

The prefrontal cortex: categories, concepts and cognition

Earl K. Miller*, David J. Freedman and Jonathan D. Wallis

Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

The ability to generalize behaviour-guiding principles and concepts from experience is key to intelligent, goal-directed behaviour. It allows us to deal efficiently with a complex world and to adapt readily to novel situations. We review evidence that the prefrontal cortex—the cortical area that reaches its greatest elaboration in primates—plays a central part in acquiring and representing this information. The prefrontal cortex receives highly processed information from all major forebrain systems, and neurophysiological studies suggest that it synthesizes this into representations of learned task contingencies, concepts and task rules. In short, the prefrontal cortex seems to underlie our internal representations of the ‘rules of the game’. This may provide the necessary foundation for the complex behaviour of primates, in whom this structure is most elaborate.

Keywords: neurophysiology; cognition; volition; attention; control

1. INTRODUCTION

Although our brains have developed exquisite mechanisms for recording specific experiences, it is not always advantageous for us to take the world too literally. A brain limited to storing an independent record of each experience would require a prodigious amount of storage and burden us with unnecessary details. Instead, we have evolved the ability to detect the commonalities among experiences and store them as abstract concepts, general principles and rules. This is an efficient way to deal with a complex world and allows the navigation of many different situations with a minimal amount of storage. It also allows us to deal with novelty. By extracting the essential elements from our experiences, we can generalize to future situations that share some elements but may, on the surface, appear very different.

For example, consider the concept ‘camera’. We do not have to learn anew about every camera that we may encounter. Just knowing that the item is a camera communicates a great deal of knowledge about its parts, functions and operations. Or consider the set of rules invoked when we dine in a restaurant, such as ‘wait to be seated’, ‘order’ and ‘pay the bill’. These rules are long divorced from the specific circumstances in which they were learned and thus give us an idea about what to expect (and what is expected of us) when we try out a new restaurant. Hearing that a ‘*coup d’etat*’ has occurred communicates the ‘gist’ of what happened without having to hear the details.

While much is known about the encoding of physical attributes and specific experiences, relatively little is known about how abstract information is encoded in the

brain. This may be because this is relatively more difficult to study than the neural correlates of physical attributes, such as shape. By definition, these categories and concepts are labels that transcend physical appearance. Think of all the wildly different-looking objects that are considered to be chairs. However, it is not always easy to disentangle encoding of categories from similarity. Let us say that we discover some neurons somewhere in a monkey’s brain that become activated whenever it views a tree. Are these neurons really encoding the category ‘tree’? They might be encoding the fact that trees happen to look more like one another than many other objects. Further, development of abstract representations requires a considerable amount of experience; learning a general principle requires a wide range of experiences so that underlying rules can be extracted.

Our laboratory has conducted experiments to establish how abstract information is represented in the brain. We trained monkeys on tasks that allowed them to group different stimuli and experiences into categories or behaviour-guiding rules. We summarize some of that work and discuss its implications for an understanding of a neural basis of high-level cognitive function. We have focused on a brain region that is central to high-level cognitive function, the PFC.

2. THE PREFRONTAL CORTEX

The PFC is an ideal place to look for neural correlates of abstract information. It occupies a far greater proportion of the human cerebral cortex than in other animals, suggesting that it might contribute to those cognitive capacities that distinguish humans from animals (Fuster 1995; figure 1). On initial examination, PFC damage has remarkably little overt effect; patients can perceive and move, there is little impairment in their memory and they can appear remarkably normal in casual conversation.

* Author for correspondence (ekm@ai.mit.edu).

One contribution of 14 to a Discussion Meeting Issue ‘The physiology of cognitive processes’.

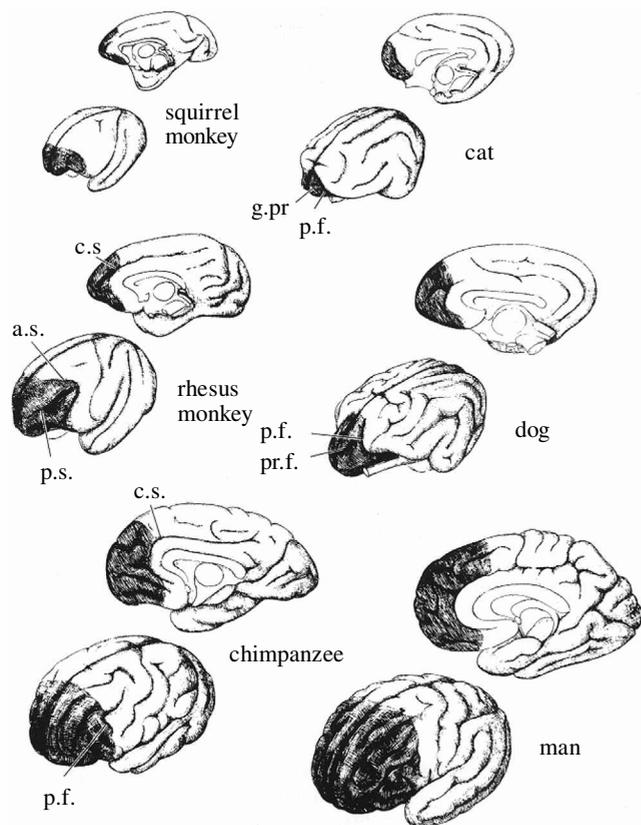


Figure 1. The relative size of the PFC in different animals. Abbreviations: a.s., arcuate sulcus; c.s., cingulate sulcus; g.pr., gyrus proreus; p.f., presylvian fissure; p.s., principal sulcus; pr.f., preoreal fissure. From Fuster (1995).

However, despite the superficial appearance of normality, PFC damage seems to devastate a person's life. They have difficulty in sustaining attention, in keeping 'on task', and seem to act on whims and impulses without regard to future consequences. This pattern of high-level deficits coupled with a sparing of lower-level basic functions has been called a 'dysexecutive syndrome' (Baddeley & Della Sala 1996) and 'goal neglect' (Duncan *et al.* 1996).

Indeed, the anatomy of the PFC suggests that it is well-suited for a role as the brain's 'executive'. It can synthesize information from a wide range of brain systems and exert control over behaviour (Nauta 1971). The collection of cortical areas that comprise the PFC have interconnections with brain areas processing external information (with all sensory systems and with cortical and subcortical motor system structures), as well as internal information (limbic and midbrain structures involved in affect, memory and reward) (see figure 2). Correspondingly, its neurons are highly multimodal and encode many different types of information from all stages of the perception-action cycle (Fuster 1995). They are activated by stimuli from all sensory modalities, before and during a variety of actions, during memory for past events, in anticipation of expected events and behavioural consequences, and are modulated by internal factors such as motivational and attentional state (see the review in Miller & Cohen 2001). Because of its highly multimodal nature and its apparent role in higher mental life, the PFC seemed like an ideal place to begin our search for neural correlates of the abstract information needed for intelligent behaviour.

