. 2002). These neurons were called parietal mirror neurons (Fogassi et al. 2005; Gallese et al. 2002).

Parietal mirror neurons, similarly to F5 mirror neurons, respond to the observation of several types of single or combined motor acts. Grasping and bimanual interactions are the most represented among the effective observed motor acts. Parietal mirror neurons respond during the execution of hand, mouth, or hand and mouth actions, and most of them present either a strict or a broad congruence between observed and executed action. Finally, in IPL there are neurons that respond only to the observation of hand actions but are devoid of motor properties. The presence of these neurons is very important, because they are similar to those described in STSa. Thus, the demonstration that in the inferior parietal lobule, connected with both the cortex of the superior temporal sulcus and the ventral premotor cortex, there are mirror neurons confirms that the parietal cortex could be the link necessary for the matching between the motor representation of motor acts and their purely visual description.

Summing up, the presence of mirror neurons in both parietal and ventral premotor cortex strongly suggests that the functional mirror neuron system is formed through the anatomical temporo-parieto-premotor circuit. As is described below, a similar circuit is also present in humans.
5. The Mirror Neuron System in Humans

The first evidence of the existence of a mirror system in humans came from a transcranial magnetic stimulation (TMS) experiment (Fadiga et al. 1995). The assumption underlying this experiment was that if observation of a motor act activates the premotor cortex, this latter could in turn facilitate, probably under threshold, the primary motor cortex, because of their strict anatomical connections. Thus, the application of a TMS pulse at threshold intensity to the motor cortex of a subject observing an action should elicit a movement congruent with that involved in the observed actions. Subjects were required to simply observe an experimenter grasping an object or performing meaningless gestures with his arm. Pure object observation and detection of an illumination change served as controls. A TMS pulse was applied over the hand representation of the motor cortex during the crucial part of hand action observation (closure of the fingers on the object) and during comparable periods of the other three conditions. Electromyographic activation (motor evoked potentials, MEPs) of hand muscles was recorded to measure the effect. The results showed that there was a specific enhancement of MEPs during grasping observation. This enhancement occurred in those muscles that subjects normally used to execute the observed actions.

Subsequent TMS, electroencephalographic (EEG), and magnetoencephalographic (MEG) investigations confirmed that action observation activates the observer’s motor cortex (Cochin et al. 1999; Gangitano et al. 2004; Hari et al. 1998; Nishitani and Hari 2000, 2002; Strafella and Paus 2000). However, this type of experiment can only provide a gross localization of the regions activated by action observation. Brain imaging experiments are more suitable to provide this information. In fact, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies demonstrated that observation of motor acts activate three regions: one around the superior temporal sulcus, one in the inferior parietal lobule, and one in the premotor cortex and the posterior part of the inferior frontal gyrus (Buccino et al. 2001; Grafton et al. 1996; Grèzes et al. 1998, 2003; Iacoboni et al. 1999; Johnson-Frey et al. 2003; Koski et al. 2002; Manthey et al. 2003; Rizzolatti et al. 1996b). The most surprising finding is that the activation of the posterior part of the inferior frontal gyrus (IFG) involves Broca’s area, previously conceived as a “speech” area. In agreement with this new finding, other experiments demonstrated that this area is activated not only by action observation, but also during the execution of manual tasks (Binkofskii et al. 1999; Bonda et al. 1994, Chollet et al. 1991; Ehrsson et al. 2000; Iacoboni et al. 1999; Koski et al. 2002) or during motor imagery of hand motor acts (Grafton et al. 1996; Parsons et al. 1995).

The activation of Broca’s area during motor acts observation could be simply caused by an internal verbalization. A recent fMRI study (Buccino et al. 2001) shows that this is not the case. Participants had to observe video clips showing motor acts performed with the mouth, hand/arm, or foot/leg. In the first condition, the acts were goal directed; in the second, they were mimed without the object. Motor acts observation was contrasted with the observation of a static
face, hand, and foot, respectively. The results showed that in both conditions there was a bilateral activation of the frontal lobe, with mouth observed motor acts represented laterally in the pars opercularis of IFG, hand/arm motor acts represented in the dorsal part of the pars opercularis (partially overlapping with the mouth field) and in the ventral premotor cortex, while foot motor acts were represented in a more medial sector of the premotor cortex. In the first condition, however, but not in the second, there was also a bilateral activation of the inferior parietal cortex (Fig. 4). Mouth actions determined rostral activation, hand/arm actions activated a more caudal region, and foot actions produced a signal increase even more caudally, as well as within the intraparietal sulcus. This experiment demonstrates first that action observation activates parietal and frontal agranular cortex in a somatotopic fashion. In the agranular frontal cortex, this somatotopy strictly parallels the classical motor homunculus. Second, this activation was bilateral. Both these findings exclude that the activation during observation can be the result of verbalization. The activation of the frontal cortex for mimed motor acts suggests a difference with the properties of monkey mirror neurons, which typically do not respond to this type of stimulus. This difference could be related to the capacity, typical of human beings, of interpreting the meaning of an act even in absence of its target object.

Summing up, the mirror system is present in humans and its localization is very similar to that found in monkeys.

6. Possible Functions Derived from the Mirror Neuron System: Imitation, Language, and Intention Understanding

The presence of mirror neurons and their basic function, understanding of motor acts done by others, constitute per se a first evidence that cognitive abilities can originate from the core organization of the motor system. However, the mirror neuron matching system can have been exploited for other social cognitive
functions, such as intention understanding, imitation, communication, and language. All these functions involve a matching mechanism. Some of them, such as imitation and language, are exclusively or mainly present in humans; others appear to be already present, at least in primitive forms, in monkeys. I will try to provide evidence and speculations about the involvement of the mirror system in all these functions.

6.1. Imitation

Imitation could be the most obvious outcome of mirror neurons, because of their property to enable the observer to immediately translate the visual information on the observed action into the motor representation that could be used to reproduce it. On the other hand, the capacity to imitate is weak or even absent in monkeys (Visalberghi and Fragaszy 2002), and it is very limited also in apes. However, mirror neurons could be the building blocks from which imitation emerged in humans. Indeed, recent experiments confirm this hypothesis.

In an fMRI experiment, Iacoboni et al. (1999) asked participants to observe or, in separate blocks, to imitate a finger-lifting movement. The results showed an activation of the left inferior frontal gyrus (IFG) during observation and, more strongly, during imitation. Koski et al. (2002) modified the previous experiment, asking the participants to imitate touching a target instead of simply imitating a finger flexion. They confirmed the activation of IFG in imitation. The crucial role of IFG for imitation was demonstrated also by Nishitani and Hari (2000) with MEG. In this experiment, participants were required to only observe another person grasping a manipulandum, in one condition, or to observe and simultaneously imitate the observed action in another condition. During both observation and imitation there was an activation of the IFG and then of precentral cortex. These activations were preceded by an activation of occipital cortex because of the visual stimulation.

In these experiments, however, the movements or motor acts to be imitated belonged to the motor repertoire of the observers. What happens if the movements to be imitated are novel, that is, in the case of true imitation or when learning by imitation is required? Note that this latter capacity is fundamental for the transmission of human culture. In a fMRI event-related study, Buccino et al. (2004) required naive participants to imitate guitar chords played by an expert guitarist. The events of the imitation condition of the task were the following: (a) observation of chords made by the expert player, (b) pause, (c) execution of the observed chords, and (d) rest. In the other three conditions, used as controls, participants had to simply observe or to perform movements different from those to be imitated. The results showed that during observation for imitation there was activation of the inferior parietal lobule and the dorsal part of ventral premotor cortex plus the pars opercularis of IFG. During imitation there was a stronger activation of the same areas, with additional activation of the somatosensory and primary motor cortex. These findings show a strong involvement of the mirror neuron system during the imitation learning task. It is interesting to note that during the pause in imitation condition, when subjects are making
a plan to reproduce the observed chord, there was a strong activation of the middle frontal cortex (area 46), which then disappeared when imitation was actually performed. The role of this region in imitation could be crucial in the light of the proposal (Byrne 1999) that imitation could involve a mechanism in which pieces belonging to the motor repertoire of the imitator could be put together in a new sequence (Rizzolatti et al. 2001). It has been hypothesized that this part of the prefrontal cortex could recombine the motor representations corresponding to the different motor acts to fit the observed model. Its activation during the pause before imitation would just reflect the activation of this machinery.

As stated before, imitation in monkeys is minimal (for the presence of imitation in monkeys capable of joint attention, see Kumashiro et al. 2003). However, there is some evidence for the phenomenon of response facilitation, that is, automatic repetition of an observed action that is already in the observer’s motor repertoire, specifically triggered when an appropriate visual (or acoustic) stimulus is presented. Studies in capuchin monkeys showed that watching group members eating food increases eating of novel food by the observer (Visalberghi and Addessi 2000). More recently, Ferrari et al. (2005b) tried to assess whether the observation/hearing of a conspecific eating familiar food would have enhanced the same behavior in observing/listening macaques. The study consisted of two experiments. In the first the experimental subjects had to observe and hear eating actions, while in the second they only heard the same actions. As a control for the second experiment, hearing paper-ripping actions was used. The results showed that observation/hearing and only hearing an eating roommate significantly enhanced eating behavior and shortened latencies to eat in the observer. In contrast, the sound of ripping paper did not increase eating behavior. Note that this enhanced behavior appeared in satiated animals. Thus, this response facilitation can be attributed to the fact that observed eating behavior causes the representation of similar actions to resonate in the motor system of the observing/listening macaque. The meaning of this eating facilitation behavior might be that of coordinating eating activities among group members, coordination being a fundamental feature of primate behavior.

Thus, in monkeys there is a kind of “resonance behavior” that could depend on the mirror neurons. However, this is not enough to also trigger true imitation. One possible reason could rely on the difficulty experienced by monkeys in building action sequences. As discussed before, this capacity could be fundamental for imitation of novel actions and, in humans, could be processed by the prefrontal cortex. This region is much more developed in the human brain in respect to that of monkeys.

6.2. From Monkey F5 to Human Broca’s Area: A Possible Pathway in Language Evolution

The mirror neuron mechanism appears to be very close to the mechanism that, during dyadic communication, enables the listener/observer to understand the meaning of the message emitted by the sender. The central point is that both sender and receiver share the same motor programs necessary to produce a
communicative message and the neural circuits that allow accessing these programs. In this perspective, in searching for a behavior that could have been the root for language evolution and which was possibly endowed with a mirror neuron mechanism, gestures appear to be better candidates than vocalization. In fact, gestures are best suitable for dyadic communication and possess a higher combinatorial power than vocalization. Although it has been shown that in some monkey species calls can be referential, vocalization is in general rigid and very much linked to emotional behavior.

In monkeys, gestures are limited to facial and postural movements. In chimpanzees, brachiomanual gestures are used for communication and can also be associated to vocalization. Thus, starting from possible progenitors similar to the presently living monkeys, one could hypothesize that mouth communicative mirror neurons and audiovisual mirror neurons were the basic structure from which more sophisticated types of communication took place. Mouth communicative mirror neurons respond to the sight of a communicative gestures and, motorically, during the execution of similar communicative gestures, but also of ingestive gestures (Fig. 5). It has been proposed that these neurons could represent the transition between transitive, ingestive motor acts and intransitive, communicative gestures (Ferrari et al. 2003a; Fogassi and Ferrari 2004). Mouth communicative mirror neurons would have provided the basis for oro-facial communication, whereas audiovisual mirror neurons would have provided that association of hand action, acoustic input, and meaning necessary to build a high number of combinations of gestures and sounds and, in particular, the combination between gestures and arbitrary sounds, a crucial step in language evolution. Of course, these evolutionary pathways require a continuous coevolution of production and perception. With the achievement of a more complex articulatory oro-facial and gestural brachiomaneal apparatus, the mirror neuron system had to be adequate to reflect, on the perceptual side, the increasing complexity of the motor structure.

This hypothesis, originally suggested by Rizzolatti and Arbib (1998), received interesting support from the proposed homology and the functional similarity between F5 and Broca’s area. The homology between the two areas is based on a similar anatomical location in the frontal cortex and on a cytoarchitectonic similarity, as both area 44 (part of Broca’s area) and area F5 are dysgranular (Nelissen et al. 2005; Petrides and Pandya 1994; Petrides et al. 2005). The functional similarity consists in the fact that (a) both Broca’s area and F5 are involved in execution and observation of hand and mouth motor acts (Rizzolatti and Craighero 2004; Rizzolatti et al. 2001); (b) both F5 and Broca’s area are activated by an acoustic input related to semantic content (Bookheimer 2002; Kohler et al. 2002; Zatorre et al. 1996); and (c) area F5 is endowed with a system for the control of laryngeal muscles and of oro-facial synergisms (Hast et al. 1974).

Recent experiments in humans suggested an involvement of the mirror neuron system in some aspects of language. Buccino et al. (2004) demonstrated that IFG (Broca’s area) was activated both when subjects observed biting action and when they observed another individual speaking silently. Biting action activated the
Fig. 5. Examples of communicative mouth mirror neurons. Neuron 76 responds when experimenter makes lip-smacking (an affiliative mouth gesture) looking at the monkey (A). The neuron does not respond when the experimenter protrudes his lips (another type of affiliative gesture) looking at the monkey (B) or when he performs an ingestive act (C). The neuron discharges also when the monkey protrudes its lips and takes the food (D). Neuron 33 activates when the experimenter protrudes his lips looking at the monkey (A) and when the monkey responds with a lip-smacking gesture (B). In each panel, the rasters and the histograms represent the neuron response. During observation or execution of ingestive motor acts, rasters and histograms are aligned (vertical line) with the moment in which the mouth of the experimenter (observation conditions) or of the monkey (motor conditions) touched the food. During observation or execution of gestures, the rasters and histograms alignment was made with the moment in which the gesture was fully expressed. Bin width = 20 ms. (Modified from Ferrari et al. 2003)

IFG bilaterally whereas observation of silent speech produced a left activation. These findings are in agreement with the presence in monkey F5 of mouth mirror neurons for ingestive and communicative actions (Ferrari et al. 2003). Similar activation of IFG for observation of silent speech was found also by other authors (Calvert and Campbell 2003; Campbell et al. 2001). Hauk et al. (2004) and Tettamanti et al. (2005) showed that reading action words or listening to
sentences related to actions, respectively, caused a somatotopically organized activation of premotor cortex. The left pars opercularis and triangularis of Broca’s area were activated by listening to sentences related to mouth actions (for example, ‘I bite an apple’). Note that this somatotopical organization strictly resembles that obtained with observation of motor acts made with different effectors (Buccino et al. 2001). It is interesting to note that in the experiment of Tettamanti et al. (2005) there was also an activation in the pars opercularis of Broca’s area, common to all action sentences, that can be interpreted as a more abstract action representation which generalizes across different effectors.

A TMS experiment by Fadiga and coworkers (2002) showed that listening to verbal material elicits a specific activation of motor cortex. In this experiment, participants were required to listen to words and pseudo-words either containing a double “r” that, when pronounced, requires a marked tongue muscles involvement, or a double “f” not requiring this strong involvement. The TMS pulse applied during listening of words or pseudo-words containing a double “r” determined a significant increase of the amplitude of motor evoked potentials (MEPs) recorded from the tongue muscles with respect to listening to words and pseudo-words containing the double “f”. Furthermore, MEPs for words were stronger than those for pseudo-words, suggesting that the phonological resonance involves also understanding of the words meaning. This experiment is, to my view, very important for the issue of the link between mirror neuron system and language, because it suggests this is the mechanism we use to understand others’ linguistic messages. This mechanism is very close to the motor theory of speech perception (Liberman and Mattingly 1985), which postulates that “... the objects of speech perception are represented by the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands. . .”

All these pieces of evidence suggest that at least phonology and, probably, the part of language semantics related to actions, can rely on the motor system and that mirror neurons can provide the matching mechanism necessary for perception and understanding of speech. If this is true, the possibility of evolution of language from a monkey mirror neuron system is not so unrealistic.

It is still to be investigated if syntax itself can have relation with the organization of the motor system. Clinical data on Broca’s patients and recent fMRI experiments suggest that Broca’s area is certainly involved in syntax processing. It is unknown if this neural organization in humans can be traced back to some more primitive primate neural circuit, because it is not clear whether some aspects of motor cortical organization can share some properties with syntactic structures.

6.3. Intention Understanding

In the last 20 years many researchers and philosophers have tried to find empirical evidence and to speculate on our mind-reading ability. Mind-reading consists of the capacity human beings have to attribute beliefs, desires, intentions, and expectations to other individuals. There are two main theoretical accounts that
try to give an explanation of mind-reading. The first, named “theory–theory,”
maintains that mind-reading is accomplished by a theoretical reasoning based on
causal laws that allow establishing a correlation between external stimuli and
internal states and among internal states. The second theory, named “simulation
theory,” attributes mind-reading to a process of internal mimicking of the other
individual mental states, occurring when we observe other individuals acting. It
is intuitive that this second theory is very close to the mirror neuron mechanism.
When we observe other individuals performing actions, observed actions automatical-
ly retrieve the action representation stored in our motor system, that is, make our internal motor system work as if we were doing those actions. Because
action representation is endowed also with the notion of goal, this automatic retrieval subserves action understanding. Of course, this process is still far from
immediately explaining how we understand, for example, others’ beliefs. However,
it can certainly come into play for understanding intentions.

There are several motor behaviors that can give us cues to understand what
other people intend to do. Brachimano-oral and oro-facial actions and gestures
are very frequently used for understanding others’ intentions. Because mirror
neurons provide a mechanism to understand the goal of motor acts performed
by others, the issue is raised on whether they can also play a role in intention
detection. Until recently a demonstration of this role was still not available,
because it was known that the discharge of mirror neurons code motor acts, but
it was not clear how this discharge related to the general action goal. It is neces-
sary to define more strictly what is the difference between a motor act and an
action. By motor acts, I mean movements that have a partial goal (e.g., grasping
a piece of food). By motor action I mean a series of motor acts that, as their final
outcome, lead to a final reward (e.g., eating a piece of food after reaching it,
grasping it, and bringing it to the mouth). In many situations, the same motor
act can belong to different actions. For example, we can grasp a glass for drinking
or for washing it. When I make one of the two actions, I have clearly in mind
which is my intention. Similarly, when I see someone else grasping a glass, I can
guess from the context which is his/her intention. Thus, the issue is whether there
are neurons the discharge of which can reflect the agent’s intention. Recently
(Fogassi et al. 2005) we addressed this issue, recording from neurons of IPL.

As reported in a previous section, IPL can be considered as a part of the motor
system. Hand motor acts are represented in its rostral half, and their property
resemble those of neurons belonging to the motor vocabulary of area F5, although
in IPL the specificity for the type of grip seems less sharp. The experiment
devised for answering the question of whether intention is reflected in the prop-
erties of motor neurons consisted in the following behavioral paradigm. A
monkey performed a task involving two main conditions (Fig. 6, upper part). In
one, starting from a fixed position, it had to reach and grasp a piece of food
located in front of it and bringing it to its mouth. In the other, the monkey had
to reach and grasp an object or a piece of food located in front of it and place it
into a container. Note that in this task the first motor act of both conditions is the
same (grasping).
Fig. 6. Examples of parietal motor and mirror neurons tested with the motor and visual task, respectively. Upper part. Top left: motor task. The monkey, starting from a fixed hand position (A), reaches and grasps a piece of food/object (B) and brings it to the mouth (condition I) or places it into a container located on the table (condition II) or near the mouth (condition III). Top right: lateral view of the monkey brain, illustrating the hand region of the inferior parietal lobule (IPL) from which neurons were recorded. Bottom: Rasters and the histograms representing the motor responses of three IPL motor neurons during grasping in conditions I and II, synchronized with the moment when the monkey touched the target. Bin width = 20 ms. Lower part. Top: visual task. The experimenter, starting from a fixed hand position, reaches and grasps a piece of food/object (B) and brings it to the mouth (condition I) or places it into a container located on the table (condition II). Bottom: Rasters and the histograms representing visual responses of three IPL mirror neurons during observation of grasping in conditions I and II, synchronized with the moment when the experimenter touched the target. (Modified from Fogassi et al. 2005)
After the monkey was trained for the task, grasping neurons were recorded from the hand representation sector of IPL. The results showed that the majority (65%) of grasping neurons discharged differently according to the intended goal of the action in which grasping was embedded (Fig. 6, upper part). Neurons coding grasping for eating discharged more strongly when grasping preceded bringing to the mouth than when it preceded placing in the container. Neurons coding grasping for placing showed the opposite behavior. The remaining (35%) grasping neurons did not show any selectivity. These data suggest that (a) the IPL contains prewired or learned chains of motor neurons, each coding a specific final goal; and (b) the discharge of IPL grasping neurons reflects the intention of the performing agent. Although this type of motor organization includes also the concept of intention, this does not mean that intention is directly coded by IPL motor neurons. In fact, these motor neurons \textit{reflect} intention, but it is possible that elsewhere, for example, in the prefrontal cortex, there are neurons discharging during the whole action, thus probably having the role of keeping in memory intention during action execution.

In the second part of the experiment described above (Fogassi et al. 2005) the visual response of parietal mirror neurons was studied in the same conditions that were used for studying the motor properties of IPL grasping neurons. Briefly, in one condition the experimenter grasped a piece of food and brought it to the mouth, and in the other he grasped the same piece of food (or an object) and placed it into a container (Fig. 6, lower part). Mirror neuron activity was recorded while the monkey observed the two types of actions.

The results showed that the majority (75%) of IPL mirror neurons discharged differently when the observed grasping motor act was followed by bringing to the mouth or by placing (Fig. 6, lower part). The remaining mirror neurons did not show any selectivity. Interestingly, the differential neuron discharge occurs before the motor act following grasping begins, therefore represents a prediction of the action final goal.

A property characterizing all mirror neurons is the congruence between their motor and visual responses. By examining, in a group of mirror neurons, the congruence between the differential motor discharge and the differential visual discharge, we could show that mirror neurons that discharged more intensely during grasping for eating than during grasping for placing discharged more intensely also during the observation of grasping for eating. Conversely, neurons selective for grasping to place discharged stronger also during the observation of this motor act.

Thus, IPL mirror neurons can discriminate among identical motor acts according to the context in which they are executed. Because the discriminated motor acts belong to specific chains, each of which leads to a specific final goal, this capacity allows the monkey to predict what is the goal of the observed action and, in this way, to “read” the intention of the acting individual. If, for example, a mirror neuron belonging to the “eating” chain fires, this means that the observed acting individual is going to bring the food to the mouth. Of course, the selection of a particular group of grasping mirror neurons depend on many factors, such
as the type of object to be grasped, the context, or the previous action made by the observed agent.

In agreement with the monkey data suggesting that the mirror system can provide a mechanism for intention understanding, a recent fMRI study in humans (Iacoboni et al. 2005) indicates that in our species the mirror neuron system also enables us to understand the intention underlying others’ actions. In this study there were three conditions. In the first, participants had to observe two different scenes representing a context of breakfast: one as if breakfast was going to begin, and the other as if breakfast were finished. In the second, they had to observe a hand grasping a cup performed in the absence of context. In the third, they had to observe the same hand grasping acts of the second condition, performed in the two contexts of the first condition. Of course, the meaning of grasping motor acts in the two contexts could easily be interpreted. Participants were divided in two groups: those of the first group were asked to purely observe the video clips (implicit task), the others to attend to the objects, type of grip, and intention of the three different conditions (explicit task). The results showed, first of all, that all conditions activate areas belonging to the mirror neuron system. More interestingly, hand actions performed in contexts, compared with the other two conditions (actions without context or context only), produced a higher activation of the inferior frontal gyrus. Finally, both groups of participants showed similar activation, with the difference that participants receiving the explicit instruction presented activations probably related to the stronger effort required by the explicit task. This latter result indicates that for understanding motor intention there is no need of top-down influences, but that intention is automatically processed. However, future experiments on intention reading using different tasks could reveal if there are activated areas, beyond the mirror neuron circuit, that can be involved in some aspect of this ability.

Summing up, the mirror neuron system in monkeys provides the first neural substrate for a primitive understanding of others’ intentions. This system probably paved the way for the evolution of the more sophisticated aspects of mindreading present in humans. However, even in humans, many of the aspects of intention reading can still rely on the automatic activation of the parietofrontal mirror neuron circuit.

7. Conclusions

The recent research in cognitive neuroscience is accumulating evidence that perception and cognition are very much grounded in the motor system. Motor representations are already present in life or are formed in the first weeks after birth before, for example, a mature visual system is developed. The vocabulary of acts present in ventral premotor cortex is a good example of how we match the external world with our internal knowledge, in which the concept of goal is fundamental. This matching constitutes a starting point not simply for operating sensorimotor transformation, but for building cognitive functions. Among these
functions there are those related to intersubjectivity, which have the mirror neuron system as the underlying structure, with its simple mechanism of comparing what others do with what we do, as if we were internally performing others’ behavior. This matching system enables us to understand actions.

In this chapter, I reviewed some of the functions that can be derived from the mirror neuron system. However, there are other social cognitive capacities that could rely on a similar mechanism, such as emotion understanding and empathy (for a discussion of this issue, see Gallese et al. 2004). In the years to come, we will probably see some implications of the mirror neuron system in the field of pathology. For example, there is now some evidence that autism could be derived from an anomaly of the mirror neuron circuit (Dapretto et al. 2006; Hadjikani et al. 2005). Thus, the present knowledge of the mirror neuron system could also become the starting point for possible rehabilitation scenarios.

References


6
Representation of Bodily Self in the Multimodal Parieto-Premotor Network

AKIRA MURATA and HIROAKI ISHIDA

1. Introduction

The consciousness of one’s own body is the fundamental process of self-recognition (Jeannerod 2003). This process is based on recognition of personal space and one’s own body using visual, somatosensory, and vestibular inputs and also intrinsic motor signals. These signals are also necessary to control action of the body. The consciousness of the body includes some dynamic processes of action control in which the sensory feedback and intrinsic motor signals update the representation of one’s own body state. The term “body image” is often used for the consciousness of the body. Although there are many definitions, we think that body image involves internal representation about the spatiotemporal dynamic organization of one’s own body that is constructed by the processes of the consciousness of the body, as Berlucchi and Aglioti (1997) claimed. In this chapter, we use the term “corporeal awareness” (Critchley 1979) for this dynamic consciousness of the body. We discuss here that corporeal awareness shares the parieto-premotor network with the sensorimotor control. This system appears to have the function of monitoring one’s own body and may be involved in the mechanisms of discrimination between self and others.

1.1. Using Somatic Sensation and Vision to Explore the Environment

In daily life, we use hands or tools to establish goal-directed action. Primates have well-developed capability of the hands. Kamakura et al. (1980) classified functional properties of the hands into searching, pressing, jointing, using as tools, grasping, manipulating objects, and forming symbols. Each requires precise sensory motor control based on sensory feedback, both somatosensory and visual. As Gibson (1962) pointed out, hands are sensory organs of active touch,
exploring the outer world, and eyes are for action control. Vision and somatic sensation influence and direct the action of the body toward the outer world.

1.2. Corporeal Awareness and Motor Control

The sensory systems are not only directed to the outer world but also are directed toward corporeal awareness in the intrinsic world. The tight correlation between action and perception is not unidirectional but bidirectional, that is, not only from the sensory system to the motor system, but also from the motor system to the sensory system. Perception is influenced by motor signals. For example, phantom limbs might be the result of the representation of the body in the brain that persists after the amputation. Some patients who feel phantom limbs can perceive vivid movement in their amputated limbs when they intended to move the limbs (Berlucchi and Aglioti 1997). This observation suggests that the experience of voluntary movement might be the effect of reafferent signal from a motor command monitored in some brain areas (Ramachandran and Rogers-Ramachandran 2000). These reafferent signal correspond to efference copy (von Holst 1953) or corollary discharge (Sperry 1950). Somatosensory and/or visual feedback are always contingent to action in usual situations. Efference copy and sensory feedback integrate to execute precise action. As we discuss later, interaction between efference copy and sensory feedback contributes to corporeal awareness. Accordingly, the process of corporeal awareness may share components with the sensory motor control process. Control of body action may translate into consciousness of the body.

2. Pathology of Corporal Awareness and the Parieto-Frontal Circuit

It is well known that damage in the parietal cortex often induces some impairment of corporeal awareness. Hemispatial neglect, anosognosia, and asomatognosia have been known to occur in patients with damage of the right parietal cortex. Anosognosic patients show an inability to recognize the presence of their motor and sensory defects, frequently concurrent with asomatognosia (Meador et al. 2000), which is impairment of the ability to recognize one’s own body parts. Very often, anosognosia and asomatognosia are accompanied by the symptoms of hemineglect. Patients with hemineglect show loss of awareness of the left side of the body, such as failing to dress on the left side or to put on makeup on the left side of the face (Beschin and Robertson 1997; Danckert and Ferber 2006), that is, personal neglect. These symptoms are observed after the damage of not only the parietal cortex but also of some other cortical regions, especially the frontal cortex (Husain et al. 2000; Pia et al. 2004). These clinical observations suggest that both the parietal and frontal cortices are involved in corporeal awareness. Given that the parieto-frontal circuit is related to spatial recognition
and motor control, it is suggested that corporeal awareness shares the motor control circuit (Pia et al. 2004). We now discuss this subject further.

3. Sense of Ownership and Sense of Agency

There are two components to bodily self-recognition. One is ownership of one’s own body parts in the sense that one’s body parts belong to the self (Gallagher 2000). The other one is the sense of agency of action, in which an executed action is recognized as being generated by one’s own body parts. This is also called a sense of authorship. These two components are not separated in usual situations, in which an action is intended and the contingency of the sensory feedback is preserved. However, when the movement is executed passively, for example, when a limb joint is moved passively, it is possible to distinguish the sense of agency from the sense of ownership (Gallagher 2000; Jeannerod 2003).

3.1. Rubber Hand Illusion

The sense of ownership is based on multimodal sensory integration. Both visual and somatosensory cues contribute to the sense of ownership (Botvinick and Cohen 1998; van den Bos and Jeannerod 2002). In human experiments, there is evidence for visual-tactile integration in the brain (Kennett et al. 2001; Press et al. 2004; Tipper et al. 2001). One example of visual-somatosensory integration in the sense of ownership is the rubber hand illusion (Botvinick and Cohen 1998). There is a feeling of ownership of a fake hand that can be seen in front of the subject, while the subject’s real hand is visually occluded. Tapping stimulation is applied synchronously on the real and fake hands. Afterward, the subject feels a sensation of being touched on the fake hand as well as ownership for the fake hand; this reflects three-way interaction among visual, tactile, and proprioceptive sensations. Furthermore, ownership of the fake hand implies that vision plays a predominant role in ownership (Jeannerod 2003). It seems that in limb position, visual information leads to proprioceptive information. However, it is also possible to perceive the rubber hand illusion with the eyes occluded, with only tactile and proprioceptive cues (Ehrsson et al. 2005). We emphasize that both visual and somatosensory inputs are crucial factors for recognizing one’s own body parts. Vision may have a predominant role, but dynamic somatosensory input also influences the level of consciousness of ownership.

3.2. Involvement of the Parietal Cortex in the Sense of Ownership

Ehrsson et al. (2004; 2005), in their functional MRI (fMRI) study, revealed that activation in the ventral premotor cortex and cerebellum had a correlation with the strength of the rubber hand illusion, implying that these areas are sites for awareness of self-ownership. However, they also found activation in the inferior
parietal cortex, although the activation did not show a correlation with the strength of ownership of the fake hand. Actually, some neurons in Area 5 of the monkey were more active when the real arm and visually presented fake arm were in the correspondent positions (see Section 7.1) (Graziano et al. 2000). Temporal synchrony of multimodal sensory information is also an important factor for the rubber hand illusion (Botvinick and Cohen 1998; Ehrsson et al. 2004). Recently, Shimada et al. (2005) revealed that the superior parietal lobule was activated with synchrony of visual and tactile sensory consequences during passive joint movement. In this experiment, the subject’s wrist joint was moved passively then visual feedback of this movement was presented on a screen with or without temporal delay. Each subject was required to judge whether the visual feedback was synchronized with his or her hand movement. The researchers evaluated activity in the parietal cortex with near-infrared spectroscopy and found that the bilateral superior parietal areas were involved in the synchronous situation whereas the right inferior parietal cortex was activated in the asynchronous condition. They suggest that a sense of ownership is mainly processed in the superior parietal lobe, whereas the right inferior parietal lobe is involved in perceiving the movements of others.

3.3. Sense of Agency and Efference Copy

As we previously described, there are two components for self–other recognition: the sense of ownership and the sense of agency (Jeannerod 2003; Tsakiris et al. 2005). The sense of agency is the sense that occurs only during voluntary movement, which implies that internal motor signals (motor commands) have a crucial role in the sense of agency.

According to computational studies, there are two internal models for motor control. One is the “inverse model,” which generates a motor command to achieve a desired trajectory of the limb. The other one is designated as the “forward model,” which acts as a neural simulator of the musculoskeletal system and environment (Wolpert et al. 2003). The copy of the motor command, that is, the efference copy, can pass into the forward model to predict the state of the motor system in response to a given motor command. Ongoing movement is monitored by actual sensory feedback and the efference copy to establish more precise movement. However, the system not only controls movement, but also discriminates who generated the action that is visually presented. If the efference copy matches with actual sensory feedback in the comparator simultaneously, the action is detected as self-generated (Blakemore et al. 2001, 2002; Jeannerod 2003).

3.4. Efference Copy, Visual Feedback, and Somatosensory Feedback for Agency Recognition

Congruency of efference copy, visual feedback, and somatosensory feedback is necessary for a sense of agency. These components support a vivid sense of
agency. For example, people cannot tickle themselves (Claxton 1975). Blakemore et al. (1999) studied the phenomenon using devices that produce tactile stimuli controlled by a subject. These devices delivered tactile stimuli to the subjects synchronized with subject movements or with some temporal delay. They found that the rate of tickling occurrence increased as the delay of tactile stimuli from actual subjects’ movement increased. They claimed that tactile sensory response of self-generated action was attenuated by the forward model using the efference copy.

The interaction between the efference copy and visual feedback also happens during self-generated action. Daprati et al. (1997) asked subjects to judge whether the hand action presented on a screen was their own real-time action or that of the experimenter. The presented actions of the experimenter on the screen were either congruent or incongruent with the subject’s own hand action. In this experiment, the subjects confused their own hand action with that of the experimenter during the presentation of congruent action, which may have been the result of the matching of the efference copy and the visual image of another’s action on the screen. However, this matching is not just for spatial contingency of action. Actually, even if some rotation of the image on the screen was provided, the subject could judge correctly in the incongruent situation. However, in the congruent situation, the subject showed confusion in the agency judgment (van den Bos and Jeannerod 2002). This result suggested that temporal congruency rather than spatial congruency between the efference copy and sensory feedback is a critical factor to distinguish self-generated action from that of others. Tsakiris et al. (2005) tried to dissociate efference information and consequent sensory feedback during self-generated action, using a setup very similar to that of Daprati’s experiment. In this experiment, the subject’s right hand was passively moved by his own left hand via an apparatus or by the experimenter’s hand. In the former condition, an intrinsic motor component existed, and in the latter condition, it did not. The researchers asked the subjects to judge whether the presented right hand action on the screen was their own real-time action or that of the experimenter. They showed that a subject could judge more correctly the agency of action during self-generated action than during the action generated by others (Tsakiris and Haggard 2005; Tsakiris et al. 2005). In usual conditions, it is difficult to dissociate efference information from proprioceptive feedback. However, the researchers were able to separate these two sources. The results suggested that the efference copy had a predominant role in agency recognition.

3.5. Inferior Parietal Lobule for the Sense of Agency

Neuropsychological and imaging experiments revealed that the inferior parietal cortex is involved in the sense of agency. Patients who had damage in the inferior parietal cortex showed difficulty of agency recognition in the same paradigm used by Daprati et al. (see previous section; Sirigu et al. 1999) in the congruent condition. Another patient who had bilateral damage involving the supramarginal and angular cortices (Schwoebel et al. 2002) executed unintended motor action in the
hand just by imagery of the hand movements without any awareness of the movement. The symptoms can be explained as follows. The patient evoked movement by just imagining the action, then the efference copy was sent to the parietal cortex, but it failed to compare the efference copy with the actual sensory feedback; this may be the reason that the patient was unaware of his own evoked action during imagery (Schwoebel et al. 2002). In conclusion, the inferior parietal cortex may contribute to the sense of agency by comparing the efference copy with actual sensory feedback.

A number of human brain imaging studies showed that the inferior parietal cortex is involved in the detection of agency of action (Decety and Grezes 2006). It was revealed that increasing activation of the right inferior parietal cortex was associated with a decreased feeling of controlling one’s own action (the researchers increased the spatial distortion of the visual feedback), then vice versa in the insula (Farrer et al. 2003). Leube et al. (2003) reported that a positive correlation between the extent of the temporal delay of visual feedback and activity was observed in the right temporoparietal junction. This area is also activated during the observation of action or biological motion (Grezes et al. 2001; Puce and Perrett 2003). These results suggested that the inferior parietal cortex detected any discrepancy between the intrinsic motor component of self-generated action and sensory feedback, which implies that the right inferior parietal lobule contributes to distinction of self-actions and other’s actions. Actually, Decety et al. (2002) revealed that the right inferior parietal cortex is activated more during observation being imitated by another person than during one’s own imitation.

3.6. Superior Parietal Lobule for the Sense of Agency

Furthermore, there is some evidence that the superior parietal cortex is related to agency recognition. A patient who had a large cyst in the left superior parietal lobule experienced fading away of the right limb (regarding sense of position and presence of the limb) within a few seconds if she could not see it (Wolpert et al. 1998). She also had a problem in performing a slow pointing movement without vision. The patient was normally able to recognize her own body parts with vision and to detect tactile and somatosensory input, except for problems in the higher somatosensory functions (astereognosis and agraphesthesia). The basic abilities of visual and somatosensory perceptions were preserved. The symptoms may result from a problem of integration of somatosensory feedback and efference copy in the absence of visual feedback, and also in the storage of information regarding the current internal state of the body. This implies that the superior parietal cortex is involved in maintaining and updating the internal state by integration of somatosensory and efference copy inputs.

Furthermore, repetitive transcranial magnetic stimulation (TMS) on the superior parietal cortex induced difficulty in judgment of asynchrony between actual hand movement and visual feedback during active movement (MacDonald and Paus 2003). The superior parietal cortex may contribute to judging temporal contingency comparing the efference copy and sensory feedback.
3.7. Sense of Agency in Schizophrenia

We should also discuss the delusional control of action in patients with schizophrenia. These patients often experience their intended action as being controlled by someone else. These patients experienced less attenuation of tickle stimulation by self-generated movements than normal subjects (Blakemore et al. 2000). An imaging study showed hyperactivity in the parietal cortex when the schizophrenic patients with the delusion of control made voluntary movements (Spence et al. 1997). Furthermore, hyperactivation in the inferior parietal cortex and cerebellum was found in the subjects under hypnosis who felt their active self-movements attributed to an external source (Blakemore et al. 2003). These results suggested that delusional control might be a problem of attenuation of sensory feedback during voluntary movement in the parietal cortex (Frith 2005).

A problem in recognition of visual feedback was also revealed in the schizophrenic patients. They were less able to correctly detect agents of spatially distorted visual feedback of their own hand movement (Dapra et al. 1997). In normal subjects, the activity of the right angular gyrus and the insular cortex appeared to be modulated by the subject’s degree of movement control with the spatially distorted visual feedback (Farrer et al. 2003). However, it was found that an aberrant relationship existed between the degree of control of the movements and regional cerebral blood flow (rCBF) in the right angular gyrus and that no modulation in the insular cortex existed in the schizophrenic patient (Farrer et al. 2004). These results suggested that a problem in the monitoring of self-generated action in the parietal cortex is responsible for the delusional control. The delusion of control may be considered as a problem in the detection of any discrepancy between the efference copy and sensory feedback or as a problem of the forward model that predicts the state of the motor system (Blakemore et al. 2002; Frith 2005). Given that the patients did not have a problem in motor control, the symptom is likely to be caused by a problem in the awareness of control of action (Frith 2005).

4. Flow of Visual Information for Action Control

The parietal and premotor cortices have strong anatomical connections with each other. As shown in Fig. 1, there are several parallel pathways between the parietal and premotor cortices. Recent physiological studies have revealed that these parallel pathways have different roles, such as arm reaching and/or hand grasping, along with corporeal awareness.

Classically, the parietal association cortex is involved in the dorsal visual stream, which is considered to be related to the processing of visual information for space perception. The dorsal visual pathway is separated into two channels: the dorsodorsal pathway and the ventrodorsal pathway (Galletti et al. 2003; Rizzolatti and Matelli 2003; Tanne-Gariepy et al. 2002) (Figs. 1, 2). The dorsodorsal pathway starts at V3/V3A and passes through V6, which is a part of the
A. Murata and H. Ishida

Fig. 1. Parieto-premotor connection in the monkey cerebral cortex (gray arrows). AIP, anterior intraparietal area; VIP, ventral intraparietal area; LIP, lateral intraparietal area; CIP, caudal intraparietal area; MIP, middle intraparietal area; dPM, dorsal premotor cortex; vPM, ventral premotor cortex; FEF, frontal eye field; MT, middle temporal area; MST, medial superior temporal area; TPOc, temporal parietal occipital (caudal); PS, principal sulcus; AS, superior arcuate sulcus; AI, inferior arcuate sulcus; CS, central sulcus; IPS, inferior parietal sulcus; PO, parieto-occipital sulcus; LF, lateral fissure; LS, lunate sulcus; STS, superior temporal sulcus. IPS, STS, and LS are opened, showing the inside of the sulci.

The dorsodorsal pathway has a collateral through VIP. Area VIP is localized in the fundus of the intraparietal sulcus on the border of the superior parietal lobule and the inferior parietal lobule. The afferent to this area originates from MT/MST. VIP has a connection with the caudal part of the ventral premotor cortex (F4) directly or via PEa, which is in part of Area 5 (Lewis and Van Essen 2000; Luppino et al. 1999; Rizzolatti and Luppino 2001). Because neurons in V6A, MIP, PEa, PF/PFG, and VIP (shaded area in Fig. 2) showed bimodal
sensory properties (visual and somatosensory) (Breveglieri et al. 2002; Colby 1998; Colby et al. 1993; Iriki et al. 1996), these areas are involved in the integration of somatosensory and visual information. Especially, the pathway from VIP to F4, including PEa and PF/PFG, is involved in the representation of peripersonal space and in hand or arm movement control.

The ventrodorsal pathway passes through the anterior intraparietal area (AIP), which is located in the anterior part of the lateral bank of the intraparietal sulcus, or PFG/PF, which is on the lateral convexity in the posterior parietal cortex. Classically, Area 7 was subdivided into sections 7a and 7b. Section 7b was further divided into PF and PFG according to cytoarchitecture and connection with other areas (Gregoriou et al. 2006; Pandya and Seltzer 1982; Rozzi et al. 2006). Area AIP and PFG/PF have connections with the ventral premotor cortex (F5) (Lewis and Van Essen 2000; Luppino et al. 1999; Rizzolatti and Luppino 2001; Rizzolatti and Matelli 2003; Tanne-Gariepy et al. 2002). In AIP and PFG, there are many neurons related to distal hand movements (Jeannerod et al. 1995; Sakata and Taira 1994; Sakata et al. 1997).
5. Dual Properties of Neurons in AIP and F5 in the Ventrodorsal Pathway

In this section, we describe the functional properties of neurons in the parieto-premotor network. The data demonstrate that the network works for the control of sensory-guided hand and arm movements. Furthermore, this system is also involved in the mechanisms of corporeal awareness. Control of body action requires awareness of the body.