3. THE PREFRONTAL CORTEX AND PERCEPTUAL CATEGORIES

Because perceptual categories often group together very different looking things, their representation must involve something beyond the sort of neural tuning that underlies the encoding of physical attributes, that is, gradual changes in neural activity as certain attributes gradually change (e.g. shape, orientation, direction). Instead, categories have sharp boundaries (not gradual transitions) and members of the same category are treated as equivalent even though their physical appearance may vary widely.

Consider a simple example of a perceptual category: crickets sharply divide a certain range of pure tones into 'mate' versus 'bat' (a predator) (Wyttanbach *et al.* 1996). Even though the input varies along a continuum, behaviour is binary. Across a wide range of lower frequencies, crickets will turn towards the sound source because it may be a potential mate. However, at a certain point (16 kHz), the behaviour suddenly flips: crickets begin to turn away from the sound source because it could be a bat. Crickets make virtually no distinction between frequencies over a wide range on either side of the boundary; they approach or avoid with equal reliability. This type of representation is illustrated schematically in figure 3. Presumably, the ability to transform the raw sensory inputs into distinct categories evolved because it is advantageous; in this case, it optimizes reproductive behaviour while minimizing fatal mistakes. Similar effects are evident in humans' perception of 'b' versus 'p' (Lisker & Abramson 1970).

The elaborate behavioural repertoire of advanced animals naturally depends on more elaborate categorization abilities. The mental lexicon of primates, for example, includes abstract categories that are characterized along multiple dimensions that are often difficult to define precisely, such as 'tool'. In addition, advanced animals have an enormous capacity to learn and adapt. Most of our categories are acquired through experience (we learn what a 'chair' is) and we can continually modify and update our categories as we learn more about them. The ability of monkeys to learn complex perceptual categories has been catalogued in studies that have taught them categories such as animal versus non-animal (Roberts & Mazmanian 1988), food versus non-food (Fabre-Thorpe *et al.* 1998), tree versus non-tree, fishes versus non-fishes (Vogels 1999a) and ordinal numbers (Orlov *et al.* 2000). Pigeons also have a remarkable ability to learn such distinctions (Bhatt *et al.* 1988; Young & Wasserman 1997).

Where such categories are encoded in the brain is unclear. In primates, they could be represented and stored in the same areas of the visual cortex that analyse form and are critical for remembering individual objects, such as the ITC. They might also be evident in the brain regions that receive the results of visual processing from the ITC and are critical for planning and guiding behaviour, such as the PFC. Both the ITC and PFC contain neurons selective for complex stimuli such as trees, fishes, faces, brushes, etc. (Desimone *et al.* 1984; Tanaka *et al.* 1991; Vogels 1999b). But whether or not this selectivity reflects category information *per se* has not been determined. With a large, amorphous category (e.g. food, human, etc.), the category boundaries are unknown. Thus, characteristics diagnostic of category representations (sharp tran-

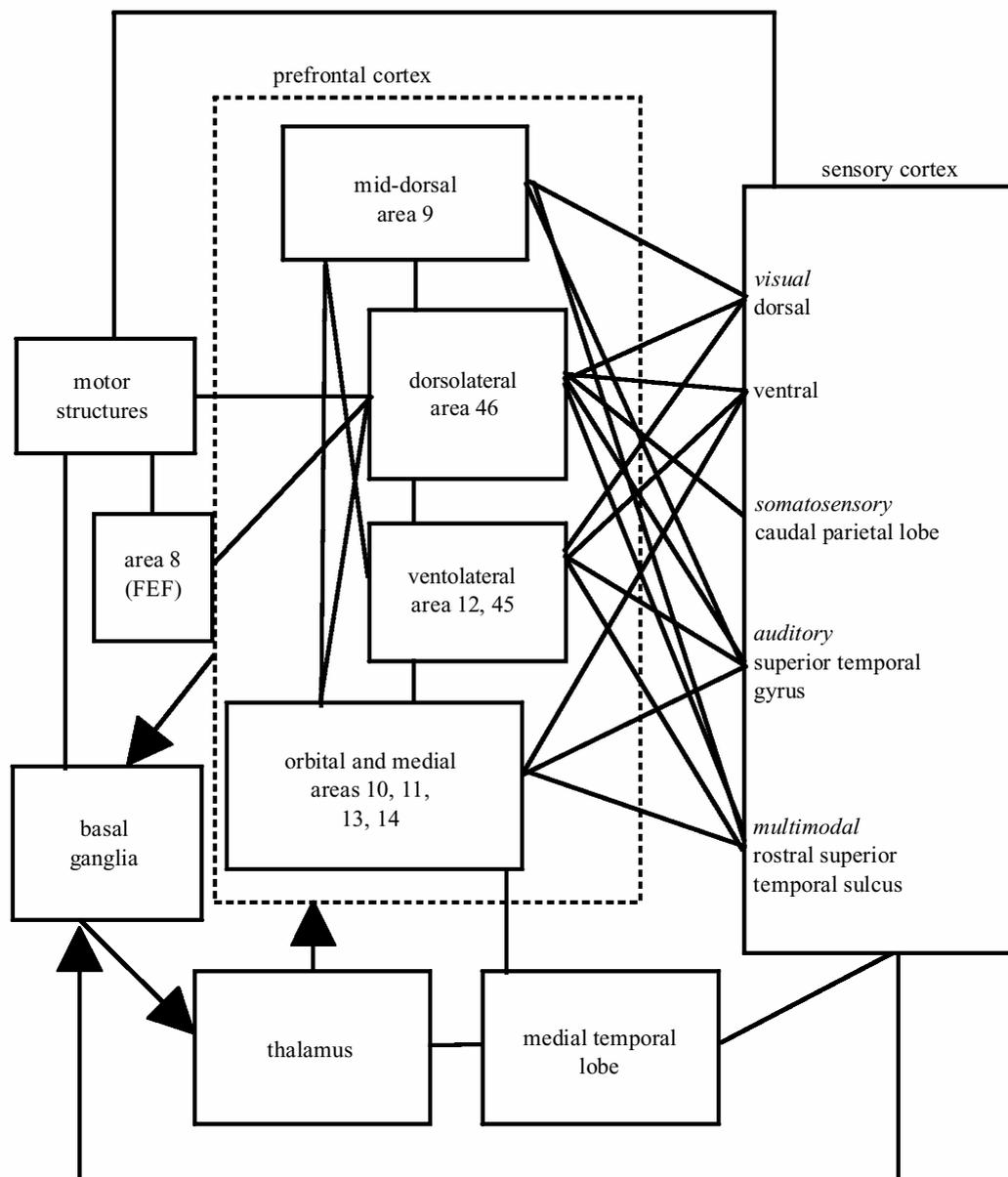


Figure 2. Schematic diagram of extrinsic and intrinsic connections of the PFC. Most connections are reciprocal; the exceptions are noted by arrows. From Miller & Cohen (2001).

sitions and little within-category distinction) cannot be tested. This is not to say that ITC neural selectivity would not make important contributions to category representation, it is just not clear whether it represents category membership *per se*.

To test for neural correlates of perceptual categories, we trained monkeys to categorize computer-generated stimuli into two categories, 'cats' and 'dogs' (Freedman *et al.* 2001; figure 4). A novel three-dimensional morphing system was used to create a large set of parametric blends of six prototype images (three species of cats and three breeds of dogs) (Beymer & Poggio 1996; Shelton 2000). By blending different amounts of cat and dog, we could smoothly vary shape and precisely define the boundary between the categories (greater than 50% of a given type). As a result, stimuli that were close to, but on opposite sides of, the boundary were similar, whereas stimuli that belong to the same category could be dissimilar (e.g. the 'cheetah' and 'house cat').

Two monkeys performed a DMC task (figure 5) that required judgement of whether successively presented sample and test stimuli were from the same category or not. For training, samples were chosen from throughout the cat and dog morph space. After training, classification performance was high (*ca.* 90% correct), even when the samples were close to the category boundary. The monkeys classified dog-like cats (60% cat, 40% dog) correctly *ca.* 90% of the time, and misclassified them as dogs only 10% of the time, and vice versa. Thus, the monkeys' behaviour indicated the sharp boundary that is diagnostic of a category representation. The dog-like cats were treated as cats, even though they were more similar in appearance to the cat-like dogs just across the category boundary than they were to the prototype cats.

We recorded in the lateral PFC, the PFC region directly interconnected with the ITC, and found many examples of neurons that seemed to encode category membership. Two examples are shown in figure 6. Note that their

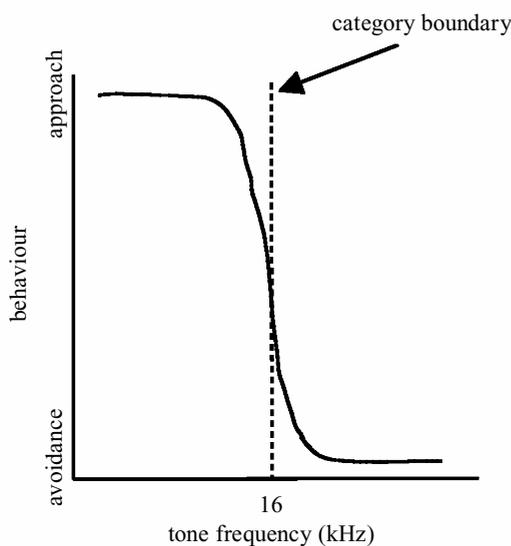


Figure 3. Schematic representation of categorical perception using crickets' responses to a continuum of pure tones as an example. Based on Wytenbach *et al.* (1996).

activity was different from dog-like (60%) cats and cat-like (60%) dogs, yet similar between these stimuli and their respective prototypes. In other words, PFC neurons seemed to make the same sharp distinctions that were evident in the monkeys' behaviour and are indicative of categorical representations, and collected different stimuli together irrespective of their exact physical experience. Category information seemed to predominate in the PFC; across the neural population, tuning was significantly shifted towards representing category rather than individual stimuli (Freedman *et al.* 2001).

As our monkeys had no experience with cats or dogs prior to training, it seemed likely that these effects resulted from training. To test for learning effects, we retrained one monkey on the DMC task after defining two new category boundaries that were orthogonal to the original boundary (figure 4). This created three new classes; each contained morphs centred around one cat prototype and one dog prototype (e.g. the cheetah and the 'doberman'). After training, we found that PFC neural activity shifted to reflect the new, but not the old, categories. An example of a single neuron is shown in figure 7. It showed a significant effect of category during the delay period when data were sorted according to the (currently relevant) three-category scheme; it distinguished one of the categories from the other two (figure 7*a*). By contrast, when the data were sorted using the old category scheme, there was no differentiation between the now-irrelevant cat and dog categories (figure 7*b*).

Our results illustrate that, with experience, category information can become incorporated at the single-neuron level, much as physical attributes of stimuli are. This did not have to be the case: in principle, categories might have been encoded in another fashion. For example, categories might have been encoded at the ensemble level, as an emergent property of neurons that represent their defining features. This ability to carve category membership into the tuning of single neurons may allow for the quick and effortless classification of familiar items.

Our results might reflect a relative specialization of the

PFC in encoding category membership. Categories, after all, are typically defined by their behavioural relevance, and the PFC plays a central part in planning voluntary behaviours. Conversely, the traditional roles of the PFC and ITC are in cognitive functions versus object vision and recognition, respectively. PFC damage causes deficits in attention, working memory and response inhibition, but usually spares object recognition, long-term memory and 'high-level' visual analysis (Fuster 1989; Miller & Cohen 2001). By contrast, ITC damage causes deficits in visual discrimination and learning (Gross 1973; Mishkin 1982) and category-specific agnosias (e.g. for faces) in humans (Gainotti 2000). It might be that the category information in the PFC was retrieved from long-term storage in the ITC for its immediate use in the task. Interactions between the PFC and ITC underlie the storage and/or recall of visual memories and associations (Rainer *et al.* 1999; Tomita *et al.* 1999). Tomita *et al.* (1999) demonstrated that top-down signals from the PFC were needed to activate long-term visual memories stored in the ITC. A similar relationship may exist for the recall of visual categories. In either case, it seems that category information is strongly represented in the PFC, a finding consistent with its role in high-level cognitive functions and in guiding behaviour. The relative contribution of the ITC remains to be determined.

4. THE PREFRONTAL CORTEX AND RULES

It is not only useful to group different sensory stimuli into meaningful categories, it is also useful to group specific experiences of our interactions with the environment along common themes, that is, as behaviour-guiding principles or rules. To this end, our brains have evolved mechanisms for detecting and storing often-complex relationships between situations, actions and consequences. By gleaning this knowledge from past experiences, we can develop a 'game plan' that allows us to extrapolate and infer which goals are available in similar situations in the future and what actions are likely to bring us closer to them.

A standard behavioural test of rule learning in monkeys is conditional associative learning (Passingham 1993). This refers to a class of tasks that require learning associative relationships that are arbitrary and extend beyond the simple one-to-one, stimulus-response mappings that underlie reflexive reactions to the environment. In conditional learning tasks, a given input does not invariably lead to a given output. Whether or not a given response is successful depends on additional, contextual, information. For example, reaching for popcorn can be rewarding, but only if one takes other information into account; if the popcorn belongs to another person, the result could be disastrous. Taking into account complex relationships in order to decide between alternative actions is, presumably, why volition evolved.