Neurons activated during grasping movement were found in the ventrodorsal stream of the dorsal visual pathway, from AIP to F5 of the monkey (Murata et al. 2000; Rizzolatti et al. 1988; Sakata et al. 1995; Taira et al. 1990). In our previous experiments, we compared neural activity during hand manipulation tasks in which a monkey was required to manipulate objects with one hand using full vision and in the dark or just fixate on the object for manipulation without grasping it (Murata et al. 1997; Sakata et al. 1995; Taira et al. 1990). We classified three different types in AIP that were activated during the task: motor dominant, visual dominant, and visual-motor neurons.

Motor dominant neurons fired during manipulation in the full vision condition and in the dark but did not show any significant difference in level of activity in both conditions. These neurons did not respond to the presentation of objects, nor to any somatosensory stimuli, and thus were considered as being related to the motor component.

Visual dominant neurons fired during manipulation in the full vision condition but not during manipulation in the dark. Visual-motor neurons were less active during manipulation in the dark than in the light. These visually responsive neurons fired during fixation on the objects, coding three-dimensional properties of the objects (object type), that is, each of these neurons was selective to a particular shape, orientation, or size of objects (Murata et al. 2000). However, the rest of the visually responsive neurons did not show any activity to the sight of objects. Because this type of neuron showed selectivity in the object to grasp (nonobject type), the activity may reflect the view of the moving hand configuration.

In the ventral premotor cortex F5, activity of neurons showed selectivity to the type of hand movement, such of precision grip, finger prehension, and whole-hand grasping (Rizzolatti et al. 1988). Rizzolatti’s group coined the term “motor vocabulary” for these different types of movements (Jeannerod et al. 1995). Functional properties of neurons of this area were similar to those of AIP (Murata et al. 1997; Raos et al. 2006). A group of grasping-related neurons in F5 was activated when the monkey fixated on the objects to be grasped. This category of neurons was active during grasping in the dark, like visual-motor neurons in area AIP. Furthermore, the other group of grasping neurons did not show any visual properties like those of motor dominant neurons in AIP.

There are also differences in the functional properties of neurons between AIP and F5 (Murata et al. 1997, 2000; Raos et al. 2006). No purely visual neurons
such as the visual dominant neurons in AIP were found in F5; there were only visual-motor neurons or motor dominant neurons. Furthermore, neurons in F5 showed sustained activity (so-called set-related activity) preceding the hand movement after the object presentation. This may reflect the process of visuo-motor transformation before execution of a particular movement. However, this set-related activity was less common in AIP neurons than in F5 neurons (Murata et al. 1996). Moreover, in F5, only visual-motor neurons showed this set-related activity (Raos et al. 2006).

These results suggest that AIP and F5 work together in visuo-motor transformation, but there is a functional difference between AIP and F5. AIP receives visual information about three-dimensional objects, such as shape, orientation, and size. The visual information about an object is necessary for the control of distal hand movement to manipulate the objects (visual dominant neurons). The origin of this visual information is the caudal intraparietal area (CIP) (Taira et al. 2000; Tsutsui et al. 2001) or the inferior temporal cortex (IT) (Uka et al. 2000), where three-dimensional visual cues activate neurons. In F5, the motor program or motor pattern that is appropriate for presented objects is selected (visual-motor neurons in F5) (Oztop et al. 2006). This motor information is then sent to the primary motor area. At the same time, this copy of motor representation, that is, the efference copy, is returned to AIP to be matched with the visual object representation (visual-motor neurons in AIP) (Sakata et al. 1995). Finally, some visual neurons (nonobject type in AIP) are related to the visual feedback signals that monitor the ongoing hand movement. In the next section, we describe the functional properties of neurons of the parieto-premotor network mainly related to corporeal awareness.

6. Parieto-Frontal Network I (PFG, AIP, and F5)

6.1. Parietal Neurons Driven by Visual Feedback

As we described in the previous section, some of the grasping neurons in AIP have strong visual properties. Some of these visual neurons in AIP and PFG did not respond to the view of the objects, suggesting sensitivity to the particular visual hand configuration during grasping. The hypothesis is that the areas may be concerned with the monitoring of ongoing hand movement. To test this hypothesis, we recorded single-cell activity from PFG and AIP of the monkey while it performed the hand manipulation task and the fixation task with the guidance of a spotlight on a monitor screen (Fig. 3). In the hand manipulation task, the monkey could not see its own hand as well as an object to be grasped directly, but their images were presented on a screen using a video camera. The monkey was required to reach and grasp the object while watching the screen. In the fixation task, the monkey was required to fixate on the screen, and we presented a movie of the monkey’s own hand movement. We found that some neurons related to the hand manipulation task in both AIP and PFG
Fig. 3. Experimental setup. The image of the hand movement and object was taken by a video camera, and the image was presented on an online screen. In the hand manipulation task, when the red spotlight was turned on, the monkey fixated on the light and pressed a key until the green light came on. With the green light, the monkey was required to release the key, then reach for and grasp the object. When the monkey released the key, the red light turned on again and the monkey held the object until the green light came on again. In the fixation task, the monkey just fixated on the green spotlight until the red light came on. In this situation, we presented a movie of the hand movements of the monkey.
responded to the movie of the monkey’s own hand movement (hand-type neurons) (Murata 2005). Many of these visual neurons (hand type), which were previously called nonobject type neurons, did not show much activity during fixation on the objects. A neuron recorded from PFG showed activity during manipulation both in the light and in the dark (Fig. 4). This neuron responded to a movie of the monkey’s own hand holding an object. The neuron remained active when we erased the image of the object from the movie, and thus the response could be considered as responding to the configuration of hand movement. As we discussed in the previous section, the motor-related activity might reflect an efference copy that might have originated from the ventral premotor cortex F5. We consider that a comparator exists in AIP and/or PFG to match the efference copy and visual feedback (see Section 3.3). Actually, TMS applied on the human homologue of AIP induced disruption of online adjustments of grasping (Tunik et al. 2005).

6.2. Monitoring of Self-Generated Action by Mirror Neurons in the Frontal and Parietal Cortices

Although both AIP and PFG were correlated to distal hand movement, there are differences in the sensory properties of the neurons. In PFG, Rizzolatti’s group found mirror neurons that were originally found in the F5 (Gallese et al. 2000). These neurons were activated during execution of hand or mouth action and also during observation of the same action made by another individual. The functions of these neurons were designated in terms of the human cognition, for example, “action recognition” (Rizzolatti and Craighero 2004), “theory of mind” (Gallese and Goldman 1998), “origin of language” (Rizzolatti and Arbib 1998), “imitation” (Rizzolatti et al. 2001; Wolpert et al. 2003), and “empathy” (Gallese 2003). However, functions of mirror neurons in the monkey remain unclear. A recent study revealed that mirror neurons in PFG contribute to recognition of observed motor acts and prediction of what will be the next motor act of the complete action, that is, understanding the intentions of the agent’s action (Fogassi et al. 2005).

Because the mirror neurons were included in the visuo-motor control system, we need to discuss their functional role with respect to the motor control. These neurons respond to the image of action. We postulated that the function of the mirror neuron to be the monitoring of self-generated action. To study if the hand-type neurons (see previous section) in PFG or AIP have properties of mirror neurons, we studied the activity of the neurons during the watching of a movie of hand action by the experimenter. We found the hand type of neurons in PFG responded to the movie of the experimenter’s hand. These results showed that some of the mirror neurons in the inferior parietal cortex were sensitive to both the view of the monkey’s own hand action and that of others’ actions as well. We suggest that mirror neurons in PFG correlated with the monitoring of self-generated action by collaboration with F5 in the ventral premotor cortex.
Fig. 4. An example of PFG neurons related to visual feedback during the hand manipulation task. Manipulation in light and manipulation in dark represent activity during the manipulation task. The vertical dashed line corresponds to the timing of holding the object. In the “Movement in light,” the monkey can see hand and objects. In the “Movement in dark,” the monkey can see only the spotlight. In both conditions, neurons were activated before and after the monkey held the object. In “Watching movie with object,” the monkey just watched a movie of its own hand movement. The neurons showed visual response to a movie of movement that was recorded from the same viewpoint in the manipulation task. The timing of peak activity was just before the holding action. The activity was considered to reflect the configuration of the hand during grasping. In “Watching movie without object,” the response remained active while the monkey was watching a movie in which the view of the object was erased.
6.3. Proposal of a Conceptual Framework for Self–Other Distinction

Figure 5 shows a conceptual framework of the parieto-premotor cortical network for bodily self-recognition. In some neurons of the parietal cortex, visual images of self-action and action of others are not distinguishable. The origin of the visual feedback signal may be the superior temporal sulcus (STS) [or EBA (extrastriate body area); this is not yet clear in the monkey brain], which represents visual body action (Astafiev et al. 2004; Keysers and Perrett 2004). Mirror neurons in PFG showed a motor component, possibly reflecting the efference copy. As we described before, F5, where mirror neurons were found, has a strong connection with PFG and AIP. Regarding the somatosensory input, some neurons in PFG showed visual-tactile bimodal properties or visual-joint bimodal properties (Leinonen et al. 1979), while neurons in AIP usually do not show any somatosensory activity. PFG may receive proprioceptive feedback from Area 5.
or SII. Matching occurs in the PFG between motor representation of action and sensory representation of action (visual and somatosensory) in the comparator; this works for feedback control of movement. If correct matching occurs in the PFG, it would be possible to distinguish the agency of action (see Section 3.3). As we describe later, PFG, VIP, and PEA form a network in the parietal cortex that would contribute to mapping of the body in the brain.

Temporal contingency of the efference copy, somatosensory feedback, and visual feedback are crucial for agency recognition (Leube et al. 2003; Miyazaki and Hiraki 2006). We found some modulation in visual response of hand-type neurons with temporally distorted visual feedback. These phenomena are congruent with the cancellation of self-tickling, showing a contribution of PFG to the sense of agency. It is possible that the efference copy influences sensory feedback in PFG, suggesting a contribution of PFG to self/other distinction.

7. Parieto-Frontal Network II (PE, PEA, and F2)

7.1. Physiological Evidence for Sense of Ownership

We have pointed out the importance of proprioceptive and tactile inputs for the sense of ownership. The superior parietal lobule is one important area for higher processing of proprioceptive and tactile information. It was reported that neurons in Area 5 were responsive to passive multijoint movements and also to tactile stimuli (Sakata et al. 1973). Many neurons in Area 5 showed the best response with a combination of joint and skin stimulations. For example, the best stimulus for a neuron was a passively moved right hand rubbing the left arm. This neuron showed response to right elbow flexion and right shoulder joint anterior elevation plus left shoulder elevation, plus tactile response on the left upper arm (distal to proximal) (Sakata 1975). Because this neuronal activity was elicited by passive joint movement, the activity seems to be related to ownership of one’s own body parts. Furthermore, these neurons may be responsive during self-generated action. Actually, this area is considered to be related to reaching control (Caminiti et al. 1996; Wise et al. 1997).

Neurons that are responsive to both visual and somatosensory stimuli are often observed in some of the parietal and premotor areas (see Fig. 2). These neurons are related to coding the egocentric reference frame and may also contribute to the monitoring of one’s own body. Area 5 sometimes showed visual and tactile bimodal properties, showing congruent directional selectivity in both modalities (Iriki et al. 1996; Sakata 1975), although Area 5 neurons were predominantly sensitive to somatosensory stimuli.

Graziano et al. (2000) reported that arm position-related neurons in Area 5 were influenced by the view of a fake hand. In this experiment, a real hand was occluded under a plate in front of the monkey and a realistic fake arm was then put on the plate. The real arm and fake arm were placed in congruent or incongruent positions. The activity of neurons in Area 5 was influenced by the real
arm position (somatosensory input) and/or by the fake hand position (visual input). The greatest activity was elicited when the positions of fake hand and the real hand matched, coding arm position by integration visual and proprioceptive input. Furthermore, the activity of these neurons was modulated dynamically as in the rubber hand illusion. The monkey was being stroked with the brush on the real hand while it was observing synchronous stroking with the brush on the fake hand. After some minutes of stroking, the sensitivity of neurons for position of the fake hand was enhanced. This plastic change of activity possibly reflected to ownership of the rubber hand, using a visual image of the fake hand, synchronous visual stimulation on the fake hand with tactile stimuli on the real hand.

Obayashi et al. (2000) also found that visual-tactile bimodal neurons of PEa correlated to sense of ownership of body parts. They found that visual receptive fields anchored on the hands followed movement of the hand. They covered the hand with an opaque plate. However, the visual receptive fields remained over the plate above the actual hand. These visual receptive fields then moved over the plate following the invisible hand. They described that the neurons played a role for updating the intrinsic representation of one’s own body state, concerning body image in the brain. The results suggested that activity of Area 5 neurons correlate to a body part-centered frame that moves with action of the body parts.

7.2. Neural Correlates of Expanded Body Image During Tool Use

Plastic modification of corporeal awareness was found during tool use (Maravita and Iriki 2004; Maravita et al. 2003). For example, when we use a stick to explore a ball in a narrow space, the stick becomes a part of our arm. Iriki et al. (1996) recorded neurons of the PEa in the anterior bank of the intraparietal sulcus of a monkey trained to use a rake to get food pellets. As described before, neurons in this area showed tactile receptive field on the hands and visual receptive fields close to the hand. Interestingly, visual receptive fields of some bimodal neurons expanded to include the rake some minutes after the monkey started retrieving food with it. They also found that visual receptive fields on the hand expanded on the virtual hand on the screen (Iriki et al. 2001). These results support that neurons in Area 5 are neural correlates of an expanded body image. For this dynamic change of activity in the neurons, dynamical visual and proprioceptive inputs must integrate in the brain. Actually, Tanaka et al. (2004) reported that neurons in the PEa were activated by passive arm movement and moving visual stimuli. Both proprioceptive and visual responses were directionally tuned and the preferred directions were spatially corresponding with arm movement. Furthermore, Tanaka et al. (2004) inverted arm posture by wrist rotation. Then, they found that the preferred direction of passive joint movement was inverted so as to match the preferred visual direction movement in space. They concluded that the neurons were related to recalibration of the intrinsic image of arm movement depending on the current posture.
Although intrinsic motor representation is very difficult to dissociate from the proprioceptive input, we would like to suggest that plastic modification of Area 5 neurons also reflects the influence of the efference copy. There is evidence that Area 5 may receive reafferent signals from motor areas. Some groups of neurons in Area 5 (PEa) showed modification of activity well before movement onset, up to 280 ms. After dorsal rizotomy from C1 to the T7 level (deafferentation), only this group of neurons was able to respond even in the absence of tactile information (Seal et al. 1982). Kalaska et al. (1983) also reported that the latency of arm movement-related neurons in Area 5 was earlier than movement onset, but later than that of the primary motor cortex. The source of reafferent signals may be the dorsal premotor cortex (F2), ventral premotor cortex (F4), or primary motor cortex (M1), because Area 5 has connections mainly with F2, F4, and M1. We suggest that Area 5 integrates dynamic visual-proprioceptive sensory information and the efference copy, and this process may contribute to plastic change of corporeal awareness during tool use and to both sense of agency and sense of ownership.

7.3. Coding of Body Parts in F2 for Planned Actions

There are many studies that have revealed a contribution of F2, which has connection with PEa, MIP and V6a, to reaching movement (Battaglia-Mayer et al. 2003; Wise et al. 1997). Furthermore, this area also showed bimodal properties. The visual responses were evoked by approaching or receding from the forelimb, although proprioceptive response was predominant (Fogassi et al. 1999). The response also contributes to body part coding. Furthermore, Hoshi and Tanji (2000) revealed that neurons in the dorsal premotor cortex coded body parts (left hand or right hand) for planned action. They emphasized that the dorsal premotor cortex may have a functional role to gather information about both the target and body parts to specify subsequent action.

8. Parieto-Frontal Network III (VIP, F4)

8.1. Coding Peripersonal Space

It is known that VIP in the fundus of the intraparietal area shows bimodal properties. The property of neurons in VIP is well studied (Colby et al. 1993). The tactile receptive fields of the neurons are usually located in the face or head and the visual receptive field is in a location congruent with the tactile receptive field. In many cases, the visual receptive fields are located very close to the body (Colby et al. 1993), namely, in the peripersonal space. These neurons are directionally selective in matching both the visual and tactile modalities. Because the location of the visual receptive fields is often independent of eye position, the neurons are considered to encode head-centered coordinates. There are also neurons that are modulated by the eye position, coding retinotopic coordinates
(Avillac et al. 2005). Therefore, VIP is considered to contribute to the head-centered coordinate system or the transformation from retinotopic to head-centered coordination.

The neural activity in this area seems to encode the location of an object in the peripersonal space (Graziano and Cooke 2006; Holmes and Spence 2004). Furthermore, electrical microstimulation in VIP induced eye blinking (Cooke et al. 2003). It is considered that VIP is related to defensive-like movement to protect the body from attack or collision (Graziano and Cooke 2006). The activity may also include body part representations, showing visual and tactile receptive fields in the different body parts. Our preliminary data revealed that some neurons in the anterior part of VIP were related to both passive arm joint movements and moving visual stimuli, showing the same directional selectivity (Murata 2002). Association of this area with limb movement was revealed by electrical stimulation that induced arm movement (Cooke et al. 2003), which suggested that VIP area may have arm representation, coding multiple frames of reference with different body parts.

8.2. F4 Neurons Coding Spatial Direction of Body Parts Movement

Neurons in the ventral premotor area F4 showed very similar activity with area VIP, which has a strong connection with F4. This area has also visual and somatosensory bimodal properties (Fogassi et al. 1996; Gentilucci et al. 1983). The visual receptive fields of many neurons are anchored on the tactile receptive fields, independent of eye position, although some neurons represent retinotopic coordinates (Mushiake et al. 1997). Furthermore, visual and somatosensory receptive fields in this area moved together with limb movements (Graziano and Cooke 2006; Graziano et al. 1994), which means that the receptive fields of neurons in this area involved multiple frames of reference (Fadiga et al. 2000; Gentilucci et al. 1988), coding body part-centered coordinates.

It is noteworthy to state that neurons in the ventral premotor cortex (F4) were found to be tuned to the spatial direction of active wrist movement even if they inverted the wrist orientation. The preferred direction of active wrist joint movement was inverted so as to match to the preferred direction movement in space (Kakei et al. 2001). As previously described, a similar property in neurons PEA was reported by Tanaka et al. (2004), although the response in PEA was elicited by passive joint movement, suggesting a strong functional correlation between F4 and PEA in coding an intrinsic image of arm movement depending on the current posture.

Finally, it is also interesting that neurons in PFG (the rostral part of 7b) showed bimodal properties around the face similar to those of VIP neurons (Hyvarinen 1981). PFG also has anatomical connections with both VIP and F4. The results suggest that the peripersonal space including body parts was represented in the network constructed in VIP, F4, and PFG.
9. Summary and Conclusions

In this chapter, we discussed multimodal bodily self-representation in the brain. We mainly emphasized that the parieto-premotor network for sensory-guided motor control also contributes to corporeal awareness. The network integrates dynamic multimodal information about the body, and this is a basis of bodily self-recognition. Sense of ownership and sense of agency are two main components of the bodily self. For the sense of ownership, somatosensory and visual feedbacks are necessary. Furthermore, comparison between actual sensory feedback and the efference copy is a core mechanism for the sense of agency. These two components share parieto-premotor networks.

As shown in Fig. 5, actual somatosensory feedback and the efference copy may converge in the parietal cortex. Multisensory parietal areas VIP, PEa, and PFG/PF have connections with each other. We suggest that the streams to the parietal areas VIP, PFG, and PEa from the premotor cortex form node processing for the recognition of bodily self. The parietal cortex receives dynamic visual hand images (visual feedback). In the inferior parietal cortex, the matching occurs among the efference copy, actual visual feedback, and proprioceptive feedback. This is a cue to the sense of agency. According to our data, this matching is mediated by the mirror neurons in the parietal cortex. This matching system may also affect self/other distinctions. In the social cognitive process, it is also important to recognize one’s own body and the bodies of others. These parieto-premotor networks may be involved in the social cognitive process.

Of course, the bodily self is just the starting point of self-recognition. Imaging studies revealed that cortical midline structures are associated with the mental self (Northoff et al. 2006). We need to obtain more experimental evidence to discuss the relationship between these areas and the parieto-premotor network.

Acknowledgments. The authors thank Prof. Yoshiaki Iwamura and Prof. Hideo Sakata for their valuable comments and criticism. Some of our studies in this paper have been partially supported by Grant-in-Aid for Scientific Research on Priority Areas “Emergence of Adaptive Motor Function through Interaction between Body, Brain and Environment” from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References


Graziano MS, Yap GS, Gross CG (1994) Coding of visual space by premotor neurons. Science 266:1054–1057


Hoshi E, Tanji J (2000) Integration of target and body-part information in the premotor cortex when planning action. Nature (Lond) 408:466–470


7
Neuronal Correlates of the Simulation, Execution, and Perception of Limb Movements

Eiichi Naito¹,²

1. Introduction

Humans and animals can control only what they can sense. We therefore cannot precisely control voluntary movements that require elaborate motor control of our body parts unless we can perceive those movements. The neuronal correlates of the mental representation of the “body image”—our perception of the size, shape, movements, and relative configuration of our body parts (Head and Holms 1911)—have long been uncertain. Because brain damage often causes one’s body image to be distorted (Berlucchi and Aglioti 1997; Berti et al. 2005; Sellal et al. 1996), the body image represented as neuronal activity in our brain must be the result of neuronal computation and the integration of multisensory information (somatosensory and visual) about our body (Graziano and Gross 1998). Because people can sense and move body parts without the aid of vision, however, the body image must largely depend on somatic input.

Recent neuroimaging techniques such as functional magnetic resonance imaging (fMRI) allow us to investigate the neuronal representations that are related to various types of our body image by measuring brain activity while people perceive limb movements (Naito 2004) or experience changing body configurations (Ehrsson et al. 2005). This chapter introduces neuronal representations underlying the somatic perception of various types of limb movements (particularly hand movements) and discusses how people perceive these movements and how this perception is related to the control of those movements.

¹National Institute of Information and Communication Technology, Kobe Advanced ICT Research Center & ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
²Graduate School of Human and Environmental Studies, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
2. Sensing Limb Movements

2.1. Sensory Afferent from the Muscle Spindle Receptor

Sensing limb movements is essential to their control (Naito 2004). Deafferented animals and human patients lacking proprioceptive input often show difficulties when performing multijoint limb movements (Bard et al. 1995; Ghez and Sainburg 1995; Sainburg et al. 1995), and deafferented patients also show severe deficits in the acquisition of new motor skills (Gordon et al. 1995).

It has been demonstrated in humans that the sensory afferents from muscle spindle receptors, cutaneous receptors, and joint receptors contribute to the signaling of limb movements to the brain by increasing their activity during passive and active limb movements (Burke et al. 1988; Collins et al. 2005; Edin 2004; Edin and Abbs 1991; Edin and Johansson 1995; Edin and Vallbo 1988, 1990), and the muscle spindle receptors are the ones most sensitive to the direction and speed of limb movements (Burke et al. 1976, 1988; Edin and Vallbo 1988, 1990; Ribot-Ciscar and Roll 1998) (Fig. 1a). The activity of the Ia afferent fibers from these receptors usually increases when the parent muscle is actually stretched (Burke et al. 1976, 1988; Edin and Vallbo 1988, 1990) (Fig. 1b) but also increases when vibratory stimuli are applied to the tendon of the parent muscle (Goodwin et al. 1972; Roll and Vedel 1982; Roll et al. 1989) (Fig. 1c). Vibrating the tendon of a muscle at about 80Hz excites the muscle spindle afferents, and the brain receives and processes their input to produce a sensation of slow movement as if the vibrated muscle were being stretched (vibrating the distal tendon of a wrist extensor muscle elicits a sensation that the hand is flexing) (Fig. 1d).

The limb is immobile during the tendon vibration, and the illusory sensation of limb movement requires no intention of movements or sense of effort. It is thus a genuine perceptual illusion (kinesthetic illusion), allowing us to selectively examine neuronal correlates of the kinesthetic processing that elicits the somatic perception of limb movements (Naito 2004).

To experience a vivid sensation of kinesthetic illusion, one has to completely relax the limb. Because tendon vibration can also evoke a tonic vibration reflex (TVR) when the vibrated muscle is not completely relaxed, the stimulation often produces actual rather than illusory movement in the opposite direction to illusion (Eklund and Hagbarth 1966). Furthermore, as the limbs are immobile during kinesthetic illusions, visual information about the static limb position attenuates the kinesthetic illusion (Hagura et al., unpublished observations). In our series of fMRI studies we therefore investigated neuronal correlates of various types of kinesthetic illusory hand movements in blindfolded and totally relaxed participants.

2.2. Cortical Processing of the Kinesthetic Information

Monkey studies have shown that the primary somatosensory cortex (SI) is one of the targets of afferent input signaling limb movements. Neurons in the cyto-
7. Neuronal Representation of Perceived Hand Movements

Architectonic area 3a (Hore et al. 1976; Huerta and Pons 1990; Philips et al. 1971; Schwarz et al. 1973) and area 2 (Burchfiel and Duffy 1972; Iwamura et al. 1994; Jennings et al. 1983; Schwarz et al. 1973) respond to input elicited by passive or active limb movements.

Motoneuronal cells in the primary motor cortex (M1) respond to the muscle spindle afferent input during tendon vibration of the arm (Colebatch et al. 1990; Porter and Lemon 1993), and some M1 cells receive input directly from the thalamic nuclei (Lemon and Van Der Burg 1979). This physiological evidence suggests that primary roles in the kinesthetic processing of limb movements are played by not only the somatosensory cortices (areas 3a and 2) but also the M1.

We investigated the kinesthetic roles of the human precentral gyrus (M1) in a right-handed male patient with focal damage to the hand region of his left precentral gyrus (Fig. 2a). He had no peripheral injuries in his hands. The damage

---

**Fig. 1.** Muscle spindle activity and kinesthetic illusion. 

- **a** Vibration applied to a tendon excites the Ia afferent fibers of muscle spindle receptors. 
- **b** The Ia afferent fibers usually increase their activity when their parent muscle is actually stretched. 
- **c** Muscle spindle receptors are also sensitive to vibratory stimuli (at around 80 Hz) applied to a tendon of their parent muscle. 
- **d** Vibrating the distal tendon of the wrist extensor muscle elicits a sensation that the hand is flexing.
Fig. 2. Movement perception and kinesthetic illusion after unilateral damage to the precentral gyrus. **a** Focal damage to the hand region of the left precentral gyrus. **b** When the left (intact) hand was passively moved (alternating extension and flexion), these movements were sensed when their speed was as low as 2.2°/s. **c** When the right (paretic) hand was passively moved, the movements at speeds below 3.2°/s were not sensed. **d** A vivid sensation of movement (hand flexion up to 45°) was perceived when the distal tendon of the left (intact) wrist extensor muscle was vibrated. **e** No movement was perceived when the distal tendon of the right (paretic) wrist was vibrated.

present 2 to 3 weeks after a stroke severely impaired voluntary movements of his right hand and fingers. With his eyes closed, he was aware of the wrist movements when the right hand was passively flexed and extended at speeds greater than 3.7°/s, but was unaware of the movements when the hand was moved at speeds below 3.2°/s. He could sense movements of his left wrist, however, when the left (intact) hand was moved at only 2.2°/s (Fig. 2b,c). Interestingly, the lower
limit speed for the perception of wrist movements on the side contralateral to the cortical damage (approximately 3°/s) corresponded to the averaged speed of the illusory movements perceived by 19 healthy participants when a vibratory stimulus (83 Hz, 2 mm; Naito et al. 2002a) was applied to the distal tendon of a wrist extensor muscle. Thus, this speed seems to be an averaged speed of the illusory wrist movements perceived when muscle spindle afferents are excited by vibrating a wrist tendon.

If the human M1 is also related to the kinesthetic processing of muscle spindle afferent input, as the MI of nonhuman primates has been shown to be, left precentral damage would be expected to cause impairment of illusory movement of the right (paretic) hand. As expected, when we subjected a tendon of the patient’s right wrist to vibratory stimuli (83 Hz, 2 mm), the damage impaired the illusory movements. The same stimuli applied to a tendon of the patient’s left (intact) wrist, in contrast, elicited a vivid sensation of illusory movements (hand flexion up to 45°) (Fig. 2d,e).

How can we interpret these neurological findings? We know that the multiple sensory afferents from the muscle spindle receptors and cutaneous receptors contribute to sensing limb movements. It is unlikely that the passive movements at the relatively low speeds (<3°/s) recruit the cutaneous receptors that can normally signal skin deformation related to limb movements to the brain (Collins et al. 2005; Edin 2004), so these receptors cannot contribute effectively to the perception of these slow movements. The perception of passive slow movements thus seems to result solely from the central processing of afferent input from the muscle spindle receptors that are most sensitive to the direction and speed of limb movements (see above). The impairment of perception of the slow movements by precentral damage, together with the impairment of illusory movements elicited by muscle spindle input, strongly suggest that the human precentral gyrus (M1) plays crucial roles in the kinesthetic processing of muscle spindle afferent input that elicits kinesthetic sensation of limb movements (kinesthesia).

Faster movement (>4°/s), on the other hand, probably recruits the cutaneous receptors signaling skin deformation associated with hand movements (Collins et al. 2005; Edin 2004), and the input from these receptors may support the kinesthetic functions in the damaged precentral gyrus. Intact processing of cutaneous input even after the precentral damage is corroborated by the finding that no significant impairment was observed even in the patient’s contralateral (paretic) hand when we tested several cutaneous discrimination tasks (two-point, roughness, shape, and size-discrimination tasks). The cutaneous input might thus be mainly processed in brain areas from the precentral gyrus (most probably in the somatosensory cortex; Areas 3b and 1) and compensate the impaired primary kinesthetic processing in the damaged precentral gyrus.

Taken all together, the neurological findings in the patient with unilateral precentral damage strongly suggest the predominant importance of the contralateral precentral gyrus (M1) in human kinesthetic perception of limb movements. This perceptual role in M1 is highly interesting because the traditional
role assigned to M1 is that of controlling voluntary movements or preparing for movements (= executive locus of voluntary motor control).

2.3. Somatotopical Regions in Multiple Motor Areas Active During Kinesthetic Illusory Limb Movement

If the precentral gyrus (M1) plays primary roles in human kinesthesia, it is more important to know whether the M1 in normal participants is also recruited during illusory limb movements. We therefore used fMRI to scan brain activity while 19 blindfolded volunteers perceived the illusory hand movements when we vibrated the tendon of either the right or left extensor carpi ulnaris. This vibration elicited an illusory movement of simple wrist flexion even though the hand remained immobile. The tendon vibration over the skin excites not only muscle spindle afferents signaling limb movements but also vibrotactile cutaneous afferents eliciting vibratory sensation (Johansson and Vallbo 1983). Thus, to investigate only brain areas related to kinesthetic illusions, we compared the brain activity during illusions with that during the skin vibration produced by applying identical stimuli to the skin over a nearby bone, which did not elicit illusions.

During fMRI scanning, all participants perceived illusory hand movements with no intention, no plan, and no sense of effort of limb movements. The illusory movements of either the right or left hand activated the hand regions of the ipsilateral cerebellum and the contralateral primary motor cortex (M1), primary somatosensory cortex (area 3a), dorsal premotor cortex (PMD), supplementary motor area (SMA), and cingulate motor area (CMA) (Fig. 3). The M1 activation was particularly robust. The somatotopical (foot) regions in the multiple motor areas were also activated when the participants perceived illusions of simple plantar-flexion elicited by vibrating a tendon of either the right or left tibialis anterior (Naito et al. 2007). Involvement of multiple motor areas in human kinesthetic processing is consistent with the electrophysiological and anatomical findings in nonhuman primates that not only M1 cells but also cells in the premotor cortex (PM) (Fetz et al. 1980) and in the supplementary motor area (SMA) and cingulate motor area (CMA) (Cadoret and Smith 1995) react to passive limb movements, and that the M1 has strong functional connections with the motor areas (PM, SMA, CMA, and cerebellum).

Most important, in this fMRI study none of these motor areas was activated when the vibratory stimuli were applied to the skin beside the tendon but not the tendon itself. Such stimuli elicited no illusions, only the sensation of skin vibration, and this sensation activated only the primary and secondary somatosensory cortices (Naito et al. 2007). Thus, these motor areas are specifically related to the kinesthetic processing of limb movements without generating actual movements (Naito and Ehrsson 2001; Naito et al. 1999, 2002a,b, 2005). During the kinesthetic processing of muscle spindle input, it is important for the brain to analyze the displacement/movement of the vibrated limb, that is, to answer the question “where is my limb?” Our findings thus suggest that this “where” function in the human somatosensory system predominantly engages a
network of multiple motor areas, along with somatosensory area 3a. The processing of cutaneous information, in contrast, such as vibrotactile, touch and pressure information from the hands, is often processed to analyze features of a provided stimulus (i.e., to determine “what” the stimulus is). The initial processing of this information from these cutaneous receptors takes place primarily in the somatotopical regions of the primary somatosensory cortex (cytoarchitectonic Areas 3b and 1) (Bodegard et al. 2001). The further analysis of the stimulus features additionally engages high-order sensory association areas. Further analysis of the curvature or shape of a touched object, for example, engages parietal association cortices (Area 2, intraparietal cortex, supramarginal gyrus, parietal operculum) (Bodegard et al. 2001; Roland 1987). This view that kinesthetic processing pre-
dominantly engages motor areas and cutaneous processing predominantly engages the somatosensory cortex also seems to generally fit with the neurological findings described above. In summary, it seems that in humans the kinesthetic processing of limb movement (the “where” function) predominantly involves a network of multiple motor areas distinct from the somatosensory cortices more specialized for cutaneous processing (the “what” function).

Although proprioceptive/kinesthetic (skeletomuscular) processing is still widely thought to occur largely in the primary somatosensory cortex (Areas 3a and 2), our series of findings in patients and normal volunteers (Naito 2004; Naito and Ehrsson 2001; Naito et al. 1999, 2002a,b, 2005, 2007) consistently demonstrates that motor areas (the M1, PMD, SMA, CMA, and the cerebellum) play predominant roles in the kinesthetic processing of limb movements. This observation also means that kinesthetic (sensory) processing is tightly coupled with motor output related to limb movements (Naito 2004).

2.4. Kinesthetic Perception in the Human Motor Cortex

Studies in normal volunteers have shown that the precentral gyrus (M1) plays primary roles in human kinesthesia and that multiple motor areas participate in the kinesthetic processing of muscle spindle afferent input. Thus, it is very important to know whether the M1 in normal volunteers also participates in the kinesthetic perception per se, in other words, to know whether the M1 activity reflects the kinesthetic percept. If this is the case, the M1 should be active when participants perceive a limb movement even when the M1 receives no direct muscle spindle afferents signaling the movement.

To address this question, we utilized a complex type of kinesthetic illusion (Naito et al. 2002b). When the tendon of one hand is vibrated while both hands are passively in contact, blindfolded participants feel that both hands are moving in the same direction (Lackner and Taublieb 1983; Naito et al. 2002b). For example, when the tendon of a person’s right wrist extensor muscle is vibrated while both hands are passively in palm-to-palm contact, that person feels both hands bending leftward. It is as if the illusion of the right hand bending were being transferred to the left hand (“transferred kinesthetic illusion”) (Fig. 4a).

Naito et al. (2002b) found that the amount of the illusory movements of the nonvibrated hand was approximately half of that of the vibrated hand, and that the illusory movements of the nonvibrated hand were well correlated with those of the vibrated hand ($r = 0.73$). These behavioral results indicate that the transfer of illusion is a passive sensory process driven by sensory input from the vibrated hand. The skin input from the palms signaling that the hands are in contact and the muscle spindle afferent input signaling that one hand is flexing at the wrist are interpreted by the brain as indicating that both hands are bending in the direction that the vibrated hand is perceived to bend.

If the M1 activity is related to the somatic perception of hand movement (kinesthesia) per se, the M1 contralateral to the nonvibrated hand must be activated, even in a situation in which the M1 does not directly receive muscle spindle
afferent input from the hand and a participant perceives that the hand is moving. To see if it is, we scanned brain activity in 12 blindfolded normal volunteers while they perceived the transfer of illusory hand movements (Naito et al. 2002b).

When the right and left hands were in contact and the participant felt that the nonvibrated hand was moving, only the M1 contralateral to that hand was activated (Fig. 4b). (When the hands were not in contact, the M1 contralateral to the nonvibrated hand was absolutely silent.) The M1 was the only area that was consistently recruited whether illusions transferred from the right hand to the
left hand or from the left hand to the right hand. We further found by using transcranial magnetic stimulation (TMS) that the excitability of the M1 contralateral to the nonvibrated hand increased when the participants perceived this transferred kinesthetic illusion, and that the level of neuronal excitability was well correlated with the perceived angle of illusory movement of the nonvibrated hand. This finding suggests that the neuronal activity in M1 reflects kinesthetic perception of limb movements per se even in a situation in which the M1 does not receive the afferent information from the muscle spindle receptors (see Kito et al. 2006).

This finding, together with the neurological evidence described above, suggests that the M1, which is the executive locus of voluntary motor control, has a perceptual role in the somatic sensation of limb movements (kinesthesia). As the M1 also plays crucial roles in the kinesthetic (sensory) processing of muscle spindle afferent input, the brain seems to process sensory afferent information by creating a perceptual representation as an outcome of the afferent processing.

The importance of M1 in the perception of limb movements is of particular significance because a similar type of complex kinesthetic illusion in which illusory hand movements affect the perception of the configuration of the body parts (other than limbs) that are attached by the two hands engages parietal association cortices, which are not primarily associated with the perception of limb movements (Ehrsson et al. 2005). When the tendon of the right biceps muscle is vibrated at the same time that a participant holds her nose between the right thumb and index finger, she feels the nose becoming increasingly elongated (Lackner, 1988). And when the tendons of the wrist extensor muscles of both hands are vibrated at the same time that the participants places the palms of both hands in contact with the sides of the waist, they not only feel that the hands are bending inwards but also feel that the waist is shrinking (body-shrinking illusion; Ehrsson et al. 2005) (Fig. 4c). These two types of complex illusions are good examples demonstrating that kinesthetic illusory limb movements transfer to the attached and nonvibrated body parts by eliciting sensations of changes of the body configurations that participants cannot voluntarily produce.

Although the body has elementary somatosensory receptors (e.g., muscle spindle receptors and cutaneous receptors), it has no specialized receptors that provide the brain with information about the size and shape of body parts. Our perception of body image seems to be determined from the pattern of sensory stimulations according to a strict perceptual logic such, so illusory movement of the hand seems to cause changes in the shape and size of other body parts (Fig. 4c). Thus, during the body-shrinking illusion the brain receives the sensory information about the skin contact between the hands and the waist and gets from muscle spindles the information that both hands are flexing. The brain interprets this multiple sensory input to mean that the waist is shrinking.

The body-shrinking illusions exclusively engaged the anterior part of the parietal somatosensory cortex (i.e., the junction between the postcentral and intraparietal sulci) (Fig. 4d), and M1 activity was not associated with this illusion. Furthermore, the level of parietal activity was well correlated with the amount
of perceived body shrinking. Thus, the parietal activity could be related to the illusory feeling that the size of the waist is changing. The parietal activation was located in the border region between somatosensory area 2 and the intraparietal sulcus. The somatosensory areas 2 and 5 in the monkey brain are thought to be higher-order somatosensory areas (Duffy and Burchfiel 1971; Iwamura 1998; Sakata et al. 1973). Cells in these areas are active when different limbs and other parts of the body are touched or moved. That is, they have complex receptive fields that include several body parts (Taoka et al. 1998). Some cells discharge when the hand, arm, or torso is touched (Taoka et al. 1998), and many cells have bilateral receptive fields (Iwamura et al. 1994). Such cells are not found in the primary somatosensory areas 3a, 3b, or 1. Thus, the neuronal populations in the higher-order somatosensory parietal areas can integrate tactile and proprioceptive information from different body parts. The shrinking-body illusion depends on the integration and interpretation of conflicting somatosensory input from different body parts: the kinesthetic input from the vibrated wrist signaling that the hands are flexing, and the tactile input from the palms signaling that the hands are in contact with the physically static waist. This conflict is resolved by recalibrating the relative size and shape of the waist, so one feels that the waist is shrinking as the hands are bending inward. The activity in the higher-order somatosensory parietal areas could reflect this integration and recalibration.

In summary, the M1 seems to have a perceptual role that is specialized for somatic sensation of limb movements (kinesthesia), whereas the somatosensory parietal cortex may participate in the neuronal integration of somatosensory input from different body parts, which integration elicits the sensation of change of body configuration.

2.5. Kinesthetic Processing, Motor Execution, and Mental Simulation of Limb Movements

The somatotopical regions in multiple motor areas (M1, PMD, SMA, CMA, cerebellum) active during illusion of a particular limb seem to well correspond to those active during voluntary movement of the limb (Fig. 3). For example, the regions active during illusory wrist flexion and the regions active during illusory foot plantar-flexion, respectively, correspond to those active during the execution of wrist movements and to those active during the execution of foot movements (Ehrsson et al. 2003; Naito et al. 2007). There are two ways in which this could be interpreted: (1) actual limb movements excite the muscle spindle afferents (see above references), and thus the parts of the motor areas active both during illusion and during execution are related to the kinesthetic sensory processing, or (2) the control of limb movements includes a neuronal process in which centrally originated signals emulate a substitutional sensory (perceptual) effect of the movement by recruiting motor areas. If the second interpretation is correct, one would expect the perception of limb movements to be influenced by a central process for mental simulation (planning) of limb movements (motor imagery) that may share common neuronal substrates related to perception of limb move-
ments. This is expected because it is now well established that motor imagery is an internal simulation (emulation) process of motor execution in which the brain centrally generates motor commands without overtly producing the movements (Jeannerod 1994) and does this by recruiting multiple motor areas that normally participate in motor execution (Ehrsson et al. 2003; Naito and Sadato 2003).

We asked blindfolded participants experiencing illusory hand flexion movements elicited by the tendon vibration of wrist extensor muscle to further imagine that their right hands were flexing or extending. The motor imagery of hand movements (wrist extension or flexion) gave directional influences to the perception of illusory hand flexion, which was enhanced by imagined flexion and attenuated by imagined extension (Naito et al. 2002a). When we used positron emission tomography (PET) to scan brain activity while participants imagined their hand movements (Fig. 5a) and while they perceived illusory hand movements (Fig. 1d), we found that mental simulation of hand movements engaged the same regions in multiple motor areas (contralateral PMD, SMA, and ipsilateral cerebellum) that were activated during the illusions (Naito et al. 2002a) (Fig. 5b,c). Indeed, somatotopical regions in multiple motor areas (M1, PMD, SMA, CMA) active during mental simulation of movements of a limb (hand or foot) (Ehrsson et al. 2003) seem to correspond to those active during illusory movement of the corresponding limb. These behavioral and neuroimaging findings indicate that the efferent processing of centrally originated signals without producing limb movements uses motor activations that are shared by the somatic perception of limb movements elicited by kinesthetic (peripherally originated sensory) processing.