To make predictions about which actions are likely to achieve a given goal in a given situation, we need to form a pattern of associations between their internal representations that describes their logical relationship (Dickinson 1980). Decades of behavioural research have illustrated that the brain has learning mechanisms that are exquisitely sensitive to behaviourally informative associations (and

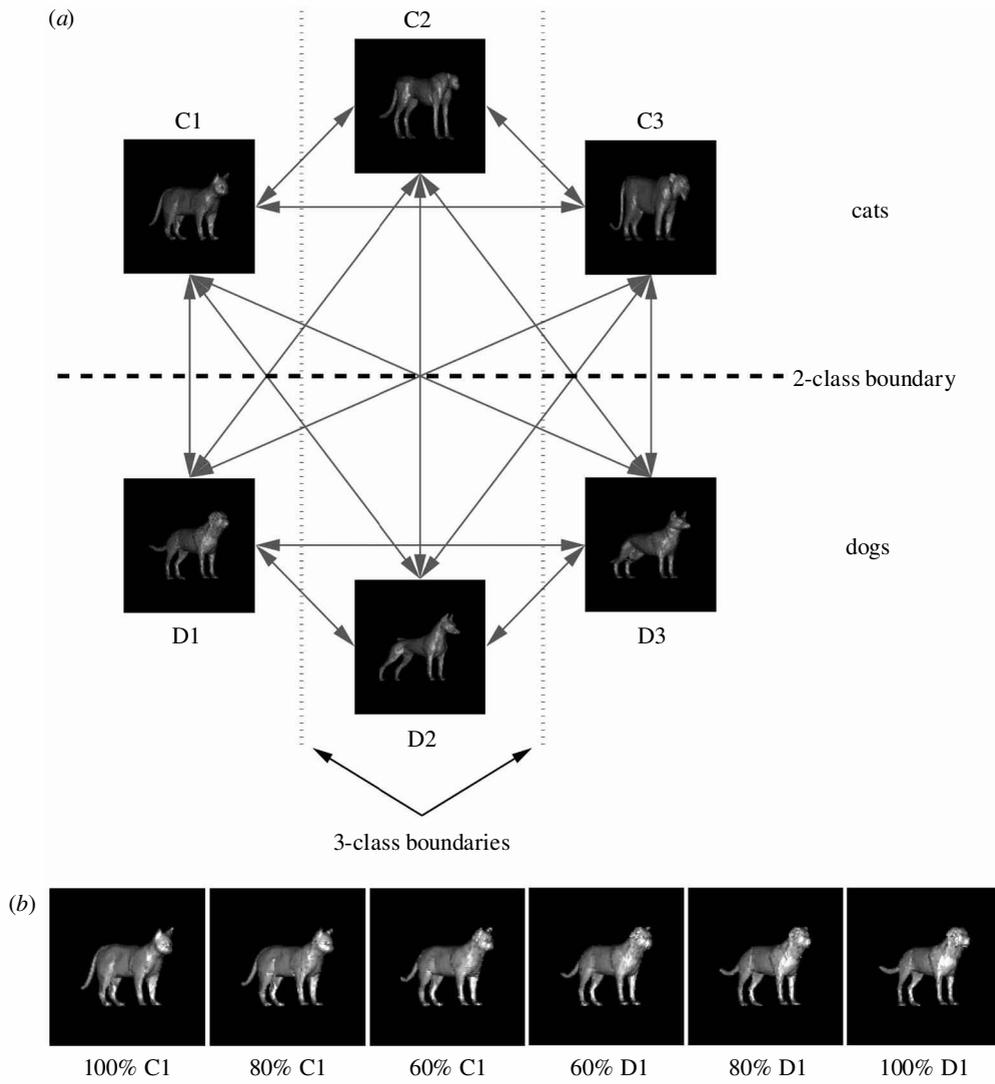


Figure 4. (a) Monkeys learned to categorize randomly generated ‘morphs’ from the vast number of possible blends of six prototypes. For neurophysiological recording, 54 sample stimuli were constructed along the 15 morph lines illustrated here. (b) Morphs along the C1–D1 line. From Freedman *et al.* (2001).

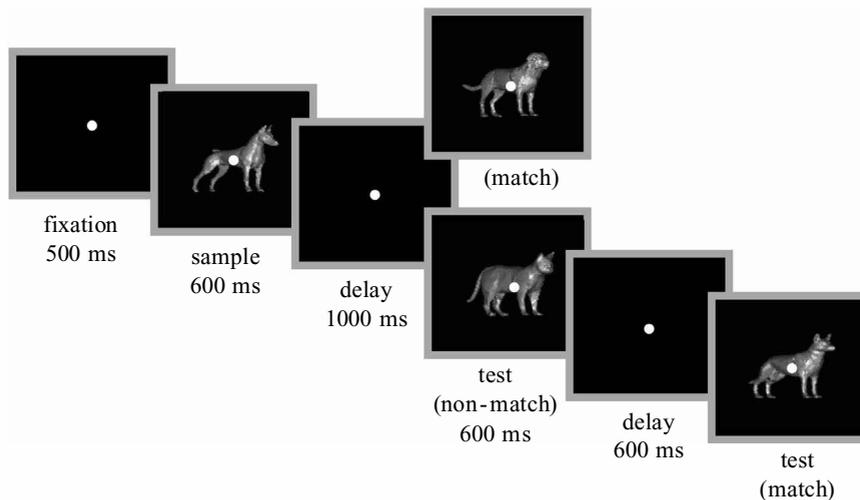


Figure 5. The delayed match-to-category task. A sample was followed by a delay and a test stimulus. If the sample and test stimulus were the same category (a match), monkeys were required to release a lever before the test disappeared. If they were not, there was another delay followed by a match. Equal numbers of match and non-match trials were randomly interleaved. From Freedman *et al.* (2001).

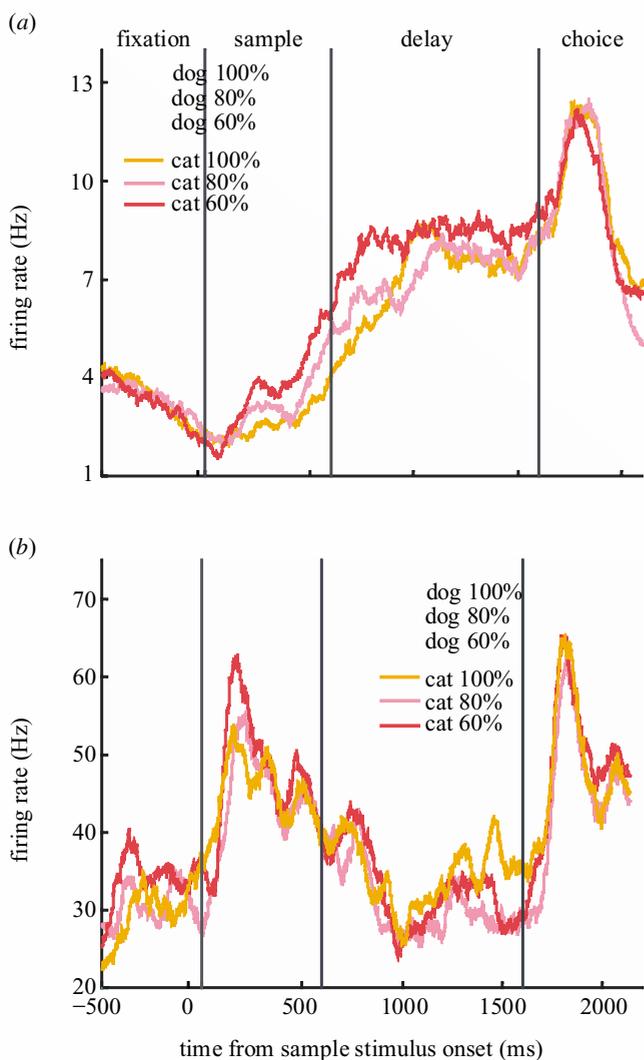


Figure 6. The average activity of two single neurons to stimuli at the six morph blends. The vertical lines correspond (from left to right) to sample onset, offset and test stimulus onset. Activity is pooled over match and non-match trials.

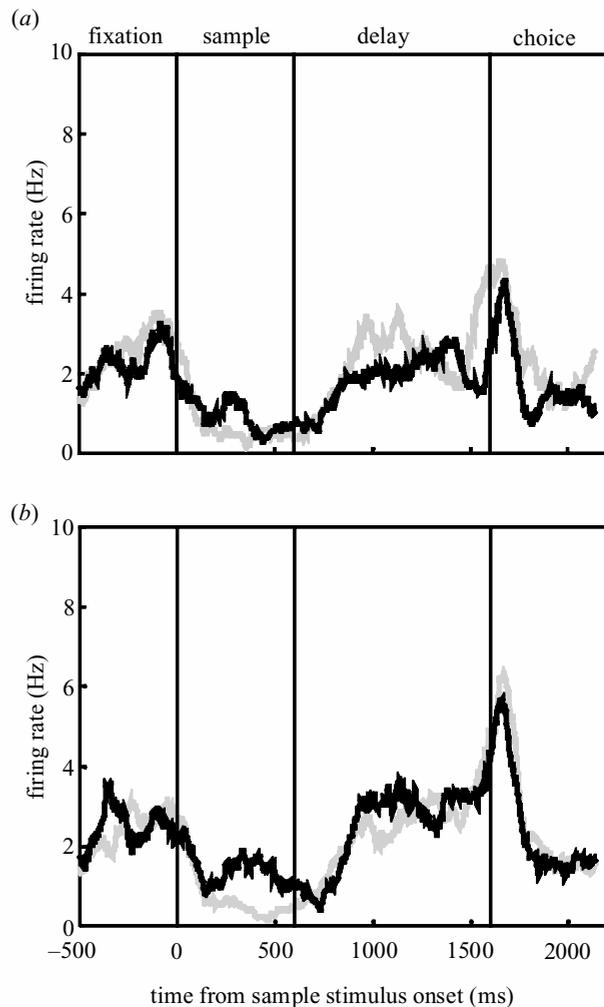
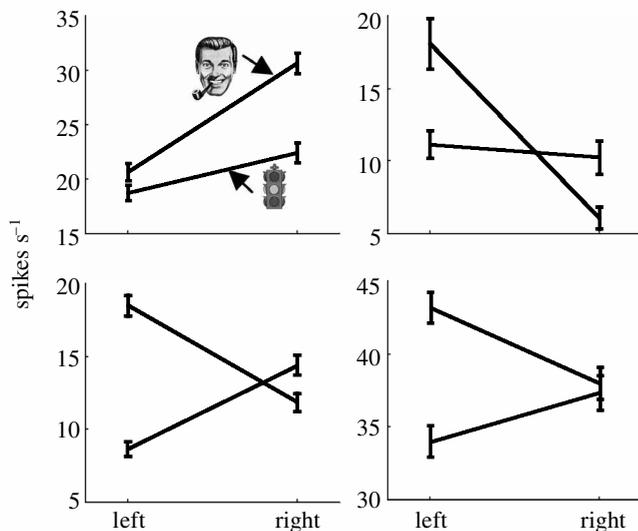


Figure 7. Activity of a single PFC neuron after the monkey was trained on the three-category scheme (see figure 4). (a) The neuron's activity when the data were sorted based on the three-category scheme (category A, black line; category B, grey line; category C, dotted line). (b) The same neuron's activity was data-sorted by the now-irrelevant two-category scheme (black line, cats; grey line, dogs).

Figure 8. The activity of four single prefrontal neurons when each of two objects, on different trials, instructed either a saccade to the right or a saccade to the left. The lines connect the average values obtained when a given object cued one or the other saccade. The error bars show the standard error of the mean. Note that, in each case, the neuron's activity depends on both the cue object and the saccade direction and that the tuning is nonlinear or conjunctive. That is, the level of activity to a given combination of object and saccade cannot be predicted from the neuron's response to the other combinations. Adapted from Asaad *et al.* (1998).

insensitive to, or even discount, associations that are not informative). The underlying neural ensemble of a goal-directed task, then, might be comprised of neurons whose activity reflects task contingencies. Many studies have shown that prefrontal neurons do have this property. This work has focused on the lateral PFC because it seems to be a site of convergence of the information needed to solve conditional sensori-motor tasks. It is directly intercon-