The claim that efferent processing of signals of central origin is associated with perception of limb movements is generally in line with the recent findings that (1) TMS stimulation of the motor cortex that recruits the preserved motor representation of missing limbs of amputees elicits the sensation of movements of the limbs (Mercier et al. 2006), (2) electrical stimulation of the SMA and CMA in awake humans can produce a sensation of limb movement (Fried et al. 1991; Lim et al. 1994), and (3) intended movements (generating motor commands or the sense of effort) of the experimentally paralyzed hands of normal blindfolded participants produce illusions of positional changes of the hands in the direction of intended flexion or extension movement (Gandevia et al. 2006). Thus it is likely that an efferent process not generating any overt limb movements contributes to the perception of limb movements, and that this efferent process includes an emulation of sensory experience substituting for the sensory feedback that would normally arise from the limb movement to be performed (Annett 1996; Frith and Dolan 1997).

It has been proposed that perceived events (perceptions) and to-be-produced events (actions) may share common representation (common coding theory, i.e., an action is coded in terms of the perceivable effects it should produce) (Hommel et al. 2001; Jackson and Decety 2004). Our finding that somatotopical regions in multiple motor areas participate in the motor execution, mental simulation, and kinesthetic processing of movements of their corresponding limbs directly supports this assumption, and has revealed that the motor activation could be a
neuronal basis of the common representation in the case of simple limb movements.

3. Perception of Hand–Object Interactive Movement

Manipulating or handling an external object or tool is one of the skillful human behaviors that require elaborate motor control of our hands. When we manipulate or handle an external object or tool, we can perceive these movements as
motor performances distinct from simple (extension-flexion) hand movements. We therefore expect the perception of hand–object interactive movements to be associated with a specific neuronal representation clearly distinct from that underlying the perception of simple hand movements.

A network of frontal motor and parietal areas plays very important roles in the manipulation of objects and tools (Lewis 2006). Electrophysiological studies in nonhuman primates have shown that neurons that play specific roles in object manipulation are located in the ventral premotor cortex (Murata et al. 1997; Rizzolatti and Luppino 2001), intraparietal cortex (Murata et al. 2000), superior parietal lobule (Iriki et al. 1996), and inferior parietal lobule (Fogassi et al. 2005). Similarly, human neuroimaging studies have revealed that inferior-frontal and parietal cortices are activated when we manipulate or explore hand-held objects (Binkofski et al. 1999; Ehrsson et al. 2000, 2001; Johnson-Frey et al. 2005; Lewis 2006; Stoeckel et al. 2004). Thus, if an action and perception may share the same brain areas, we may expect the perception of hand–object interactive movement to be associated with activity in some of the frontoparietal areas that are specifically active during the manipulation of objects.

We therefore investigated neural correlates of the somatic perception of hand–object interactive movements that occur when a touched object is moving with the hand touching it (Naito and Ehrsson 2006). We did this by using a new type of kinesthetic illusion produced by having blindfolded participants place the palm of the right or left hand on an object (a ball) while we vibrated the tendon of a wrist extensor muscle; this elicited the illusion that the wrist is flexing and the touched object is moving along with the hand (hand–object illusion) (Fig. 6a). The illusory wrist flexion is elicited by muscle spindle kinesthetic input. The skin contact between the palm and the object may stimulate slowly adapting receptors in the skin (Johansson and Vallbo 1983) that may provide information about the

Fig. 6. Hand–object illusion. a Blindfolded participants placed the palm of the right or left hand on a ball while the tendon of a wrist extensor muscle was vibrated; this elicited the illusion that the wrist was flexing and the touched object was moving along with the hand (hand–object illusion). b The hand–object illusion activated only the left inferior parietal lobule (IPL) and area 44 (pink, left hand–object illusion; green, right hand–object illusion). The yellow dot indicates a region active during right and left hand–object illusions.
object’s stereognostic properties. The hand-object illusion thus depends on the central integration of kinesthetic and haptic information from the hand.

We used fMRI to scan brain activity while 12 right-handed blindfolded participants perceived the hand–object illusion (Fig. 6a). In this study, we vibrated the tendon of a wrist extensor muscle to elicit the illusion of hand flexion either when the hand was in contact with the ball (CONTACT-TENDON) or when it was not (FREE-TENDON). In two additional control conditions (CONTACT-BONE and FREE-BONE), we vibrated the skin surface over the nearby bone, which does not elicit illusions. Because the participants perceived hand–object illusions only in the CONTACT-TENDON condition, we could identify activity associated with the hand–object illusion by examining the interaction term in the factorial design [(CONTACT-TENDON—CONTACT-BONE)—(FREE-TENDON—FREE-BONE)].

The hand–object illusion activated the multiple motor areas that are also activated during illusions elicited when the hand is free. It also activated the inferior parietal lobule (IPL; supramarginal gyrus and parietal operculum) and Area 44 in the left hemisphere (Fig. 6b). These areas seem to correspond to the frontoparietal cortices that are activated when we actually manipulate hand-held objects or explore them (see above references). This result supports our hypothesis that the perception of hand–object interactive movements engages the brain areas that are active during the execution of these movements. Also, specific parts of the frontoparietal cortices activated during the perception of hand–object interactive movement may be key neuronal substrates that differentiate the sensory experience of this particular movement from the perception of simple hand movement.

A region in the left IPL was activated during right and left hand-object illusions (Fig. 6b). Because the IPL can be anatomically and functionally designated a high-order sensory association area, it may have a function in the integration of kinesthetic and stereognostic haptic information from the hand. Indeed, sensory (perceptual) roles of parietal regions in hand–object interactive movements, which might differ from the relatively motor-dominant roles of the frontal cortex (area 44), have been suggested by the results of studies in humans (Nickel and Seitz 2005) and nonhuman primates (Murata et al. 1997; Murata et al. 2000). Because left hemisphere lesions that include the IPL often cause ideomotor apraxia (i.e., deficits in association between the internal representation of hand movement and the external representation of an object or tool) (Johnson-Frey 2004), the left IPL activity (Fig. 6b) is probably involved in the integration of kinesthetic (motor) representation of hand movement and haptic information about the object. This view seems to be supported because the coupling of the activity in the left IPL with activity in the ipsilateral (left) intraparietal area (IPS), where the properties of various types of objects are represented in humans (Binkofski et al. 2001; Stoeckel et al. 2004), was enhanced only when the participants perceived the hand–object illusions (not during illusory movements of free hands) (Fig. 7). Thus, the left IPL, in concert with the IPS, plays crucial roles when the human brain incorporates an object or tool into the body image (Naito
and Ehrsson 2006; see also Maravita and Iriki 2004). A specific perception that a hand is moving together with a hand-held object could be the result of activations in multiple motor areas eliciting a perception that a hand is moving and of left IPL activation that links the motor representation of hand movements with the external representation of an object or tool.

We have not yet directly tested the hypothesis that an efferent process specific to a hand–object interactive movement shares common neuronal substrates related to the perception of this particular movement. In the literature, however, the imagining of hand–object interactive movements (i.e., of tool use) seems to specifically engage frontoparietal cortical regions (area 44, IPL, and superior parietal lobule including IPS) that are active during the actual manipulation of frequently used tools (Johnson-Frey et al. 2005). This finding means that the motor imagery (planning) involved in the manipulation of tools activates these areas significantly more than random hand movements do. These frontoparietal regions in the left hemisphere are activated during the planning of both right-hand and left-hand manipulation. These lines of evidences suggest that the fron-
topoparietal activation could be a common representation that is specific to the execution (action), mental simulation, and somatic perception of the hand–object interactive movement. Thus, an efferent process specific to a hand–object interactive movement engages frontoparietal cortices and may emulate the perception of this particular movement without generating any overt movements.

4. Summary

The evidence reviewed in this chapter demonstrates that human motor areas are associated with the central processing of peripherally originated kinesthetic (muscle spindle) input that elicits the somatic perception of limb movements. The fact that the primary motor cortex plays crucial roles in the kinesthetic afferent processing has revealed a critical importance of this motor area in human kinesthesia and has extended our conventional understanding that the somatosensory cortex is a primary contributor to the sensation of limb movements.

A common neuronal process in the brain, one that plays specific and essential roles in a particular type of limb movement (multiple motor areas in simple limb movements or frontoparietal cortices in hand–object interactive movements), is likely to be shared between the execution, mental simulation, and somatic perception of the corresponding movement. The activation common to each type of movement and its perception could be related to the neuronal processing of centrally originated signals that emulates a sensation of the corresponding movement, and thus provides a substitute for the sensory feedback that would arise from the movement to be performed. This neuronal coding common to an action and its perception may allow us to voluntarily control the neuronal representation of our body image (perceived body) to perform a variety of body movements requiring elaborate motor control of our body parts.

Acknowledgements. E.N. was supported in part by a Grant-in-Aid for young scientists from the MEXT Japan and by a grant from the 21st Century COE Program (D-2 to Kyoto University) of the MEXT Japan.

References


7. Neuronal Representation of Perceived Hand Movements


Goodwin GM, McCloskey DI, Matthews PBC (1972) The contribution of muscle afferents to kinesthesia shown by vibration induced illusions of movement and by the effects of paralysing joint afferents. Brain 95:705–748


Head H, Holmes G (1911) Sensory disturbances from cerebral lesions. Brain 34:102–254


Huerta MF, Pons TP (1990) Primary motor cortex receives input from area 3a in macaque. Brain Res 537:367–371


Kito T, Hashimoto T, Yoneda T, Katamoto S, Naito E (2006) Sensory processing during kinesthetic aftereffect following illusory hand movement elicited by tendon vibration. Brain Res 1114:75–84


Lemon RN, Van Der Burg J (1979) Short-latency peripheral inputs to thalamic neurons projecting to the motor cortex in the monkey. Exp Brain Res 36:445–462


Schwarz DWF, Deecke L, Fredrickson JM (1973) Cortical projection of group I muscle afferents to area 2, 3a and the vestibular field in the rhesus monkey. Exp Brain Res 17:516–526
8

Neural Basis of Saccadic Decision Making in the Human Cortex

SAMSON FREYERMUTH¹, JEAN-PHILIPPE LACHAUX², PHILIPPE KAHANE³, and ALAIN BERTHOZ¹

1. Review of Cortical Control Mechanisms of Ocular Saccades

One of the most fundamental decisions in human behavior is the decision made to look at a particular point of space. This basic process is subserved by the saccadic system, which is present in all animal species with an oculomotor system. Numerous studies have shown that several mechanisms have been developed through evolution to produce saccades.

First, reactive saccades to visual or acoustic stimuli can be produced by a very fast pathway from the retina, or the ear, to the superior and inferior colliculi. In the case of visually driven saccades, the neuronal organization of this first pathway is now well documented in animals.

A second pathway for generating visually guided saccades involves a projection from the retina to the visual cortex, followed by transformations occurring in the parietal cortex. From there, several cortical frontal areas, in which the saccadic eye movement originates, contribute to preparing the motor command. Then this command is sent down to the superior colliculus or the brainstem saccadic generator. The different cortical areas involved in saccade generation have been studied in the primates, but in the present chapter we concentrate upon the study of the contribution of different cortical areas in humans (Fig. 1).

Recent advances concerning the recording of neuronal activity in the human brain in epileptic patients have allowed us to identify the role of each of these areas in the control of eye movements. In this chapter, we briefly summarize a few of the brain imaging studies on cortical control of saccades and describe recent new findings obtained with intracranial recordings of brain activity in

¹Laboratoire de Physiologie de la Perception et de l’Action, CNRS, Collège de France, Paris, France
²Inserm U280. Mental processes and brain activation. Centre Hospitalier Le Vinatier, 69500, Bron, France
³Department of Neurology and JE2413 Research Unit, Grenoble Hospital, Grenoble, France
epileptic patients. We particularly focus on the problem of decision making in saccadic control by describing experiments during which subjects had to decide to make a horizontal saccade to a particular direction in space.

1.1. Occipital and Temporal Cortex

The visual cortex is obviously essential for the control of visually guided saccades, and it is also influenced in return by the saccadic system; for instance, even V1 seems to be activated in relationship with the position of the eye (Andersson et al. 2006). However, the visual cortex is traditionally more associated to the transmission of information concerning the retinotopic coordinates of the visual stimulus. The fact that V1 activity can be modulated by static eye position, even without visual input (Law et al. 1998), suggests that the efferent signals from the saccadic system to V1 may contribute to the stability of visual perception (Bodis-Wollner et al. 2002).

In the temporal medial cortex the visual area V5, traditionally thought to be involved in visual motion perception, is also activated during predictive saccades (O’Driscoll et al. 2000). The medial temporal cortex seems also to be implied in the learning and memorization of sequences of saccades, as suggested by the fact that lesions in this area induced a deficit in the reproduction of sequences of saccades toward fixed targets (Ploner et al. 2001).

Moreover, temporal areas are activated during the visual analysis of motion. Indeed, in response to the transition from incoherent to coherent motion, the medial temporal area and the superior temporal sulcus become activated, showing a significant dependence of response strength and latency on motion coherence (Aspell et al. 2005).
1.2. Parietal Cortex

Within the parietal lobe, a large posterior cluster was observed bilaterally during reflex saccades, including superior parietal lobules and posterior segments of the intraparietal sulcus, with an extension to the intersection of the intraparietal sulcus and the transverse occipital sulcus (Simon et al. 2002, 2004). It has been suggested that there is a parietal oculomotor field (PEF) located in the posterior part of the parietal cortex (which corresponded to the monkey lateral intraparietal area, LIP) (Berman et al. 1999). From a strict point of view, specific activity related to saccades in humans within the parietal lobe is localized in the intraparietal sulcus (Grobras et al. 2001; Perry et al. 2000; Petit et al. 1996).

The PEF seems to be principally implied in the triggering of reflex saccades. It is also necessary for visuospatial integration. The study of parietal neglects shows that the PEF contributes to memorize locations of different saccades (Husain et al. 2001). Lesions of the PEF induced a small increase in the latency of saccades (Heide et al. 1998) as well as a decrease in the precision of saccades (Gaymard et al. 2003). In addition, visual attention can be decreased following lesions of this area. Last, during the execution of a paradigm of a memorized sequence of saccades, activity of the precuneus in the parieto-occipital sulcus has been found (Petit et al. 1996; Heide et al. 2001), which seems to be equivalent to area V6a in monkeys. The precuneus also seems to be involved in detection of visual targets (Simon et al. 2002) and in visual attention (as shown by pro-saccades or no-go saccades paradigms) (Brown et al. 2006).

1.3. Frontal Cortex

Two oculomotor frontals areas are crucial for the final configuration of the saccadic motor commands: the frontal eye field (FEF) and the supplementary eye field (SEF).

1.3.1. Frontal Eye Fields

The FEF is located in the dorsolateral premotor cortex, as first demonstrated by Petit et al. (1996), at the intersection between the precentral sulcus and the superior frontal sulcus (Brodmann area 6). It is divided in two areas: the deep FEF and the lateral FEF, as shown first by Lobel et al. (1996) (Grobras et al. 2001; Lobel et al. 2001).

A great number of neuroimaging studies (Petit et al. 1997), lesion effects (Pierrot-Deseilligny et al. 2002), transcranial magnetic stimulation (Olk et al. 2006; Tobler et al. 2002), and focal electric stimulations (Blanke et al. 2003) have demonstrated the functional role of the FEF in the generation of accurate and fast contralateral saccades. For a review, see also Paus et al. (1996).

More specifically the FEF is implied in the triggering of intentional saccades (visually guided, memorized, or predictive) and not so much in the triggering of
reflexive saccades. In addition, it seems that it contributes to the suppression of inappropriate reflex saccades (Cornelissen et al. 2002). It is also implied in ocular pursuit, but it has been actually shown that there may be a separation between the part of the FEF implied in saccade and the part implied in pursuit (Rosano et al. 2002). The FEF is also engaged in ocular fixation (Petit et al. 1995).

Not only is the FEF an area that should be considered as involved in motor functions, it is also involved in cognitive processes such as visual spatial attention (Astafilev et al. 2003) or some aspects of working memory (Curtis et al. 2006).

For example, when the subject moves his or her attention without eye movement, the FEF can be activated, as first demonstrated by Lang et al. (1994) and further confirmed by Nobre et al. (2000). More generally, as suggested by the so-called motor theory of attention (Rizzolatti et al. 1987, 2001), it seems that the same structures are involved in the generation of saccades and in shifts of attention or imagined saccades (Beauchamp et al. 2001; Corbetta et al. 1998a; Corbetta et al. 1998b; Schall et al. 2004; Grosbras et al. 2005). It is therefore seems that the FEF has a fundamental part in the preparation of the saccades and codes both the motor preparation and the intention to make a saccade (Conolly et al. 2002); this is implied in the computation of the retinotopic coordinates of forthcoming saccades, as shown by a study using transcranial magnetic stimulation (TMS) (Tobler et al. 2002). Another study has shown that transcranial stimulations could provoke, when applied at the level of FEF, an increase of visual attention (Grosbras et al. 2002). These suggestions have also been confirmed by neuropsychological data, from patients having lesions in this area, which show that they may not have a major impairment in the execution of saccades. However, these FEF lesions tend the patient toward difficulties in retaining in memory the location of visual targets (Pierrot-Deseilligny et al. 2002).

Last, the FEF is associated with spatial working memory (Postle et al. 2000), being tied to the selection and prospective coding of saccade goals (Curtis et al. 2006). Thus, the FEF is implicated in the control of endogenous allocation and maintenance of visuospatial attention (Corbetta et al. 2002).

1.3.2. Supplementary Eye Field

The supplementary eye field (SEF) is located at the level of the medial wall of the premotor cortex, very close to the paracentral sulcus (Grosbras et al. 1999). The functional role of the SEF has been found both by brain imaging and by neurological evidence. Petit et al. (1996) have shown that it is implied in spontaneous voluntary saccades in darkness. It is also largely implied in the motor programming of saccades, with so-called vestibular “contingent” saccades and with the preparation of sequences of saccades, as shown by Pierrot-Deseilligny et al. (2004) and Isoda et al. (2003) in monkeys. Thus, the SEF essentially controls the triggering of memorized saccadic sequences (Heide et al. 2001).

It has been suggested that during the learning of sequence of saccades two subregions of the SEF are involved: an anterior area would be activated during sequences of visual simulation, whereas a more posterior subarea of SEF would
be activated just before the execution of the motor sequence (Grosbras et al. 2001).

In addition, lesions studies (Parton et al. 2006) have shown a significant deficit of saccade generation when the patient must switch from pro-saccade to anti-saccade, and when he must select the appropriate saccade in a conflict situation. The SEF could therefore play a key role in saccadic control when in a conflicting situation where different oculomotor responses are in competition.

Furthermore, in addition to FEF and SEF, a third area in the frontal cortex is involved in the control of eye movement: the cingulate eye field (CEF). The contribution of this area to eye movement was first found by Petit et al. (1996), followed by Gaymard et al. (1998). This area is recruited in a number of oculomotor tasks when the subject must integrate saccades in general behavior of orientation or decision: hence, the subject plays a role in the conversion of motivations and impulses into actions. It does not seem to be involved in the control of visually guided saccades but rather in the suppression of unwanted saccades (Milea et al. 2003).

Last, some evidence has been provided of involvement of the ventrolateral premotor cortex during the execution of saccades (Heide et al. 2001).

1.4. Prefrontal Cortex

Saccadic decision making also implies the prefrontal cortex. In general, the prefrontal cortex is essentially implied in complex behavior requiring specific abilities. One of the fundamental principles of the operation of the prefrontal cortex seems to be the adaptability of the neural code, tied with the working memory processes (Funahashi et al. 2006). Most of the neurons in this area modify their properties depending upon the specific information that they carry, producing a dense and distributed network of the different inputs, action, reward, and other relevant information (Duncan 2001). Thus, this distributed network permits establishing the proper mappings between inputs, internal states, and outputs needed to perform a given task, such as tapping a sentence on prefrontal cortex on your computer (Miller et al. 2001). The prefrontal is therefore a source of top-down control signals, permitting the development of goal-directed behaviors (Miller et al. 2005), following a temporal sequence control (Knutson et al. 2004; Koechlin et al. 2003).

Another principle of the functionality of the prefrontal cortex is modifying or inhibiting future responses, which requires the temporal integration of events aimed at by intentions of actions (Berthoz 2003). In monkeys it has been shown, by recording in the dorsolateral prefrontal cortex (Constantinidis et al. 2002), that some inhibitory interaction exists between prefrontal cortex neuronal populations activated at different temporal moments. The inhibitory aspect of the prefrontal cortex would therefore play a role in the control of the temporal activity of neurons as well as the temporal flux of information to be treated.

An oculomotor area located just anteriorly with respect to the SEF seems to play a distinct functional role: it is the pre-SEF. The hypothesis of a pre-SEF
area has flowed from the data of Luppino et al. (1993) on sequences of hand movements, showing the activation of a pre-supplementary motor area (pre-SMA) for the learning of sequences of actions.

The location of the human pre-SEF has been shown by Grosbras et al. (1998). This area is involved in the oculomotor preparation and in the inhibition of saccades (Coull et al. 2000; Yamamoto et al. 2004). This pre-SEF seems to be specially associated with the learning process of this motor preparation (Pierrot-Deseilligny et al. 2003). In addition, the pre-SEF plays a role in covert reorienting and inhibition of return (Lespien et al. 2002).

The decision to make a saccade seems also to involve the dorsolateral prefrontal cortex (DLPFC). The DLPFC is located in the medial frontal gyrus (Brodmann Area 46). It plays a central decision role in oculomotor behavior (Pierrot-Deseilligny et al. 2005). It is also essential in the combination between visual information and more general sensory information coming from the parietal cortex and knowledge and the aims and general goals for the generation of an action (Schall et al. 2001). It has therefore been suggested that there is a prefrontal eye movement field (PFEF), which is the term that we use in the following sections of this chapter. The involvement of the prefrontal cortex in the control of eye movement has been revealed by experiments concerning the anti-saccade task (Guitton et al. 1985) and the inhibition of unwanted saccades, mediated by the superior colliculus.

The PFEF showed increased activity during the instruction of anti-saccades as compared with pro-saccades (De Souza et al. 2003). The PFEF is therefore activated during memorized saccades, playing a role in memorization and restitution of oculomotor movements (Ozyurt et al. 2006).

2. Cortical Contribution to the Production of Saccade: From Decision to Execution

The execution and decision to make an eye movement can be studied by functional magnetic resonance imaging (fMRI), but these techniques do not give the precise dynamics of the neurons underlying the production of saccade.

The essential effort is currently to understand the temporal coordination of this network, that is to say, the time-course of neural activations during saccade generation. Oculomotor execution and decision can be studied only by integrating dynamic component. This dynamics links can only be observed with (extremely rare) direct recordings of the neural activity (for example, in the human FEF and SEF). Very few studies have benefited from such circumstances to report the time-course of neural activation in the frontal eye fields during saccades in humans (Lachaux et al. 2006; Sakamoto et al. 1991; Yamamoto et al. 2004).

For ethical reasons, this dynamical network cannot be elucidated by simultaneous single-cell electrophysiological recordings and imaging techniques in humans. However, intracerebral stereo-electroencephalography (SEEG) offers a valuable tool to measure integrated electrical phenomena at the millimeter scale (i.e.,
comparable to the fMRI spatial resolution), and a detailed analysis of its temporal dynamics could shed some light on some components of the neurovascular coupling in humans performing an oculomotor task. In fact, the SEEG signals clearly reveal the multidimensionality of the ensemble neuronal responses, which consist of event-related potentials (ERPs), induced synchronizations and desynchronizations in distinct frequency bands. In the following study, the local neural activity was recorded in epileptic patients who had been stereotactically implanted with multicontact depth electrodes to monitor intractable epileptic seizures (we describe here the method used in the Department of Neurology at Grenoble Hospital, France).

2.1. Stereotaxic Electroencephalography (SEEG): Methods and Data Analysis with Time–Frequency Maps

The patients suffered from drug-resistant partial epilepsy and were candidates for surgery. MRI was done for every patient, and the areas of interest showed no lesions. Intracerebral exploration allowed us to localize precisely the epileptic focus, with the help of stereotactically implanted multilead depth electrodes (Kahane et al. 2004), a dozen of semirigid electrodes (0.8 mm diameter), each including 10 to 15 recording dots (2 mm long, separated by 1.5 mm) (Dixi, Besançon, France). The recording dots are identified by a stereotaxic method, anatomically localized with the help of the Talairach and Tournoux proportional atlas (Talairach and Tournoux 1988) (Fig. 2). The coordinates of each electrode contact are given following these references: origin (anterior commissure), anteroposterior axis (anterior commissure-posterior commissure), and vertical axis (interhemispheric plane).

This depth electrode records the spontaneous activity of the cerebral areas of interest; selection of sites to implant the electrodes is made entirely for clinical purposes, with no reference to experimental protocols. During the 2 or 3 weeks of observation (video-SEEG), the patients can give their informed consent to participate in the experiment (all the selected patients had normal vision without corrective glasses). After data processing, the results are shown with a time–frequency map (Fig. 3).

2.2. Neuronal Dynamic of FEF and SEF Activity During a Visually Guided Horizontal Saccade

The purpose of the study (Lachaux et al. 2006) was to explore brain activations during saccade generation. The analysis was focused on the recording sites exploring the FEF and SEF regions, as assessed by several functional imaging studies (Beauclamps et al. 2001; Grobras et al. 1999; Heide et al. 2001; Lobel et al. 2001).

The time–frequency (TF) analysis (see Fig. 3) revealed that the generation of saccades was simultaneous with energy modulations in several frequency bands.
Those modulations extended in a broad frequency range up to 200 Hz (upper limitation of the analysis, 512-Hz sampling rate). Following a previous study (Crone et al. 1998), several frequency bands are considered separately: beta (15–30 Hz), low gamma (30–45 Hz), high gamma (60–90 Hz), and very high gamma (110–140 Hz). The event-related potentials (ERP) were also computed to examine EEG modulations phase-locked to the saccade onset. Given the short time span of interest in this study (mostly the 100–200 ms separating the cue from the saccade onset), the alpha band was not examined.

The timing of the responses in the very high gamma band (VHG), compared to the timing in the other bands, seemed more reproducible within each brain region and more closely related to the saccade timing (gradual increase of activity peaking at the saccade onset) (Fig. 4). In fact, the generation of pro- and anti-saccades coincided with focal energy increases in the gamma range in FEF and SEF regions (Fig. 5). The time-course of the increase matched well the timing of the saccades, particularly in the high portion of the gamma range (typically above 100 Hz, i.e., VHG). This activation shows no lateralization effect.

In the FEF, we could distinguish (a) VHG activations that responded to the direction of the saccades and (b) VHG activations that responded to the direction of the cue. The results also suggest that SEF activity begins before the appearance of the instruction to execute the saccade, whether FEF activation is evoked at the onset of the presentation of this instruction. SEF seems to be particularly
activated before anti-saccades, in accordance with the data obtained in the monkey (Amador et al. 2004; Munoz et al. 2004).

These results also confirmed the fact that each frontal lobe is involved in the production of saccades both to the left and to the right (Herter et al. 2004).

2.3. Involvement of the Frontal and Prefrontal Cortex in Saccadic Decision

To study the role of the frontal and prefrontal cortex in decision processes, the experimental task (Fig. 6) was divided into three conditions (A, B, C), as follows.

---

Fig. 3. Schematic representation of stereo-electroencephalography (SEEG) data processing and analysis. (From Lachaux et al. 2006)
Fig. 4. Energy profiles during the execution of a saccade. Recordings on frontal eye field (FEF) and supplementary eye field (SEF) contact point electrodes. ERP, event-related potentials; sac, saccade. (From Lachaux et al. 2006)

Fig. 5. Time–frequency representation of a saccade in FEF and SEF contact point electrodes. RH, right hemisphere; LH, left hemisphere. (From Lachaux et al. 2006)
A: In one condition, the three targets are shown on the screen, a central fixation target and two lateral targets at an eccentricity of about 7%. In the first task, the subject has to fixate the central target, then at the place of this central fixation target appears an arrow, which was the central cue indicating to the subject to make a saccade either to the left or the right target. The subject was instructed to keep the fixation while the arrow was appearing, and therefore not to make the instruction for saccade. Then, the fixation point reappeared for a duration of 3750, 5750 or 7750 ms; during this period, the subject had to keep in mind, in working memory, the direction of the instructed saccade, but not yet make the saccade. Then the arrow reappeared and the subject had to execute the saccade. The task was thus subdivided into four components: fixation, instruction, fixation, execution.

B: In the second condition, a decision was required. First, the subject was shown the three targets and had to fixate the central target, as in the first condition; then, instead of a fixation target a lozenge appeared, telling the subject that he had to decide to make a target either to the left or to the right, but the subject was instructed to not yet move the eyes. Then, the fixation target appeared, and the subject was instructed to keep in memory the decision he had made during the preparation phase. Then, a double arrow appeared and the subject had to execute the saccade in the direction that he had decided.

C: Two control tasks: (a) The subject first fixes, then executes a saccade to the left or to the right following the appearance of a central arrow. There is no preparation phase. (b) The subject fixates a central target that suddenly changes color, but to the subject is instructed to keep the fixation, without making a saccade.

This paradigm allowed us to distinguish between simple fixation, preparation, execution, and decision making.

In previous studies in which this paradigm was used, during fMRI recordings of brain activity (Milea et al. 2005; Pierrot-Deseilligny et al. 2005), several areas of the prefrontal cortex (frontopolar, DLPFC) were shown to be involved in the preparation and decision of horizontal saccades.
We could apply this paradigm to two epileptic patients who had electrodes located in bilateral SEF, bilateral FEF, right DLPFC, right rolandic operculum, and bilateral cingulate gyrus. The data were analyzed according to the methods already described; time–frequency maps were constructed and the amplitude of the activity was measured. Samples of the recorded data are shown in Figs. 7–10. During execution, we observed an activation in all conditions of the bilateral

**Fig. 7.** Time–frequency map of the right-FEF activity during the execution of a left saccade (0 represents beginning of the saccade)

**Fig. 8.** Time–frequency map of right-FEF activity during the preparation of an imposed left saccade (nonexecuted) (0 represents extinction of the cue)
8. Saccadic Decision Making

Fig. 9. Time–frequency map of right-FEF activity during the preparation of a free left saccade (nonexecuted) (0 represents extinction of the cue)

Fig. 10. Time–frequency map of right-DLPFC activity during the preparation of a free left saccade (nonexecuted) (0 represents extinction of the cue)
FEF, the bilateral SEF, the cingular gyrus, and the rolandic operculum. During the preparation condition, we observed a bilateral activation of the FEF and the cingular gyrus; however, in the decision condition we observed in addition a bilateral activation of the FEF of the cingular and of the PFEF. Therefore, this preliminary work shows a specific activation of the PFEF in this oculomotor task when the decision is needed. These preliminary results in SEEG seemed to confirm the previous study in fMRI.

References


Part III
Memory as an
Internal Representation
1. Introduction

Neural representations of spatial information are substrates for behaviors that range from simple limb movements and basic locomotion to sophisticated navigation through complex environments. The processing of different types of spatial information, including the storage and recall of related neural representations, is integral to the ability to navigate through and interact with the external environment. Finding food, shelter, and potential mates requires an animal to develop an understanding of the spatial relationships between itself and numerous objects and goals within its environment. Two forms of information necessary for spatial navigation are the knowledge of one’s location within an environment and directional heading, or orientation. This information is represented by neural activity distributed over several nuclei within the limbic system and neocortex. Furthermore, the ability to integrate, store, and recall these representations is essential for long-term survival strategies. This chapter discusses the neural representations of spatial location and orientation and how they can contribute to a spatial memory system.

The study of spatial navigation and spatial memory has a rich history that reflects the ubiquitous nature of this neural construct (for review, see O’Keefe and Nadel 1978; Redish 1999). Advances in neuroscience have allowed experimenters to study spatial representations as model systems for information transformation in the brain. Arrays of chronically implanted electrodes have been developed to monitor the activity of multiple neurons simultaneously in freely behaving animals (Wilson and McNaughton, 1993; for review, see Buzsaki 2004). Developments in functional magnetic resonance imaging (fMRI) technology, as well as neurophysiological recording in humans, have provided access to human neural activity during the exploration of virtual environments (Burgess et al. 2002; Ekstrom et al. 2003). More recently, polysynaptic tracers have been used to identify synaptically connected neural circuits supporting directional

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
orientation (Brown et al. 2005a), and analyses of the temporal components of gene transcription have identified neuronal circuits activated during specific spatial behaviors (Guzowski et al. 2005).

Representations of spatial information in the brain are organized in frameworks that are centered on the body (egocentric) or the outside world (allocentric). Egocentric representations include head- or limb-centered representations such as those found in parietal cortex (Colby and Duhamel 1996), while allocentric representations are organized relative to the environment and therefore do not change as the animal locomotes. Both allocentric and egocentric information can be used in support of spatial behavior (Wang and Spelke 2002), although allocentric information is thought to be the basis for most spatial navigation. The focus of this chapter is the neurophysiology of three different allocentric neural representations within the limbic system, two of which encode location-specific information and one that encodes directional heading.

Two types of neurons in the limbic system have been identified whose activity encodes allocentric spatial information in support of spatial memory and navigation. In 1971, hippocampal pyramidal cell activity was found to be highly correlated with the location of a freely behaving animal in space (O’Keefe and Dostrovsky 1971). These cells are referred to as “place cells,” and their discovery fostered intense interest because they could provide the basis for a neural representation of a “cognitive map” (Muller 1996; O’Keefe and Nadel 1978; Tolman 1948). Over time, place cell activity has been found to encode more than just the spatial location of the animal. Indeed, these cells have been shown to encode nonspatial information such as contextual information related to the geometric and behavioral aspects of the experimental environment. Recently discovered neurons in the rat entorhinal cortex, called “grid cells,” encode allocentric spatial information (Hafting et al. 2005) in a complementary fashion such that complete external environments may be represented by topographically related neural ensembles.

The representation of directional heading in the brain is accomplished by “head direction cells,” which are found in several nuclei throughout the limbic system (Ranck 1984; Taube et al. 1990a). Head direction (HD) cell activity encodes the direction in which an animal’s head is pointing, independent of the animal’s location or body position. This activity is thought to represent the animal’s perceived directional heading and provides an animal with a sense of direction that, although sensitive to external allocentric information, is constructed primarily from idiothetic, or internally generated, information (Bassett and Taube 2001; Sharp et al. 2001a).

The extent to which these complementary neural representations are utilized in support of spatial navigation and memory is a subject that generates considerable interest. Understanding how the brain accomplishes complex signal transformation from one reference frame to another and how this information is subsequently processed to enable accurate navigation has been a major goal in the spatial cognition field. Navigation represents a high level cognitive process that is amenable to experimentation in animal models and should provide key
insights into the neural mechanisms of other high-level cognitive processes. Furthermore, memories of spatial locations are fundamental to accurate navigation, linking the study of navigation with memory. The spatial representations discussed here (place, grid, and HD cells) are not tied to only one form of memory, as both semantic and episodic memories can incorporate such spatial information. This chapter briefly reviews our current understanding of the place, grid, and HD cells systems and discusses the interactions of spatial representations with memory systems.

2. Place Cells

It is generally accepted that hippocampal place cell activity represents more than simply the current location of the animal, although the exact correlates of this activity remain unclear (Eichenbaum et al. 1999; Leutgeb et al. 2005a; O’Keefe 1999). Recording CA1 pyramidal cells in freely moving rats led to the finding that the peak firing rate of some cells occurs when a rat is located in a specific area of an environment, whereas the firing rate is very low when the rat is located elsewhere (O’Keefe and Dostrovsky 1971). Figure 1 illustrates the location-specific activity of place cells; the areas of the environment in which the place cell’s firing rate is the highest is referred to as the cell’s “place field” or “firing field.” Initial characterization of the firing patterns of hippocampal pyramidal cells was performed by O’Keefe (1976), who suggested that the location-specific activity may be related to (1) a simple sensory stimulus or some combination of sensory information in the environment, (2) a combination of sensory stimuli and some aspect of the animal’s behavior, or (3) all the environmental stimuli that converge on the animal at a particular location in an environment. This final view gave rise to the cognitive map theory of hippocampal function (O’Keefe and

Fig. 1. Firing rate versus location plot for a typical place cell in the CA1 hippocampus. The firing rate for each location is depicted using the gray scale shown on the right. (From Muller 1996, with permission)
Nadel 1978), which postulated that hippocampal activity provides a neural representation of allocentric space in support of navigation and spatial memory. Systematic study of hippocampal activity in freely behaving rats exploring cue-controlled environments has provided evidence that the activity can be correlated with all three possibilities (Muller 1996). Other neurons throughout the hippocampal formation and limbic system show place related activity, including granule cells in dentate gyrus (Jung and McNaughton 1993), CA3 pyramidal cells (Lee et al. 2004), subicular pyramidal cells (Barnes et al. 1990; Sharp and Green 1994), parasubiculum (Taube 1995b), entorhinal cortex (Frank et al. 2000, Fyhn et al. 2004; Hafting et al. 2005; Hargreaves et al. 2005; Quirk et al. 1992), and postrhinal cortex (Burwell and Hafeman 2003; Hargreaves et al. 2005), as well as medial prefrontal cortex (Hok et al. 2005). In all brain areas where place cell activity has been reported, there does not appear to be any readily identifiable topographical organization of the place cell representation among physically proximal place cells (Redish et al. 2001; but see Hampson et al. 2002).

2.1. Basic Properties

Location-specific activity of place cells is evident upon a rat’s entry into an environment, and a place cell’s firing field is relatively stable over long periods of time so long as the surrounding environmental cues remain stable (Thompson and Best 1990). A recorded hippocampal pyramidal cell in CA1 has an approximately 1 in 3 chance of exhibiting a place field in a given environment (Wilson and McNaughton 1993). Although a place cell may exhibit multiple place fields in an environment, the usual number of place fields per cell is one (Muller et al. 1987). There has been no systematic relationship demonstrated between the multiple place fields of a given place cell. The animal’s present location is represented by a population of cells with overlapping place fields whose quantity depends on the number of place cells active in any given environment (Harris et al. 2002). This population activity can be used to predict the animal’s precise location (Wilson and McNaughton 1993). Further, Muller and Kubie (1987) demonstrated that place fields were optimized when spikes predicted where the animal would be about 100 ms in the future, suggesting that place cell activity may encode where the animal anticipates it will be in about 0.1 s.

Each environment is thought to be represented by different combinations (or subsets) of place cells depending on the internal behavioral state or sensory context (Markus et al. 1995; Quirk et al. 1990) and is discussed further below. One important property of place cell discharge is that cell activity is closely associated with the hippocampal theta rhythm, a prominent 8- to 12-Hz oscillation observed in the hippocampus during active exploration. As an animal traverses a cell’s place field, cell firing occurs increasingly earlier in the theta-phase cycle; this phenomenon is known as phase precession O’Keefe and Recce (1993). More recently, Huxter and colleagues (2003) reported that the place cell rate code (firing frequency) was independent of the cell’s temporal code (phase precession), suggesting that the time of firing and rate of firing are dissociable and
may encode variables such as location within a place field and running speed through the field, respectively.

The influence of environmental features on place cell activity has been widely studied. Place cells recorded in rats exploring a radial arm maze were often found to have directional properties; some cells fired only when the rat was moving down one arm in one direction (McNaughton et al. 1983a,b; Muller et al. 1994; O’Keefe 1976; O’Keefe and Recce 1993; but see Battaglia et al. 2004). This finding indicates that the hippocampal representation can incorporate the rat’s direction of movement as well as its location. An experimental paradigm introduced by Muller and colleagues (1987) allowed for precise control over the sensory information available to the animal, while maintaining relatively constant behavior across the entire environment. The paradigm, which allowed manipulations of distinct and sparse environmental cues as animals foraged for randomly placed small food pellets, became standard in many subsequent experiments with freely behaving animals. This setup also automated the tracking and recording of rat location and neural activity as the animals explored a fixed, high-walled, cylindrical environment. A single visual cue was introduced into the environment as a white cue card that covered ~100° of arc along the cylinder wall. Rotations of the cue card resulted in corresponding and predictable rotations of place field locations, indicating that the hippocampal representation of location was influenced by visual landmark information (Muller et al. 1987). Removing cues did not abolish place cell activity in the hippocampus (Muller and Kubie 1987; O’Keefe and Speakman 1987), nor did turning off the lights while rats foraged for food (Markus et al. 1994). When multiple cues were available in an environment, hippocampal neurons tended to be more sensitive to distal cues than local cues (Hetherington and Shapiro 1997). Distal cues appear to be more stable as an animal locomotes through an environment and are therefore more reliable for navigation. Experiments that included cue-conflict situations have produced a wide range of complex place cell responses that appear to depend on the particular manipulation (Brown and Skaggs 2002; Knierim and Rao 2003; Tanila et al. 1997; Yoganarasimha et al. 2006).

In addition to being affected by simple visual cues, place cell activity is sensitive to the geometric features of an environment. Increasing the size of the recording cylinder caused place field size to scale up to a certain point, after which the location-specific activity “remapped,” or changed unpredictably to a new pattern of place fields (Muller and Kubie 1987). [Note: place cells are said to “remap” when the pattern of place fields among a population of cells is altered as a result of a change in environment or context.] Changing the shape of the environment from a cylinder to a square also elicited remapping (Muller and Kubie 1987; Sharp 1997), whereas adjusting the relationships between the walls of a square environment resulted in a range of responses from minor alterations to complete remapping (O’Keefe and Burgess 1996; Wilson and McNaughton 1993). “Morphing” the shape of an environment from a circular to a square-shaped arena via several intermediate shaped enclosures results in a smooth change between the representations observed in the circular and square environments (Leutgeb et al.
2005c), although Wills et al. (2005) reported only abrupt transitions between square and circular-shaped enclosures. The authors of the latter study interpreted these results to suggest that hippocampal processing involved attractor-level dynamics. Differences in protocols between the two studies may have led to the different results. Although Wills et al. first familiarized rats to both the circular and square-shaped environments and then presented random intermediary shapes, Leutgeb et al. used a protocol in which the animal was presented with a gradual and continuous transition over eight steps between the two differently shaped environments. Finally, place cell representations are also sensitive to the environmental context, and this aspect is discussed further below.

2.2. Sensory Contributions

Place cells are located in high-level association areas that are several synapses removed from incoming sensory information. Various sensory modalities contribute to the place cell signal, supporting the role of the hippocampus as a high-level association system (Poucet et al. 2000). Blind rats (from birth) exhibit place fields upon recognition of an environment via other sensory information (Save et al. 1998), but the visual cortex is clearly an important contributor to the representation in intact animals (Paz-Villagran et al. 2002). Interestingly, lesions of the vestibular labyrinth abolish normal hippocampal place cell firing (Russell et al. 2003; Stackman et al. 2002), whereas place cell activity is less affected in rats that are simply disoriented (for review, see Smith et al. 2005). Both tactile and olfactory cues can also influence the stability and location of place fields (Save et al. 2000; Shapiro et al. 1997). Consistent with the view that place cells are influenced by a wide variety of sensory stimuli, recent experiments showed that a tilted surface could also provide a reference cue for place cells, indicating that combined tactile and otolith signal information can influence cell activity (Jeffery et al. 2006).