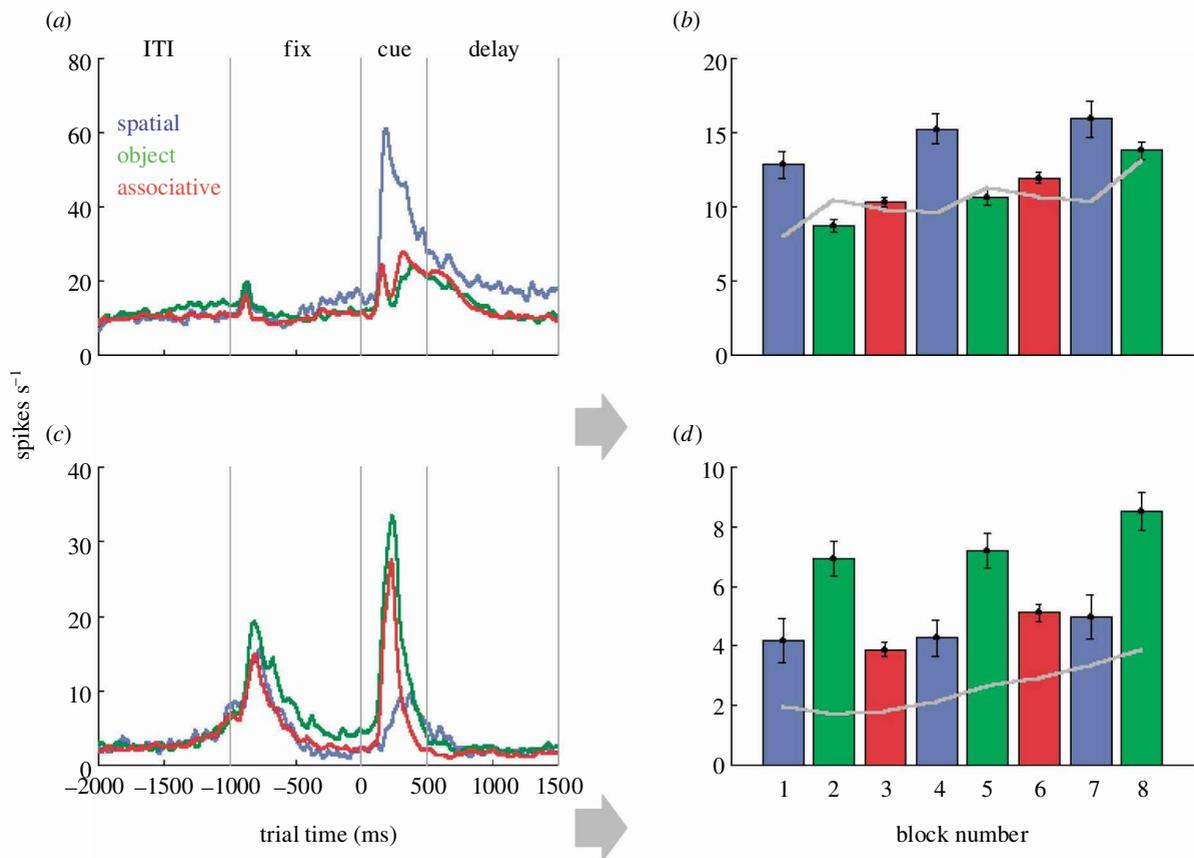


Figure 9. Spike rate versus time histograms for two neurons, each sorted by task. The final second of the three-second ITI is represented by the first 1000 ms (–2000 to –1000 ms). Fixation occurs soon after (*ca.* –1000 to –800 ms). Cue presentation occurs at the time-point marked 0 ms. Task-related differences in baseline firing rate were generally observed to begin in the fixation period. While the activity of some neurons diverged almost coincident with initial fixation (c), the activity of others diverged progressively as the appearance of the cue became more imminent (a). The bar graphs (b,d) demonstrate the reproducibility of these small task-specific changes in activity across multiple repetitions of the same task. The mean fixation-period firing rate (with standard errors) for each block of trials is shown for the two neurons in (a) and (c). The bars are colour-coded to reflect the task being performed in each block, and the colours match those in the histograms to the left. The light grey line superimposed over these bars shows the activity of these neurons during the second immediately preceding the ITI.

nected with the higher-order sensory and motor cortex and indirectly connected (via the ventromedial PFC) with limbic structures that process ‘internal’ information, such as memory and reward (Goldman-Rakic 1987; Pandya & Barnes 1987; Fuster 1989; Barbas & Pandya 1991). The neural activity in the lateral PFC reflects this; many of its neurons exhibit multimodal responses (Vaadia *et al.* 1986; Watanabe 1992; Rao *et al.* 1997; Rainer *et al.* 1998*a,b*; White & Wise 1999). Further, the lateral PFC is critical for normal learning of conditional associations between sensory cues and voluntary actions (Petrides 1985*a*, 1990; Gaffan & Harrison 1988; Eacott & Gaffan 1992; Parker & Gaffan 1998*b*). Indeed, following training on conditional learning tasks, as many as 50% of the neurons in the lateral PFC show conjunctive tuning for learned associations between cues, voluntary actions and rewards.

For example, Watanabe (1990, 1992) trained monkeys to perform tasks in which visual and auditory cues signalled, in different trials, whether a reward would or would not be delivered. The majority of lateral PFC neurons were found to reflect the association between a cue and a reward. A given neuron might be activated by a cue, but only when it signalled a reward. By contrast, another

neuron might be activated only by a cue that signalled ‘no reward’. In our own experiments, we have trained monkeys to associate, in different blocks of trials, each of two cue objects with an eye saccade to the right or to the left (Asaad *et al.* 1998). We found that the activity of 44% of randomly selected lateral PFC neurons reflected associations between objects and the saccades that they instructed (figure 8). Other neurons had activity that reflected the cues or the saccades alone, but they were fewer in number. Fuster *et al.* (2000) have recently shown that PFC neurons can also reflect learned associations between visual and auditory stimuli.

Importantly, these changes do not require a prodigious amount of experience. Changes in PFC neural properties are evident after one day’s experience and can be detected after just a few minutes of training. For example, Bichot *et al.* (1996) studied the FEFs, part of Brodmann’s area 8 that is important for voluntary eye movements. Normally, neurons in this area fire selectively to saccade targets appearing in certain visual field locations. However, when monkeys were trained to search for a target defined by a particular visual attribute (e.g. red), the neurons in the FEFs acquire sensitivity to that attribute (Bichot *et al.*

1996). When monkeys were trained to search for a different target every day, neurons not only discriminated the current target but also distracting stimuli that had been a target on the previous day relative to stimuli that had been targets even earlier (Bichot & Schall 1999). Monkeys were also more likely to make errors in choosing that distracting stimulus. It was as though the previous day's experience left an impression in the brain that influenced neural activity and the monkey's behaviour.

We also observed evidence for rapid plasticity in our study of learning of conditional object-saccade associations in the PFC (Asaad *et al.* 1998). Initially, the monkeys chose their responses at random, but learned the correct cue-response pairing over a few (5–15) trials. As they learned the association, neural activity representing the forthcoming saccadic response appeared progressively earlier in successive trials. In other words, the initiation of response-related delay activity gradually shifted with learning—from a point in time just before the execution of the response and reward delivery to an earlier point in time, nearly coincident with the presentation of the cue.

Further support for a role for PFC neurons in representing task demands comes from training monkeys to alternate between different task rules. This adds another level of complexity beyond the conditional tasks described above. Now, there is more than one rule assigned to each cue and another cue tells the monkey which rule to use in a given trial. For example, following a given cue, monkeys can learn to direct a response to either the cue's location (spatial matching rule) or an alternative location associated with the cue (associative rule), depending on which rule is currently in effect. When tested in this fashion, many PFC neurons show rule-specific activity. For example, a PFC neuron might respond to a given visual cue when the monkey is using an associative rule, but exhibit weak or no activity under identical sensory and attentional conditions that differed only in that the monkey was using a spatial rule instead (White & Wise 1999; Asaad *et al.* 2000). Also, when monkeys switch between different tasks, many PFC neurons show shifts in baseline activity that communicates which task is currently being performed (Asaad *et al.* 2000; figure 9). Such effects have been found in the PFC for associative versus spatial rules, for object matching versus spatial matching versus associative rules, and for shape matching versus object matching rules (Hoshi *et al.* 1998; White & Wise 1999; Asaad *et al.* 2000).

In all of these cases, however, the rules are relatively literal or concrete. A certain cue, or set of cues, always signals a specific response. As noted, knowledge from our past experiences can be applied to a wider range of future circumstances if we abstract general principles or rules rather than specific cue-response contingencies. The ability of the PFC to represent abstract rules, those not tied to specific stimuli or actions, was recently addressed in our laboratory (Wallis *et al.* 2001).

Monkeys were trained to use two abstract rules: 'match' versus 'non-match'. They faced a computer screen and viewed two successively presented pictures. If the match rule was in effect, the monkeys released the lever if the pictures were the same and continued to hold the lever if the pictures were different. If the non-match rule was in effect, the reverse was true; monkeys released if the pic-

tures were different and held if they were the same. The rule was randomly instructed in each trial by the presentation of a cue at the same time as the first picture was presented. To disambiguate neuronal responses to the physical properties of the cue from responses to the rule that the cues instructed, cues signifying the same rule were taken from different modalities, while cues signifying different rules were taken from the same modality (figure 10). The monkeys could perform this task well above chance levels even when they were seeing a stimulus for the very first time. This 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 11 shows a good example of a rule-selective neuron that exhibited greater activity when the match rule was in effect than when the non-match rule had been indicated. This activity cannot be explained by the physical properties of the cue or the picture, since activity was the same regardless of which cue was used to instruct the monkey, and regardless of which picture the monkey was remembering. It cannot be related to the upcoming response since the monkey did not know whether the second-presented picture would require a response. Nor could it be related to differences in reward expectation, since 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 is that the differences in activity reflected the abstract rule that the monkey was currently using to guide its behaviour.

What function does the ability to abstract a rule serve? It is a form of generalization that permits a shortcut in learning, thereby 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 stimuli). For example, the monkey might learn that whenever the chef is presented with a drop of juice, then at the test phase the correct response is to choose the chef. But 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 the monkey 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.

This shortcut that abstraction of the rule permits is also reflected in the neuronal encoding. 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 non-match rule when the cues are presented in the auditory

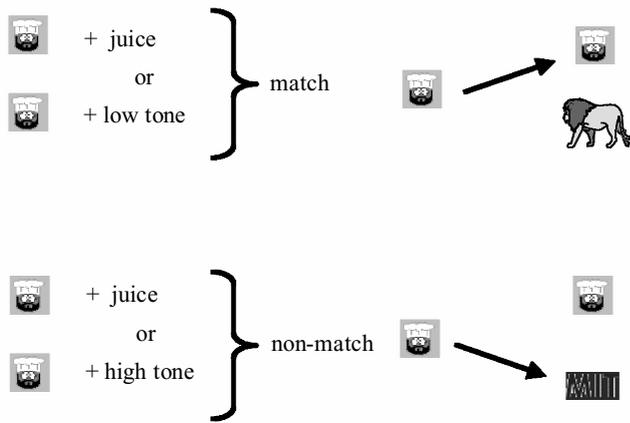


Figure 10. Schematic diagram of the abstract rules task. Monkeys switched between choosing a test stimulus that did or did not match the sample depending on whether the match or non-match rule was in effect. The cues that signalled these rules are shown on the left.

modality, and one encoding this information in the taste modality. But such a solution is computationally expensive, since if a third modality was introduced a third population of cells would be required. It is more efficient to abstract a 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 loss of flexibility that is observed after prefrontal damage in both monkeys and humans. It is not inconsistent with studies emphasizing the role of the PFC in working memory or planning, but indicates that an important component of these processes might be the use of behaviour-guiding rules.

5. PREFRONTAL CORTEX AND COGNITIVE CONTROL

The results presented above suggest that PFC activity reflects categories and rules. This seems consistent with conjectures that a cardinal PFC function may be the acquisition and representation of the formal demands of tasks, the guiding concepts and principles that provide a foundation for complex, intelligent behaviour. In order to understand this, we must turn to theories of cognitive control—the ability of the brain to coordinate processing among its millions of neurons in order to direct them toward future goals.

(a) *Controlled versus automatic behaviours*

In order to understand what we mean by ‘cognitive control’, it is important to understand the distinction between controlled and automatic behaviours. Much of our behaviour is automatic, that is, direct reactions to our immediate environment that do not tax our attention. For example, if someone suddenly throws a baseball at your face, you might reflexively duck. You may not have willed this behaviour; it just seems to happen. Many such reflexive, automatic processes are ‘wired’ into our nervous systems by evolution. However, others can be acquired through a great deal of experience, as learning mechanisms gradually imprint highly familiar behaviour. If you are walking a highly familiar route and traffic is light, you may traverse a great distance (and even negotiate turns) with little

awareness of having done so. In these cases, your behaviour is driven in a ‘bottom-up’ fashion: largely determined by stimuli in the immediate environment and their ability to trigger behaviours with which they are strongly associated. In neural terms, they are dependent on well-established neural pathways waiting to be fired off by the correct input.

However, if something unexpected happens on your walk, you need to ‘take charge’ of your actions. You pay attention to your surroundings and try to anticipate and accommodate the action of others; you may even decide to take an alternative route. In this case, your behaviour is not governed by simple input–output, stimulus–response relationships. You use your knowledge of the world: the current objective (arriving at work intact) and results from previous experiences to weigh the alternatives and consequences. Because these behaviours tax your attention and seem to be driven by ‘internal’ information (knowledge about unseen goals and how to achieve them) and thus are initiated by us and not by the environment, they are referred to as ‘controlled’. The same basic sensory, memory and motor processes that mediated automatic behaviour can be engaged. However, now they are not simply triggered by the environment, they are shaped and controlled in a top-down fashion, by our knowledge of how the world works. The observations that PFC neurons encode acquired information about task contingencies, categories and rules suggest it as a source of top-down signals.