2.3. Correspondence Between Neural Activity and Spatial Behavior

If the location information encoded by hippocampal place cells supports spatial navigation and memory, then it would be expected that their activity would be related to navigational decisions and performance on spatial tasks (O’Keefe and Nadel 1978; Muir and Taube 2002a). Place cell activity observed during error trials provided strong evidence that cells encoded the place where the rat perceived it was located, rather than firing as result of a certain conjunction of cues (O’Keefe and Speakman 1987). Lenck-Santini et al. (2001a,b) found a tight correspondence between accurate performance and place cell activity. When place fields were made to be out of register relative to their originally established position, the animal’s performance was poor. In a second paradigm, these authors (Lenck-Santini et al. 2002) again tested the extent to which performance corresponded to the spatial information derived from hippocampal place cells. Using
the cylinder with the single prominent visual cue card described above, animals were first trained to retrieve a reward that was opposite and far away from the cue card. Following recording of place cells in the cylinder the authors rotated the visual cue in the presence of the animal and then monitored their subsequent search behavior and place cell activity. The authors observed few correct responses to the card-referred correct area, and found that animals tended to search in areas that were in register with the spatial information derived from the recorded place cells (i.e., place field referred to areas).

In contrast, Jeffery et al. (2003) reported preserved performance in a hippocampal-dependent spatial task despite complete place cell remapping that was brought about by changing the overall environmental background from a black box to a white box. The authors reasoned that if place cells showed remapping, then behavioral performance should be poor, because the spatial maps that were learned in the first box were no longer operative in the second box. Given that remapping of place fields occurred, it was surprising to see that performance was unaffected. Perhaps the best demonstration of how place cell activity is tied to spatial memory showed that the long-term stability of place fields was correlated with successful spatial task performance (Kentros et al. 2004). Such results led these authors to conclude that place cell activity is a neural correlate of spatial memory.

The relationship between navigational decisions and place cell activity depends on the experimental manipulation and the animal’s previous history (Bower et al. 2005; Ferbinteanu and Shapiro 2003; Frank et al. 2000; Kobayashi et al. 2003; Wood et al. 2000; cf. Lenck-Santini et al. 2001a). For example, Wood et al. (2000) recorded from place cells as rats performed a continuous alternating T-maze task where they alternated between left and right turns to retrieve a reward. The authors reported that two-thirds of the cells that discharged on the central stem (before turning left or right) responded differentially depending on which way the animal was going to turn at the subsequent choice point. Other place cells recorded in this study did not show this distinction and always fired in the stem, regardless of the turn response that was evoked. Ferbinteanu and Shapiro (2003) extended these studies. Using a plus maze they showed that place cells could be characterized according to whether their firing patterns were related to where the rat was about to go (prospective coding) or where the rat just came from (retrospектив coding). Prospective coding cells discharged in the start arm, but only when the rat subsequently turned one direction and went down one particular arm of the maze; when the rat turned the other direction and traveled down the alternative arm, the cells did not fire in the start arm. Retrospective coding cells discharged on a goal arm, but only when the rat was coming from one of the two start arms; when the rat came from the other start arm, the cell did not fire in the goal arm. Thus, place cells appeared to encode information regarding recently entered locations as well as locations about to be entered, indicating that memory and planning are important aspects of the place cell representation.

These differences in place cell activity may account for some aspects in the high degree of variability reported in the frequency of firing for place cells
(Fenton and Muller 1998; but see Leutgeb et al. 2005b). These authors performed a quantitative analysis of place cell firing as an animal made repeated passes through the cell’s place field. Cells fired more randomly than predicted by a Poisson model, indicating that additional factors must influence the rate of firing. Finally, although place cells have often been found to encode goal locations, or at least be concentrated around goals (Breese et al. 1989; Hollup et al. 2001; Poucet et al. 2004), this relationship remains unclear as some studies that have changed goal locations have not found an affect on place cell activity (Speakman and O’Keefe 1990). The importance of having a specific neural representation for a goal has been described by Poucet and colleagues (2004). Recently, Hok et al. (2005) identified cells in the medial prefrontal cortex with location-specific correlates; interestingly, a much higher percentage of these cells than predicted by chance had their place fields located near the goal. It is possible that the prefrontal cortex interactions with the hippocampus serve to guide the animal, navigationally and/or motivationally, from its current location to a goal location. Clearly, more work is needed to develop our understanding of the critical relationship between representation and behavior.

2.4. Hippocampal Representations of Spatial Information in Nonhuman Primates

Reconciling rat, nonhuman primate, and human data regarding hippocampal representations has been understandably difficult because of technical and ethical constraints. Recording techniques in freely moving monkeys are being developed (Ludvig et al. 2001; Sun et al. 2002) that should provide a wealth of new data regarding correlates of hippocampal activity. Although location specific responses of hippocampal neurons in freely moving monkeys have been observed that are similar to rat place cells (Ludvig et al. 2004), other experimental manipulations have only uncovered hippocampal neurons whose activity is highest when the monkey is looking at a certain location in space. This type of cell has been referred to as a spatial view cell (Rolls and O’Mara 1995; Rolls 1999). Hippocampal cells observed while the monkey is “driving” itself around a small room in a cart will discharge when a monkey is in a particular place and looking in a particular direction; these are similar to spatial view cells (Matsumura et al. 1999; Ono et al. 1993). Ono’s group also reported that hippocampal neurons recorded in restrained monkeys were sensitive to location in both real and virtual environments (Hori et al. 2003, 2005).

2.5. Hippocampal Representations of Spatial Information in Humans

Several techniques have been developed to illuminate human hippocampal spatial representations. Invasive neurophysiological recordings in human patients who navigated through a virtual environment showed that hippocampal cells
responded at specific locations within a virtual environment as well as to specific views (Ekstrom et al. 2003). However, the nature of the location-specific firing was not very robust and there did not appear to be any systematic decrease in firing away from the peak firing rate in surrounding areas as occurs in place fields recorded in rats. On a broader scale, fMRI studies have examined hippocampal activity as subjects perform tasks requiring a range of spatial information processing. In these cases, activation of the hippocampus coinciding with locations in virtual place navigation or processing of spatial information has been observed (Burgess et al. 2002). Although the hippocampus appears to be activated in these visual spatial tasks, the nature of these tasks is far different than actual navigation in which a subject actively locomotes or turns its head and/or body. Navigation involving locomotion and movement engages the use of idiothetic cues, whereas a visual spatial task performed when looking at a computer screen does not engage these types of cues.

2.6. Generation of the Place Cell Signal

Given that place cells have been identified throughout the hippocampus proper (dentate gyrus, CA3, CA1), subicular complex, and some parahippocampal areas, the question arises as to where this signal is first generated. Table 1

<table>
<thead>
<tr>
<th>Lesion area</th>
<th>Recording area</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentate gyrus</td>
<td>CA1, CA3</td>
<td>Intact place cell activity</td>
<td>McNaughton et al. 1989</td>
</tr>
<tr>
<td>CA3</td>
<td>CA1</td>
<td>Intact place cell activity</td>
<td>Mizumori et al. 1989</td>
</tr>
<tr>
<td>CA3</td>
<td>CA1</td>
<td>Intact place cell activity</td>
<td>Brun et al. 2002</td>
</tr>
<tr>
<td>Septal area</td>
<td>CA1</td>
<td>Intact place cell activity, but less reliability in place fields</td>
<td>Leutgeb et al. 1999</td>
</tr>
<tr>
<td>Septal area</td>
<td>CA1</td>
<td>Disrupted place cell activity</td>
<td>Brazhnik et al. 2003</td>
</tr>
<tr>
<td>Entorhinal cortex</td>
<td>CA1</td>
<td>Disrupted place cell activity</td>
<td>Miller and Best 1980</td>
</tr>
<tr>
<td>Perirhinal cortex</td>
<td>CA1</td>
<td>Place cell activity intact, but some instability in place field location</td>
<td>Muir and Bilkey 2001</td>
</tr>
<tr>
<td>Postsubiculum or ADN</td>
<td>CA1</td>
<td>Place cell activity intact, but unstable; poor cue control with visual landmarks</td>
<td>Calton et al. 2003</td>
</tr>
<tr>
<td>Retrosplenial cortex</td>
<td>CA1</td>
<td>Place cell activity intact, but remapping occurs</td>
<td>Cooper et al. 2001</td>
</tr>
<tr>
<td>Prefrontal cortex</td>
<td>CA1</td>
<td>Place cell activity intact, but greater reliance on local cues</td>
<td>Kyd and Bilkey 2003, 2005</td>
</tr>
<tr>
<td>Vestibular labyrinth</td>
<td>CA1</td>
<td>Disrupted place cell activity</td>
<td>Stackman et al. 2002; Russell et al. 2003</td>
</tr>
<tr>
<td>Dorsal hippocampus</td>
<td>Entorhinal cortex</td>
<td>Place cell activity intact</td>
<td>Fyhn et al. 2004</td>
</tr>
</tbody>
</table>

ADN, anterodorsal thalamic nucleus
summarizes the results of several lesion studies that have addressed this issue. In general, lesions of the entorhinal cortex, the area immediately afferent to the hippocampus, disrupted place cell activity in the hippocampus (Miller and Best 1980). Disrupting individual parts of the hippocampus proper usually left place cell activity intact in downstream hippocampal areas, as well as in the entorhinal cortex. Lesions of related areas farther away from the hippocampus (e.g., septal area, postsubiculum, anterior thalamus, retrosplenial cortex) generally did not disrupt the hippocampal place signal per se, but rather altered the stability or reliability of activity or how place fields responded to manipulations of salient visual landmarks.

Determining the nature of spatial representations upstream of the hippocampus is useful in understanding how the hippocampus processes information and generates its representations. Projections from the superficial entorhinal cortex comprise the main cortical input to the hippocampal formation, and early recording studies produced evidence for location-specific activity in the entorhinal cortex (Frank et al. 2000; Quirk et al. 1992). This location-related activity was later localized primarily to the medial subdivision of entorhinal cortex (Hargreaves et al. 2005), and specifically in the dorsolateral band of medial entorhinal cortex (Fyhn et al. 2004). Hippocampal lesions do not affect place cell activity in the entorhinal cortex (Fyhn et al. 2004), suggesting that the place cell signal in hippocampus is supported by spatial information conveyed by the medial entorhinal cortex.

3. Grid Cells

Recently, further exploration of the firing properties of place cells in the dorso-lateral medial entorhinal cortex revealed the existence of a novel form of spatial representation (Fyhn et al. 2004; Hafting et al. 2005). The activity of some neurons in this area displayed location-specific firing at multiple locations that were organized in a repeating grid-like pattern that covered the entire apparatus. The points of increased firing formed equilateral triangles that could also be represented as a repeating hexagonal grid (Fig. 2). One neuron, or “grid cell,” therefore encodes location in a grid-like representation that covers the entire environment and is not influenced by the organization of external landmarks. Different grid cells contained different preferred orientations of their grids and each grid cell had a characteristic, repeating metric spacing between points of high activity. Unlike the activity of hippocampal pyramidal cells, this representation is topographically organized, as the activity of nearby grid cells is similar with respect to the spacing and orientation of the repeated place fields (reminiscent of cortical columns, etc). In theory, and of major significance to navigation, the distance traveled, velocity, and directionality of the animal can be derived over time from each grid cell’s activity. Indeed, further investigation of the deeper layers of medial entorhinal cortex has revealed the organization of populations of grid cells, HD cells, and “conjunctive” grid-by-HD cells whose activity
9. Spatial Navigation and Memory

is weakly modulated by the running speed of the animal (Sargolini et al. 2006). Neurons whose activity represents both location and heading are of particular importance and are discussed in the next section.

Grid cells located in the superficial layers of entorhinal cortex are anatomically situated to project the regular, topographical representation of space to the hippocampus because they reside in the cortical layers projecting directly to the dentate gyrus via the perforant pathway (as well as to CA1 via the alveolar pathway). The HD cells and conjunctive grid-by-HD cells observed in the medial entorhinal cortex are confined to the deeper cortical layers, suggesting that the superficial and deep entorhinal layers work together in processing spatial and movement-related information (Sargolini et al 2006). The precise contribution that grid cells provide to hippocampal place cell activity is yet unknown, although the regularity of the grid cell activity can clearly provide information regarding the animal’s movements through its environment. The location at which directional information first merges with the location and grid representations appears to be the entorhinal cortex, as the HD cell circuit via the postsubiculum projects
prominently to the superficial layers of the entorhinal cortex (see next section). Taken together, this information could clearly contribute to the neural substrates of path integration and spatial navigation (for review, see O'Keefe and Burgess 2005).

4. Head Direction Cells

The neural representation of directional heading, or orientation, is encoded by head direction (HD) cells. The neurons exhibit their highest rate of firing when the animal’s head is facing a narrow range of directions, and fall silent when the head is pointing elsewhere. HD cells have been observed in mice (Khabbaz et al. 2000), rats (Ranck 1984; Sharp et al. 2001a; Taube 1998; Taube et al. 1990a), chinchillas (Muir and Taube 2002b), guinea pigs (Rilling and Taube, unpublished observations), and monkeys (Robertson et al. 1999).

4.1. Basic Properties

Cells in the rat postsubiculum (PoS; dorsal presubiculum) were found to be most active when the animal’s head faced a specific direction in its environment, such that all the vectors representing the animal’s HD when the cell fired were parallel (Taube et al. 1990a,b). The orientation of this vector at the peak of activity is described as the “preferred firing direction,” and the discharge of individual HD cells encode different directions, such that the preferred firing directions from all cells appears to be equally represented over a 360° range. Directional sensitivity is consistent over several recording sessions indicating that, in a given environment, each cell’s representation of direction remains stable over time. The range of firing typically covers ~90°, and the firing rate decreases approximately linearly from a maximum as the rat moves its head to either side of the cell’s preferred direction (Figure 3). Subsequent discovery of HD cells in the anterodorsal thalamic nucleus (ADN; Taube 1995a) and the reciprocal connections between

![Figure 3](image-url)

Fig. 3. Three typical head direction (HD) cells from three different brain areas. Note the difference in peak firing rate for each cell. Peak firing rates vary across different HD cells from each brain area and range from about 5 to 120 spikes/s. (From Taube and Bassett 2003, with permission)
these ADN and PoS (van Groen and Wyss 1990) indicated that the HD signal is synaptically distant from primary sensory inputs, and thus can be considered a highly processed signal (Taube 1998). Other areas where a significant number of neurons exhibiting directional selectivity have been found include the lateral dorsal thalamic nucleus (Mizumori and Williams 1993), dorsal striatum (Wiener 1993), retrosplenial cortex (Chen et al. 1994; Cho and Sharp 2001), lateral mammillary nucleus (LMN; Blair et al. 1998; Stackman and Taube 1998), medial precentral cortex (Mizumori et al. 2005), and medial entorhinal cortex (Sargolini et al. 2006). Figure 3 displays examples of HD cells from three different brain areas. Note the differences in peak firing rate among the three cells, although each of the three brain areas contains HD cells with peak firing rates over this entire range.

HD cell activity has been found to be robust and reliable, provided the animal maintains a perception of environmental stability (Knierim et al. 1995). It is independent of body position relative to the head, as well as the animal’s location within an environment. HD cells are active whether the animal is moving or still, and there is no indication that adaptation of firing rate occurs if the animal continues to face a preferred direction (Taube et al. 1990a). HD cells maintain direction-specific firing when the animal locomotes in the vertical plane (Stackman et al. 2000), but appear to lose their directional sensitivity when the animal locomotes inverted on a ceiling (Calton and Taube 2005) or in the vertical or upside-down planes when exposed to 0g (Taube et al. 2004).

Visual input plays an important role in controlling the cell’s preferred firing direction, although it is unnecessary for the actual generation and maintenance of HD cell activity (Goodridge et al. 1998). The directional specificity of these cells remains consistent within and between recording sessions in the light, although there is evidence that drift may occur in the dark over extended periods of time. The preferred firing direction of a HD cell within an environment can be manipulated by rotating or moving salient visual cues, but the cues themselves are not required for HD cell activity (Taube et al. 1990b). Moreover, the shift in the preferred firing direction following rotation of background visual cues can be very rapid; the reorientation process can occur within 80ms (Zugaro et al. 2003). The directional selectivity of HD cells appears to be in register with the place-specific firing of hippocampal place cells (Knierim et al. 1995). This finding suggests that HD cells in the rat are part of a larger, cognitive mapping system that is based on representations of both place and heading (McNaughton et al. 1996; Taube 1998; see next section).

Several studies have shown that HD cell discharge can be correlated with the speed of a head turn, with faster head turns leading to a higher firing rate in the preferred direction than when the animal is turning slowly or is still (Taube 1995a; Zugaro et al. 2002). Similarly, HD cells discharge at higher firing rates when the rat is actively moving, as opposed to when it is passively turned through the cell’s preferred firing direction (Taube et al. 1990b; Zugaro et al. 2002). Other studies have shown that the HD signal in the ADN and LMN, but not the PoS, actually anticipates the animal’s future heading by about 25–50 ms (Blair and
Sharp 1995; Taube and Muller 1998). Among the hypotheses proposed to explain this predictive quality is that ADN HD cells are informed of future motor commands that control head movement by corollary discharge (motor efference copy). One prediction of this hypothesis is that when the rat does not play an active role in moving its head through the preferred firing direction, the amount of anticipatory time is reduced or abolished because the motor efference signal is absent. Bassett et al. (2005) tested this hypothesis by loosely restraining the animal and passively rotating it back and forth through the cell’s preferred direction. Contrary to predictions, they found a consistent increase in the amount of anticipation in each HD cell during the passive sessions compared to the freely moving sessions. These results suggest that anticipatory properties of ADN HD cells cannot be accounted for solely by a motor corollary discharge signal.

4.2. Generation and Maintenance of Head Direction Cell Activity

The organization of the neural circuitry that drives and maintains the HD signal has been extensively studied over the past decade. Various sensory signals have been thought to participate in HD cell responses and the shaping of a cell’s tuning curve. For example, as with place cells, rotations of visual cues result in corresponding rotations of the preferred firing direction in HD cells (Taube et al. 1990b). When familiar visual landmarks are unavailable to the animal, such as when the rat is placed in a novel environment, HD cells will maintain their directional firing responses, but the preferred firing directions will often shift to a new orientation. However, the preferred firing directions will be similar to those in the familiar environment if the animal is allowed to self-locomote under its own volition into the novel environment (Brown et al. 2005b; Taube and Burton 1995). The maintenance of HD cell activity in the absence of familiar visual information suggests that internally generated (idiothetic) cues such as vestibular, motor, and proprioceptive signals contribute to the internal representations of location and heading. Importantly, lesions of the vestibular periphery abolish the HD signal in rats (Stackman and Taube 1997; Stackman et al. 2002). When an animal is deprived of normal proprioceptive and motor efference cues by passively moving it on a wheeled cart, then the preferred firing directions usually shift a significant amount when the animal is moved from a familiar to novel environment (Stackman et al. 2003). Thus, despite the presence of an intact vestibular system, the animal was unable to update its perceived directional heading without proprioceptive and motor efference cues. Recent studies have also shown that optic flow can also influence HD cell activity (Wiener et al. 2004). Taken together, these data indicate that HD cells function as part of a system that incorporates internally generated information. When visual landmark information is pitted against spatial information derived from internal cues, the information from landmarks usually dominates over the information from idiothetic cues so long as the landmarks remain stable (Blair and Sharp 1996; Goodridge and Taube 1995; Knierim et al. 1995, 1998; Zugaro et al. 2000).
The densest populations of HD cells in the rat brain are found in the LMN, ADN, and PoS. Several studies have used combined lesion and recording techniques to determine which brain areas are critical in generating the HD cell signal (Table 2). Although the HD signal was originally discovered in the PoS, other areas afferent to the PoS are viewed as the origin of the signal (Goodridge and Taube 1997). Both Sharp and Taube consider the connections between the dorsal tegmental nucleus and the LMN as crucial to its generation (Sharp et al. 2001a; Taube and Bassett 2003). This view is based on findings that (1) lesions of either the dorsal tegmental nucleus (Bassett et al. 2007) or the LMN (Blair et al. 1998) disrupt the HD signal in the ADN, and (2) there are cells within the dorsal tegmental nucleus that encode angular head velocity (Bassett and Taube 2001; Sharp et al. 2001b). Angular head velocity cells are presumably important for updating an animal’s perceived directional heading following a head turn, as a mathematical integration in time of angular head velocity yields the amount of angular head displacement. This displacement signal could then be summed with the HD signal to yield the new directional heading. Thus, current theories view the HD signal as being generated within the tegmento-mammillary circuitry and then projected rostrally to the ADN and PoS. Other areas where the HD signal has been identified, such as the dorsal striatum and retrosplenial cortex, are believed to rely upon this circuitry for the directional signal, although direct evidence supporting this view remains forthcoming. Vestibular projections to the tegmento-mammillary circuit, possibly traveling from the vestibular nuclei via the nucleus prepositus or supragenual nucleus (Biazoli et al. 2006; Brown et al. 2005), appear critical for the HD signal because lesions of the vestibular periphery disrupt direction-specific firing in ADN cells (Stackman and Taube 1997).

Table 2. Effect of lesions to various areas on head direction cell activity

<table>
<thead>
<tr>
<th>Lesion area</th>
<th>Recording area</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoS</td>
<td>ADN</td>
<td>Normal HD cell activity; but poor cue control</td>
<td>Goodridge and Taube 1997</td>
</tr>
<tr>
<td>ADN</td>
<td>PoS</td>
<td>No directional activity</td>
<td>Goodridge and Taube 1997</td>
</tr>
<tr>
<td>LMN</td>
<td>ADN</td>
<td>No directional activity</td>
<td>Blair and Sharp 1998</td>
</tr>
<tr>
<td>LMN</td>
<td>ADN</td>
<td>No directional activity</td>
<td>Bassett et al. 2007</td>
</tr>
<tr>
<td>DTN</td>
<td>ADN</td>
<td>No directional activity</td>
<td>Bassett et al. 2007</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>ADN</td>
<td>Normal HD cell activity</td>
<td>Golob and Taube 1997</td>
</tr>
<tr>
<td>Lateral dorsal thalamus</td>
<td>PoS</td>
<td>Normal HD cell activity</td>
<td>Golob et al. 1998</td>
</tr>
<tr>
<td>Retrosplenial cortex</td>
<td>ADN</td>
<td>No directional activity</td>
<td>Wang and Taube, unpublished observations</td>
</tr>
<tr>
<td>Vestibular labyrinth</td>
<td>ADN</td>
<td>No directional activity</td>
<td>Stackman and Taube, 1997</td>
</tr>
<tr>
<td>Parietal cortex</td>
<td>PoS</td>
<td>No directional activity</td>
<td>Stackman et al. 2002</td>
</tr>
<tr>
<td></td>
<td>ADN</td>
<td>Normal HD cell activity</td>
<td>Calton and Taube 2001</td>
</tr>
</tbody>
</table>

ADN, anterodorsal thalamic nucleus; HD, head direction; LMN, lateral mammillary nucleus; DTN, dorsal tegmental nucleus; PoS, postsubiculum
Based on electrophysiological, anatomical, and theoretical data, several neural network models have been proposed to account for the generation and maintenance of HD cell activity (Goodridge and Touretzky 2000; Hahnloser 2003; McNaughton et al. 1989, 1995; O’Keefe 1990; Rubin et al. 2001; Skaggs et al. 1995; Redish et al. 1996; Song and Wang 2005; Touretzky and Redish 1996; Worden 1992; Xie et al. 2002; Zhang 1996). For the most part, these models use attractor-level dynamics to generate a “bump” (or hill) of activity to represent the HD signal, and, except for Rubin et al. (2001), they all depend on prominent inhibition for opposing head directions and recurrent excitation for nearby head directions. Most models also assume that the signal is generated in a single brain area, although one recent proposal by Song and Wang (2005) used the neural architecture across the tegmento-mammillary connections, as well as a more realistic neuron spiking model, to generate the HD cell signal.

4.3. Functional Role of Head Direction Cell Activity

Several experiments have examined the causal relationship of HD cells with behavior. Lesions of brain areas containing HD cells have been found to consistently disrupt spatial behavior (ADN; Wilton et al. 2001; Froharadt et al. 2006; LMN: Vann 2005; PoS: Taube et al. 1992). Although under most conditions HD cell activity is related to behavior on spatial tasks (Dudchenko and Taube 1997), the information from HD cells has not always been found to consistently guide spatial behavior (Golob et al. 2001; Muir and Taube 2002a). Such results support the notion that spatial navigation is supported by a flexible array of neural systems upon which the animal may draw to reach its goal. If HD cell activity represents more of a sensory/propiroceptive signal indicating the direction the animal is pointing its head, instead of a motor signal that drives the head to a particular orientation, then it follows that HD cells may serve the functional role of providing information on the animal’s current directional heading, which could then be used in navigational processes.

4.4. Place and Head Direction Cell Interactions

Simultaneous recordings of place cells and HD cells in freely behaving animals suggest that the two representations are functionally coupled. In a study by Knierim and colleagues (1995), CA1 pyramidal cells and ADN HD cells were recorded simultaneously while rats explored a high-walled cylinder with one cue card on the wall. Rotation of the cue card resulted in equal and corresponding shifts of both the CA1 place field location and preferred firing direction of the HD cell. The coupling of these two representations suggests that they participate in a system that use such integrated information in support of a broader representation. This notion is supported by the recent discovery of location- and direction-specific activity in the entorhinal cortex (Hafting et al 2005; Sargolini et al 2006).

Experimental manipulations using conflicting cue information often show that ensembles of place cells split their representations to follow different cue sets (Knierim 2002; Tanila et al. 1997; but see Brown and Skaggs 2002). While record-
ing both CA1 place cells and ADN HD cells, Yoganarasimha et al. (2006) found that the ADN HD cell ensembles never split their representation of head direction in more than 200 recording sessions, whereas the place cell ensembles split to follow either local or distal cues in approximately 75% of these sessions. This result indicates that the HD cell system is more strongly bound to one set of cues (in this case, the distal landmarks) than is the place cell representation.

A novel class of cell has been discovered whose activity appears to represent both location and direction. These “place-by-direction” cells sometimes referred to as “topographic cells”, found in the presubiculum and parasubiculum, encode both location and direction in an open field environment, such that the place field activity is significantly higher when the rats traverses the area in a certain direction (Cacucci et al. 2004). The activity of these cells appears to be similar to HD cells, but they only fire when the rat is in a place field. Furthermore, the apparent coupling between direction and location is disrupted, or remaps, when the rats are moved between different environments, while the preferred orientation of these cells remains the same. In contrast to HD cells, the discharge of these place-by-direction cells was strongly modulated by the animal’s theta rhythm.

Taken together, these data suggest that each of these representations not only contribute to the normal function of the other, but that they interact in such a way that supports a larger and more comprehensive representation of spatial information (McNaughton et al. 1996; Taube 1998).

5. Path Integration

Animals use different navigational strategies depending on the sensory information available to them; these strategies are based on allocentric landmark navigation and path integration (Gallistel 1990; Redish 1999). Path integration involves the integration of internally generated, idiothetic cues (e.g. vestibular, proprioception, motor copy) and the use of these cues to track location and heading from a known starting point. This behavior requires continually updating representations for the distance traveled and the directional heading of the animal, and can be described as the ability to make a beeline return to a starting point after traveling in a circuitous path around an environment. For example, mother rats can head straight to their nest after searching a dark, open field for pups in the absence of olfactory cues (Mittelstaedt and Mittelstaedt 1980).

Path integration supports spatial navigation but does not require explicit representations of spatial maps. The path integration strategy requires knowledge of one’s starting location, an ability to determine and track one’s directional heading, and the ability to monitor the distance traveled in that direction. Such information may be integrated and represented by the running speed-modulated activity of grid cells, HD cells, and conjunctive cells found in the deep layers of medial entorhinal cortex (Sargolini et al 2006).

In the absence of visual information, animals can use self-motion signals and knowledge of a starting location to effectively navigate through a novel environment (Mittelstaedt and Mittelstaedt 1980). The internal representation of the
sense of distance traveled, or a spatial metric, is not well understood, although the regularity in the spacing of the entorhinal grid cell activity suggests that distance information may be incorporated into the representation at that point. Spatial scaling is observed in grid cells from the postrhinal border to more ventral areas (Hafting et al. 2005). Similar scaling is also apparent in the hippocampal representation, such that location specific activity varies systematically from the septal pole (high resolution: place field size, ~25 cm) to the temporal pole (low resolution: place field size, ~42 cm; Maurer et al. 2005).

Path integration is of particular interest because although it is a behavior that can rely entirely on internally generated cues, such information is clearly incorporated into the representations of spatial location and direction. The vestibular system is a significant source of idiothetic cues, especially those related to head orientation. It has therefore been widely theorized that vestibular signals are important in navigation and path integration (Brown et al. 2002; Sharp et al. 2001a; Taube 1998).

The extent to which hippocampal lesions disrupt path integration has been inconsistent. On one hand, Whishaw and colleagues (Maaswinkel et al. 1999; Whishaw and Gorny 1999; Whishaw and Maaswinkel 1998) and Golob and Taube (1999) produced experimental results supporting a role for the hippocampus in path integration. In contrast, Alyan and McNaughton (1999) demonstrated that hippocampal lesions did not impair spatial accuracy in a digging task that clearly incorporated path integration. Lesions of the parietal cortex (Save et al. 2001), retrosplenial cortex (Whishaw et al. 2001), or dorsal tegmental nucleus (Frohardt et al. 2006) have also been shown to disrupt path integration. Additional investigation into the relationships between spatial representations and path integration are necessary to develop our understanding of navigation.

6. Spatial Representations as Substrates of Memory or Context

6.1. Relationship Between Spatial Navigation, Representations, and Memory

The hippocampus is preferentially involved in declarative/episodic memory (Sco-ville and Milner 1957; Squire 1992) and allocentric spatial cognition (Burgess et al. 2002; O’Keefe and Nadel 1978), and over the past few decades many studies have explored the interaction between these two types of information. A key issue that remains unclear is whether encoding spatial representations comprises one aspect of hippocampal memory processing (Eichenbaum et al. 1999) or whether this process is a specialized function of the hippocampus (O’Keefe 1999). The discovery of the grid cell representation of allocentric space in the entorhinal cortex (Hafting et al. 2005) relieves the expectation that the hippocampus must be a solitary player in mapping space. Furthermore, observations that hippocampal representations concomitantly include spatial and contextual
information suggest that a spatial contextual mapping system may be central to episodic memory encoding (Jeffrey et al. 2004). At the very least, spatial navigation is supported by the interplay of representations of current location, previous experiences, self-movement, and goal information. The neural substrates of spatial navigation and episodic memory therefore appear to be linked by representations of allocentric space and context.

In addition to location, hippocampal place cells have been found to encode several nonspatial parameters. Wiener et al. (1989) demonstrated that a hippocampal cell could fire in a location-specific manner when the animal ran between four corners of a box to receive a reward, but then discharged in a time-locked manner to specific behaviors in an olfactory discrimination task that was conducted in the same room and in the same apparatus. Markus et al. (1995) also showed how the place cell population remapped when a different foraging task was employed, even though the different tasks were conducted in the same environment using the same apparatus; the place cell population appeared to have unique representations for the different tasks. In a related experiment, Kobayashi et al. (1997) recorded the same place cells in the same environment under different conditions. Rats first learned to receive intracranial self-stimulation at random locations. After monitoring place cell activity in this task, rats were then trained to shuttle back and forth between two locations to receive the intracranial self-stimulation. Many place cells altered their firing patterns under this condition and fired in new locations. When the animals were returned to the random location task, the cells returned to their previous pattern of firing, except when the animals’ behavior was such that they expected to find rewards at the previously valid locations; when this behavior occurred, the firing patterns reverted to those observed during the shuttling task.

More recently researchers have postulated that these altered patterns of firing represent nothing more than context. Along this line, Anderson and Jeffery (2003) recorded hippocampal cell activity while rats foraged for food in the same environment but under different background contextual cues. Cues comprised either black or white cue cards that were paired with either lemon or vanilla odors. Some cells responded only to a single cue, while other cells responded to a particular combination of cues (e.g., white card with vanilla odor). Smith and Mizumori (2006) monitored hippocampal cells while animals approached one of two reward locations depending on the temporal context. Rats were given 20 trials: during the first block of 10 trials they were required to go to one location, and for the second block of 10 trials they had to go to a second location. A 30-s period of darkness between the two trial blocks signaled when the animals had to switch strategies. Thus, the environment and task demands were kept constant, and the only thing that changed was which place the rats went for reward, which was determined by the temporal context. They found that while some hippocampal cells fired at the same location during both trial blocks, the majority of cells had different representations for the two trial blocks. The authors considered these data indicate that this latter group of cells may have represented the contextual situation.
6.2. Temporal Coding and Reactivation of Spatial Activity During Sleep

One key to integrating different types of information into a cohesive neural representation appears to lie in the encoding of the temporal relationships of cell firing (Harris et al. 2002; Buzsaki 2005). This subject has been addressed in studies of hippocampal population activity during both awake and sleep states in rats. Hippocampal activity is known to reflect sequences of both spatial and nonspatial information in rats and humans (Dragoi and Buzsaki 2006; Fortin et al. 2002; Kumaran and Maguire 2006). Some aspects of hippocampal place cell activity can be reactivated independently of behavior suggesting that the hippocampus is contributing to the formation of neocortical memory traces (see Vertes 2004 for an opposing view). Specifically, a growing body of work has provided evidence that the firing patterns of hippocampal place cells active during a behavioral episode are reflected during subsequent episodes of sleep. CA1 neurons exhibiting place fields during exploration were found to exhibit increased activity during sleep relative to hippocampal neurons that were inactive in that environment (Pavlides and Winson 1989). Wilson and McNaughton (1994) observed that hippocampal place cells with overlapping place fields in an environment were more likely to fire together during subsequent slow-wave sleep. This reactivation was then shown to reflect the temporal order of the place cell activity that had occurred during the exploration (Skaggs and McNaughton 1996; Lee and Wilson 2002), suggesting that the hippocampus “replays” behavioral experiences to facilitate memory processing. This effect has also been observed in REM sleep (Louie and Wilson 2001). Recently, the replay of temporal sequences of spatial behavior has been shown to occur during quiet, awake periods immediately following a behavioral experience, although in this case the reactivation of hippocampal place cells occurs in the reverse temporal order (Foster and Wilson 2006). Taken together, these data suggest that there is a strong relationship between hippocampal activity and memory formation, especially in the realm of spatial information.

7. Conclusions

During the past 35 years, considerable progress has been made in our understanding of the neural representations that support spatial navigation and memory. The recent discovery of allocentric mapping in the entorhinal cortex represents a major advance in our understanding and ability to model representation of spatial information. The clean, easily accessible neural signals of place cells, grid cells, and HD cells lend themselves to experimental study that is now only limited by our ability to develop appropriate paradigms. Thus, as data collection and processing techniques improve, so must our ability to create experiments that elucidate the precise relationships between physiology and behavior.
References

Alyan S, McNaughton BL (1999) Hippocampectomized rats are capable of homing by path integration. Hippocampus 113:19–31
Cooper BG, Mizumori SY (2001) Temporary inactivation of the retrosplenial cortex causes a transient reorganization of spatial coding in the hippocampus. J Neurosci 21:3986–4001
Dudchenko PA, Taube JS (1997) Correlation between head direction cell activity and spatial behavior on a radial arm maze. Behav Neurosci 111:3–19
Fenton AA, Muller RU (1998) Place cell discharge is extremely variable during individual passes of the rat through the firing field. Proc Natl Acad Sci U S A 95:3182–3187
9. Spatial Navigation and Memory


Hetherington PA, Shapiro ML (1997) Hippocampal place fields are altered by the removal of single visual cues in a distance-dependent manner. Behav Neurosci 111:20–34


McNaughton BL, Barnes CA, Meltzer J, Sutherland RJ (1989) Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. Exp Brain Res 76:485–496


O’Keefe (1976) Place units in the hippocampus of the freely moving rat. Exp Neurol 51:78–109


Taube JS (1995a) Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. J Neurosci 15:70–86
Taube JS (1995b) Place cell activity recorded from the parasubiculum in freely moving rats. Hippocampus 5:569–583
1. What Can Uncover the Secret of the Brain’s Power?

Let’s imagine there was a big and amazing company with a tremendous number of employees. That company could accept a great variety of difficult and ambiguous orders and could quickly deliver the best products—sometimes even surprising the customers who had placed the orders. Rival companies eagerly wanted to know the secret of the super company and employed many spies to find it. After a long investigation, one of the spies found the secret at last. He reported: “The secret of the excellent power of the company is that there are many sections specialized for different tasks.” Of course, he was soon fired.

A similar story could be made concerning modern neuroscience research (Fig. 1). The brain is an amazing organ with a tremendous number of cells and can quickly produce high-quality responses to a wide variety of ambiguous orders and questions. Many neuroscientists have made great efforts to find the secret of the brain’s power, and at last they found it. That is, there are many different regions to do different work in the brain. Of course, the neuroscientists are not fired. The functional maps in the brain are one of the greatest discoveries in modern neuroscience. Such functional maps, however, have told us nothing about the secret of the brain’s power to carry out its excellent and mysterious strategies for information processing. For instance, how can an almost unlimited amount of information be stored in the brain? How can information be automatically associated with other information? How can information be categorized on the basis of similarities? How can different information be processed in parallel ways? These mysterious “how” problems cannot be answered even if we solve “where” problems; that is, where are the functional maps in the brain? Only uncovering the mode by which the brain actually processes information can solve the how problems and let us know the secret of the brain’s power.
The first step in elucidating the “how” questions is to detect the “neuronal code” for information processing. Therefore, we must ask what kinds of neuronal activity encode information in the brain. This chapter will explain the background, experimental strategies, and essential points to be considered in investigating the most likely coding, that is, ensemble coding by cell assemblies. Comprehensive reviews with a complete bibliography of population ensemble coding including cell-assembly coding can be found elsewhere (Sakurai 1996b, 1998a,b, 1999).

2. Not Single But Ensemble Neuronal Coding

2.1. The Single Neuron Is Super But Its Activity Cannot Be a Basic Code

The easiest answers to questions about the basic neuronal code could lie at the level of the activity of single neurons, the basic structural components of the brain. An individual neuron, indeed, is well known to have the capability to communicate and process signals. Recent findings have also shown that spikes from a soma can be conducted not only to its axon terminal but also back to its dendrite trees (Stuart et al. 1997) and a single neuron can release more than one type of neurotransmitter at its terminal, that is, denial of the traditional Dale’s law (Reyes et al. 1998). The well-known phenomena of long-term potentiation (LTP) and depression (LTD) indicate various memory-like functions proper to individual neurons. All these facts suggest that a single neuron has an ability to communicate and process many signals and information in a wide variety of ways and can be regarded as a microprocessor with memory functions.
However, the activity of an individual neuron alone poses several problems for the encoding of information. The first and the most serious is that neuronal firing activities are unstable. Textbooks of neuroscience sometimes illustrate neurons as simple balloon-like structures but, in fact, real neurons are not so simple and each neuron has many, probably several thousand, synaptic contacts on it; this means that a single neuron receives at least several thousand inputs per second (Legény and Salcman 1985; Paisley and Summerlee 1984). Such a large number of synaptic inputs to a neuron generates large fluctuations in the membrane potential and makes the single-neuronal activity fluctuate continuously (Steriade et al. 1993). Indeed, many neurons have interspike interval distributions that mostly fit the low-order gamma distribution, and their standard deviation is roughly the same order as the mean interval (Snowden et al. 1992; Softky and Koch 1993). These facts inherent in synaptic events of neuronal networks make individual neuronal activity inadequate as a basic code of information.

Many experiments recording neuronal responses from animals have also shown that a single neuron is not activated uniquely by one specific complicated or simple stimulus (Abeles 1988; Braitenberg 1978). Even the famous “face neurons” in the temporal cortex do not respond to single unique faces but to several faces or to several features comprising the several faces (Young and Yamane 1992). Thus, activation of single neurons alone cannot specify any unique information in a situation. With regard to functional transmissions between neurons, individual single neurons have only very weak effects on other neurons and cannot produce a sufficiently strong transmission to generate spikes in the next neurons (Espinosa and Gerstein 1988; Toyama et al. 1981). Moreover, synaptic transmission of activity among neurons seems to be very labile and can be changed dynamically by many factors in the animals and their environments (Aertsen and Braitenberg 1996; Sakurai 1993, 1996a).

Theoretical arguments suggest that the number of neurons in the brain is insufficient to represent the tremendous amount of information that an animal processes in a lifetime. The number of information items is virtually unlimited, because combinations and configurations of items produce new items, that is, there is a combinatorial explosion (von der Malsburg 1988). Representing such items by single neurons is also not a convenient way for associating and distinguishing among information items, for representing the degree of similarity or difference among items, or for constructing new concepts and ideas from individually separated items of information (Wickelgren 1992). Moreover, representation by a single neuron is not an efficient way to realize the redundancy of brain functions. Functional compensation is often seen when brain structures are lesioned in spite of the fact that a limited number of neurons are involved.

All these experimental and theoretical points clearly argue against the single-neuron coding hypothesis (Barlow 1972), which holds that information is uniquely encoded and processed in the activity of individual neurons. Therefore, it is highly likely that the basic neuronal code is the ensemble activity of a population of neurons, that is, ensemble coding by many neurons.
2.2. What Is Ensemble Coding?

Ensemble coding by a population of neurons does not necessarily imply that information is represented by abstract diffuse states of equivalent neurons or diffuse properties of the state of the brain (John 1972; Pribram 1991), because all recording experiments so far have shown that the properties of individual neurons more or less vary with each other. Ensemble coding also does not imply that many neurons with equivalent properties construct a population and summate their firings simultaneously just to overcome the unreliability of single neuronal activities (Shaw et al. 1982). This type of population coding should be regarded as “giant-neuron coding” (Wickelgren 1992) and inevitably has some of the same weaknesses as single-neuron coding. Ensemble coding by many individual neurons with more or less different individualities is plausible and should be experimentally explored.

Concerning the way neurons comprising the ensemble code cooperate with one another, the time-domain/correlational hypothesis (Abeles 1988, Gerstein et al. 1989; Singer 1990) proposes functional neuronal circuits by temporally correlated spike activity among neurons, which can realize cooperation and integration among distributed neurons with different properties. There is indeed an increasing amount of experimental data directly showing that correlated neuronal activity among neurons represents external and internal information when animals are performing behavioral tasks (Abeles et al. 1993; Ahissar et al. 1992; Sakurai 1993, 1996a; Schoenbaum and Eichenbaum 1995; Vaadia et al. 1995). These ensemble activities of populated neurons can be thought to reflect dynamically changing neuronal correlations involved in several forms of information processing.

3. Ensemble Coding by Cell Assembly

3.1. Modern Concept of Cell Assembly

The correlated activity of neurons is thought to be caused by selective enhancement in synaptic connections among neurons, that is, functional synaptic connections (Aertsen and Braitenberg 1996). Such selective enhancement refers not to a permanent change in synaptic efficacy between neurons but the transient dynamics of a neuronal interaction that sustains it during activity. Therefore, the event- and behavior-related dynamic modulation of the correlated firing in the neurons, introduced above, supports the notion that neurons can associate rapidly into a functional group to process the necessary information, at the same time becoming dissociated from concurrently activated competing groups. The competition among the neuron groups is reliant on a property of differences in the dynamics of competing groups.

Such a functional group or microcircuit of neurons with flexible functional synapses, which implies that changes in neuronal efficacy are part of the short-term dynamics of sustaining activity in the neuronal group, was proposed by the
psychologist Donald Hebb as the “cell assembly” (Hebb 1949). Hebb’s original model postulates rate-coding, by which an assembly is differentiated from others when the neurons comprising the assembly show higher rates of activation than the others. This assumption, however, makes it impossible to differentiate several coactivated assemblies because it is impossible to determine which neurons belong to which particular assembly (the so-called superposition catastrophe) (von der Malsberg 1986). To solve this problem, the modern concept of the cell assembly is defined by the temporally correlated firing of neurons (Abeles 1982; Engel et al. 1992; von der Malsberg 1986).

3.2. **Major Properties of Cell-Assembly Coding**

Coding by cell assemblies has several important properties (Wickelgren 1992). (i) Overlapping set coding of information items. The same neuron is a part of many different assemblies, suggesting that cell assemblies are overlapping sets of neurons that encode different information. (ii) Sparse coding of information items. Any individual cell assembly contains a small subset of all of the neurons in the cortex. Overlapped portions among assemblies contain not very large and not very small numbers of neurons to efficiently encode many different information items. (iii) Dynamic construction and reconstruction. Cell assemblies are temporal sets of neurons interconnected by flexible functional synapses. (iv) Dynamic persistence property. Activation of a cell assembly will persist for a time via feedback from the excitatory synapses among the neurons. (v) Dynamic completion property. Activation of a large enough subset of a cell assembly results in activation of the complete cell assembly.

These properties of cell-assembly coding can be arranged and divided into two major sets (Fig. 2). The first one is the overlap of neurons from (i) and (ii). There can be more groups of neurons in a finite brain than there are single neurons when the groups are allowed to overlap, indicating that there can be more

![Fig. 2. Two major properties of the cell-assembly coding. (From Sakurai 1999, with permission)'](attachment:image)
information items represented in assemblies than in single neurons. The overlapping also means that the same neurons may be used in different kinds of processing at the same time. This is a good way of realizing the parallel organization of the cortex where parallel and distributed processing is more the rule than the exception. The second major set of properties is dynamic construction and reconstruction of assemblies by the connection dynamics from (iii), (iv), and (v). The correlation-based concept of the cell assemblies explained above is also responsible for this second major property. As functional synaptic connections among the neurons comprising an assembly can be defined as activity correlations, the generation of spike activities of neurons within an assembly should fire in phase, whereas spike generation of neurons in different assemblies should be out of phase (Arieli et al. 1995). The property of dynamic connections predicts that such functional connections of neurons detected by the firing correlations should change dynamically when the information being processed changes.

4. Examples of Experiments Detecting the Cell-Assembly Coding

4.1. An Experimental Strategy to Detect the Cell-Assembly Coding

We could obtain experimental evidence indicating the cell-assembly coding by detecting its two major properties. The first property, partial overlapping of neurons among assemblies, is the ability of one neuron to participate in different types of information processing. Therefore, functional overlapping of individual neurons in multiple tasks should be detectable experimentally. Some studies have already found neurons that were related not only to multiple events in a task but also to different information processing between tasks (Miyashita and Chang 1988; Montgomery et al. 1992; Vaadia et al. 1989). The second property, connection dynamics of assemblies, is the capability for functional synaptic connections, detected by activity correlations of the neurons, to change among different types of information processing. Therefore, correlation dynamics among multiple neurons in multiple tasks are expected in experiments. The strategy in these experiments can be called multiple-task comparison of individual neuronal activities and multiple neuronal correlations (Fig. 3). The key point here is that the same multiple neurons are recorded during different behavioral tasks. Adequate behavioral tasks for such experiments are those that require information processing to which cell-assembly coding can strongly contribute. Cell-assembly coding has advantages especially for the representation of an extremely large number of items, the association/separation of items, the representation of the degree of similarity/difference among items, and the construction of new concepts from the associated, separated, and individual items. All these features are required in memory processing, suggesting that the most reasonable tasks to use for examining cell-assembly coding are memory tasks (Palm 1990). Consequently,
experiments should examine functional overlapping of individual neurons and correlation dynamics among the multiple neurons when the animal is performing multiple different memory tasks. An example of a series of such experiments is introduced in the following.

4.2. The Major Properties of Cell-Assembly Coding Were Found in Working Memory and Reference Memory Processes

The first experiments (Sakurai 1992, 1993, 1994) compared different processing of memory for identical stimuli, that is, working memory (WM) and reference memory (RM) tasks with identical to-be-remembered stimuli. Male albino rats were tested on the memory tasks that took place in a small operant chamber, one wall of which had a translucent response panel. The panel was covered by a guillotine door during intertrial intervals. A loudspeaker to present tone stimuli was set above the chamber. The task for WM was an auditory continuous non-matching-to-sample task (Sakurai 1987, 1990a,b). On each trial, a high or low tone was presented and the guillotine door opened to make the response panel available. The rat made “go” (pressing the panel) responses to indicate that the current tone was different from the tone on the immediately preceding trial (nonmatch). Remembering the stimulus with its temporal context, that is, working memory, was necessary for correct performance. The task for RM was continuous discrimination. The physical situation of the task was the same as that of the WM task. The rat, however, was required to make go responses only on high-tone trials throughout a session. The temporal context of the to-be-remembered stimulus was not necessary. The essential point is that, in both the WM and RM tasks, the apparatus, rats, and to-be-remembered stimuli were identical and reinforcement rates for motivation and task difficulty were almost identical because of longer training, and only the types of processing of memory required for correct performance differed from one another. Therefore, any differences
in neuronal activity between the tasks should be caused by differences in the type of processing being utilized.

Single neuronal activities from the hippocampal formation (CA1, CA3, and dentate gyrus) and the auditory temporal cortex (AC) were recorded when the rat was performing the WM and RM tasks. Firing rates of each neuron were statistically analyzed to see what parameters in the tasks (discriminative stimulus, motor response, etc.) were related to changes of neuronal activity to determine to which functions each neuron was related. With regard to the neurons related to sensory discrimination, which showed differential activation between the high and low tones only during the tasks, the hippocampal formation had such neurons in only one of the RM and WM tasks. The AC, on the other hand, had some of the neurons related to sensory discrimination in only the WM task, in only the RM task, and in both the WM and RM tasks. With regard to the neurons related to motor control, which showed activity increment before go responses, some neurons in the hippocampal formation were found in only the WM task, in only the RM task, and in both the WM and RM tasks.

Correlation dynamics based on functional synaptic connections among the neurons were evaluated by cross-correlation analysis (Perkel et al. 1967), in which correlations in the activity between neurons reveal the number of instances in which the discharge of one neuron is followed by a discharge in another neuron. This analysis was carried out on each pair of the simultaneously recorded activities of neurons and gave a set of difference cross-correlograms (Ahissar et al. 1992) with a statistical test of revealed correlations among the neurons (Abeles 1982). Some neuronal pairs showed dynamic changes of the functional synaptic connections between WM and RM tasks. Figure 4 shows an example of differ-

Fig. 4. An example of difference cross-correlograms from a pair of neurons (hippocampal CA1) showing task-dependent changes of the functional synaptic connection. A correlogram represents the number of spikes per bin (0.5 ms) that occurred in cell 1 before (left half) and after (right half) spikes in cell 2. To test that peaks in the difference cross-correlograms are statistically significant, the band of confidence limits (Abeles 1982) is shown by broken lines.
ence cross-correlograms from a pair of neurons (hippocampal CA1) showing stronger synaptic connections when the rat was performing one of the tasks, the WM task in this example.

The results from the analysis of single neuronal activities showed that some neurons of the hippocampal formation and auditory temporal cortex are involved in only WM, in only RM, and in both WM and RM. This result suggests a partial and functional overlapping of the individual neurons involved in different processing of memory, that is, WM and RM. The results from the cross-correlational analysis of neuron pairs suggested correlation dynamics among the neurons, that is, some functional synaptic connections among the neurons changed between WM and RM tasks. These data clearly indicate the two major properties of cell-assembly coding.

4.3. Cell Assemblies Are Composed of Task-Related Single Neurons: A Possibility for Dual Coding by Cell Assemblies and Their Single Neurons

In contrast to the first experiment, the second experiment (Sakurai 1996a, 1998c) compared identical processing of memory for different types of stimuli, that is, discriminations of simple auditory, simple visual, and configural auditory-visual stimuli. The apparatus was identical to that in the first experiment, except that light bulbs to present visual stimuli were added and placed on the immediate right and left sides of the chamber. The task for auditory stimuli was a simple auditory discrimination. On each trial, either a high (A) or a low (B) tone was presented and then the guillotine door opened to make the response panel available. Only go responses on A trials were correct and rewarded. The task for visual stimuli was a simple visual discrimination. Either the right (X) or left (Y) light was illuminated on each trial. The rat was required to make go responses only on X trials. The task for configural stimuli was configural auditory-visual discrimination. On each trial, simultaneously presented high tone and right-side illumination (AX), high tone alone (A), right-side illumination alone (X), or simultaneously presented low tone and left-side illumination (BY) were presented. Only go responses on AX trials were correct. In these tasks, the apparatus, rats and processing of memory were identical. Reinforcement rate for motivation and task difficulty assessed by the rat’s correct performance were almost identical among the tasks; only the types of stimuli to be processed for correct performance differed. Therefore, in contrast to the tasks in the first experiment, any differences in neuronal activity among the tasks should be caused by differences in the type of stimulus being processed.

Single neuronal activities of the hippocampal formation and the temporal cortex were recorded when the rat was performing all the tasks. Neurons that showed behavior-correlated differential activity, that is, statistically different activity between the discriminative stimuli, in each memory task were judged to be “task related” and to be involved in the processing of the task. Of the total
number of neurons recorded from the hippocampal formation and temporal cortex, about one-fourth of the neurons showed task-related activity in only one of the tasks, in two of the tasks, or in all three tasks (Fig. 5). This result, that is, the fact that the proportions of one-task-related, two-tasks-related, and all-tasks-related neurons are almost the same, supports the functional overlapping property of cell-assembly coding, which predicts that each individual neuron is more generally involved in several different memory processings. Figure 6 illustrates a plausible model of the cell assemblies, each of which encodes the auditory, visual, or configural discrimination tasks. The neurons comprising each of the assemblies are somewhat overlapped with each other. In the overlapped and nonoverlapped portions, all types of the task-related neurons are expected to be found and were in fact found in the present experiment.

If the model in Fig. 6 is correct, activity correlations, that is, functional synaptic connections, among the neurons within each of the assemblies are expected to be found. Specifically, the neurons which showed task-related activity in the auditory task (A, AV, AC, and AVC in Fig. 6) are part of the assembly encoding the auditory memory process and should show activity correlations when the rat
is performing the auditory task. The task-related neurons in the visual task (V, AV, VC, and AVC) are part of the assembly encoding the visual memory process and should show activity correlations when the rat is performing the visual task. The task-related neurons in the configural task (C, AC, VC, and AVC) are part of the assembly encoding the configural memory process and should show activity correlations when the rat is performing the configural task. The neurons which showed task-related activity in all the tasks (AVC) contribute to all the assemblies and should show activity correlations in all tasks. These hypotheses can be summarized as follows: a pair of neurons that have overlapped related tasks should show activity correlations when the overlapped tasks are being performed. The hypothesis depicted in Fig. 6 also means that a pair of neurons that have no overlapped related tasks to each other or both of which have no related tasks should show no activity correlation in all tasks. Figure 7 is an example of the difference cross-correlograms that fit one of these expectations from the hypothesis of Fig. 6. A summary of the results showed that most pairs of the hippocampal neurons supported the hypothetical model, which predicted that neurons should show activity correlations during performance of the tasks when they are related to the same tasks and no activity correlation during performance of the tasks when they are related to different tasks or no task.

These results suggest a possible way in which the hippocampal individual neurons and cell assemblies encode different aspects of memory processes, that is, dual coding (Cook 1991; Krüger and Becker 1991). Individual neurons did show differential activation among discriminative stimuli in each task, which means that the types of presented discriminative stimuli can be coded by single neurons. On the other hand, neuron pairs showed differential correlations of activities among the different tasks; thus, types of task, that is to say, types of information process or context, can be coded by cell assemblies. A possible
interpretation is that cell assemblies are comprised of task-related single neurons and give contextual meaning to the single neuronal activities.

This speculation is supported by recent studies. Arieli et al. (1995) reported that collective ensembles of activity of many neurons in the visual cortex occurred not only in the stimulus-evoked periods but also in the spontaneous nonstimulus periods. The work also suggests that the ongoing spontaneous activity of populated neurons may reflect the processing of the "context" by cell assemblies, which affects processing of incoming sensory stimuli or motor responses. In the present second, as well, the correlated activity of neurons was recorded during the nonstimulus intertrial intervals and encoded the context, that is, what task the animal was currently engaged in. Other more recent studies (de Oliveira et al. 1997; Riehle et al. 1997) reported that synchronized activation among many neurons in the primary motor cortex (MI) and the middle temporal cortex (MT and MST), which represented coding by cell assemblies, reflected the animal’s cognitive processes, for example, expectancy or attention, needed to perform the task. On the other hand, changes of firing rates of individual neurons were time-locked to external stimuli and events and are regarded as their neuronal codes.

4.4. The Modes of Cell-Assembly Coding Are Task Dependent, Especially When Temporal Information Is Processed

The conclusion from the second experiment as to the mode of dual coding by cell assemblies and single neurons, cannot be general for all other types of infor-
Dynamics of functional synaptic connections among neurons, might play especially important roles in the processing of temporal events, because cell-assembly coding must be dynamic and task dependent, that is, dependent on types of information to be processed.

The third and fourth experiments compared identical processing of memory for different types of stimuli with or without temporal events, that is, discrimination of specific time duration of auditory stimuli by rats (Sakurai 2002) and working memory for different time duration of visual stimuli by monkeys (Sakurai 2001; Sakurai et al. 2004). In the third experiment, the proportions of task-related neurons were very different from those of the second experiment. Neurons related to both tasks for discrimination of tone duration and tone pitch were many, and there were no neurons related to the discrimination of tone duration alone, but some neuron pairs showed task-dependent activity correlation between the different tasks. Figure 8 illustrates a plausible model of the coding by cell assemblies, each of which encodes each of the duration and pitch discrimination tasks (Sakurai 2002). The cell assembly encoding the duration discrimination task was completely contained within the cell assembly encoding the pitch discrimination task. These results suggest that the modes of dual coding proposed in the second experiment should be changed, especially for tasks in which processing of temporal events is needed. Dynamics of neuronal correlation, that is, the dynamics of functional synaptic connections among neurons, might play especially important roles in the processing of temporal events.

At present, no final and general conclusion can be offered about the modes of cell-assembly coding. It can be said, however, that the modes of cell-assembly coding must be dynamic and task dependent, that is, dependent on types of information process in the working brain. Therefore, behavioral studies of multiple tasks (Sakurai 2000) are the key to the studies of cell-assembly coding, and the multiple-task comparison of individual neuronal activities and multiple neuronal correlations should be continued to see what types of cell assembly encode what types of information process.
5. For Present and Future Research

5.1. Different Cell-Assembly Coding in Different Information Processing in Different Structures of the Brain

The concept of dual coding by cell assemblies and single neuronal functions, as indicated in the second experiments above, might integrate two traditional views, as Eichenbaum (1993) has suggested. On the one hand, there is the view of simple hierarchical organization of “feature detectors” that encode simple and complex stimulus features by the activity of single neurons. On the other hand, there is the view of the fully distributed representation that encodes each item by distinct spatiotemporal actions of similar arrays of neurons. The integration of the organizational hierarchy and the distributed interactions could explain the functional integration and cooperation between sensory/motor cortical processes as the former and association-cortical/limbic processes as the latter (Abeles et al. 1993; Braitenberg 1978; Eichenbaum 1993). However, as Riehle et al. (1997) has indicated, stimulus- and behavior-dependent modulations of activity correlations, which reflect cooperative interactions among neurons constituting an assembly, have already been observed in various sensory cortical areas including visual, auditory, and somatosensory areas. On the motor cortex, Riehle et al. (1997) has reported dual coding by activity synchronization among neurons (cell-assembly coding) and activity modulation of the single neurons. The former encodes purely cognitive processes (e.g., expectancy) and the latter encodes external events. This mode of dual coding in the motor cortex resembles the dual coding suggested in the second experiment above for the hippocampal formation (see Section 4.2). Synchronized neuronal activities encoding cognitive processes (e.g., attention) have also been reported in the extrastriate cortex in the visual pathway (de Oliveira et al. 1997). Therefore, it might not be possible to simply regard sensory/motor cortical processes as having hierarchical organizations with single neurons and association-cortical/limbic processes as having fully distributed organization with cell assemblies. All structures of the brain may have their own characteristic modes of dual coding by cell assemblies and single neurons.

A current hypothetical framework (Sakurai 1996b, 1998a,b, 1999) is the following. In the early sensory and the late motor stages of processing, a cluster of neurons with similar properties makes an assembly and represents an elemental event or movement. Individual properties of neurons can make relatively strong contributions to the cell-assembly coding there. A cluster of functional organization in the inferotemporal cortex (Fujita et al. 1992; Tanaka 1996; Wang et al. 1996) and a neuronal group constituting a population vector in the motor cortex (Georgopoulos 1995) are good examples of such cell assemblies. A small difference among properties of the neurons in an assembly there can encode a small difference among elemental stimuli and movements. This orderly or small-scale localized organization of cell assemblies changes gradually to parallel, distributed, and larger organizations as higher stages of processing are required. When a higher integrative process is required to encode more complex events, new concepts, new associations, etc., a larger assembly of neurons with different
properties begins to work by temporally correlated activity among neurons and among smaller assemblies (nesting of assemblies). Besides the dynamic functional connections among neurons, overlapping of neurons among the assemblies also changes dynamically to represent the degree of similarity and difference among information items and processings. The dynamic correlational constitution of assemblies is most apparent in memory processing, because most abstract and dynamic modes of coding are needed there. These features of cell assemblies of the sensory, higher integrative, and motor processing are thought to change gradually, not suddenly, among the processing and brain structures. The difference there is a matter of degree.

5.2. More Technical Improvements Are Necessary

The cell assembly dynamically changes its size and functional connections to encode various types of information. If correlated and synchronized firing among the neurons truly reflects cell-assembly coding, it should show dynamic changes that depend not only on tasks and events being processed but also on the distance among the neurons. The latter assumption, however, has not been fully tested yet because of a technical problem, that is, it is difficult to separate extracellular activities from closely neighboring neurons. Their spike waveforms sometime overlap on a common electrode when they fire coincidently, and ordinary spike-sorting techniques cannot separate such overlapping spikes (Lewicki 1994), making it impossible to detect precise correlation and synchrony in a population of neighboring neurons.

We recently have solved, however, the problem and have introduced a unique method of spike sorting using a combination of independent component analysis (ICA) and k-means clustering (Takahashi et al. 2003a,b). This method solves the spike-overlapping problem and the “nonstationary waveform problem” (Fee et al. 1996), and can sort activities of closely neighboring neurons in behaving animals (Fig. 9). In the following study, we also have developed a system that can sort the neighboring neuronal activities in real time (Takahashi and Sakurai 2005). These methods are expected to show detailed and real features of interactions in and among cell assemblies in the working brain. Recently, we employed these unique spike-sorting and multi-neuronal recording methods and showed that some closely neighboring neurons have dynamic and precise synchrony to represent certain situations (tasks) for working memory (Sakurai and Takahashi 2006).

5.3. More Theoretical and Computational Considerations Are Necessary

Although more progress in experimental techniques is needed for the future research, the recording of a great amount of raw data alone can say nothing about the actual features of the cell-assembly coding. To analyze the data adequately, to make valid hypothetical frameworks for experiments, and to reveal actual performance of neuronal networks from the experiments, considerations from
that match and include the experimental data are needed. The best meeting place between experimental facts and theoretical models is currently the research on microcircuits (Douglas and Martin 1991), especially on the cell assembly (Amit et al. 1994; Palm 1993). Computer simulation studies that use sophisticated cell-assembly models (Frieses and Friesen 1994; Fujii et al. 1996; Rieke et al. 1997) are required, and theoretical and computational approaches to cognitive modeling based on newly developed concepts of associative neuronal networks (Yufi 1998) are encouraging. It is needless to say that, to conduct such sophisticated simulation studies, further advances in basic physiological experiments are required to provide more detailed information and to establish parameters about the properties of real neurons and synapses.

One of the most sophisticated theoretical considerations in computational neuroscience concerns how the brain represents and stores a great amount of information. “Sparse-coding” (Kanerva 1988; Meunier et al. 1991; Palm 1990; Wickelgren 1992) is a convincing hypothesis to explain the representation of an almost unlimited number of information items in the brain. There can be many
more groups of neurons than there are single neurons when the groups are allowed to overlap. Although more extensive overlapping allows more groups of neurons, there would be difficulty in discriminating different information items based on proper subsets of the groups representing items. However, a great many and, at the same time, easily discriminated groups of neurons are obtained especially when the overlapping is sparse, that is, sparse coding. Such coding means that the degree of allowed overlap is important (Palm 1990). Besides computational simulations to see what degree of sparse overlapping would be ideal, experimental approaches are needed to see what degree of overlapping is actually employed in the working brain. The proportions of the functionally overlapped neurons, that is, the neurons related to two or three of the tasks, as found in the first and the second experiments above (see Sections 4.1 and 4.2), suggest that around one-third of the task-related neurons are overlapped. It might therefore be assumed that the degree of overlap of neurons for the sparse coding is around 30%.

Another important theoretical consideration concerns the role of cortical neurons, that is, functioning as integrators or coincidence detectors. This discussion centers on whether neurons can act effectively as temporal integrators and transmit temporal patterns with only low reliability or, in contrast, whether they act as coincidence detectors which relay synchronized input and the temporal structure of their output is a direct function of the input pattern. König et al. (1996) has introduced the discussion of coincidence detector versus integrator from a comprehensive point of view and has claimed that the former might be a prevalent operation mode of neurons. From a theoretical perspective, coincidence detection allows much richer spatiotemporal dynamics of the system, more speed of processing, and smaller size of neuronal populations. These advantages are the same as those of cell-assembly coding. In other words, cell-assembly coding implies that neurons act as coincidence detectors. Indeed, coincidence detection can realize more elegantly and in a highly economic way the binding of distributed neurons into functionally coherent groups of neurons, that is, cell assemblies. Watanabe et al. (1998) investigated how cell assemblies behaved if they received various kinds of input coincidentally arrived and showed sophisticated performance of the assemblies to encode various information. In addition to such theoretical support, there is physiological and behavioral experimental evidence for coincidence detection. The time constants of postsynaptic membranes for excitatory postsynaptic potentials are very short (<10 ms), which means that neurons are naturally sensitive only to coincidence inputs from many neurons, and some animals, that is, some brains, can detect and utilize timing differences of the order of microseconds between external and internal events (König et al. 1996). Coincidence detection, which is highly supportive for cell-assembly coding, is not in conflict with the assumption that the firing rates of individual neurons are also carriers of information in the brain. Some effective codes of individual neuronal activities can be propagated by neuronal assemblies that carry other kinds of information in a highly distributed manner, that is, dual coding.
6. Conclusion

Individual neurons are surely super dynamic microprocessors with memory (see Section 2.1). Therefore, a functional group with many individual neurons, that is, cell assembly, must be extraordinarily “super” and really dynamic in its nature. Such assemblies can realize the excellent and dynamic information processing by the brain. Uncovering the actual features of the cell assemblies in the brain can lead us to uncovering the secrets of the amazing power of the brain, not only for the relatively small-scale dynamics of neuronal networks but also for large-scale dynamics underlying reorganization in global structures of the brain, such as phantom limbs (Melzack 1989; Ramachandoran 1992) and others.

In research on cell assemblies, whatever theoretical frameworks are employed and whatever techniques are used to observe neural activity, it is most important to define the types of information processes in the working brain. The reason is that the cell assembly is dynamic in its nature and its features must be able to vary with the types of information item and processing, that is, task dependent. Thus, using working brains in behaving animals in behavioral tasks is essential, and psychological consideration of the tasks should not be neglected.

In his famous *Textbook of Psychology* (Hebb 1972), in which the present author encountered the concept of cell assembly for the first time, Hebb suggested that “Cell assembly refers to a bridging conception, relating the mediating process—known from behavior—to brain function” (p. 280). Modern neuroscience may now be gradually approaching some firm results concerning the nature of this “bridging conception,” the cell assembly, which must be the key to scientifically understand the real mechanisms of our mind.

Acknowledgments. This work was supported by the Research for the Future Program, a Grant-in-Aid for Target-Oriented Research in Brain Science, the 21st Century COE Program, and the CREST program.

References

Sakurai Y (1990a) Hippocampal cells have behavioral correlates during the performance of an auditory working memory task in the rat. Behav Neurosci 104:253–263
Sakurai Y (1990b) Cells in the rat auditory system have sensory-delay correlates during the performance of an auditory working memory task. Behav Neurosci 104:856–868
Sakurai Y (1996a) Hippocampal and neocortical cell assemblies encode memory processes for different types of stimuli in the rat. J Neurosci 16:2809–2819
Sakurai Y (1999b) Elemental, configural, and sequential memory processes in the rat can be tested in a single situation in one day. Psychobiology 27:486–490
11

Representation of Numerical Information in the Brain

ANDREAS NIEDER

1. Introduction

Numbers can be used most flexibly to quantify, rank, and identify virtually everything that is imaginable. Although true counting and mathematics are cultural achievements that are bound to language, it became evident during the past decades that basic numerical skills are independent from language; they arise earlier in phylogeny and in ontogeny. Human adults, infants, and animals are able to nonverbally and approximately grasp the numerical properties of objects or events. Such nonverbal numerical systems are the precursors on which verbal numerical representations are built, and their neural foundation can be studied in animal models.

This chapter reviews the progress that has been made in our understanding of the neural underpinnings of numerical competence in human and nonhuman primates. It is structured according to the basic concepts numerical quantity, which refers to the empirical property cardinality (the size of a set, also termed numerosity), and numerical rank, which refers to the property serial order (Wiese 2003).

2. Numerical Quantity (Cardinality)

2.1. Behavior

Animals can discriminate the cardinality of sets. The German zoologist Otto Koehler was the first to demonstrate convincingly numerical competence in several species of birds and mammals (reviewed in Koehler 1951). During the past six decades, roughly half a dozen bird and mammal species have successfully been trained to perform several numerical tasks in different modalities (reviewed in Davis and Perusse 1988). Importantly, animals can transfer their numerical
knowledge to numerosities they have not been trained on (Brannon and Terrace 1998), which indicates a true understanding of the concept of cardinalities and their sequential arrangement. Moreover, numerical intelligence is not induced by extensive training in the laboratory, but can be demonstrated spontaneously in the wild (Hauser et al. 1996, 2000; McComb et al. 1994). The data collected for different species indicate that quantity information is indeed exploited by animals in their natural habitat to draw informed choices, for example, in foraging situations or social interactions.

Also, preverbal human infants of several months of age have already the capacity to represent cardinality, in both the visual and the auditory domain (Feigenson et al. 2004). Infants can also engage in rudimentary arithmetic, which was first shown in experiments in which 5-month-old infants were shown basic addition and subtraction operations on small sets of objects (Wynn 1992).

Human adults can reliably compare the cardinality of sets under conditions that prevent or discourage verbal counting (Barth et al. 2003; Whalen et al. 1999). In contrast to precise verbal counting, however, nonverbal discrimination performance is only inaccurate, or approximate. Some indigenous human cultures that lack number words (Gordon 2004) or have a very restricted concept of verbal counting (Pica et al. 2004) rely completely on nonverbal cardinality assessment (so-called one-two-many system of “counting”). Thus, humans without a linguistic number concept can only estimate the number of items by means of a nonverbal quantification system, very much like infants and animals. Together, these studies provide evidence for an evolutionary ancient quantification system that operates independent from language and can thus be studied in animal models (Nieder 2005).

### 2.2. Numerosity-Selective Neurons

Recordings in monkeys demonstrated the capacity of single neurons to encode numerical quantity (Nieder et al. 2002, 2006). In monkeys performing a visual delayed match-to-numerosity task (Fig. 1a), a high proportion of numerosity-selective neurons (Fig. 1c) was found in the lateral prefrontal cortex (PFC), irrespective of covarying nonnumerical parameters. In the posterior parietal cortex (PPC), numerosity-selective neurons were most abundant in the fundus of the intraparietal sulcus (IPS); they were few in other PPC areas or the anterior inferior temporal cortex (aITC) (Nieder and Miller 2004) (Fig. 1b). Neurons in a somatosensory-responsive region of the superior parietal lobule (part of area 5) have been reported to keep track of the number of movements (Sawamura et al. 2002), but in a movement type-dependent manner (that is, the neurons responded differently whether the monkey’s movement was “push” or “turn”). Area 5 neurons were not encoding numerosity in visual displays (Nieder and Miller 2004).

Numerosity-selective neurons in the PFC and IPS were tuned for the number of items on a visual display, that is, they showed maximum activity to one of the five presented quantities—a neuron’s “preferred numerosity” (Fig. 1c); all numerosity-selective neurons together formed a bank of overlapping numerosity
11. Numbers in the Brain

Fig. 1. Representation of visual cardinality in rhesus monkeys. a Behavioral task. Monkeys performed a delayed match-to-numerosity task. They were required to extract the numerosity in a visual array, memorize it briefly, and match it to the one of two alternative test displays that showed the same numerosity (but with a different visual pattern). b Lateral view of the brain of a monkey shows the recording sites in LPFC and posterior parietal cortex (PPC). The proportions of numerosity-selective neurons in each area are coded according to the gray color scale. As, arcuate sulcus; Cs, central sulcus; IPS, intraparietal sulcus; LF, lateral fissure; LS, lunate sulcus; Ps, principal sulcus; Sts, superior temporal sulcus. c Responses of a single neuron recorded from the PFC. The neuron showed graded discharge during sample presentation (500–1300 ms) as a function of numerosities 1 to 5. The inset in the upper right corner shows the tuning of the neuron and its response to different control stimuli. The preferred numerosity was “4” for this PFC neuron. d Behavioral numerosity discrimination functions of two monkeys. The curves indicate whether they judged the first test stimulus (after the delay) as containing the same number of items as the sample display. The function peaks indicate the sample numerosity at which each curve was derived. Behavioral filter functions are skewed on a linear scale (not shown), but symmetrical on a logarithmic scale. e The averaged single-cell numerosity-tuning functions (from PFC) are also asymmetrical on a linear scale (not shown), but symmetrical after logarithmic transformation.
filters (Fig. 1e), thus mirroring the animals’ behavioral performance (Fig. 1d). Interestingly, the neurons’ sequentially arranged overlapping tuning curves preserved an inherent order of cardinalities. This point is important because numerosities are not isolated categories, but exist in relationship to one another (for example, “3” is greater than “2” and less than “4”); they need to be sequentially ordered to allow meaningful quantity assignments.

The fact that IPS neurons require shorter latencies to become numerosity selective than PFC neurons suggests that the IPS might be the first cortical stage that extracts visual numerical information (Nieder and Miller 2004). As PPC and PFC are functionally interconnected, that information might be conveyed directly or indirectly to the PFC where it is amplified and maintained to gain control over behavior.

The response properties of numerosity-selective cortical cells can explain basic psychophysical phenomena in monkeys, such as the numerical distance and size effect (Fig. 1d). The numerical distance effect states that it is easier to discriminate quantities that are numerically remote from each other (say, 2 versus 6 is easier than 5 versus 6), whereas the numerical size effect captures the finding that pairs of numerosities of a constant numerical distance are easier to discriminate if the quantities are small (for example, 2 versus 3 is easier than 5 versus 6). The numerical distance effect results from the fact that the neural filter functions that are engaged in the discrimination of adjacent numerosities heavily overlap. As a result, the signal-to-noise ratio of the neural signal detection process is low, and the monkeys would make many errors. On the other hand, the filter functions of neurons that are tuned to remote numerosities barely overlap, which results in a high signal-to-noise ratio and, therefore, good performance in cases where the animal has to discriminate sets by a larger numerical distance.

The numerical size effect is based on the finding that neuronal tuning obeys Weber’s law: the widths of the tuning curves (or neuronal numerical representations) increase linearly with preferred numerosities (that is, on average, neurons become less precisely tuned as the preferred quantity increases). Hence, rather selective neural filters that do not overlap a lot are engaged if a monkey has to discriminate small numerosities (say, 1 and 2), which results in high signal-to-noise ratios and few errors for the discrimination. Conversely, if a monkey has to discriminate large numerosities (such as 4 and 5), the filter functions would overlap considerably. Therefore, the discrimination would show a low signal-to-noise ratio, which leads to poor performance.

2.3. The Scaling of Numerical Representations

As mentioned above, the neurons’ overlapping tuning curves are arranged along a “number line” in an orderly fashion (Fig. 1e). But what is the scaling scheme of such a number line: are neuronal numerical representations best described on a linear, or a nonlinear, possibly logarithmically compressed scale? The latter would be predicted if Fechner’s law holds. Fechner’s law states that the perceived magnitude \( S \) is a logarithmic function of stimulus intensity \( I \) multiplied by a
modality and dimension specific constant \((k)\). As the behavioral discrimination and single-unit tuning functions might be regarded as the monkeys’ behavioral and neural numerical representations (Fig. 1d,e), the crucial question then concerns which scaling scheme would provide symmetrical (that is, Gaussian) probability density distributions. Both the performance and the single unit data for numerosity judgments are better described by a compressed scale (Fig. 1d,e), as opposed to a linear scale (Nieder and Miller 2003). Therefore, single-neuron representations of numerical quantity in monkeys obey Fechner’s law.

2.4. Functional Imaging in Humans

In humans, the prefrontal cortex and the parietal lobe, in particular, the IPS, have long been regarded as the prime source for numerical competence (Dehaene et al. 1999). This view, however, was almost exclusively based on verbal and symbolic number tasks, which are only vaguely comparable to numerical magnitude estimation. If anatomical and functional similarities between the brains of monkeys and humans do exist, nonverbal cardinality should be processed in the human brain in equivalent areas as seen in monkeys. Indeed, such a corresponding blood oxygenation level-dependent (BOLD) activation in the IPS of humans has been found using functional magnetic resonance imaging (fMRI) adaptation with numerosities by Piazza et al. (2004) (Fig. 2a). Subjects were repeatedly presented with several visual displays of a fixed numerosity (for example, 16 dots), without the requirement to discriminate them. The rationale of this protocol is the following: if any region of the brain contains a population of numerosity-selective neurons that are tuned to a specific number of dots and automatically detect numerical information, such a population of detectors should habituate (that is, decrease its discharge) when numerosities are repeatedly presented, while neurons tuned to other numerosities should not be affected. Such a habituation effect was then “read out” by recording the event-related fMRI activation to a single deviating numerosity (for example, 32 dots) that was presented at the end of a display sequence. In fact, the only region in the brain that is significantly habituated to numerosity was the horizontal segment of the IPS (Fig. 2a). Using the same fMRI adaptation protocol, the IPS was also activated in 4-year-old children (Cantlon et al. 2006). With this fMRI adaptation protocol, the average numerosity-tuning curve of the underlying neural population could be traced indirectly in humans. Similar to single-neuron responses in monkeys, fMRI tuning curves in humans seem to be well described on a logarithmic scale, which suggests that the nonverbal “number line” in all primates is nonlinearly compressed.

Language allows humans to use symbolic representations, and numbers are the symbols that are employed when dealing with numerical information. Interestingly, numerical values that are cued by number symbols and set size have been shown to activate corresponding structures in the brain in humans. The IPS, which has been suggested to play a central role in basic quantity representations, is sometimes the only area specifically activated in simple number detection or
comparison tasks. Eger et al. (2003) performed fMRI measurements on subjects who were asked to merely detect either numerals, letters, or colors, which were presented visually (written words and symbols) or acoustically (spoken words). To avoid confounds by response selection and associated cognitive states (such as attention), the authors analyzed the presentation of nontarget numerals (numerals that were not required to be detected) and compared it to nontarget letters or colors. The IPS was the only region that exhibited higher activation for numerals, both visually and acoustically (Fig. 2b). Therefore, numerical activa-
tion in the IPS seems to be automatic (task independent), supramodal (visual and auditory), and notation independent (irrespective of whether numerals are spoken or written, presented in arabic notation or spelled-out form). The IPS, however, is also engaged in more general magnitude judgments. Pinel et al. (2004) scanned subjects with fMRI while they compared arabic numerals for luminance, font size, and numerical value. They observed strong overlap in the neural substrates in the three tasks. Number and size, but not luminance, activated a common parietal region, pointing toward an intimate relationship between spatial and numerical representations. This finding might mean that number-coding neurons are intermingled with other magnitude-coding neurons along the IPS.

Beyond mere encoding of quantity information per se, verbal numerical competence requires additional cognitive components. Dehaene and coworkers (2003) suggest that verbal counting and calculation engages two more parietal regions: a posterior dorsal parietal area that is activated by shifts in spatial attention whenever subjects count, and a left angular gyrus area that is related to linguistic processing. Moreover, simple calculation tasks (such as subtraction) typically activate a distributed network that involves parietal, prefrontal, and premotor cortices.

Developmental studies confirm that impairments of arithmetical abilities correlate with abnormalities in the organization of parieto-frontal networks, and the IPS in particular. Using voxel-based morphometry, Isaacs et al. (2001) compared the density of gray matter in adolescents who were born at equally severe grades of prematurity. Half of these subjects (with otherwise normal IQ) suffered from dyscalculia, and the only region in the brain that showed reduced gray matter associated with arithmetical deficits was the left IPS. Therefore, dyscalculia in children might be the result of specific disabilities in basic numerical processing, rather than the consequence of deficits in other cognitive abilities. Arithmetic deficits are also found in certain genetic conditions, such as the Turner syndrome (X monosomy) (Molko et al. 2003), fragile X syndrome (Rivera et al. 2002), and velocardiofacial syndrome (Eliez et al. 2001). In these conditions, fMRI hypoactivation (decreased BOLD activation) was found in the IPS and wider parieto-frontal networks.

2.5. Neuropsychological Studies

Brain damages in humans can cause relatively selective impairments in dealing with numbers, called acalculia. The most frequent lesions yielding acalculia involve the left inferior parietal area or the left parieto-occipito-temporal junction (Dehaene and Cohen 1997). In cases of left inferior parietal lesion, acalculia is frequently associated with agraphia, finger agnosia, and left-right confusion in a tetrad of deficits called “Gerstmann’s syndrome” (Gerstmann 1940). In addition, calculation deficits have also been observed following left medial frontal, left and right frontal, and temporo-occipital lesions. A major issue in neuropsychological studies is the dissociation between deficits in language and numerical
functions, particularly when the left hemisphere is affected. A recent study indicates that syntactical mathematical operations (such as calculation rules, e.g., solving equations according to brackets) are largely independently represented from linguistic syntax, at least in an adult brain (Varley et al. 2005).

3. Numerical Rank (Serial Order)

3.1. Behavior

Numerical rank is the second major numerical concept that shows biological precursors. List learning—the ability to encode and then retrieve an arbitrary list of items in their correct order—opened a window to study how the ordinal rank of objects is learned and stored by animals. Terrace and coworkers (Chen et al. 1997; Schwartz et al. 1991) performed a series of experiments showing that rhesus monkeys learned several up to seven-item lists (e.g., lists of photographs). This procedure provided the opportunity to evaluate the monkeys’ knowledge of the ordinal position of list items with “derived lists.” Derived lists are novel lists that are constructed by picking individual items from previously learned lists and reassembling them. Interestingly, the derived lists on which each item’s ordinal position was maintained were acquired rapidly by the monkeys. However, the monkeys needed as much time to learn the derived lists on which each item’s ordinal position was changed as they did for completely novel lists. This result indicates that the monkeys acquired knowledge of each item’s original ordinal position, and they could exploit this knowledge on the order-maintained derived lists.

In a study addressing monkeys’ strategies for list recall, macaques were trained to report three-item lists with the sample items displayed in temporal sequence (Orlov et al. 2000, 2002). After presentation of such a list, a test stimulus that showed the three items and a distractor item was presented, and the monkeys were required to touch the three items in the order that was previously shown to them without touching the distractor. The distractor was taken from one of the other learned lists and, therefore, had a fixed ordinal position within its own list. Interestingly, the monkeys primarily mixed up a list item with a distractor if the distractor had the same ordinal position (in its own list) as the correct item. This finding indicates that monkeys intuitively categorized images in such lists by their ordinal number.

3.2. Single-Unit Studies of Serial Order

Ordinal categorization of visual items requires both information about the rank of an item (for example, based on temporal order) and its identity. The single-neuron correlate of temporal rank order information in visual lists has recently been studied in monkeys trained to observe and remember the order in which three visual objects appeared, so that the animals could plan a subsequent triple-reaching movement in the same order (Ninokura et al. 2003, 2004) (Fig. 3a).
Neurons in the ventrolateral PFC were selective for visual object properties, whereas neurons in the dorsolateral PFC were selectively tuned to the rank order of the objects irrespective of the visual properties of objects; for example, a rank-order selective neuron would be active whenever the second item of the shuffled lists appears (Fig. 3b). A third class of neurons, found in the ventrolateral PFC, showed the most complex response that was characterized by integrating the objects’ sensory and order information; such a neuron would only discharge whenever a certain object appeared at a given position in the sequence.
The lateral PFC is an ideal region in the brain to encode both sensory object properties and rank-order information because it receives massive sensory input from the temporal and parietal lobes, and projects to pre-motor and motor areas of the frontal lobe (Miller and Cohen 2001). As a result, neurons that encode the ordinal position of task-related hand or eye movements have been found frequently in prefrontal (Barone and Joseph 1989; Funahashi et al. 1997) and a variety of motor-related cortical areas in trained monkeys, such as the frontal eye field (FEF) (Isoda and Tanji 2003), the presupplementary (pre-SMA) and supplementary motor area (SMA) (Shima and Tanji 2000), the caudate nucleus (Kermadi and Joseph 1995), the anterior cingulate cortex (CGa) (Procyk et al. 2000), and even the primary motor cortex (M1) (Carpenter et al. 1999). Motor-related areas such as M1, SMA, pre-SMA, and FEF may receive numerical information that has been computed on earlier stages of the cortical hierarchy to perform appropriate serial-order actions.

3.3. Human Studies

Based on patient studies, the lateral prefrontal cortex has been implicated in maintaining temporal order information, which is an integral aspect of episodic memory (McAndrews and Milner 1991). It is well known that damage to the human frontal cortex causes impairment in tasks that require recall of the temporal order of stimuli (Petrides and Milner 1982). A similar ordering impairment has resulted from lesioning the dorsolateral frontal cortex in monkeys (Petrides 1995), which supports the view that the dorsolateral PFC in primates is important in maintaining information about the order of events. Functional MRI studies of humans show that prefrontal and parietal cortices are more strongly activated for order information (e.g., the order of words in a list) than for item information (presence of words in a list) (Cabeza et al. 1997; Marshuetz et al. 2000). In addition, the lateral frontoparietal areas, the basal ganglia and the cerebellum were preferentially involved in ordinal control of hand movements.

4. Conclusion

Together, the reviewed literature indicates that language-based numerical competence does not emerge de novo in evolution but arises from biological predispositions. This hypothesis is supported by studies on prelinguistic representations of numerosity in human infants and nonverbal numerical competence in animals. Recent studies in primate neurophysiology and human functional imaging indicate the posterior parietal cortex in close association with the prefrontal cortex as key structures for numerical intelligence. The intraparietal sulcus in particular is consistently activated in both nonverbal and verbal numerical tasks and seems to host an evolutionary ancient quantification system that operates supramodally and independently from language. A key question for future research will be the relationship between mechanisms underlying cardinality and serial order judg-
ments, the two major aspects of numerical competence that have been studied largely independently so far.

References

Gerstmann J (1940) Syndrome of finger agnosia, disorientation for right and left agraphia and acalculia. Arch Neurol Psychiatry 44:398–408
Koehler O (1951) The ability of birds to “count.” Bull Anim Behav 9:41–45
Part IV
Manipulation of Internal Representation
Prefrontal Representations Underlying Goal-Directed Behavior

Jonathan D. Wallis

1. Introduction

The prefrontal cortex (PFC), the brain area directly behind our forehead, remains enduringly mysterious. It has dramatically expanded in size and complexity across evolution, reaching its pinnacle in humans, where it accounts for about 30% of our total cortical area. Despite this, damage often appears to have little effect. The patient’s language, memory, sensorimotor abilities, and intelligence remain intact. Nevertheless, their lives dramatically fall apart. One famous PFC patient is Elliot, a young man whose PFC was damaged during an operation to remove a brain tumor (Eslinger and Damasio 1985). He had been a successful, happily married individual with a high-paying job, but within months of the operation, he had quit his work, lost a large sum of money to a fraudster, divorced his wife, lost contact with family and friends, and married a woman he had known for only a short while. In summary, after making a series of excellent life choices, within months of PFC damage he had made a series of catastrophic ones. This chapter explores the neuronal representations within PFC that are so important for us to operate effectively in our everyday lives.

Despite the clear behavioral deficits that PFC patients exhibit, until relatively recently investigators had focused almost exclusively on how sensory or motor information is represented in PFC (Goldman-Rakic 1987). Specifically, PFC neurons represent sensorimotor information in working memory as a pattern of ongoing, selective electrical activity (Funahashi et al. 1989; Fuster and Alexander 1971; Kubota and Niki 1971). However, it was later found that the precise information represented by a PFC neuron is highly dependent on the demands of the task (Asaad et al. 2000; Rao et al. 1997), and focus switched to understanding how behaviorally relevant information is encoded. A common theme that has emerged is that PFC represents such information in a hierarchical manner.

Anatomically, the PFC sits at the apex of the perception–action cycle (Barbas and Pandya 1991; Pandya and Yeterian 1990). It receives inputs from all sensory

Helen Wills Neuroscience Institute and Department of Psychology, University of California at Berkeley, Berkeley, CA 94720, USA
modalities, but it is highly processed information. For example, its main visual input is from the inferior temporal cortex, which processes complex object information. Its connections become fewer as one moves to progressively lower level visual areas. Indeed, PFC does not connect at all with V1, the lowest level visual cortical area. Likewise, PFC connects with high-level motor areas, such as premotor cortex, rather than lower-level motor areas such as primary motor cortex.

This anatomical hierarchy seems to give rise to a functional hierarchy. PFC damage spares simple, reflexive behaviors such as appetitive and orienting responses to salient, rewarding objects (Grueninger and Pribram 1969). However, behaviors that are more complex depend on the integrity of the PFC (Gaffan et al. 2002; Roberts and Wallis 2000; Wallis et al. 2001b). In the remainder of this chapter we focus on three such behaviors, beginning with the simplest and moving to the progressively more complex. First, we examine the involvement of PFC in the control of reward-based responses. We then examine its role in the representation of arbitrary stimulus–response associations. Finally, we examine how the PFC permits behavioral control using high-level, abstract rules and strategies.

2. Reward-Based Responses

Much of our behavior focuses on minimizing or maximizing a particular goal state. For example, animals seek to maximize food intake and minimize energy expenditure (Stephens and Krebs 1986). The ventromedial region of the PFC (vmPFC) is particularly important for processing information required to accomplish this. Its anatomical connections are ideal for this purpose. It connects with all sensory modalities (the only PFC region to do so), including the olfactory and gustatory cortices that are critical for monitoring food intake (Carmichael and Price 1995b). It also connects with limbic regions such as the nucleus accumbens and amygdala, which process whether something is rewarding (Carmichael and Price 1995a).

In the remainder of this section, we examine the neuronal representations encoded by vmPFC that allow it to control behavior using reward information. First, we examine the evidence that vmPFC is a specialized region within the PFC for processing reward information. Second, we examine ways in which the vmPFC reward representation contributes to behavioral flexibility. We then look at how vmPFC contributes to decision making. Finally, we examine evidence that vmPFC encodes an abstract representation of reward, and we discuss the implications this has for behavioral control.

2.1. Representation of Reward

The vmPFC responds to a wide variety of rewards. Results from functional magnetic resonance imaging (fMRI) show that human vmPFC is activated by pleasant and unpleasant smells, sights, and sounds (Anderson et al. 2003; Royet et al. 2000), touches (Rolls et al. 2003), and abstract rewards, such as receiving money
(O'Doherty et al. 2001). At a single-neuron level, however, the representation of reward information is very specific. For example, vmPFC neurons in the monkey distinguish between the taste of blackcurrant or banana (Thorpe et al. 1983), different concentrations of glucose (Thorpe et al. 1983), and different types of fat (Rolls et al. 1999), whereas vmPFC neurons in the human discriminate pleasant from unpleasant scenes and different facial expressions (Kawasaki et al. 2001). Neurons in vmPFC also integrate multiple aspects of reward. For example, vmPFC neurons integrate both taste and smell information, potentially producing a neuronal representation of flavor (Rolls et al. 1996). They can also rapidly learn that a reward is predicted by a specific visual stimulus (Tremblay and Schultz 2000a,b; Wallis and Miller 2003a) or olfactory stimulus (Schoenbaum et al. 1998, 1999). Lesion studies support these neurophysiological findings. Lesions of vmPFC impair the ability to learn which responses maximize reward, both in monkeys working for food rewards (Baylis and Gaffan 1991; Dias et al. 1996) and in humans working for monetary rewards (Bechara et al. 1994).

There is, however, evidence that challenges the notion that vmPFC is a specialized region within PFC responsible for representing reward information. For example, damage to the dorsolateral PFC (dlPFC) also impairs the ability of humans to maximize reward on a gambling task (Fellows and Farah 2005). Neurons that encode reward outcomes occur throughout the frontal lobe, not just vmPFC (Leon and Shadlen 1999; Roesch and Olson 2004; Watanabe 1996). Furthermore, if a neuron responds more vigorously when a monkey expects a reward, it is not necessarily encoding the expected reward per se. For example, if a monkey expects a large reward it will pay more attention, be more motivated, and work harder, all of which might be driving the neuron’s response (Maunsell 2004). Distinguishing between these different possibilities behaviorally is extremely difficult. Our approach has been to record simultaneously from multiple regions of PFC, using the pattern of anatomical connections to direct our hypotheses. Our rationale is that by determining the latency at which neurons encode reward information, we can distinguish between neurons that are directly encoding the reward and those that are then using this information to drive other processes such as attention.

We trained monkeys to choose between different pictures associated with delivery of “payoffs”: specifically, different amounts of fruit juice (Wallis and Miller 2003b). Two pictures appeared to the left and right of a central fixation point, and monkeys were required to make a saccade to the location of the picture they wanted to choose. Monkeys soon learn to maximize their reward by selecting pictures associated with larger payoffs. We recorded simultaneously from dlPFC and vmPFC and found neurons in both areas that encoded the expected payoff. We predicted that neurons in vmPFC would encode the expected reward more quickly than neurons in dlPFC, because vmPFC receives direct projections from gustatory cortex. Indeed, this is precisely what we observed (Fig. 1). Furthermore, vmPFC neurons tended to encode the payoff alone, whereas dlPFC neurons encoded a combination of the payoff and upcoming motor response (Fig. 2). In sum, we interpreted these results as showing that reward information enters the frontal lobe through vmPFC and then passes to
The time-course of selectivity for the expected payoff (the amount of reward) across the dorsal prefrontal cortex (dIPFC) (gray) and ventromedial prefrontal cortex (vmPFC) (black) population of neurons. The thick line indicates the mean selectivity of the neurons; the error bars indicate the standard error of the mean. Both populations begin to encode the expected payoff at about the same time, but selectivity reaches its peak value in the vmPFC before the dIPFC population. The measure of selectivity is derived from the receiver operating characteristic (ROC) of each neuron’s firing rate. The ROC is the probability that an independent observer could correctly identify the payoff given the firing rate of the neuron. No selectivity equates to an ROC value of 0.5 (in practice it is slightly higher than this, because we rectify the ROC value during its calculation; small fluctuations from noise push the value to about 0.52). Maximal selectivity equates to a value of 1.0. Distribution of peak selectivity across the population of dIPFC and vmPFC neurons. The vmPFC population reaches its peak selectivity approximately 60ms before the dIPFC population (Wilcoxon’s rank-sum test, \(P < 0.05\)).
Fig. 2. Spike histograms from two single neurons encoding the expected payoff and/or the monkey’s response (a left or right saccade). Inset bar graphs indicate the mean neuronal firing rate (±standard error) during the presentation of the reward-predictive cue (the first 500 ms). Black indicates that the cue predicted the delivery of eight drops of juice, dark gray, four drops, and light gray, two drops. A vmPFC neuron encoding the predicted reward in a parametric fashion irrespective of reward direction. It showed a depression in its firing rate that was greatest for eight drops of juice, less for four drops, and least of all for two drops. Its firing rate, however, was the same irrespective of whether the monkey would make a left or right saccade. Significantly more vmPFC neurons (28%) showed this pattern of selectivity compared to dlPFC neurons (13%, chi squared = 9.8, $P < 0.005$). A dlPFC neuron that showed a complex pattern of selectivity which encoded a combination of the reward and the upcoming saccade. During the cue epoch, the neuron only discriminated between the different expected reward amounts when the monkey would make a rightward saccade (showing a high firing rate when eight drops of juice were expected). In contrast, during the subsequent period the same neuron was reward selective only when the monkey would make a leftward saccade. Significantly more dlPFC neurons (43%) encoded a combination of the reward and the response compared to vmPFC neurons (19%, chi squared = 19, $P < 0.0005$)
dlPFC, which uses the information to control behavior. Our results are in line with both prior neurophysiological and neuropsychological investigations of PFC. Similar to previous neurophysiological studies, we found neurons that encoded the expected payoff in both dlPFC and vmPFC. Furthermore, our results suggest that vmPFC may be the source of reward signals to dlPFC, which may explain why neuropsychological studies have pointed to the relative importance of vmPFC for tasks that require assessment of reward value.

2.2. Flexibility of Reward Representations

One could argue that vmPFC is simply receiving payoff information from lower-level sensory cortices. However, vmPFC differs from the sensory cortex in not only representing reward information but also representing it flexibly. For example, the response of vmPFC to payoffs is dependent on whether those payoffs satisfy a motivational requirement. Thus, vmPFC neurons that respond to a particular taste will lose that response if the monkey is satiated with that taste (Rolls et al. 1989). In contrast, gustatory neurons of the insular cortex are not affected by satiety (Yaxley et al. 1988). Furthermore, lesions of vmPFC impair the ability of monkeys to alter their choice behavior depending on their motivational state (Baxter et al. 2000). Similarly in humans, vmPFC activity associated with the taste or smell of a particular food is reduced when the subject is satiated with that food, while the vmPFC response to nonsatiated foods remains (Gottfried et al. 2003).

A second way in which vmPFC contributes to behavioral flexibility is in allowing us to respond efficiently to changes in the value of things in our environment. A simple test of this behavior is the stimulus–reward reversal. Here the subject learns that one of two pictures predicts a reward. Without warning, the experimenter switches which picture is associated with reward, and the subject has to switch the choices accordingly. Damage to vmPFC impairs the ability to do this in rats (Schoenbaum et al. 2003), monkeys (Dias et al. 1996), and humans (Rolls et al. 1994). In all three species, the problem is inhibiting the previously correct stimulus–reward association rather than in learning a new association. Thus, vmPFC seems to be more important for flexibly controlling behavior than learning about reward outcomes per se. Neuronal activity in vmPFC supports these findings. For example, if picture A, but not picture B, is followed by a reward, a vmPFC neuron may respond to picture A. When the contingencies are reversed, such that now picture B, but not picture A, predicts reward, the neuron likewise reverses its response, responding to picture B (Schoenbaum et al. 1999; Thorpe et al. 1983; Wallis and Miller 2003a).

2.3. Decision Making and Reward

So far, we have considered very simple decisions where the choice varies along a single dimension (reward size). However, the reasons behind our choices in real life are complex, often requiring us to consider many variables. For example,
deciding whether you want a coffee or cold soft drink requires considering a whole range of factors, not just their taste. You might consider their relative price, your energy level, and even the weather. There is evidence that vmPFC is important for integrating these multiple variables to derive an overall “value” for an action that can then guide choice behavior. Assad and colleagues trained monkeys to make simple choices between different drinks while they recorded the electrical activity of neurons in the monkeys’ vmPFC (Padoa-Schioppa and Assad 2006). It was not a simple decision, however, because the monkeys had to weigh not only the taste of the drink but also the volume that was available.

For example, a thirsty monkey might prefer the taste of fruit juice to water. If this is the case, then if the choice is between equal volumes of both, he will obviously choose the juice. However, increasing the volume of water can compensate for its less-desirable taste. If the volume of water is sufficiently large, relative to the juice, then the monkey will pick the water. At some point, the volume of water will compensate for its less-desirable taste exactly, and the monkey will be indifferent between the two choices. This indifference point in effect measures the value of the juice relative to the water. For example, if the monkey equates four drops of water with one drop of juice, we know that the monkey considers juice four times more valuable than water.

This task provided a sensitive behavioral measure of the monkey’s subjective preferences as well as the objective physical properties of the rewards. The key question was what would vmPFC neurons encode? Would they encode the subjective value of the drink or its objective physical properties? It turns out that most vmPFC neurons encode one of three aspects of the task. When the monkey first encounters the choice, many neurons encode the monkey’s subjective value of one or other of the drinks on offer. Then when the monkey is preparing its choice, vmPFC neurons encode the subjective value of the drink that the monkey will eventually choose. Finally, when the monkey actually receives the reward, vmPFC neurons encode the drink’s taste. These results suggest that vmPFC neurons are important for deriving the value of an outcome and using this value signal to guide choice behavior.

The authors focused on gustatory stimuli, but these processes could easily apply to more high level decision making. A common feature of decisions involves weighing one attribute of a choice against another. For example, let us return for a moment to Elliot, whom we met at the beginning of this chapter. After his series of poor life choices, he returned to live with his parents. He found an accountancy job, but it involved a 200-mile round-trip commute. After a few weeks, the company fired him for a lack of punctuality. Elliot had failed to determine the value of the job properly by not integrating all the factors relevant to the decision. Indeed, a recent systematic study of patients with vmPFC damage, found that they had a specific difficulty in integrating multiple attributes pertaining to a decision (Fellows and Farah 2005).

To make a choice effectively, one must consider not only the outcome. Economists and behavioral ecologists emphasize the importance of two other variables in making a decision: the cost of obtaining the payoff (in terms of time and
energy) and the risk that the payoff will not be obtained (Kahneman and Tversky 2000; Stephens and Krebs 1986). Neurons in vmPFC encode some of these different variables. For example, they encode time costs, specifically, whether a reward will occur following a short or long delay (Roesch and Olson 2005). Lesions of the medial region of vmPFC in rats produce deficits in the ability to integrate different-sized payoffs (different sizes of food reward) with different sizes of cost (the effort required to clamber over ramps of different sizes) to make efficient choices (Walton et al. 2002).

Extending these findings, we examined whether PFC neurons encode an abstract representation of value by integrating the major decision variables of payoff, cost, and risk (Kennerley et al. 2005). We trained monkeys to choose between pictures while we simultaneously recorded from dIPFC and vmPFC. Each picture was associated with a specific outcome. Some pictures were associated with a fixed amount of juice, but only on a certain proportion of trials (risk manipulation). Other pictures were associated with varying amounts of juice (payoff manipulation). Finally, some pictures were associated with a fixed amount of juice, but the subject had to earn the juice by pressing a lever a number of times (cost manipulation). Neurons encoded the choice outcomes in a variety of ways, including encoding a single variable, a combination of two variables, or all three variables. The majority of the selective neurons were located in vmPFC where about half encoded the outcome in some way. The neurons encoded a very specific representation of a wide variety of choice outcomes. Across the population, they encoded the value of the choice irrespective of how we manipulated its value. At a single neuron level, however, neurons would often show selectivity when we manipulated value in one way but not in another. This elaborate representation of the choice outcome is what we would expect from an area involved in reward-based decision making.

2.4. Abstract Reward Representation

In summary, PFC neurons encode a somewhat abstract representation of reward. This encoding scheme offers distinct computational advantages. When faced with two choices, A and B, one might imagine it would be simpler to directly compare them rather than going through an additional step of assigning them an abstract value. The problem with this is that as the number of available choices increases the number of direct comparisons increases exponentially. Thus, choosing between A, B, and C would require three comparisons (AB, AC, and BC), whereas choosing between A, B, C, and D requires six comparisons (AB, AC, AD, BC, BD, and CD). The solution quickly suffers from combinatorial explosion as the number of choices increases. In contrast, valuing each choice along a common reference scale provides a linear solution to the problem.

Incidentally, this is also the solution that society has adopted to the problem of valuing goods and services. Initially, we relied on barter, which involved directly comparing goods and was a system that suffered combinatorial explosion. Societies gradually switched to using a reference good against which to
compare other goods. Most commonly, this was gold, but people have used other reference goods, such as shells on the Solomon Islands, and cigarettes in prisons. Ultimately, the reference good came to have no intrinsic value, as is the case with paper money. The parallel between the mechanisms that humans use to value things, and the mechanisms that the brain uses, has been noticed by neuroscientists, who have termed the neuronal representation of abstract value a “neuronal currency” (Montague and Berns 2002).

An abstract representation provides important additional behavioral advantages, such as flexibility and a capacity to deal with novelty. For example, suppose an animal encounters a new food type. To determine whether it is worth choosing relative to other potential food sources, the animal must determine the value of that food. Following the barter system, the animal can only make this comparison by iteratively comparing the new food with all previously encountered foods. Following the currency system, on the other hand, the animal has only to perform a single calculation. By assigning the new food a value on the common reference scale, it knows the value of this foodstuff relative to all other foods. Second, it is often not clear how to compare directly very different outcomes. For example, how does a monkey decide between grooming a compatriot and eating a banana? Valuing the alternatives along a common reference scale helps with this. For example, although I have never needed to value my car in terms of bananas, I can readily do so because I can assign each item a dollar value (for the interested, at the time of writing a new compact car works out at approximately 30,000 bananas in the United States).

To conclude, there is accumulating evidence that PFC neurons encode an abstract representation of value that can then drive choice behavior appropriately. The advantage of this type of representation is that it provides flexibility and a capacity to deal with novelty, both of which are hallmarks of PFC function (Dias et al. 1997; Knight 1984; Norman 1986). However, although the ultimate goal of all behavior links to reward or punishment, our behavior is often under the control of intervening, arbitrary rules, and relationships. We may need to drive to the grocery store to obtain our food rewards, but in so doing, we must obey many arbitrary rules, such as driving through green lights but stopping at red lights. There is evidence that PFC also encodes such arbitrary stimulus–response associations.

3. Stimulus–Response Associations

Acquiring stimulus–response associations involves learning that an arbitrary response is associated with an arbitrary stimulus. For example, we learn that when the word “PUSH” is on a door, it instructs a different motor response than “PULL.” Reward is important for acquiring the stimulus–response association because it indicates which response is correct to each word (in this example, our reward is to get to the other side of the door). However, unlike reward-based learning, just knowing which response led to reward is insufficient, because
pushing and pulling are correct equally often and consequently are rewarded equally often. Instead, to solve the problem one must know both the stimulus and the response. Tasks that present two or more competing stimulus–response associations are referred to as conditional tasks because they follow the logic of “If stimulus A, then response B.”

Although reward-based learning is associated with the vmPFC, stimulus–response learning is more associated with the lateral PFC. The anatomical connections of this region are ideal for such learning. It is only weakly connected with areas processing reward but strongly connected with sensory and motor regions (Petrides and Pandya 1999). Neurons in the lateral PFC have properties that are appropriate for representing conditional rules. Their activity codes both sensory cues and the behavioral responses instructed by them (Sakagami and Niki 1994a,b; Watanabe 1986a,b). Miller and colleagues have explored these neuronal properties in detail (Asaad et al. 1998; Pasupathy and Miller 2005). They taught monkeys that two objects (A and B) instructed whether to make a leftward or rightward eye movement. However, once the monkeys had solved the problem, the experimenters reversed the contingencies (Fig. 3). For example, suppose the monkeys initially learned that object A instructed a leftward saccade and object B a rightward saccade. When the contingencies reversed, they now had to learn that object A instructed a rightward saccade and B a leftward saccade. The monkeys were able to learn several of these reversals in a single session. Crucially, this enabled the experimenters to determine to which aspect of the task a given neuron was responding. As in previous studies, some PFC neurons responded to the visual properties of the objects (Fig. 4a) whereas others encoded the motor response (Fig. 4b). However, there were also neurons that encoded the combination of cue and response (Fig. 4c). For example, a neuron might respond to object A, but only when object A instructed a leftward saccade. These findings strongly suggest that PFC neuronal activity represents the conditional rule.

![Fig. 3. Schematic of the visuomotor conditional task. The animal learns to associate a foveally presented visual cue with a saccade to the right or left. Once the monkey learned the association, the cue–response pairings switched; the cue that had required a rightward saccade now required a leftward saccade and vice versa. This enabled the investigators to determine whether a neuron responded to the visual properties of the cue, the motor response, or the cue–response association](image-url)
Neuropsychological evidence also points to the involvement of the frontal lobes in the performance of conditional tasks (Petrides 1985). Using experimentally induced lesions in monkeys, investigators have shown that performance of conditional tasks is dependent on interactions between PFC and inferior temporal cortex, a region responsible for processing objects. Monkeys with a disconnection between lateral PFC and the temporal lobe have great difficulty in learning conditional rules, although they are still capable of acquiring stimulus–reward associations, which presumably depends on the intact vmPFC (Parker and Gaffan 1998).

The obvious route by which PFC and the temporal lobe might interact is via the uncinate fascicle, a direct corticocortical tract between the regions (Ungerleider et al. 1989). However, transection of the uncinate fascicle does not affect
conditional learning (Eacott and Gaffan 1992). Another possible route of interaction is through the striatum. Both the temporal lobe and the frontal lobe project to the striatum, which then feeds back to both structures via the thalamus (Middleton and Strick 1996). Lesions of the striatum impair the performance of conditional tasks (Reading et al. 1991), whereas neurons in the caudate show very similar properties to those in PFC during performance of such tasks (Pasupathy and Miller 2005).

However, one critical difference between PFC and the striatum is the speed at which the two structures learn stimulus–response associations (Fig. 5). Neurons

![Diagram](image)

**Fig. 5.** Population strength of direction selectivity shown as a function of correct trials since the contingency reversal and time from cue onset for (a) PFC and (b) Cd (caudate). The measure of selectivity in this case is the proportion of explainable variance in the neuronal firing rate that is attributable to the direction factor. *White dots* indicate “rise time” (time for the neuronal population to reach half the value of its maximum selectivity). Selectivity strength increases and appears earlier in both areas as learning takes place. However, changes appear earlier and reach an asymptote sooner in the striatum than the PFC. c Rise times for PFC (inverted triangles) and Cd (circles). *Dashed lines* show sigmoids of best fit. d Behavioral performance immediately following a reversal of the contingencies, smoothed using a window of five trials. The rate at which the monkey learns closely parallels the rate of learning in PFC, whereas the Cd learns substantially quicker
in the caudate nucleus learn to encode new visuo-motor associations very rapidly. Indeed, they learn the association before the monkey’s behavior reflects this learning. In contrast, PFC neurons learn the associations more gradually, and in fact their rate of learning closely parallels the monkey’s learning rate. This observation suggests that the striatum is a specialized brain region for learning conditional rules and the output of this learning “trains” slower learning mechanisms in PFC. It also suggests that PFC controls behavior more directly than the striatum, as it is not until the learning is represented in PFC that behavior changes. However, although subcortical structures can contribute to the learning of these simple stimulus–response associations, if the task is more complex cortical mechanisms take over, as we discuss in the next section.

4. High-Level Behavioral Control

4.1. Abstract Rules

Consider a ringing telephone. This auditory stimulus instructs a specific motor response: to pick up the phone and answer it. However, this response is only appropriate if it is our phone. If we are guests in someone else’s house, we should wait for our host to answer the phone. In other words, the correct motor response depends not only on the sensory stimulus, but also on the context in which it occurs. PFC is responsible for implementing these more complex rules. For example, a patient with PFC damage would answer the phone irrespective of whether they were in a guest’s house or at home (Lhermitte et al. 1986). It is the capacity to use these high-level, overarching behavior-guiding rules, such as the social rules that govern answering telephones, which seems to be most heavily dependent on the integrity of PFC.

One task that directly tests this ability is the Wisconsin Card Sorting Test (WCST). The patient is required to sort a deck of cards on which there are a number of colored shapes. They have to determine by trial and error the correct property by which to sort the cards. For example, if the “color” rule were in effect, the subjects would have to sort the cards according to the color of the shapes. Unbeknownst to the subject, the experimenter can switch which is the correct rule, and the subjects need to detect this and modify their behavior accordingly. Patients with PFC damage, particularly lateral PFC, have difficulty doing this (Milner 1963). Using more-focal, experimentally induced lesions in monkeys, experimenters have confirmed that dIPFC is more important for rule or strategy implementation than vmPFC (Dias et al. 1996; Roberts and Wallis 2000; Wallis et al. 2001b).

We recently devised a task that enabled us to examine abstract rule use in a neurophysiological study (Wallis et al. 2001a). We trained monkeys to use two abstract rules: “match” versus “nonmatch” (Fig. 6). They faced a computer screen and viewed two successively presented pictures. For the match rule, the monkeys released the lever if the pictures were the same and continued to hold the lever if the pictures were different. For the nonmatch rule, the reverse was
true; monkeys released if the pictures were different and held if they were the same. A cue presented at the same time as the first picture instructed the monkey as to the correct rule. To disambiguate neuronal responses to the physical properties of the cue from responses to the rule that the cue instructed, cues signifying the same rule were from different modalities, while cues signifying different rules were from the same modality. The monkeys could perform this task well above chance levels even when seeing pictures for the very first time. This result indicates that they had abstracted two overarching principles of the task that could then be applied to novel stimuli—the minimal definition of an abstract rule.

The most prevalent activity across the PFC was the encoding of the current rule. Figure 7 shows a rule-selective neuron that exhibited greater activity when the match rule was in effect as opposed to the nonmatch rule. The activity cannot be explained by the physical properties of the cue or the picture, because activity was the same regardless of which cue was used to instruct the monkey, and regardless of which picture the monkey was remembering. It does not reflect the upcoming response, because the monkey did not know whether the second-presented picture would require a response. Nor does it reflect differences in reward expectation, as the expectation of reward was the same regardless of which rule was in effect. Furthermore, the performance of the monkeys was virtually identical for the two types of rules (error rates differed by less than 0.1% and reaction times by less than 7 ms). Thus, the most parsimonious explanation

Fig. 6. Schematic of the behavioral task. Monkeys grasped a lever and maintained central fixation. A sample object was followed by a brief delay, and then by a test object. Illustrated are two trial types for each rule (bifurcating arrows). For the match rule, the monkeys released a lever if the test object matched the sample. For the nonmatch rule, they released the lever if the test object did not match. Otherwise, they held the lever through a second delay until appearance of a second test object that always required a response. A cue, presented simultaneously with the sample picture, instructed the monkey as to which rule was in effect.
is that the differences in activity reflected the abstract rule that the monkey was currently using to guide its behavior.

What function does the ability to abstract a rule serve? Abstraction is a type of generalization that permits a shortcut in learning, allowing the animal to maximize the amount of reward available from a particular situation. To illustrate this, consider the above task. The monkey could potentially solve the task as a series of paired associates (in fact, 16 associations, consisting of four different
pictures each paired with four different cues). For example, the monkey might learn that whenever a drop of juice accompanies the presentation of the chef picture, then the correct response at the test phase is to choose the chef. However, notice that this type of learning tells the monkey nothing about which response is appropriate to a lion appearing with a drop of juice. In other words, unless the monkey abstracts the rule that juice indicates that he should match, then each time new pictures are used the monkey would have to learn an entirely new set of 16 associations by trial and error. The problem with this trial-and-error learning is that errors are lost opportunities for reward. Given that the monkeys performed well above chance when they encountered novel pictures, it is clear that they are not engaging in trial-and-error learning, but rather have abstracted two rules that they can then apply as required.

The neuronal encoding also reflects this shortcut that abstraction of the rule permits. It would be entirely possible for the monkey to solve the task without single cells encoding the rule. For example, there might be two populations of cells, one encoding the match and nonmatch rule when the cues are from the auditory modality and one encoding this information in the taste modality. However, such a solution is computationally expensive; if a third modality was introduced, a third population of cells would be required. It is more efficient to abstract the rule that cues presented in different modalities commonly instruct, and indeed this is the solution that the brain uses. The prevalence of neurons encoding such rules in the PFC is consistent with the impaired ability to switch between rules flexibly that occurs after PFC damage in both monkeys and humans.

4.2. Task Set

The concept of a behavior-guiding rule closely relates to the notion of a “task set.” German experimental psychologists at the start of the 20 century were the first to discuss task sets or “Einstellung,” and the concept was later invoked by Lurchins, who observed that subjects trained to solve a problem with a complex strategy would later fail to notice when a simpler strategy would achieve the same goal (Lurchins 1942). This phenomenon, called the Einstellung effect, arose because the subjects developed a task set to solve the problems, essentially attending to certain aspects of the task at the exclusion of other aspects. The most obvious behavioral evidence that a subject has developed a task set is their tendency to show a reaction time cost when switching from one task to another, the so-called switch cost. This action is thought to reflect a kind of mental “gear changing” as the subject switches between the different stimulus attributes, conceptual criteria, goal states, and action rules necessary to solve different tasks (Monsell 2003).

There is evidence that the implementation of task set depends on the PFC, particularly the lateral regions. For example, neuroimaging studies have consistently revealed activation of lateral PFC during task set switching (Dreher and Berman 2002; Sohn et al. 2000). Furthermore, a study of patients with frontal
lobe damage revealed that the more extensive the damage to right ventrolateral PFC (vIPFC), the greater their difficulty in switching between tasks (Aron et al. 2004). Neuronal properties in PFC are also consistent with representing a task set. Many of the sensory, mnemonic, and motoric responses of PFC neurons are dependent on the precise task that the subject is performing (Asaad et al. 2000; Hoshi et al. 1998; Hoshi et al. 2000; White and Wise 1999). For example, a neuron might respond to a visual object when the monkey had to hold that object in working memory, but not when the same object instructed a conditional motor response (Asaad et al. 2000). Furthermore, some neurons showed a tonic increase in firing rate whenever the monkey was performing one task relative to another, consistent with encoding a task set (Asaad et al. 2000).

4.3. PFC Interactions with Other Brain Areas

Irrespective of whether the high-level control of behavior occurs by the implementation of specific behavior-guiding rules, or whether it involves the construction of task sets of relevant sensory and conceptual information, these representations must interact with other brain areas to influence behavior. One possibility is that this interaction takes place through the striatum, similar to behavioral control by stimulus–response associations. We have recently explored whether this is the case by simultaneously recording from multiple brain regions including the striatum, inferior temporal cortex, and multiple areas within the frontal lobe (PFC and premotor cortex) during the performance of the abstract rule task (see Fig. 6). We found a lot of overlap in neuronal properties: every area encoded two or more task variables (the rules, the pictures, the match/non-match status of the test picture, and the behavioral responses). However, there were differences. Unlike the encoding of stimulus–response associations, which were equally encoded by PFC and the striatum (albeit more gradually acquired by PFC), the frontal lobe clearly encoded abstract rules more strongly than the striatum (Fig. 8), which implies a limit on the learning capabilities of the striatum. Specifically, it suggests that as the control of behavior becomes more abstract it also becomes more dependent on the cortex. In addition, the PFC was more of a “crossroad” for this task than other areas; it was the only area to represent all major task variables. This, of course, makes sense because the PFC is at an anatomical crossroad. It is one of the most well connected brain areas, directly connected with most of the cerebral cortex, including premotor cortex and inferior temporal cortex, and many subcortical structures, including the dorsal striatum.

Investigators have also examined the interaction between brain regions that occurs during the implementation of task sets. For example, functional connectivity studies using fMRI show that activation in the most anterior regions of PFC closely correlates with different PFC regions depending on which task the subject is preparing to perform (Sakai and Passingham 2003). For example, anterior PFC interacts strongly with dlPFC, a region associated with spatial working memory, when subjects prepare to perform a spatial task. In contrast, it interacts with an
area implicated in phonological working memory, posterior vIPFC, when subjects were required to perform a verbal task.

These same authors have recently observed similar results for tasks involving the processing of phonological or semantic information (Sakai and Passingham 2006). Anterior PFC activity correlated with anterior vIPFC in the semantic task, whereas it correlated with more posterior ventral frontal regions during the phonological task. Furthermore, the higher the anterior PFC activity, the faster the subjects performed the task, suggesting that anterior PFC is causal in implementing the improvement in performance observed during the development of a task set.
5. Functional Organization

In summary, PFC controls behavior at several different levels. However, different PFC regions are somewhat specialized with regard to which operations they control. There appears to be a dorsal–ventral gradient, whereby ventral regions control behavior driven by the satisfaction of low-level motivational states, whereas more dorsal regions control behavior according to more arbitrary relationships among sensory stimuli and motor responses. There also appears to be an anterior–posterior gradient whereby more concrete relationships (such as specific stimulus–response associations) are encoded by more posterior PFC regions and more abstract relationships (task sets) are encoded by more anterior PFC regions.

A recent fMRI study by Koechlin and colleagues explored this distinction more directly (Koechlin et al. 2003). They trained subjects on a task that involved three distinct levels of control. At the most basic level the subjects simply had to perform specific stimulus–response associations. At the next level subjects had to reconfigure stimulus–response associations dynamically according to specific instructional contextual cues. Finally, the subjects had to select task sets (sets of stimulus–response associations evoked by a specific contextual cue) according to the temporal order in which they occurred (an episodic cue). The authors discovered that as the subject was required to use these progressively higher levels of control there was an increased recruitment of progressively more anterior frontal regions. Thus, stimulus control recruited premotor cortex, contextual control recruited premotor cortex, and posterior lateral PFC, and episodic control recruited both anterior and posterior lateral PFC as well as premotor cortex.

6. Conclusion and Future Directions

The past decade has seen a move away from investigating the PFC in purely sensorimotor terms toward understanding how the PFC represents behaviorally relevant variables. This has proven challenging. When investigators focused on understanding how the PFC encoded sensorimotor variables, they were able to draw on decades of psychophysical research, our understanding of how the posterior brain encodes this information, and an intuitive grasp of the important variables underlying perception and action. Unfortunately, we lack integrative models of how behavior might be controlled. Consequently, progress has been slower as we struggle to determine the important behavioral variables that might form a framework for the functional organization of PFC.

Although it often appears that large swathes of PFC are responsible for encoding a specific function (Duncan and Owen 2000; Wallis et al. 2001a), a broad organization of function can be discerned. There seem to be two functional gradients in PFC. The first is a temperature gradient, which runs from ventral PFC to dorsal PFC. The “hot” emotional and reward processes take place in ventral PFC, while the “cold” cognitive processes take place in dorsal PFC. The second
is a gradient of abstraction and runs along the anterior–posterior axis. Posterior PFC processes specific and concrete associations, whereas progressively more anterior areas process abstract, high-level rules and strategies.

This model of the functional anatomy of PFC suggests avenues of future research. For example, the functional organization of ventral PFC remains largely unknown. However, if the abstraction gradient were a general feature of PFC organization, then one would expect posterior ventral areas to encode specific reward associations, such as knowledge that a specific sensory stimulus predicts a food reward. In contrast, anterior ventral PFC would encode the more abstract valuation schemes, such as integrating multiple attributes of a decision. Future research can also examine how these different areas interact together during behavior. For example, one might expect a back-and-forth recruitment of anterior and posterior PFC regions as behavior requires more or less complex control.

In conclusion, the notion of a function and anatomical hierarchy, which has proven useful in understanding the organization of sensory and motor systems, appears to extend to PFC. As befits its place at the apex of the perception-action cycle, the PFC is particularly involved when abstraction and generalization are required. These capacities enable the organism to behave flexibly and cope efficiently with novel situations, which are hallmark features of PFC function.

References

Bechara A, Damasio AR, Damasio H, Anderson SW (1994) Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50:7–15
Lurchins AS (1942) Mechanization in problem solving. Psychol Monogr 54


Watanabe M (1986b) Prefrontal unit activity during delayed conditional go/no-go discrimination in the monkey. II. Relation to go and no-go responses. Brain Res 382:15–27
13
The Prefrontal Cortex as a Model System to Understand Representation and Processing of Information

Shintaro Funahashi

1. Introduction

Working memory is a mechanism for short-term active maintenance of information as well as for processing of maintained information (Baddeley 1986; Baddeley and Hitch 1974). Working memory is known to be a fundamental mechanism for higher cognitive functions, such as thinking, reasoning, language comprehension, and decision making. When we perform these cognitive functions, the brain monitors the external world continually, pays attention to the stimuli important for achieving the particular goal, receives necessary information from the external world to achieve the goal, retrieves related information from long-term memory, maintains this information temporarily for processing, then sends the outcome of information processing to brain areas where it is utilized. Among these processes, temporarily active maintenance of information and processing of maintaining information are essential components to achieve cognitive function. These two components are also the main components of working memory. Because working memory is known to be a fundamental mechanism for higher cognitive functions, understanding of neural mechanisms for working memory could provide important evidence to understand underlying mechanisms for thinking, planning, and decision making.

In the late 1980s, Goldman-Rakic (1987, 1998) proposed the idea that working memory is an important concept to understand functions of the dorsolateral prefrontal cortex (DLPFC). This idea has been supported by a variety of experiments including lesion studies (see reviews by Fuster 1997; Goldman-Rakic 1987; Petrides 1994), neurophysiological studies using nonhuman primates (see reviews by Funahashi and Kubota 1994; Funahashi and Takeda 2002; Fuster 1997; Goldman-Rakic 1999), and brain imaging studies using human subjects (see Stuss and Knight 2002). The close relationships between working memory and

Department of Cognitive and Behavioral Sciences, Graduate School of Human and Environmental Studies, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
functions of the DLPFC indicate that neurophysiological investigations of DLPFC neurons would provide important evidences for essential neural mechanisms of working memory, such as how information is represented in neural activity and how information is processed in the nervous system. For example, by examining how a variety of information is maintained by neural activity observed in the DLPFC, we could understand how the nervous system represents a variety of information. In addition, by examining the temporal change of maintaining information in DLPFC neurons along the progress of a trial, we could obtain a hint to understand how the information is processed in the nervous system.

Neurophysiological studies in the DLPFC revealed that one type of neural activity (e.g., tonic sustained activation during the delay period or delay-period activity) is a neural correlate of the mechanism for active maintenance of information (Funahashi 2001; Funahashi and Takeda 2002; Fuster 1997; Goldman-Rakic 1998; Miller 2000). The characteristics of delay-period activity have been studied in detail (Funahashi et al. 1989; Fuster 1973; Kojima and Goldman-Rakic 1982; Takeda and Funahashi 2002). Although evidence of the neural mechanism for temporary active maintenance of information has been accumulated in the DLPFC, little is known about the mechanisms for manipulating, integrating, or processing information. To understand how information is processed in the DLPFC, we first need to know what information neural activities represent while a subject performs working memory tasks. Processing information can be considered as a change of the information represented in neural activities from one kind to another. Therefore, to understand the neural mechanism for information processing, we need to know how information represented in neural activities changes along the progress of the working memory task and what mechanisms produce this change.

When we consider the mechanisms for information processing, we may need to focus on the behavior of a population of neural activities, because single particular information is always encoded by a population of neurons. For example, a large number of prefrontal neurons responded to the visual stimulus in the same manner consistently when the visual stimulus was presented at the same place in the visual field (Funahashi et al. 1990; Mikami et al. 1982; Suzuki and Azuma 1983). However, the best responding position, the spatial tuning, the magnitude of the activation, and its temporal pattern all differ from neuron to neuron (Funahashi et al. 1990; Takeda and Funahashi 2002). In spite of these differences, all these neurons encode the position where the visual stimulus is presented in the visual field by changing their activities systematically. Therefore, to understand how visual information is represented and processed in the nervous system, we need to consider not only the information that each single-neuron activity represents but also the information that a population of neural activities represents. In this chapter, we describe evidence of neural mechanisms related to how the information is represented in activities of DLPFC neurons by examining characteristics of delay-period activity and how the information is processed in the DLPFC by examining a temporal change of information represented in a population of DLPFC activities.
2. Prefrontal Cortex, Delayed-Response Task, and Spatial Working Memory

It has been known that monkeys with bilateral lesions in the DLPFC show severe impairment in delayed-response performance since Jacobsen (1936) first reported this observation. For an ordinary experiment using the delayed-response task, the Wisconsin general test apparatus (WGTA) has been used. In the WGTA, the monkey is in a small cage and the table is placed between the monkey’s cage and the experimenter. An opaque screen is set in front of the cage. After the intertrial interval of several seconds, the opaque screen is raised. The experimenter puts the reward at one of two possible baiting places on the table and covers both baiting places with identical opaque plates, while the monkey is watching the experimenter’s behavior. Then, an enforced delay of several seconds or minutes is introduced by lowering an opaque screen. At the end of the delay period, the opaque screen is raised and the monkey can now select the place where the reward is hidden. Because the two baiting places are covered with identical opaque plates, the monkey is required to select a correct place based on his memory. If the monkey selects a correct place, the reward is given. If the monkey selects a wrong place, the opaque screen is dropped immediately and an additional time-out period is introduced.

To perform the delayed-response task correctly, the subject is required to internally maintain the position where the reward was baited during the delay period (internal representation of spatial position). The subject is also required to utilize maintained information for the behavior to select a correct position (processing of internally represented information). In addition, because the experimenter randomly selects the position where the reward is baited, the subject is required to update the information regarding the reward-baited position from trial to trial (manipulation of internally represented information). Thus, to perform the delayed-response task correctly, the subject needs to use the spatial working memory mechanism, by which spatial information is internally represented, maintained temporarily, and processed and manipulated to perform appropriate behavior.

The performance of the delayed-response task is closely related to the function of the DLPFC. Since Jacobsen (1936) first reported that large bilateral lesions of the lateral part of the prefrontal cortex in monkeys produced permanent impairment of delayed-response performances, subsequent lesion studies using monkeys repeatedly confirmed Jacobsen’s observation (see Fuster (1997). Butters et al. (1971) had shown that the middle one-third of the principal sulcal area is the crucial area for correct performances of the delayed-response task, because the bilateral lesions of this portion of the DLPFC produced severe impairment in delayed-response performance. Although bilateral lesions of the DLPFC were thought to be necessary to produce an impairment of the delayed-response task, Funahashi et al. (1993a) showed that monkeys having only a unilateral lesion in the DLPFC exhibited a specific deficit (mnemonic scotomas) in an oculomotor...
version of the delayed-response task, such that monkeys performed incorrect memory-guided saccades with wrong directions only when the visual targets were presented within the visual field contralateral to the hemisphere with the lesion. In addition to animal studies, Freedman and Oscar-Berman (1986) showed that human patients with bilateral frontal lobe disease exhibited selective delayed-response deficits. Recent noninvasive brain imaging studies showed significant activation in the DLPFC while human subjects performed delayed-response tasks and tasks that require temporary maintenance of spatial information (Brown et al. 2004; Curtis and D’Esposito 2006; Curtis et al. 2004; Postle et al. 2000). Therefore, it is reasonable to predict that neural mechanisms necessary for correct performance of the delayed-response task, such as internal representation, processing, and manipulation of spatial information, are present in the DLPFC.

3. Oculomotor Delayed-Response Task

We have used an oculomotor version of the delayed-response task (oculomotor delayed-response task or ODR task) (Funahashi et al. 1989, 1990, 1991, 1993a,b; Takeda and Funahashi 2002) to understand neural mechanisms of spatial working memory in the DLPFC. The ODR task is a modification of a memory-guided saccade task that was developed and used extensively for oculomotor studies (Bruce and Goldberg 1985; Hikosaka and Wurtz 1983). In this task, the monkey sat quietly in a monkey chair in a dark sound-attenuated room. The monkey faced a TV monitor on which a fixation point and visual cues were presented. The monkey’s eye positions were monitored by a magnetic search coil technique. In the ordinary ODR task (Fig. 1, top), the monkey was required to make a memory-guided saccade to the location where the visual cue had been presented.

![Fig. 1. Schematic drawings of oculomotor delayed-response tasks. Upper row: an oculomotor delayed-response (ODR) task, in which the monkey was required to make a saccade to the direction where the visual cue had been presented. Lower row: a rotary ODR (R-ODR) task, in which the monkey was required to make a saccade to the direction 90° clockwise from the visual cue direction.](image-url)
After a 5-s intertrial interval, a fixation point (FP; small white circle) was presented at the center of the TV monitor. If the monkey looked at the FP for 1 s (fixation period), a visual cue (a white circle) was presented for 0.5 s (cue period) randomly at one of eight predetermined peripheral positions around the FP. The monkey was required to maintain fixation at the FP throughout the 0.5-s cue period and subsequent 3-s delay period. At the end of the delay period, the FP was extinguished. This was the go signal for the monkey to make a saccade within 0.4 s (response period) to the position where the visual cue had been presented. If the monkey made a correct saccade, a drop of liquid reward was given. Because the cue position was selected from eight predetermined peripheral positions randomly by a computer, the monkey could not predict the position where the visual cue would appear in the next trial.

Maintenance of spatial information of the visual cue during the delay period is essential for ODR performances. Therefore, a hint regarding neural mechanisms necessary for the internal representation of the spatial information can be obtained by examining how spatial information is encoded by neural activities during the delay period and how different spatial information is encoded by neural activities. In the ODR task, spatial information is initially informed to the subject by presenting the visual cue. However, this spatial information is eventually needed to transform into motor information because saccadic eye movement toward the visual cue direction is needed to obtain the reward. Therefore, activity observed in some DLPFC neurons would encode the spatial information of the visual cue, whereas activity observed in other neurons would encode the motor information for forthcoming saccadic eye movements. Thus, questions regarding how sensory or motor information is encoded by neural activities can be answered by examining what information the activity of each neuron encodes. In addition, spatial information is eventually transformed into motor information in the brain while the subject performs the ODR trial. Therefore, a hint regarding a neural mechanism for processing internally representing information can be obtained by examining how the information encoded by neural activities changes along the progress of the ODR task. Thus, an analysis of neural activity observed in DLPFC neurons during ODR performances would provide important evidence to understand neural mechanisms regarding how information is represented internally and how internally represented information is processed.

4. Task-Related Activities in the DLPFC

Several task-related activities have been observed in the DLPFC while monkeys performed the ODR task. One of these activities is cue-period activity, which is a phasic excitatory response occurring about 120 ms after the visual cue presentation (Funahashi et al. 1990; Takeda and Funahashi 2002). Most cue-period activity exhibited spatial selectivity, such that cue-period activity was observed only when the visual cues were presented within a certain area of the visual field (Funahashi et al. 1990). The same response was also observed in a visual probe
task, in which visual stimuli were presented at the same positions as positions where the visual cues were presented in the ODR task but had no behavioral significance (Funahashi et al. 1990). Neurons in the DLPFC are known to respond to simple visual stimuli (e.g., spot or bar stimuli) and have visual receptive fields (Mikami et al. 1982; Suzuki and Azuma 1983). Therefore, cue-period activity is the visual response to the presentation of the visual cues. The presence of the directional selectivity in cue-period activity would correspond to the presence of the visual receptive field in DLPFC neurons having cue-period activity.

Another task-related activity observed in the ODR task is response-period activity (Funahashi et al. 1991; Takeda and Funahashi 2002). As saccadic eye movements were performed during the response period, response-period activity can be considered saccade-related activity. Response-period activity can be classified into two groups based on whether response-period activity starts before or after the initiation of the saccadic eye movement (presaccadic or postsaccadic activity, respectively). Although neurons exhibiting presaccadic activity were in the minority, response characteristics of presaccadic activity observed in the DLPFC were very similar to those observed in the frontal eye field (Bruce and Goldberg 1985). Therefore, presaccadic activity observed in the DLPFC could participate in the generation and control of saccadic eye movements.

On the other hand, much of the response-period activity was postsaccadic in the DLPFC (Funahashi et al. 1991; Takeda and Funahashi 2002). Because postsaccadic activity was initiated about 100ms after the initiation of the saccadic eye movement, this activity could not participate in the initiation, the execution, or the control of the saccadic eye movements. Almost all postsaccadic activities showed directional selectivity, such that postsaccadic activity was observed only when the monkey performed saccadic eye movements toward particular directions, mainly contralateral to the hemisphere where the neuron was located (Funahashi et al. 1991). In addition, the timing of the initiation of postsaccadic activity coincided with the termination of delay-period activity (Goldman-Rakic et al. 1990). Therefore, it is concluded that postsaccadic activity is a feedback signal from oculomotor control centers and participates in the control of the maintenance of delay-period activity by playing as a reset signal of delay-period activity (Funahashi and Kubota 1994).

The most important task-related activity observed in the DLPFC is delay-period activity. Most delay-period activity was a tonic sustained activation during the delay period (Fig. 2A), although some were gradually increasing or decreasing types of activity (Funahashi and Takeda 2002). Delay-period activity was usually initiated several hundred milliseconds (ms) after the presentation of the visual cue, maintained during the delay period, and ended just after the saccadic eye movement was initiated (Funahashi et al. 1989). No visual stimulus was presented during the delay period except the FP, and the monkey was only gazing at the FP without any gross movement. Therefore, delay-period activity is thought to be a neural correlate of the mechanism for temporary maintenance of information necessary to perform the ODR task (Funahashi 2001; Funahashi and Kubota 1994; Goldman-Rakic 1987).
A ODR task

B Tuning curve

C Polar plots of best directions

Fig. 2. Characteristics of delay-period activity observed in dorsolateral prefrontal cortex (DLPFC) neurons. A An example of directional delay-period activity. B An example of a tuning curve of directional delay-period activity. C Polar distribution of the best directions of delay-period activity. A majority of the best directions were directed toward the contralateral visual field.
5. Delay-Period Activity and Temporary Maintenance of Information

The important feature of delay-period activity is that much of this activity exhibits directional preferences (Funahashi et al. 1989; Rainer et al. 1998; Takeda and Funahashi 2002). Delay-period activity was observed only when the visual cue was presented at one or a few adjacent positions in the visual field (see Fig. 2A). The preferred direction can be determined as the direction at which the maximum delay-period activity was observed in the tuning curve constructed by fitting mean discharge rates of delay-period activity across all target directions on a Gaussian function (Fig. 2B). Although preferred directions of delay-period activity differed from neuron to neuron, the majority of preferred directions were directed toward the contralateral visual field of the hemisphere where neurons were located (Funahashi et al. 1989; Takeda and Funahashi 2002) (Fig. 2C).

An essential feature of working memory is temporarily active maintenance of information necessary to perform a cognitive task (Baddeley 1986). As explained above, delay-period activity showed a tonic sustained activation during the delay period and exhibited directional selectivity. The duration of delay-period activity was prolonged or shortened depending on the length of the delay period (Funahashi et al. 1989; Fuster 1973; Kojima and Goldman-Rakic 1982). Delay-period activity was observed only when the monkey performed the correct responses (Funahashi et al. 1989; Fuster 1973). When the monkey made an error, delay-period activity was not observed or was truncated. These results indicate that delay-period activity observed in the DLPFC is a neural correlate of the mechanism for temporarily active maintenance of spatial information (Funahashi and Kubota 1994; Funahashi and Takeda 2002; Fuster 1997; Goldman-Rakic 1998). Therefore, it can be concluded that neural mechanisms of internal representation and temporary maintenance of information correspond to the tonic activation of a neuron or a population of neurons during the delay period. Different information would be represented or maintained by the tonic activation of different neurons or a different population of neurons, because every DLPFC neuron having delay-period activity exhibited a different directional preference.

6. Information Represented by Delay-Period Activity

If delay-period activity is a neural correlate of the mechanism for temporarily active maintenance of information, the next question would be what information delay-period activity maintains. To answer this question, Funahashi et al. (1993b) examined characteristics of delay-period activity using two kinds of the ODR tasks (delayed pro-saccade task and delayed anti-saccade task). In the delayed pro-saccade task, the monkey was required to make a memory-guided saccade toward the cue direction, whereas in the delayed anti-saccade task, the monkey was required to make a memory-guided saccade toward the opposite to the cue
Directional selectivity of delay-period activity recorded from the same DLPFC neuron was compared between these two tasks. The result showed that a great majority (about 70%) of delay-period activity represented information regarding visual cue direction, whereas a minority represented information regarding the saccade direction (Funahashi et al. 1993b). Similar results were obtained by Takeda and Funahashi (2002). They used the ODR task and a rotatory ODR (R-ODR) task, in which monkeys were required to perform memory-guided saccades to the direction 90° clockwise from the cue direction (Fig. 1, bottom). By comparing the best directions of delay-period activity of the same neuron between the ODR and R-ODR tasks, they showed that a large proportion (86%) of directional delay-period activity represented information regarding the visual cue direction, whereas a small proportion (13%) represented information regarding the saccade direction. Thus, both studies showed that the majority of directional delay-period activity represented information regarding where the visual cues were presented. As preferred direction of delay-period activity differed from neuron to neuron, Funahashi et al. (1989) proposed that DLPFC neurons showing directional delay-period activity have mnemonic receptive fields (memory fields) in the visual field. This idea was supported by Rainer et al. (1998) Thus, using spatial working memory tasks such as the ODR task, we could show that delay-period activity observed in the DLPFC is a neural correlate of the mechanism for maintaining information temporarily. In addition, we could also show that, in the DLPFC, the majority of delay-period activity represents retrospective information (e.g., visual information), whereas the minority represents prospective information (e.g., forthcoming motor information). Similar results have been observed recently by Genovesio et al. (2006).

Experiments using nonspatial working memory tasks, such as delayed matching-to-sample tasks and delayed conditional tasks, have also revealed that delay-period activity is a neural correlate of the mechanism for maintaining nonspatial information, such as faces (O’Scalaidhe et al. 1999; Wilson et al. 1993), and object shapes, patterns, or colors (Freedman et al. 2001; Inoue and Mikami 2006; Lauwereyns et al. 2001; Miller et al. 1996; Rainer and Miller 2000, 2002; Rainer et al. 1999; Rao et al. 1997; Sakagami and Niki 1994; Sakagami et al. 2001). These studies show that delay-period activities of different DLPFC neurons exhibit different preferences in stimulus modality and quality. These results indicate that each DLPFC neuron exhibiting delay-period activity represents different stimulus modality or quality. However, Romo et al. (1999) showed that discharge rates of delay-period activity observed in a DLPFC neuron varied as a monotonic function of the vibration frequency in a somatosensory discrimination task, in which monkeys were required to discriminate the difference in frequency between two mechanical vibrations applied to the fingertips. Based on this result, they concluded that this monotonic stimulus encoding could be a basic strategy to represent the quality of one-dimensional sensory stimulus in working memory (Brody et al. 2003). In addition, White and Wise (1999) found rule-dependent neural activity in the DLPFC. They showed that DLPFC neurons maintain rules of behavioral tasks by modulating the magnitude of task-related activity. Similar
results have been reported by Wallis et al. (2001), Wallis and Miller (2003), and Mansouri et al. (2006). Further, Hoshi et al. (2000) showed that prefrontal delay-period activity reflects motor selections. Saito et al. (2005) and Mushiake et al. (2006) showed that prefrontal delay activity reflected multiple steps of future events in action plans (e.g., immediate goal or final goal of sequential behavior).

Recently, Fukushima et al. (2004) showed that most of the delay-period activity observed in the DLPFC encoded spatial target representations (e.g., direction of the saccade) when the monkeys performed an ODR-like task but were required to update the target position sequentially by the presentation of nonspatial target-shift cues. Similar observation was also reported by Amemori and Sawa-guchi (2006). These results suggest that, although the majority of delay-period activity represented retrospective information when monkeys performed the ODR task, the ratio of delay-period activity representing retrospective information changes depending on the task demands. If the visual information plays a major role to perform the task (e.g., the ODR and the R-ODR tasks), DLPFC neurons tend to maintain retrospective information as delay-period activity.

In summary, delay-period activity observed in DLPFC neurons can be considered a neural correlate of the mechanism for temporarily maintaining information (Funahashi and Kubota 1994; Funahashi and Takeda 2002; Fuster 1997; Goldman-Rakic 1998). A wide range of information can be temporarily maintained as delay-period activity in DLPFC neurons, including visuospatial information, motor information, nonspatial visual features, quality differences of one stimulus modality, or task rules. What information is maintained as delay-period activity can be changed flexibly depending on the task demand. Regional differences in information processing within the DLPFC have been proposed; for example, neurons in the middorsolateral prefrontal cortex participate mainly in visuospatial information processing, whereas neurons in the midventrolateral prefrontal cortex participate in nonspatial visual information processing (Goldman-Rakic 1998; O’Scalaidhe et al. 1999; Wilson et al. 1993). In addition, Hoshi and Tanji (2004) showed area-selective neuronal activity in the DLPFC, such that neurons in the ventral sector responded preferentially to the visuospatial properties of the cue, whereas neurons in the dorsal sector were involved primarily in retrieving information from the cue (e.g., the location of the target or which arm to use). Hoshi (2006) summarized functional specialization within the DLPFC. However, both neurons encoding spatial information and neurons encoding nonspatial visual information can be observed within the same region in the DLPFC (Carlson et al. 1997; Quintana and Fuster 1999; Rainer et al. 1999; Sakagami and Tsutsui 1999), suggesting that interactions are present among neurons encoding spatial information and neurons encoding nonspatial visual information. In addition, many DLPFC neurons exhibit similar task-related activity in both spatial and nonspatial visual working memory tasks (Rao et al. 1997). Therefore, each DLPFC neuron could represent not just one modality of information but could represent a range of modalities of information depending on the task condition and the task demand.
7. Information Processing in the DLPFC

Recent neurophysiological studies show that information represented by a population of prefrontal activities changes along the progress of a trial. For example, Quintana and Fuster (1999) observed a group of neurons attuned to the cue color and another group of neurons attuned to response directions in the DLPFC while monkeys performed working memory tasks using color cues. They found that the discharge of neurons attuned to the cue color tended to diminish along the progress of the delay period, whereas the discharge of neurons attuned to response directions tended to be accelerated along the progress of the delay period. These results indicate that the temporal modulation of the firing patterns along the trial reflects the alteration of the strength of represented information by a neuron. Therefore, a gradual change of the firing strength observed in a neuron or a population of neurons and a gradual change of an activated neural population with the progress of the trial can produce a gradual change of information represented by a population of neurons. While monkeys performed a delayed paired-associate task and a delayed matching-to-sample task, Rainer et al. (1999) observed that neurons whose activity encoded information regarding sample objects gradually decreased their magnitude of activity during the delay period, whereas neurons whose activity encoded information regarding anticipating target objects gradually increased their magnitude of activity during the delay period. A population analysis of DLPFC activities revealed that a population of DLPFC activities tended to reflect information regarding the sample object during the sample presentation and the early part of the delay period. However, a population of activities tended to reflect information regarding the anticipated target object toward the end of the delay period. These results indicate that the information represented by a population of DLPFC activities changes with the progress of the trial. These results also indicate that the temporal change of the represented information can be seen more clearly by looking at the behavior of a population of neural activities than by looking at the behavior of each neuron. Thus, the temporal change of the information represented by a population of neurons with the progress of the trial might correspond to the neural correlate of information processing occurred in the DLPFC.

8. Information Processing in the DLPFC Revealed by a Population Vector Analysis

A large number of DLPFC neurons exhibited a particular type of task-related activity with similar temporal pattern, with similar magnitude of activity, and with similar directional preference, while the subjects performed the ODR task (Funahashi et al. 1989, 1990, 1991). This observation suggests that a particular piece of information (e.g., a spatial position of a visual cue) is encoded by a population of DLPFC neurons having a particular type of task-related activity
(e.g., cue-period activity). In addition, as we have already seen, the temporal change of the represented information can be seen more clearly by looking at the behavior of a population of neural activities than by looking at the behavior of each neuron. Therefore, to understand how information is processed in the DLPFC while the monkey performs working memory tasks, we need to consider not only the information represented by each single-neuron activity but also the information represented by a population of neural activities and its temporal change with a progress of a trial.

Takeda and Funahashi (2002) had examined prefrontal single-neuron activity using the ODR and R-ODR tasks and showed that the majority of delay-period activity represented the location of the visual cue, whereas the minority of delay-period activity represented the direction of the saccade. However, they also showed that the majority of response-period activity represented the direction of the saccade. These results indicate that the transformation from visual information to saccade information is executed in the DLPFC. Most task-related activities representing either sensory or motor information exhibited directional preferences, suggesting that temporal changes of directional selectivity obtained by a population of DLPFC activities during ODR and R-ODR performances could visualize a neural process related to the transformation from visual information to saccade information operated during task performances.

Takeda and Funahashi (2004) used population vectors to show the information represented by a population of DLPFC activities and illustrate its temporal change along a trial of the ODR and R-ODR tasks. The analysis using population vectors was first introduced to evaluate information processing occurring in the primary motor cortex by Georgopoulous et al. (Georgopoulos et al. 1986, 1988; Kettner et al. 1988; Schwartz et al. 1988). To calculate a population vector under one particular trial condition, first a “cell vector” should be calculated from neural activity for every neuron. The length of the cell vector represents the neuron’s discharge rate at that trial condition and the direction of the cell vector corresponds to the neuron’s preferred direction. Each neuron’s preferred direction is usually determined from the tuning curve obtained by fitting mean discharge rates across all examined directions on the cosine function or Gaussian function. Then, a population vector is calculated by a weighted sum of cell vectors for all neurons calculated under the particular trial condition. Takeda and Funahashi (2004) calculated multiple population vectors using all DLPFC neurons that exhibited directional preferences to illustrate the temporal change of preferred directions of a population of DLPFC activities with a progress of a trial in both ODR and R-ODR tasks. They set a 250-ms time-window along the trial, calculated the population vector using a population of neural activities that occurred during this time-window, then moved this time-window in 50-ms time-steps from the cue onset to the end of the response period.

Figure 3A shows population vectors calculated by a population of DLPFC activities at the 180° trial of the ODR task. Directions of population vectors were mostly directed toward the 180° direction, indicating that the same directional preference was maintained by a population of DLPFC activities during the delay
Fig. 3. A Temporal change of the directions of population vectors along the 180° trial of the ODR task. Most of the population vectors were directed toward the 180° direction. B Temporal change of the directions of population vectors along the 180° trial of the R-ODR task. The direction of the population vector gradually rotated from the 180° direction to the 90° direction during the delay period. C The difference between the vector direction and the cue direction along the ODR trial. The population vector was directed toward the cue direction during the delay period. D The difference between the vector direction and the cue direction along the R-ODR trial. The direction of the population vector gradually rotated from the cue direction to the saccade direction during the delay period.
period of the ODR task. To confirm this observation, the temporal change of mean differences between the direction of the population vector and the direction of the visual cue was calculated across all cue conditions (Fig. 3C). This result showed that the mean differences were close to 0° across all periods. In the ODR task, the monkey was required to maintain information regarding either the direction of the visual cue or the direction of the saccade during the delay period. The direction of the visual cue and the direction of the saccade were the same in this task. Therefore, this result indicates that the directional information is maintained along the delay period of each cue condition by a population of DLPFC neurons.

Figure 3B shows population vectors calculated by a population of DLPFC activities at the 180° trial of the R-ODR task, in which the visual cue was presented at the 180° direction and the direction of the correct saccade was the 90° direction. Directions of population vectors were directed toward the 180° direction during the cue period and at the beginning of the delay period. However, directions of population vectors began to rotate in the middle of the delay period, continued to rotate slowly from the 180° direction to the 90° direction during the late half of the delay period, and finally directed toward the 90° direction at the response period. The mean differences between the directions of the population vectors and the directions of the visual cue gradually changed from close to 0° to almost 90° during the delay period (Fig. 3D). The vector began to rotate at approximately 2 s after the delay onset (1 s before the Go signal) and rotation was maintained at a constant speed (about 90°/s) until the vector was in a direction similar to that of the saccade target. This result indicates that the information represented by a population of DLPFC activities changes from visual information to motor information during the delay period while the monkey performed the ODR tasks. Fuster (1997) has proposed the mediation of cross-temporal contingency as an important function of the DLPFC. He considered the delay period of the task as the period for the cross-temporal bridging of sensory-motor information, which is a dynamic process of internal transfer as well as a process of cross-temporal matching. The present result that population vectors rotate gradually from a sensory information domain to a motor information domain during the delay period supports this notion that the DLPFC plays a significant role for mediating the cross-temporal contingency. This result also suggests that delay-period activity observed in DLPFC neurons plays a role for a dynamic process of internal information transfer in working memory performances.

9. Interactions Among DLPFC Neurons as a Possible Mechanism of Information Processing

The population-vector analysis using directionally selective task-related activities showed that the information represented by a population of DLPFC activities changes with the progress of a trial, such that the information represented by a population of DLPFC activities changed from visual information to motor infor-
mation during the delay period while the monkey performed the ODR tasks. What neural mechanisms in the DLPFC enable such information processing? A possible neural mechanism would be functional interactions among DLPFC neurons, each of which exhibits different kinds of task-related activity or exhibits similar activity but representing different information.

One such example observed in the DLPFC is the interaction between neurons exhibiting delay-period activity and neurons exhibiting postsaccadic activity. Many DLPFC neurons exhibit saccade-related activity while the monkey performed visually guided or memory guided saccade tasks (Boch and Goldberg 1989; Funahashi et al. 1991; Takeda and Funahashi 2002). However, most saccade-related activity is postsaccadic (Funahashi et al. 1991; Takeda and Funahashi 2002) and starts several ten of milliseconds after the initiation of the saccade. Therefore, postsaccadic activity is not directly related to the execution and the control of saccadic eye movements. Rather, because a great majority of postsaccadic activities showed directional selectivity (Funahashi et al. 1991; Takeda and Funahashi 2002), this activity is considered to be an activity fed back from the oculomotor centers (Funahashi et al. 1991). On the other hand, the delay-period activity usually ended just after a saccade was performed. Histograms constructed by a population of delay-period activities and a population of postsaccadic activities both aligned at the initiation of the saccadic eye movement reveal that the termination of delay-period activity coincided with the initiation of postsaccadic activity (Goldman-Rakic et al. 1990). Delay-period activity is no longer necessary once the response has occurred, regardless of whether the response is correct. It is important to remove and replace unnecessary information maintained in working memory. Therefore, a coincidence between the termination of delay-period activity and the initiation of postsaccadic activity could reveal that neurons having postsaccadic activity interact with neurons having delay-period activity to terminate delay-period activity.

A cross-correlation analysis of neural activities simultaneously recorded from a pair of neurons has been used to examine whether functional interactions are present between these two neurons (Perkel et al. 1967a,b; Rao et al. 1999) used the cross-correlation analysis to examine functional interactions between pyramidal and nonpyramidal neurons in the DLPFC and found feed-forward excitatory interactions as well as inhibitory interactions between putative pyramidal neurons and adjacent putative nonpyramidal interneurons. Funahashi and Inoue (2000) also applied cross-correlation analysis to simultaneously isolated DLPFC activities while monkeys performed the ODR task. Among 168 pairs of simultaneously isolated single-neuron activities, 18% showed excitatory sharp and symmetrical peaks at time 0 (central peak), indicating that these pairs of neurons tend to fire simultaneously. Therefore, it is expected that functional interactions are present between these pairs of neurons. On the other hand, 23% had excitatory sharp but asymmetrically distributed peaks displaced from the time 0 (displaced peak), indicating that one of the paired neurons sends excitatory outputs to other neuron of the pair. A similar observation was recently reported by Sakurai and Takahashi (2006).
The cross-correlation analysis also revealed the presence of the information flow from neurons having cue-period activity to neurons having presaccadic activity through neurons having delay-period activity (Funahashi and Inoue 2000). In addition, pairs of neurons, both of which exhibited delay-period activity, tended to have either significant excitatory central peaks or displaced peaks, indicating that interactions between neurons having delay-period activity are widely present in the DLPFC. In pairs of neurons whose cross-correlograms had central peaks, the best directions of same task-related activity were almost identical or similar between two neurons. In pairs of neurons whose cross-correlograms had displaced peaks, the best directions of same task-related activity were also similar between two neurons. Interactions between neurons that had different directional preferences were more often observed in pairs of neurons both of which showed delay-period activity. These results suggest that interactions among neurons playing different functional roles (e.g., neurons having different task-related activity or exhibiting different directional preferences) play an important role for processing information. Frequently observed interactions among neurons having delay-period activity suggest that these interactions play an important role for transforming information from sensory aspects to motor aspects (Takeda and Funahashi 2004), for integrating different kinds of information, or for creating a more-complex representation, such as delay-period activity representing a combination of visual cue positions and the temporal order of their presentation (Barone and Joseph 1989; Funahashi et al. 1997).

10. Dynamic Interactions Among DLPFC Neurons

Functional interactions drawn from either central peaks or displaced peaks in cross-correlograms were overall interactions between particular neuron pairs, because these cross-correlograms were calculated using activities recorded throughout the block of 100–150 trials. However, each neuron changed the magnitude of activity depending on the trial conditions (e.g., the position where the visual cue was presented or the direction where the saccade was directed), the temporal context of the trial (e.g., cue period, delay period, or response period), or trial events (e.g., cue presentation or saccade performance). Therefore, it is expected that the strength of the functional interaction can be modulated depending on trial conditions of the task as well as the temporal context of the trial.

To examine whether the strength of the functional interaction is modulated depending on trial conditions of the task and the temporal context of the trial, we calculated cross-correlograms of neuron pairs for each cue condition in the ODR task (Funahashi 2001). Figure 4 shows an example of cross-correlograms calculated from simultaneously isolated pairs of activities for every cue condition. The height of the central peak of the cross-correlogram was modulated depending on the cue conditions, such that a higher central peak was observed in trials in which the visual cue was presented at the 90°, 225°, and 315° positions. The modulation of the peak height of the cross-correlogram was observed in most
13. Prefrontal Cortex in Information Processing

These results indicate that many DLPFC neurons have functional interaction with neighboring neurons and that the strength of functional interaction changes dynamically depending on the context of the task.

An analysis used a joint peristimulus time histogram (j-PSTH) was used to examine the temporal modulation of functional interaction between a pair of simultaneously recorded single-neuron activities (Aertsen et al. 1989; Gerstein and Perkel 1969, 1972). The important features of this method are that the temporal change of functional connectivity between two neurons can be seen in a temporal change of the height of the PST coincidence histogram and that this
change is independent of the task-related modulation of the neurons’ firing rates during the trial (Aertsen et al. 1989). The validity of this method and the importance of the temporal change in functional connectivity between prefrontal neurons have been demonstrated by Vaadia et al. (1995). Thus, the j-PSTH method is a useful method to examine the temporal modulation of functional interaction between DLPFC neurons along the progress of the trial.

Figure 5 shows examples of jPSTH and a PST coincidence histogram calculated from activities of two neurons (neurons y05702a and y06702b) during ODR performances. In this example, both neuron y05702a and neuron y05702b exhibited tonic excitatory delay-period activity. The cross-correlogram calculated from activities of these two neurons had an excitatory symmetrical peak at time 0 (Fig. 5). The PST coincidence histogram shows that the probability of simultaneous firing of two neurons (an index of the strength of functional connectivity) changed along the progress of the trial and kept a high value during the delay period. This result suggests that functional connectivity between two neurons was stronger during the delay period. Although the temporal patterns of PST coincidence histograms varied from neuron pair to neuron pair, stronger functional connectivity was observed only during the cue period or the response period in some neuron pairs, during the entire delay period in some neuron pairs (e.g., Fig. 5), or during the early half of the delay period or during the late half of the delay period in other neuron pairs.

Fig. 5. An example of the temporal modulation of functional connectivity between two simultaneously isolated prefrontal neurons during ODR performances. The color map on the left is a joint-peristimulus time histogram (PSTH), indicating the strength of correlated firing between two neurons. The diagonal histogram on the right is a coincidence histogram, indicating the temporal modulation of the probability of coincidental firing of two neurons. C, D, R, cue, delay, and response period, respectively.
Thus, a large number of DLPFC neuron pairs exhibited significant central peaks or displaced peaks around time 0 in cross-correlograms, suggesting that a large number of DLPFC neurons have excitatory interactions with neighboring neurons. A pair of neurons with significant peaks in cross-correlograms often had different kinds of task-related activity or different spatial selectivity in the same task-related activity, suggesting that information processing occurs in the DLPFC through these interactions between neighboring neurons. In addition, the strength of the functional interaction between DLPFC neurons was dynamically modulated depending on trial conditions and the behavioral context of the trial, indicating that dynamic changes of the functional interaction among DLPFC neurons could be neural mechanisms of information processing.

11. Dynamic Interactions Between the Prefrontal Cortex and Other Cortical Areas

Cortico-cortical and cortico-subcortical connections with the DLPFC have been examined extensively [see reviews by Pandya and Barnes (1987) and Fuster (1997)]. The DLPFC has reciprocal cortico-cortical connections with the posterior parietal cortex, the inferior temporal cortex, the superior temporal polysensory areas, the anterior cingulate, the retrosplenial cortex, and the parahippocampal gyrus. The DLPFC also has strong reciprocal connections with the mediadorsal nucleus of the thalamus. In addition, the DLPFC has connections with the frontal eye field, the presupplementary motor area, the premotor cortex, and the caudate nucleus. Morris et al. (1999) reported that the mid-DLPFC (areas 46 and 9) projects to the retrosplenial area 30 and the posterior presubiculum, and suggested that this system is the anatomical substrate of functional interaction between the DLPFC and the hippocampus. Thus, the DLPFC has anatomical connections with various cortical and subcortical areas. These observations strongly suggest that the DLPFC can receive various kinds of information including sensory, motor, and emotional information as well as information stored in long-term memory and is therefore in a good position to obtain any kind of information necessary to perform any cognitive task. These observations also suggest that the DLPFC can send various information and control signals to cortical and subcortical areas and is therefore again in a good position to supervise information processing operating in other areas.

Recently, top-down modulatory influences by the prefrontal cortex have been demonstrated. Using a visual associative memory task, Hasegawa et al. (1998) showed by a partial split-brain paradigm in monkeys that retrieval of visual images from long-term memory was under the executive control of the prefrontal cortex. In addition, by neurophysiological experiments using monkeys with callosal resection, Tomita et al. (1999) showed that top-down signals from the prefrontal cortex regulated episodic memory retrieval in the temporal cortex. These observations indicate that the prefrontal cortex manages information processing occurred in other cortical areas using top-down signals.
Similarly, Lumer and Rees (1999) showed that, in bistable viewing conditions, covariation of activity was observed in multiple extrastriate areas, the parietal cortex, and the prefrontal cortex by fMRI. Coordinated activation of these areas was not linked to either sensory or motor events, but rather reflected perceptual events or internal changes of perception. Therefore, they concluded that functional interactions between the visual cortex and the prefrontal cortex could contribute to conscious vision. Recently, Windmann et al. (2006) used normal subjects and patients with unilateral prefrontal damages and examined whether the prefrontal cortex influences bistable vision by maintaining the dominant pattern while protecting it against the competing representation or by facilitating perceptual switches between the two competing representations. Their results suggest that the prefrontal cortex initiates perceptual reversals by withdrawing top-down support from the dominant representation without boosting the suppressed view.

On the other hand, enhancement or suppression of the visual response was observed in inferior temporal neurons when two visual stimuli (one preferred and the other nonpreferred) were presented simultaneously in the neuron’s receptive field. Chelazzi et al. (1998) proposed that this phenomenon may be explained by a “biased competition” model of attention, such that objects in the visual field compete for representation within the cortex and that this competition is biased in the pattern of activity by the “top-down” feedback from structures involved in working memory. Although this is just a hypothesis, according to the results of Hasegawa et al. (1998), Tomita et al. (1999), and Lumer and Rees (1999), a top-down signal from the prefrontal cortex could bias the processes occurred in the visual system. Recently, Johnston and Everling (2006) showed that the DLPFC sends task-specific signals directly to the superior colliculus. This result supports the notion that the DLPFC sends the top-down signal to regulate the activity of other brain areas in accordance with task requirements.

As was suggested by Miller and Cohen (2001), these results indicate the presence of top-down modulation by the prefrontal cortex and its importance in performing cognitive tasks. These results also suggest that top-down modulation by the prefrontal cortex affects a wide variety of cognitive activities including sensory perception, attention, and episodic memory encoding and retrieval through dynamic functional interactions of cortico-cortical and cortico-subcortical pathways.

12. Conclusions

Working memory includes temporary active maintenance of information as well as processing of maintained information. Neurophysiological examination of the neural mechanism of working memory provides important insights to understand how a variety of information is represented in the nervous system and how the information is processed in the nervous system. An examination of task-related
activity in the DLPFC while monkeys performed various working memory tasks reveals that delay-period activity is a neural correlate of the mechanism for short-term active maintenance of information. Analyses of characteristics of delay-period activity reveal that a variety of information including spatial and nonspatial visual features, forthcoming behavioral responses, the quality as well as the quantity of expected reward, the difference of the tasks, or the rule of the task are encoded as positional or directional preferences of this activity, or visual feature specificity of this activity, or by difference in the magnitude of this activity, depending on the reward or task conditions.

However, little is known about neural mechanism for processing information. To understand this mechanism, we need to know what information each task-related activity represents and how information represented in the nervous system changes with the progress of the trial. Using two ODR tasks (ODR and R-ODR tasks), we found that a great majority of delay-period activity represents retrospective information (e.g., the location of the visual cue) whereas a minority of delay-period activity represents prospective information (e.g., the direction of the forthcoming movement). In addition, using population vector analysis, information represented by a population of prefrontal activity can be visualized as the direction of the population vector, and the temporal change of information represented by a population of prefrontal activity can be visualized as the temporal change of the direction as well as the length of the population vector during the delay period. The mechanism participating in the gradual change of information represented by a population of activities remains unresolved. However, functional interactions among neighboring neurons representing different information and dynamic modulation of these interactions depending on the context of the trial could be a mechanism of this process (Constantinidis et al. 2001; Funahashi 2001; Funahashi and Inoue 2000). Further analyses of dynamic and flexible interactions among neighboring neurons and their temporal modulation are needed to understand neural mechanisms of processing information.

Acknowledgment. This work was supported by Grant-in-Aids for Scientific Research (14380367, 17300103) and Grant-in-Aid for Scientific Research on Priority Areas: System study on higher-order brain function (18020016) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. This work was also supported by the 21st Century COE Program (D-10 at Kyoto University) by MEXT of Japan.

References


Funahashi S, Bruce CJ, Goldman-Rakic PS (1990) Visuospa
tial coding in primate pre-
Funahashi S, Bruce CJ, Goldman-Rakic PS (1991) Neuronal activity related to saccadic
eye movements in the monkey’s dorsolateral prefrontal cortex. J Neurophysiol 65:
1464–1483.
Funahashi S, Bruce CJ, Goldman-Rakic PS (1993a) Dorsolateral prefrontal lesions
and oculomotor delayed-response performance: evidence for mnemonic “scotomas.” J
Neurosci 13:1479–1497
Funahashi S, Chafee MV, Goldman-Rakic PS (1993b) Prefrontal neuronal activity in
rhesus monkeys performing a delayed anti-saccade task. Nature (Lond) 365:753–756
Funahashi S, Inoue M, Kubota K (1997) Delay-period activity in the primate prefrontal
cortex encoding multiple spatial positions and their order of presentation. Behav Brain
Res 84:203–223
Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance:
Fuster JM (1997) The prefrontal cortex: anatomy, physiology, and neuropsychology of
the frontal lobe, 3rd edn. Lippincott-Raven, Philadelphia
goals by separate neural populations in prefrontal cortex. J Neurosci 26:7305–7316
Gerstein GL, Perkel DH (1969) Simultaneously recorded trains of action potentials:
Gerstein GL, Perkel DH (1972) Mutual temporal relationships among neuronal spike
movements to visual targets in three-dimensional space. II. Coding of the direction of
movement by a neuronal population. J Neurosci 8:2928–2937
Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of
movement direction. Science 233:1416–1419
Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behav-
Handbook of physiology, section 1: the nervous system, vol V. American Physiological
Society, Bethesda, MD, pp 373–417
Goldman-Rakic PS (1998) The prefrontal landscape: implications of functional architec-
ture for understanding human mentation and the central executive. In: Roberts AC,
Robbins TW, Weiskrantz L (eds) The prefrontal cortex: executive and cognitive func-
memory and disordered cognition in schizophrenia. Biol Psychiatry 46:650–661
Goldman-Rakic PS, Funahashi S, Bruce CJ (1990) Neocortical memory circuits. Cold
prefrontal cortices: cognitive interaction to retrieve long-term memory. Science 281:
814–818
nigra pars reticulata. III. Memory-contingent visual and saccade responses. J Neuro-
physiol 49:1268–1284
Hoshi E (2006) Functional specialization within the dorsolateral prefrontal cortex: a
review of anatomical and physiological studies of non-human primates. Neurosci Res
54:73–84


Jacobsen CF (1936) Studies of cerebral function in primates. I. The functions of the frontal association areas in monkeys. Comp Psychol Monogr 13:1–60


Large-Scale Network Dynamics in Neurocognitive Function

Anthony Randal McIntosh

1. Overview

The study of human mental function is, without a doubt, at the edge of a new frontier, thanks largely to neuroimaging [e.g., functional magnetic resonance imaging (fMRI), magneto- and electroencephalography]. Access to human neurobiology can potentially provide the critical link between psychological theories of various cognitive functions and the concomitant physiology. Indeed, many psychological theories find their ultimate verification where the constructs appear to have a direct neural instantiation, yet few believe in an isomorphic relation between brain and mind. As such, there is a gap between how psychological constructs are represented and how the operations of the nervous system fit with such representations. The fundamental challenge is captured by a quote from William James (1890: p 28): “A science of the relations of the mind and brain must show how the elementary ingredients of the former correspond to the elementary functions of the latter.” The modern neuroscientist, and particularly the cognitive neuroscientist endowed with the new insight into human neurophysiology, must determine what features of the human brain are central in translating the biological representations to mental phenomena.

The purpose of this chapter is to present a framework for the exploration of critical linking features between brain and mind. Cognitive functions are the outcome of the complex dynamical interactions within distributed brain systems. These features arise from the anatomy and physiology of the brain. The anatomy enables a system that has a maximal capacity for both information segregation and integration. The physiological property of response plasticity, where optimal neural responses can change depending upon stimulus significance or internal network states, modifies the information as it is passed to different levels of the system. The distribution of information in the brain, here meant to indicate that information which is carried by signals among neurons (e.g., firing rate, temporal synchrony), allows several parts of the brain to contribute to a broad range of

Rotman Research Institute-Baycrest, University of Toronto, Toronto, Ontario M6A 2E1, Canada
mental function. The combination of anatomical architecture and physiology forms the basis for a *neural context*, wherein the potential contribution of one neural element to an operation is sculpted by its interactions with other elements. Such contextual influences emphasize the dynamic nature of neural function, wherein the relevance of that element changes as the mental function unfolds. Examples of contextual effects are presented and serve to develop a second linking feature, the *behavioral catalyst*, where certain neural elements are critical for a particular mental function when they enhance the transition between mental states.

2. Anatomical Considerations

Neurons are connected to one another both locally and at a distance. Most other systems in the body show some capacity for cell-to-cell communication, but the nervous system appears to be specialized for rapid transfer of signals. This means that a single change to the system is conveyed to several parts of the brain simultaneously and that some of this will feed back onto the initial site. There are obvious extremes to just how “connected” a system can be, and the nervous system occupies some intermediate position. Local cell networks are highly interconnected, but not completely so, and this means that adjacent cells can have common and unique connections. As a more explicit illustration, consider the three networks in Fig. 1. The leftmost network has nodes that are completely connected and can be considered a redundant network. The rightmost network has disconnected elements, resulting in subnetworks that will operate independently. The central network most closely resembles the nervous system, where certain nodes are densely connected while others are more sparse. This configuration has been referred to as *degenerate* by Tononi and colleagues (Tononi

![Diagram of wiring diagrams for three networks](image)

**Fig. 1.** Diagram of wiring diagrams for three networks. The *Independent* network (*far left*) has two sets of independent nodes that can exchange information within subnetworks but cannot between them. The *Redundant* network (*far right*) has each node connected to every other node, enabling information exchange among all nodes, but with little capacity for the information to change (i.e., all nodes behave the same). The *Degenerate or Semi-connected* system (*center*) shows a balance between extremes and thus has the potential for optimal segregation and integration of information.
et al. 1994, 1999), and I have used a term borrowed from graph theory, *semiconnected*, to designate this particular property of local cell networks (McIntosh 2000).

Although the local networks in different cortical areas show cytoarchitectonic variation, the cellular components and internal connectivity of cortical circuits are generally similar throughout the cortex. What distinguishes the function of any local cortical network is its topological uniqueness, that is, its particular pattern of interconnectivity with other networks. The unique set of local networks with which a given local cortical network is directly connected has been called its “connection set” (Bressler 2003) or “connectional fingerprint” (Passingham et al. 2002). The implication, stated simply, is that regions can only process information that they receive. Cells in the medulla tend not to respond to visual stimuli because they are not connected to visual structures. Anatomy determines whether a given ensemble is capable of contributing to a process.

The connections between local neural ensembles are sparser than the intraensemble connectivity. Estimates of the connections in the primate cortical visual system suggest that somewhere between 30% and 40% of all possible connections between cortical areas exist (Felleman and Van Essen 1991). Large-scale simulation studies show that this sparseness is a computation advantage for the nervous system in that it allows for a high degree of flexibility in responses at the system level even when the responses of individual units are fixed (Tononi et al. 1992). Additional analyses of the anatomical connections of large-scale networks in the mammalian cerebral cortex have demonstrated a number of distinct topological features that lead to systems with maximal capacity for both information segregation and integration (Sporns and Kotter 2004; Sporns and Zwi 2004).

Neural network theories of brain function have, to varying degrees, emphasized these two basic organizational principles: segregation and integration. At each level in a functional system, there is a segregation of information processing such that units within that level will tend to respond more to a particular aspect of a stimulus (e.g., movement, color). At the same time the information is *simultaneously* exchanged with other units that are anatomically connected, allowing first for units to affect one another and second for the information processed within separate units to be integrated.

### 3. Functional Considerations

Neural plasticity is an established phenomenon. With maturation and following damage, there are massive changes in the functional responses of neural elements as the system adapts to the external and internal environment. These changes can take place over a few days or years. There is a type of plasticity that is more short-lived. Cells can show a rapid shift in response to afferent stimulation that is dependent on the behavioral situation in which they fire. This *transient response plasticity* occurs over a much shorter timescale compared to recovery from damage. Physiological investigation has consistently shown transient plasticity in
the earliest parts of the nervous system, from single cells in isolate spinal cord preparations to primary sensory and motor structures. The changes can occur within a few stimulus presentations (Edeline et al. 1993; Dorris et al. 2000) and are likely a ubiquitous property of the central nervous system (Wolpaw 1997).

The generality of this feature in the brain is often neglected in considerations of the mind–brain link. Transient response plasticity is a rudimentary feature of an adaptable system. The degree of this plasticity is constrained by anatomy where, as noted above, a neural element has a restricted range of information to which it can respond. A consequence of the anatomical structure is that adjacent neurons have similar response properties (e.g., orientation columns in primary visual cortex), whereas neurons slightly removed may possess overlapping, but not identical, response characteristics. These broad tuning curves are characteristic for most sensory system cells and those in motor cortices. The broad tuning curves result from anatomy, where cells share some similar and some unique connections. Interestingly, anatomy also ensures that response plasticity also has a graded distribution, which has important implications for the representational aspects of the brain. Rather than having each neuron code sharply for a single feature, the distributed response takes advantage of a division of labor across many neurons, enabling representations to come from the aggregate of neuronal ensembles.

Electrophysiological studies of in motor and sensory cortices have provided some examples of aggregate operations achieved through population coding (Averbeck et al. 2006; Georgopoulos et al. 1986; Pasupathy and Connor 2002; Young and Yamane 1992). Population coding is likely used for higher-order cognitive functions, but on a larger scale than for sensory or motor functions. When neural populations interact with one another, these rudimentary functions combine to form an aggregate that represents cognitive processes. Cognitive operations are not localized in an area or network of regions, but rather emerge from dynamic network interactions that depend on the processing demands for a particular operation.

4. Neural Context

As noted in the opening section of this chapter, the anatomical and functional properties of the brain combine to enable a contextual dependency in how specific neural elements impact on a mental function. This neural context is the local processing environment of a given neural element that is created by modulatory influences from other neural elements (Bressler and McIntosh, in press). Neural context allows the response properties of one element in a network to be profoundly affected by the status of other neural elements in that network. As a result of neural context, the relevance of a given neural element for cognitive function typically depends on the status of other interacting elements (Bressler 2003; McIntosh 1999). By this definition, the processing performed by a given brain area may be modulated by a potentially large number of other areas with which it is connected. Because brain areas are most often bidirectionally
connected, the neural context of each connected area emerges spontaneously from its interactions. Furthermore, to the extent that basic sensory and cognitive operations share similar brain constituents, they experience similar neural contextual influences.

Neural context refers only to the context that arises within the brain as a result of interactions between neural elements. Another contextual influence that interacts with neural context comes from the environment: situational context (Bressler and McIntosh, in press). Unlike neural context, situational context represents a host of interrelated environmental factors, including aspects of the sensory scenes and response demands of both the external and internal milieus. A red light presented to a person in isolation usually means nothing, but a red light presented to that person while driving an automobile elicits a situationally specific response. Situational context is most often what researchers have in mind when they examine “contextual effects” on the brain (Bar 2004; Beck and Kastner 2005; Chun 2000; Garraux et al. 2005; Hepp-Reymond et al. 1999).

In most normal circumstances, neural context is shaped by situational context. The environments in which animals and humans must survive have a high degree of structural complexity, an important consequence of which is a fundamental uncertainty in the organism’s perceptuo-motor interactions with those environments. Complete information is never available to allow total certainty about the state of the environment and the optimal course of action in it. The limited information that the organism has about its environmental situation usually renders ambiguous its perceptual interpretation of environmental entities and the appropriate actions to be directed toward them. The ability to utilize and manipulate information about the organism’s situational context can dramatically reduce uncertainty, thereby enhancing the organism’s interactions with the environment and lending survival advantage to its species.

Because the complexity of the environment’s structure spans multiple structural and temporal scales, situational context must affect all types of cognitive function, including sensation, perception, emotion, memory, planning, decision making, and action generation. It is reasonable to infer therefore that neural context should also be of primary importance in the implementation of those functions by the brain. In other words, just as situational context can have effects at multiple scales and across multiple behaviors, so too is neural context expected across all spatial and temporal scales in the brain and across all behaviors.

5. Empirical Examples

As a principle of brain function, neural context can be most easily demonstrated in relatively simple nervous systems, such as those of invertebrates. While these systems admittedly do not have the broad behavioral repertoire of primates, if contextual effects are indeed central to neural network operation, they should be present in simpler organisms. Neural context has been demonstrated in the Aplysia abdominal ganglion that the same neurons fire during performance of quite different behaviors (Wu et al. 1994). What appears to differentiate these
behaviors is not the activity of a particular neuron, or group of neurons, but rather the overall activity patterns of an entire network. Such observations have been made in other invertebrate species across highly dissimilar behaviors (Popescu and Frost 2002), suggesting that the observed behavioral variation resides in the large-scale dynamics of entire networks rather than dedicated circuits (Kristan and Shaw 1997).

It has been hypothesized (Bressler 2004) that cortical context emerges from multiple recurrent interactions among cortical areas. An important prediction from this hypothesis is that the representation of categorical information in the cortex should be reflected by patterns of activity distributed across large cortical expanses rather than by the activity in a single specific area. Category specificity has rapidly become a major focus in human neuroimaging research, exemplified by studies demonstrating face category-specific responses in the fusiform gyrus (Kanwisher et al. 1997). However, a drawback of many such studies is that they employ very strict univariate statistical criteria that conceal all but the largest amplitudes in the activity patterns. Nonetheless, studies that have characterized the distributed response to faces have reported that greater category specificity is revealed by the entire activity pattern in occipital and temporal cortices than by any specific area (Haxby et al. 2001; Ishai et al. 1999). Importantly, these studies have determined that the specificity of the distributed response is not dramatically altered if the regions typically associated with the category of interest are excluded. Further analysis of the distributed patterns emphasize that the most effective category specificity is captured by the specific pattern of voxel activity across the entire ventral temporoccipital region than by any single region (Hanson et al. 2004; O’Toole et al. 2005).

Although these data seem to contradict the evidence that one can identify category-specific effects in the brain (Spiridon and Kanwisher 2002), this underscores an important issue in the study of representations in the brain. The brain operates at many spatial scales, synaptic, cellular, ensemble, population, and what can be considered mesoscopic, involving multiple populations (Freeman 2000; McIntosh, in press). Within each of these there can be multiple time scales (Breakspear and Stam 2005; Breakspear et al. 2006). Thus, signals that may appear discrete and carrying a very narrow piece of information (e.g., a category) become part of a larger collection of interacting units that together convey far more precise information, similar to what was presented early as population coding. It is, therefore, imperative to consider very carefully what signals exist across several spatiotemporal scales in the determination of the optimal mapping of perceptual, behavioral, or cognitive representations to brain dynamics.

6. Large-Scale Interactions and Neural Context

As contextual effects depend on the interactions between neural elements, the optimal manner to assess neural context is by estimating functional or effective connectivity. Functional connectivity emphasizes pairwise interactions, often in
terms of correlations or covariances. Effective connectivity incorporates additional information, such as anatomical connections, and considers an interaction of several neural elements simultaneously to explicitly quantify the effect one element has on another. Both functional or effective connectivity were introduced in the context of electrophysiological recordings from multiple cells (Aertsen et al. 1987) and have been used in reference to neuroimaging data (Friston 1994; Friston et al. 1993; Horwitz 2003).

One well-established functional distinction in the brain is between object and spatial visual pathways. The foundation for this dual organization can be traced at least as far back as Kleist in the 1930s (Kleist 1935). One of its strongest expressions to date is in the dorsal and ventral cortical processing streams, described by Ungerleider and Mishkin (Ungerleider and Mishkin 1982), which correspond to spatial and object processing pathways, respectively.

A similar duality was identified in humans with the aid of positron emission tomography (PET) (Haxby et al. 1991). In this experiment, a match-to-sample task for faces was used to explore object vision. For spatial vision, a match-to-sample task for the location of a dot within a square was used. The results from the right hemisphere analysis are presented in Fig. 2 (left hemisphere interactions did not differ between tasks). Effects along the ventral pathway from cortical area 19v extending into the frontal lobe were stronger in the face-matching model whereas interactions along the dorsal pathway from area 19d to the frontal lobe were relatively stronger in the location-matching model. Among posterior areas, the differences in influences were mainly in magnitude. Occipitotemporal interactions between area 19v and area 37 were stronger in the face-matching model while the impact of area 17/18 to 19d and the occipitoparietal influences from area 19d to area 7 was stronger in the location-matching model.

The model allowed for interactions between the dorsal and ventral pathways with connections from area 37 to area 7 and from area 7 to area 21. In the right hemisphere, the interactions among these areas showed task-dependent differences in magnitude and sign. The temporoparietal influence of area 37 on area 7 was relatively stronger in the location-matching model. The parietotemporal influence of area 7 on area 21 showed a difference in sign between the two functional models. These results show that although the strongest positive interactions in each model may have been preferentially located within one or the other pathway, the pathways did not function independently, but exerted contextual modulatory influences on one another.

Another important result of this study is that, although the prefrontal cortex (PFC) did not show a difference in mean activity between tasks, processes involving the PFC shifted depending on the task. The influence of the dorsal and ventral pathways on the frontal cortex was similar in magnitude for the two tasks, but the origin of the positive and negative influences differed, implying that the qualitative nature of influence on the frontal lobe was different (positive influences in the location-matching model were from areas 7 and 19d, and that in the face-matching model was from area 21). In terms of neural context, these results demonstrate that it is not an area’s activity per se that is the key to understanding
Object vision

Spatial vision

Fig. 2. Effective connectivity between cortical areas in the right hemisphere for object and spatial vision operations. The numbers on the cortical surface refer to Brodmann areas (d, dorsal; v, ventral). The arrows represent the anatomical connections between areas; the magnitude of the direct effect from one area to another is proportional to the arrow width for each path. (Adapted from McIntosh et al. 1994)
its contribution to a task, but rather its pattern of interaction with other areas in large-scale networks.

Network interactions that underlie cognitive operations are observable as differences in the effective connections between elements of the network. As illustrated above, if visual attention is directed to the features of an object, effective connections among ventral posterior cortical areas tend to be stronger, whereas visual attention directed to the spatial location of objects leads to stronger interactions among dorsal posterior areas. Another way that cognitive operations may be observed is through the modulation of effective connections that occurs when one area provides an enabling condition to foster communications between other areas. Such enabling effects may represent a primary mechanism whereby situational context is translated into neural context.

7. Attention

In a study of anterior cingulate cortex (ACC) effective connectivity, Stephan et al. (2003) examined whether hemispheric functional asymmetry was determined by a word stimulus (short words, with one letter colored red) itself or by the task, that is, the situational context. In one instance, subjects judged whether the word contained the letter “A,” ignoring the red letter, and in another instance, they made a visuospatial judgment indicating whether the red letter was right or left of center. A direct comparison of the activity (measured with functional magnetic resonance imaging, fMRI) revealed strong hemispheric differences. The letter task produced higher activity in the left hemisphere whereas the visuospatial task produced higher activity in the right hemisphere. The ACC was similarly active in both tasks relative to baseline but showed distinctly different patterns of effective connectivity between tasks. Specifically, during the letter task, the ACC was coupled to the left prefrontal cortex (PFC); during the visuospatial task, the ACC was linked with the right posterior parietal cortex (PPC). These data are a compelling example of how situational context (in this case, task demands) can modulate the neural context within which a cortical area (i.e., the anterior cingulate) operates.

8. Working Memory

Although working memory is often considered a unique psychological construct, another perspective emphasizes its close relationship to attention (Bressler and Tognoli 2006; Deco and Rolls 2005; McElree 2001). Both working memory and sustained attention involve activity in overlapping regions of PPC, PFC, and ACC. In an fMRI study of the relationship between attention and working memory, Lenartowicz and McIntosh (2005) used two variants of a two-back working memory task: a standard version with strong attentional demands, and a cued version that more strongly promoted memory retrieval. Activation of
ACC was found in both tasks, although it was more sustained in the standard condition. However, the regions functionally connected to the ACC, and the relationship of the connectivity patterns to memory performance, differed completely between tasks. In the standard task, the observed pattern was related to a speed–accuracy tradeoff, with strong functional connection of ACC to PFC and PPC. In the cued task, the connectivity pattern was related only to better accuracy, and involved functional connections with middle and inferior PFC and inferior temporal cortex. By virtue of these different patterns of functional connectivity, the contribution of ACC to attention- and memory-driven performance was similarly changed. In other words, each task invoked a different neural context within which the ACC interacted, resulting in two very different behavioral profiles. Conversely, the difference in neural context reflected the difference in the functional role that ACC fulfilled.

9. Awareness and Neural Dynamics

In two studies of sensory associative learning, we obtained evidence that both learning cross-modal association and the awareness of such associations related to interactions among distributed cortical regions. In the first study, two tones were used having differential relations to the visual stimuli. One tone was a strong predictor of the presentation of a visual stimulus (Tone+) and the other tone a weak predictor (Tone−) (McIntosh et al. 1999). In this positron emission tomography (PET) rCBF (regional cerebral blood flow) study, brain activity was measured in response to isolated presentations of the Tone+ and Tone− as the subjects learned. Much to our surprise, subjects in our sample divided perfectly in half into those who were aware of the stimulus associations and those who were not. The index of awareness came from debriefing questionnaires. Furthermore, only “Aware” subjects learned the differentiation between the tones, while “Unaware” subjects showed no behavioral evidence of learning.

In examining the underlying brain activity that supported learning and awareness, we observed that the strongest group difference in brain activity elicited by the tones was in the left prefrontal cortex (LPFC) near Brodmann area 9. In aware subjects, LPFC activity showed progressively greater activity to Tone− than Tone+. Ventral and medial occipital cortices and right thalamus showed progressively greater activity to Tone+ than to Tone−. In Unaware subjects, no consistent changes were seen in LPFC or in any of the other regions. At first, these results seem to confirm the prominent role of PFC in monitoring functions (Burgess and Shallice 1996; Stuss and Benson 1987), and especially its putative role in awareness (Knight et al. 1995). However, PFC activation has also been found in tasks where there was no overt awareness, such as in the previous sensory learning task, and in implicit novelty assessment (Berns et al. 1997). It was thus possible that interactions of PFC with other brain regions, present in Aware but not in Unaware subjects, would better describe the neural system underlying awareness in this task.
When the interactions of LPFC were assessed between the two groups, we observed a remarkable difference in the strength and pattern of functional connections among several brain areas, including the right PFC, bilateral superior temporal cortices (auditory association), occipital cortex, and medial cerebellum. These areas were much more strongly correlated in Aware, than Unaware, subjects. To explore some of the network interactions within an anatomical reference, effective connectivity was assessed with structural equation modeling (McIntosh and Gonzalez-Lima 1994) were constructed for a subset of regions identified in the functional connectivity analysis.

There were significant changes in the effective connections for Aware subjects, including robust interactions involving LPFC (Fig. 3). During the last Tone+ scan, feedback to occipital cortex was positive from temporal and prefrontal cortex, which may reflect implicit and explicit expectancy of the upcoming visual discrimination (McIntosh et al. 1998). In the Tone– scan, this feedback switched

**Fig. 3.** Functional networks from late phases of training in a differential sensory conditioning task. Networks from two groups are shown. Aware subjects showed strong difference in effective connections involving left prefrontal area 9 and other regions that distinguished between the two tones. Conversely, the network for the unaware (Not Aware) subjects did not differ between tones and showed no strong left prefrontal involvement. *Line thickness* may be interpreted according to the legend at the bottom of the figure.
to negative, which could reflect the knowledge that there would be no visual event following presentation of the Tone+. The functional network for Unaware differed from Aware subjects, but there were no significant changes in effective connections across the experiment for the Unaware group. There were non-zero interactions in the functional network, but the involvement of LPFC was weak, which confirms that the LPFC was not interacting systematically across subjects in the Unaware group.

10. Reversal Learning and the Medial Temporal Lobe

A more salient demonstration of neural context came from the follow-up study, looking again at differential sensory associative learning, but where the contingencies between the tones and visual stimuli were reversed partway through the study (McIntosh et al. 2003). As before, subjects were classified as Aware or Unaware based on whether they noted that one of two tones predicted a visual event. Only Aware subjects acquired and reversed a differential response to the tones, but both groups showed learned facilitation.

When we related brain activity (index by blood flow measured with PET) to behavior in each group, we observed that medial temporal lobe (MTL) activity related to facilitation in both groups. This finding was curious given the suggestion the MTL is critical for learning with awareness but not when learning proceeds without awareness (Clark and Squire 1998, 2000). Given the principle of neural context, it was possible that this common regional involvement in the two groups was an expression of contextual dependency. We then examined the functional connectivity of the MTL and observed completely different interactions of the MTL between groups (Fig. 4). In the Aware group, dominant MTL interactions were observed for prefrontal, occipital, and temporal cortices, while in Unaware subjects, MTL interactions were more spatially restricted to infero-temporal, thalamus, and basal ganglia. The functional connectivity of each region from the spatial pattern for each group with the MTL, with other regions in the pattern, as well with behavior, are presented in Fig. 4 as pseudo-colored images to illustrate two important points. First, the regions unique to each group showed strong overall functional connectivity, and second, the regions related strongly with behavior only for the group from which the regional pattern was obtained. The MTL was thus part of a large-scale interactive system related to learning in both groups, but only in one case was learning accompanied by awareness resulting from different constituents of the interacting systems. In other words, the differences in neural context serve as a possible explanation for the involvement of the MTL in both groups.

As stated earlier, there are many scales at which brain function can be captured and related to behavior. The MTL shows a distributed pattern of interactions that related to learning with or without awareness, but this focuses only on how one region (albeit rather large) interacts at a mesoscopic scale. Neuroimaging data, with its broad spatial coverage, can give a comprehensive depiction of the
dependency between interregional interactions and behavior. In mapping brain activity to behavior, we were able to identify two reliable spatial patterns in the Aware group that related (1) to general associative learning (Facilitation) and (2) to the differential association of the two tones (Discrimination). For the Unaware group, a single spatial pattern different from the Aware group, related to general associative learning, was identified (Facilitation), with a second pattern showing robust relationships to behavior only in the early phase of the experiment (Facilitation 2, Face). As with the assessment of MTL functional connectivity above, we explore the functional connectivity between the two behavioral patterns within each group and with the regional pattern for the MTL analysis. This analysis examined whether there was an even broader pattern of functional connectivity that could be related the learning for the two groups. These interpattern correlations, computed at each of eight time points during the experiment, are displayed as pseudo-colored images in Fig. 5. For the Aware group, interpattern correlations for the MTL and Facilitation pattern and MTL and Discrimination pattern are strong across all eight time points. However, the correlations

Fig. 4. Summary of the dominant functional connections of the left medial temporal lobe (MTL) in learning with (A) and without (B) awareness. Panel A shows dominant interactions, indicated by bidirectional arrows, between MTL and occipital, temporal, and prefrontal cortices when learning proceeds with awareness. Panel B shows dominant interactions, between MTL with thalamus, basal ganglia, and contralateral MTL when learning proceeds without awareness. The key feature is the common involvement of the MTL, but because the regions to which it is functionally connected differ. Correlations [color coded for the range of +1 (deep red) to −1 (deep blue) as per legend] highlight the specificity of the functional connectivity to each group. The first column represents the correlation of left MTL (LMTL) activity with the other voxels. The middle columns represent the correlations among the voxels. The last two columns represent the correlation values for the voxels with behavior.
between the two behavior patterns (i.e., Facilitation and Discrimination) became progressively stronger with time. For the Unaware group, only the MTL and Facilitation patterns maintained strong correlations across the experiment, with the other two sets of correlations becoming progressively weaker across time. The impression from the correlation images can be quantified by computing the norm (first eigenvalue) of each correlation matrix. These norms are presented in Fig. 6 and show an increase for all three sets of correlations for the Aware group, whereas for the Unaware group, the correlation between MTL and the Facilitation pattern is stable across time and the norms decline for the other two sets, indicating overall weaker correlations.

The demonstration of large-scale correlation patterns related to learning and awareness brings back an earlier point about the scales at which brain functions can occur. Had the analysis stopped at the exploration of the MTL, the likely interpretation would have emphasized the role of MTL functional connectivity in learning with awareness, much the same as the preceding study looking at...
prefrontal functional connectivity. This ignores the richness of the data, however, where it is obvious that the pattern is far more distributed, engaging a wide expanse of the brain. Thus, just as with visual category perception, awareness and learning are embodied in the combined actions and interactions across the brain. It is not simply whether an area is active or not, but how its activity contributes to the broader neural pattern that needs to be the emphasis in considering the mind–brain link.

11. Critical Functions and Behavioral Catalysts

One problem for the notion of a neural context is the observation that the expression of certain behaviors or cognitive states appears to rely critically on specific brain areas. The overt expression of declarative memory, for example, depends on the integrity of the MTL. Such dependencies have led many researchers to speculate that such regions form a part of the neurocognitive system whose
constituents subserve that function, but this implies a rather static view of brain function.

An alternative perspective, and one consistent with the idea of neural context, is to consider neural dynamics as a critical feature to understanding functional dependencies such as the MTL and memory. Most often studied at the cellular level, such dynamics are thought to be vital in enabling neurons to code for rapid temporal shifts in the environment and to make rapid adjustments of effectors at time scales much smaller than any single cell can achieve (Milton and Mackey 2000). It is quite likely that such dynamics are at play across many levels of organization in the brain, with a similar general outcome: that it allows for rapid integration of information and responding. These same dynamics are also likely to underlie contextual effects between interacting populations that will manifest as changes in similarly large-scale behaviors, such as attentional states, perceptions, memory types, and quite likely consciousness (Bressler and Kelso 2001). We have speculated that shifting between behavioral states may require the integrity of certain key regions, which when damaged would result in a deficit in that state. Such region may not necessarily participate in the processing within that particular state, but rather enable the transition—it is a behavioral catalyst (McIntosh 2004).

The likely feature of such catalysts is their anatomical relation to regions that are processing the primary information in the state in question. In the awareness study, the MTL was engaged in learning with or without awareness, interacting with regions that seemed to be related to learning in either attentive state. The MTL is anatomically connected with regions that were part of both patterns, providing the potential for the MTL to catalyze the transition between two different networks and thereby the movement from learning without awareness to learning with awareness. The critical point at which the MTL is needed is when learning moves to the conscious state. Before this, the MTL can be engaged, by virtue of its anatomical links, but not be critical for behavioral expression.

Considering regions that are critical for the expression of a function as potential catalysts emphasizes the dynamic nature of brain function. The temporal expansion of any behavior or cognitive function can be viewed as a series of transitions that require specific regions to be intact (Haken 1996; Kelso 1995). In some cases, this dependence may reflect a network node that transmits information between regions (e.g., the lateral geniculate nucleus in the visual system). In other cases, areas enable the change in dominant interactions from one set of regions to another. These are the catalysts.

The temporal dimension of catalyst operations is illustrated in a very simplistic schematic (Fig. 7). A ten-node network contains three subnetworks (A, B, C) and a catalyst node (X). Anatomically, X bridges between A and C, but shows no connection with B (although the precise anatomy is less important for the present illustration). The graph in Fig. 7 presents hypothetical activity profiles for the three subnetworks and the catalyst node X. Subnetwork B is active from the outset, and its impact is entirely on subnetwork A. As subnetwork A continues to be active, it begins to activate node X, which, after a threshold is reached,
would in turn activate subnetwork C, enabling it to process signals from subnetwork B; this may come through modulatory effects from X that reduce thresholds on the recipient node of subnetwork C. At the same time, feedback from X to subnetwork A would reduce the efficacy of signals from subnetwork B, resulting in a decline of its activity. As subnetwork B continues to send signals, subnetwork C maintains activity and a new function is instantiated.

The physiological plausibility of such state changes may be suspect, but I would again emphasize that effects may look different across spatial and temporal scales. A local increase in synaptic excitation may actually manifest as a general decrease of ensemble activity (van Vreeswijk and Sompolinsky 1996). In terms of effective connectivity, at the mesoscopic level, the same region can have a facilitatory effect or suppressive effect depending on situational or neural context. Indeed, we have shown that feedback effects can move from suppression to

---

Fig. 7. Hypothesized action of a catalyst node (X) on the activity and interactivity of three subnetworks. The top of the figure depicts the connections between the three subnetworks: A (blue), B (magenta), and C (orange), and the catalyst node (green). The activity (hypothetical units) across time of the three subnetworks and the catalyst node is plotted in the graph (hypothetical units). At a critical point in time, the catalyst node facilitates the transition of interactions from A and B to C and B.
facilitation as a function of learning (McIntosh et al. 1998), while others have shown that one mutually antagonist processing stream can facilitate another if the situational context requires it (Buchel et al. 1999). All this is meant to emphasize that, although there may be general principles that govern brain function, their expression may be different across spatial and temporal scales.

12. Conclusion

I have made the argument that the anatomical structure of the brain enhances segregation and integration of information, and that this, paired with transient plasticity and population coding, makes for a whole system that is specialized for a broad range of mental functions. There are other features that I have not presented here which are necessary to consider in developing the mind–brain link, such as predictive modeling (Dayan et al. 1995; Friston 1997; Hinton and Dayan 1996) and nonlinear dynamics (Breakspear 2004; Freeman and Holmes 2005; Kelso 1995; Jirsa and Kelso 2000). With these general features, the same neural mechanisms that underlie perceiving a face, or hearing a siren or symphony, also give rise to attention, memory, and consciousness. Such a general mechanism produces a formidable, but not insurmountable, challenge for cognitive neuroscience theory.

A direct response to this challenge is to revise the concepts that link the brain to cognition. The notion of neural context and catalysts is an attempt to put forward such new concepts. Although psychological theories tend to differentiate cognitive processes from operations such as sensation and action, it is important to recognize that sensation, perception, attention, and memory are usually intimately intertwined. The inconvenient overlap in activation patterns for putatively different cognitive functions (Cabeza and Nyberg 2000) is an indication that there are common neurophysiological mechanisms that support these diverse cognitive functions. Thus, the division between psychological constructs is not absolute in the brain.

Another important new concept that embodies general brain principles is evident in Tononi's (Tononi 2004, 2005) hypothesis that the highest form of cognition, consciousness, can be directly linked to a neural system’s capacity to integrate information. This capacity is a direct result of the physiology and anatomy of the nervous system. Formulated this way, conscious, and therefore most of, cognition reflects a potentiality of a nervous system, rather than a special property that only exists in the primate brain. Indeed, the potential can be measured quantitatively (Tononi and Sporns 2003), enabling a more direct assessment of the dynamics of a nervous system and the richness of the behavioral functions it generates. The interesting implication of considering cognitive functions from this perspective is that it dissociates the basic biological capacity from the phenomenology of our personal experience. It seems somewhat dissatisfying to conceive of the mind–brain link in this way, but in doing so, it opens a new venue for investigation of information integration, contextual effects, and poten-
tial catalysts without having to label such endeavors as looking only at sensation, or attention, or memory. The positive consequence is that one can then measure the brain using the same quantitative tools and neural principles in characterizing transition between different cognitive functions. This effort necessarily yields a picture of human mental function that is more integrated across cognitive and behavioral domains and, at the same time, grounded in emerging principles of brain operation.

References


Bressler S, McIntosh AR (in press) The role of neural context in large-scale neurocognitive network operations. In: Jirsa V, McIntosh AR (eds) Handbook of brain connectivity. Springer,


James W (1890) The principles of psychology. Dover, Boston


Subject Index

a
abstract 294
abstract rule 299, 300
abstraction 301, 306
acalculia 277
action goal 141
action observation 134
action representation 140
action sequence 137
active touch 151
activity 312
agranular frontal cortex 124
AIP 159, 160
allocentric 220, 235–237
ambiguous visual input 14
amygdala 88
anatomical hierarchy 306
angular cortices 155
angular head velocity 233
animal models 21
anosognosia 152
an antagonistic pattern of modulations 12
anterior cingulate cortex 345
anterior frontal area 114
anterior PFC 115
anterior–posterior gradient 305
anterior prefrontal cortex 109
anterodorsal thalamic nucleus 230
arbitrary stimulus–response 295
area 2 179
area 3a 179
area 5 154, 158, 166
area 7 159
area 44 191, 192
area-selective neuronal activity 320
arithmetic 272
asomatognosia 152
associative memory 28
associative neuronal network 264
attention 64–66, 79–88, 262
attention shifting 29
attention window 27
attractor 234
“audio-visual” mirror neuron 130
auditory discrimination 257
auditory temporal cortex 256
awareness 346

b
β 38
basal ganglia 348
behavioral catalyst 352
biased competition 79–88
bimodal 169
bimodal property 168
bimodal sensory property 158
binding memory view 110
binding-related activity 115
biological effector 132
biological motion 132, 156
biting action 138
blood oxygenation level-dependent
(BOLD) activation 34, 275
bodily self 170
body image 151, 177, 193
body part-centered coordinates 169
body-shrinking illusion 185
cross-correlogram 257
cross-oriented 8
cross-temporal bridging of sensory-motor
cue-period activity 315
information 324
cuttaneous receptors 181
Broca’s area (BA) 10 109, 114
correlation discharge 1
cutaneous receptors 181
Broca’s patient 140
corollary discharge 1
cue-period activity 315
cortical damage 181
cutaneous receptors 181
cortical damage 181
cutaneous receptors 181
cortical surface 6
cutaneous receptors 181
cost 293
cutting 241
counting 271
cutaneous receptors 181
cross-correlation analysis 256, 325
cutaneous receptors 181
delayed anti-saccade task 318
delayed matching to sample 32, 105
delayed pro-saccade task 318
delay-response 313
delay-period 312
delay-period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
dynamic interaction 329
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
delayed antecedent of 318
delayed response 313
delay period 312
delay period activity 316
delay period activity 316
delayed pair associate task 318
delayed pro-saccade 318
delayed anti-saccade 318
delayed matching to sample 32
decoding 321
deception 321
decision making 292
declarative/episodic memory 236
degenerate 338
delayed pro-saccade task 318
delayed response 313
delay period 312
delay period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
delayed antecedent of 318
delayed response 313
delay period 312
delay period activity 316
delayed pair associate task 318
delayed pro-saccade task 318
delayed response 313
delay period 312
delay period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
delayed antecedent of 318
delayed response 313
delay period 312
delay period activity 316
delayed pair associate task 318
delayed pro-saccade task 318
delayed response 313
delay period 312
delay period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
delayed antecedent of 318
delayed response 313
delay period 312
delay period activity 316
delayed pair associate task 318
delayed pro-saccade task 318
delayed response 313
delay period 312
delay period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
delayed antecedent of 318
delayed response 313
delay period 312
delay period activity 316
delayed pair associate task 318
delayed pro-saccade task 318
delayed response 313
delay period 312
delay period activity 316
delusiveness 321
dentate gyrus 256
difference cross-correlogram 256
directed preference 318
discrimination tasks 181
distributed representations 52–61
dorsal premotor cortex (PMD) 182
dorsal thalamic nucleus 233
dorsal-ventral gradient 305
dorsal visual stream 157
dorsodorsal pathway 157
dorsolateral prefrontal cortex (DLPFC)
drug-resistant epilepsy 205
dual coding 257
dyadic communication 137
dynamic changes of the functional
dynamic completion property 253
dynamic construction and reconstruction
dyscalculia 277
dysgranular 138
eating behavior 137
EBA 165
eccentricity 6
effective connectivity 343
efference copy 152, 154, 161, 168
efferent process 188
egocentric 220
einstellung 302
electrical stimulation 32
electroencephalography 134, 337
emotion understanding 145
empathy 145
encoding 105
ensemble 339
entorhinal cortex 227, 228, 235
episodic memory 237, 280
epoch-related design 115
event-related design 115
event-related paradigm 34
event-related potentials (ERPs) 205
executive locus 186
expanded body image 167
expectancy 262
explicit detection of change 116
explicit scene understanding 104
extensor carpi ulnaris 182
external object 189
extrastriate visual areas 33

f
face 342
face expression 71, 72, 90–93
face neuron 251
face representation 50
face-selective neuron 50–58
feature binding 108, 112
feature-based memory system 116
feature-bound representation 111
feature combination 51, 52
feature detectors 262
feature memory view 110
Fechner’s law 275
feedback 23
feedback loop 38
feed-forward 23
FEF 201, 210
F4 159, 168, 169
filter functions 274
F5 159, 160
flexibility 292, 295
flexibly 306
foot 182
forward model 154
frontal eye field (FEF) 18, 32, 108
frontoparietal cortices 193
frontoparietal network 107, 115
F2 168
function 306
functional connectivity 327, 342, 35
functional gradient 305
functional interaction 325
functional magnetic resonance imaging (fMRI) 3, 33, 177, 227, 275, 337, 345
g
generalization 306
Gerstmann’s syndrome 277
gesture 138
giant-neuron coding 252
grappling 126
grappling for eating 143
grappling for placing 143
grid cell 228, 235, 236

h
hand manipulation 160
hand–object illusion 190
hand–object interaction 127
hand–object interactive movement 189, 190, 193
hand region 180
haptic 191
head direction (HD) cell 230, 232, 235
headcentered coordinate 169
hemispatial neglect 152
hierarchical organization 22
hierarchy 288
hippocampus 221, 227, 233, 236, 256
homology 138
“How” problem 249
human 177, 226
human motor area 193
hypothesis 30

i
identification 30
ideomotor apraxia 191
idiothetic 236
idiothetic cues 232, 235
illusory contours 36
imitation 129, 136
imitation learning 136
implicit binding 116
in barter 294
independent component analysis (ICA) 263
inferior frontal gyrus (IFG) 17, 109, 115, 134
inferior frontal sulcus 17
inferior parietal cortex 156
inferior parietal lobule 108, 116, 133, 191, 192, 201, 275, 280
inferior temporal cortex 28, 47, 50, 51, 56, 57, 59, 61, 62, 64, 65, 67–74, 79, 80–83, 85, 86, 93, 94, 303
information flow 326
information shunting 29
ingestive motor act 129
integration 339
integration of multisensory information 177
integrator 265
intention 140, 141
intention understanding 136, 144
intentional saccade 201
internal model 154
internal motor repertoire 130
internal simulation 188
intersubjectivity 145
intraparietal and intracocipital sulci (IPS/IOS) 106
invariance 61–64
invariant representation 47–78, 55, 64, 71, 73, 78
inverse model 154
invertebrate 341
iso-eccentricity bands 6
iso-oriented 8

j
Jacobsen 313
joint peristimulus time histogram (j-PSTH) 327

k
kinesthesia 186
kinesthetic illusion 178
k-means clustering 263
Koehler 271

l
language 130, 136
language evolution 138
late process 25
lateral interaction 4
lateral mammillary nucleus 231
layout 103
learning 69, 70
left hemisphere 191
limb movements 177
LIP (lateral intraparietal area) 158, 201
long-term associative memory 29
long-term potentiation (LTP) 250
LTD 250

m
magnetoencephalographic 134
maintenance 105
manipulation-related activity 115
many-to-one mapping 23
match 299
matching system 129, 132
meaningless gestures 134
meaning 14
medial temporal lobe 348
MEG 38
memory field 319
memory for feature binding 113
memory load 110, 112
memory process 18
mental representation 177
mental simulation 187
microcircuit 264
middle frontal cortex 137
middle frontal gyrus 17
mind-reading 140
MIP (middle intraparietal area) 158
mirror neuron 127, 163
mirror-symmetrical stimuli 11
mnemonic-symmetrical stimuli 11
M1 168
monkey 126
monotonic stimulus encoding 319
mosaic organization 124
motoneuronal cell 179
motor act 124
motor homunculus 135
motor imagery 188
motor intention 144
motor knowledge 126
motor neuron 126
motor output 184
motor repertoire 136
motor representation 126
mouth communicative gesture 129
mouth mirror neuron 129
MST 158
MT 158
multidimensional MOPT 113
multiple frames of reference 169
multiple neuronal correlation 254
multiple object permanence tracking (MOPT) 112
multiple object tracking (MOT) 107
multiple-task comparison 254
muscle spindle receptor 178

naming task 15, 18
natural scene 64–69
navigation 220, 224, 227, 234–236
navigational decision 225
n-back task 107
neighboring neurons 263
network 338
neural context 340
neural plasticity 339
neuronal code 250
neuronal currency 295
nonhuman primate 226
nonmatch 299
non-reward 88–93
nonspatial working memory task 319

nonstationary waveform problem 263
novelty 295
number line 275
number 271
numeral 276
numerical 272
numerical competence 280
numerical distance effect 274
numerical intelligence 280
numerical rank 278
numerical size effect 274
numerosity 274
numerosity-selective 274

object identification 25
object processing 343
object representations 63, 79
object working memory 106
object-properties-processing 27
observed agent 144
observed model 137
occipital cortex 15, 28
occipitoparietal 343
occipitotemporal 343
oculomotor delayed-response task 314
ODR task 314, 322
optic flow 232
orbitofrontal cortex 17, 88–93
ordinal rank 278
orientation selectivity 10
overlapping set coding 253

parietal cortex 15, 111, 124
parietal mirror neuron 133
parietal operculum 183
parietotemporal mirror neuron circuit 144
passive viewing 18
path integration 235, 236
PEa 158, 167
PEF 201
perception 21
perception–action cycle 287, 306
perceptual categorization 18
perceptual grouping 3
peripersonal space 168
peripersonal stimulus 126
personal neglect 152
PF/PFG 158
PFEF 204, 212
phantom limbs 152
phase 13
phonetic gestures 140
phonological resonance 140
pictorial object description 127
place cell 221, 228, 231, 234, 237, 238
place cell remapping 225
population coding 28, 340
population vector 262, 322
posterior parietal area 114, 124
posterior parietal cortex 272, 345
postsaccadic 316
postsaccadic activity 316
postsubiculum 230
precentral gyrus 180, 181
prediction 143
prefrontal cortex (PFC) 15, 17, 111, 133, 137, 272, 343, 345, 346
premotor cortex 303
presaccadic activity 316
pre-SEF 203
primary motor cortex (M1) 134, 179
primary somatosensory cortex (SI) 178
primary visual cortex (V1) 3, 9, 11
processing subsystem 26
projective test 14
prospective information 319
proximal motor act 126
pseudo-word 140
PST coincidence histogram 328

rate-coding 52–61, 253
reaching 126
recalibration 187
receiver operating characteristic 290
receptive field 23, 187
receptive field size 64–69
recognition 30
reference memory 255
reflexive top-down processing 36
regions of interest (ROI) analyses 114
relationships (task sets) 305
remapping 223
resonance behavior 137
response facilitation 137
response modulation 11, 13
response-period activity 316
retina 22
retinotopic area 6
retinotopic localization 6
retinotopic mapping 5
retrieval 105
retrospective information 319
reversal 296
reversal learning 88–93, 348
reverse hierarchy theory 104
reversible cooling 32
reward information 288
right hemisphere 183
rise time 298
risk 294
R-ODR task 324
rolandic operculum 210
Rorschach inkblot test 14
rotatory ODR (R-ODR) task 319
rubber hand illusion 153, 167
rule-dependent neural activity 319

s
saccade 199
scene gist 103
schizophrenia 157
SEEG 205
SEF 202, 210
segregation 339
self-other distinction 165
semantic content 138
semiconnected 339
sense of agency 153, 154
sense of ownership 153, 154, 166
sensing limb movement 178
sensorimotor transformation 126
sensory feedback 152, 188
sequence 137
serial order 271
set-related activity 161
short-term associative memory 28
signal detection theory 38
silent speech 139
simulation theory 141
sleep 238
slow-wave sleep 238
social behaviour 88–93
social cognitive capacity 145
social environment 129
somatic perception 188
somatosensory feedback 154
somatotopical organization 140
somatotopy 133, 135
sparse coding 253
spatial 238, 343
spatial contingency 155
spatial frequency 13
spatial information 224
spatial view cell 226
spatial working memory 106
spatial-properties-processing 28
spatiotemporal cortical map 4
spatiotemporal manipulation cost 112
spatiotemporal modulation 8
specializations in inferior temporal cortex 48–50
speech area 134
speech perception 140
speed of processing 61, 70, 71
spike histograms 291
spike-overlapping problem 263
spike-sorting 263
SPL 18
stimulus-dependent synchrony 56–58
stimulus–response 295, 303, 305
stimulus–reward reversal 292
strategic top-down processing 30
striate cortex 22
striatum 298, 303
strictly and broadly congruent mirror neuron 129
structural equation modeling 347
STS 165
superior colliculi 199
superior frontal gyrus (SFG) 109
superior frontal sulcus (SFS) 106, 108
superior parietal cortex 156
superior parietal lobule (SPL) 108
superior temporal cortices 347
superior temporal sulcus 132
superposition catastrophe 253
supplementary motor area (SMA) 109, 182
suppressive interaction 13
suppressive modulation 10
supramarginal 155
supramarginal gyrus 183
surround suppression 6, 10, 13
sustained delay activity 105
synchronization 205
synchrony 263

t
tactile receptive field 125
task-related neuron 259
task set 302, 304
temporal change of the information represented 321
temporal change of directional selectivity 322
temporal congruency 155
temporal contingency 156, 166
temporal correlation analysis 15
temporoparietal junction 156
tempro-parieto-premotor circuit 133
tendon vibration 179
thalamus 348
theory of attention 202
theory–theory 141
theta rhythm 222
time-course 290
time-domain/correlational hypothesis 252
time–frequency (TF) analysis 205
tonic vibration reflex (TVR) 178
tool 132
tool action 132
tool use 192
tool-responding mirror neuron 132
top-down control 203
top-down modulatory influence 329
top-down processing 24
top-down signal 330
transcranial magnetic stimulation (TMS) 134, 186
transferred kinesthetic illusion 184, 185
trial-and-error learning 302
type-identification paradigm 113

u
uncinate fascicle 297
understanding intention 141
understanding of the goal 129
Subject Index

v
value 293
vector completion 39
ventral premotor cortex 124, 126
ventral stream 22
ventrodorsal pathway 157
ventromedial 288
very high gamma band (VHG) 206
vestibular system 232
vestibular 233
VIP (ventral intraparietal area) 158, 168
virtual environment 226
visual attention 33, 202
visual buffer 27
visual discrimination 257
visual feedback 154
visual imagery 31
visual paired associates 31
visual processes 25
visual system 22
visual working memory 103
visuomotor neuron 127

“vocabulary” of motor acts 126
vocalization 138
voluntary movements 180
V2 9, 11
V3/V3A 157
V3/VP 9, 11
V5 200
V6 157
V6A 158
voxel-based morphometry 277

w
Weber’s law 274
“what” function 184
“where” function 184
Wisconsin Card Sorting Test 299
Wisconsin general test apparatus (WGTA) 313
with novel 306
words 140
working memory 202, 203, 255, 287, 303, 304, 311, 345