The observation that humans with PFC damage seem impulsive makes sense in light of this distinction between controlled and automatic behaviours. Without the PFC to provide top-down signals about expectations of goals and required behaviours, the patient simply reacts to their environment with whatever behaviours are strongly associated with the cues that are immediately present. Shallice & Burgess (1991) examined this by using a ‘shopping test’. They described frontal lobe-damaged patients who are able to execute simple routines in which clear sensory cues could elicit a familiar action (e.g. buy a loaf of bread). However, they were unable to carry out an errand that involved organizing a series of such routines because they kept going ‘off task’. They would, for example, enter shops that were irrelevant to the errand, just because they happened to be passing them. Another example is utilization behaviour. A patient with PFC damage will impulsively use items placed in front of them such as a comb or, in (hopefully) one case, a urinal. It seems that the basic elements of behaviour are intact, but that the patients override prepotent, reflexive, responses to coordinate behaviour in accord with an unseen goal.

A classic test of the ability to learn and follow goal-orientated rules is the Wisconsin Card Sorting Test (Milner 1963). Subjects are instructed to sort cards according to the shape, colour or number of symbols appearing on them. They start with one rule (e.g. colour) and, once that is acquired, the rule changes until all of the cards have been sorted using all possible rules. Normal humans have little difficulty with this task. By contrast, humans with prefrontal damage can learn the first sorting criterion (a relatively simple mapping between a stimulus attribute and a response) but then are unable to escape it; they cannot override the previous behaviour and do not

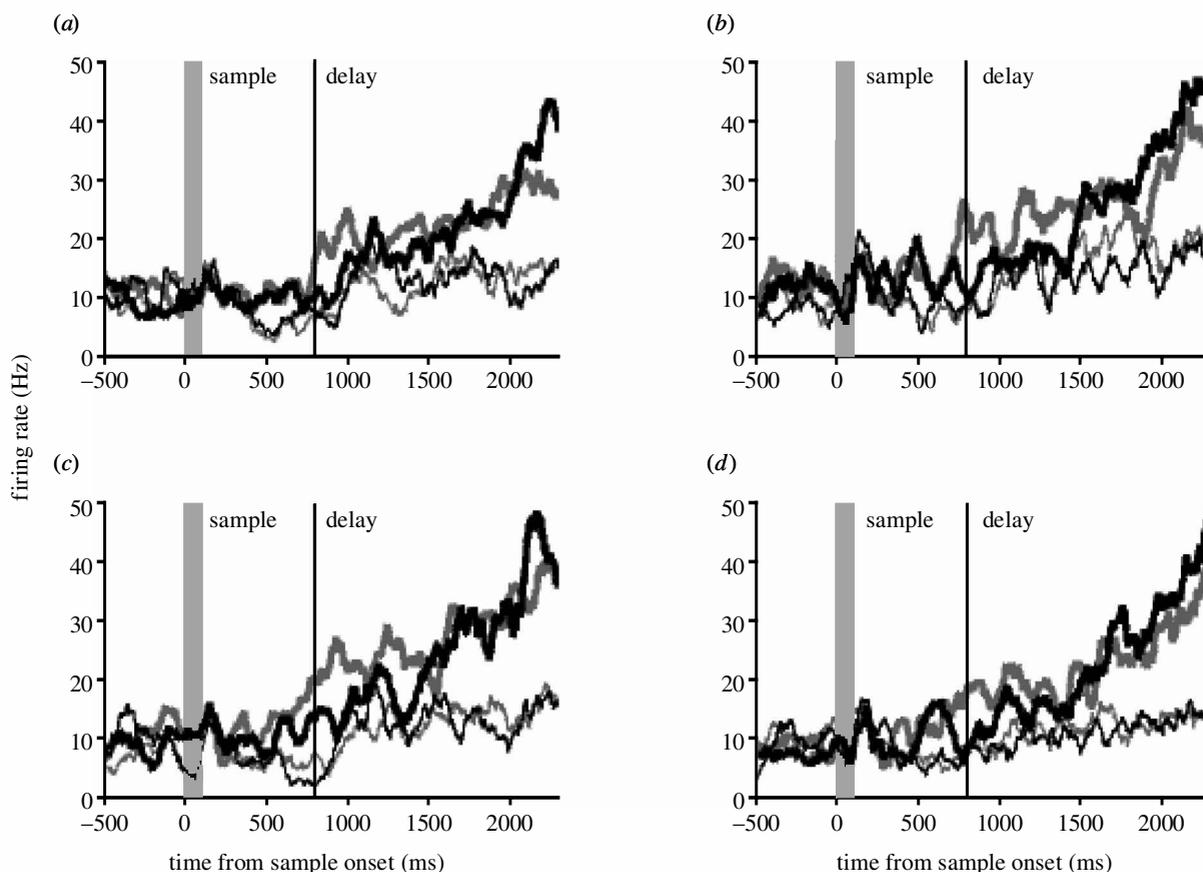


Figure 11. A neuron exhibiting rule selectivity. The neuron shows greater activity during *match* trials, regardless of which cue signified the rule or which object was remembered. (a) Sample object 1; (b) sample object 2; (c) sample object 3; (d) sample object 4. The vertical grey bar marks the cue epoch. Match (juice) and match (low tone) are represented by thick black and grey bars, respectively. Non-match (juice) and non-match (high tone) are represented by thin black and grey bars, respectively.

realize that when the rule changes and they must learn a new one (Milner 1963). Monkeys with PFC lesions are impaired in similar tasks (Dias *et al.* 1996). PFC damage in humans or monkeys, or disconnecting the PFC from its sensory inputs in monkeys, also produces deficits in a standard test of rule learning, the aforementioned conditional learning tasks (Petrides 1985*a,b*; Gaffan & Harrison 1988, 1991; Murray & Wise 1997; Parker & Gaffan 1998*a*; Murray *et al.* 2000).

(b) *A theory of the prefrontal cortex and cognitive control*

In summary, we have seen that PFC damage seems to disrupt cognitive control, the ability of animals to direct action toward unseen goals, and leaves them at the mercy of the environment. We discussed neurophysiological studies that indicate that the PFC represents task-relevant knowledge such as categories and rules. How these properties are acquired, and how they are used for cognitive control, have been addressed in a model of PFC by Miller and Cohen (Miller & Cohen 2001).

In this view, the ability to form representations of the formal demands of behaviour stems from the position of the PFC at the top of the cortical hierarchy (Fuster 1995). The PFC is a network of neural circuits that is interconnected with cortical regions that analyse virtually all types of sensory inputs, and with regions involved in generating motor outputs. It is also in direct contact with a wide array

of subcortical structures that process, among other things, 'internal' information such as motivational state. The PFC thus provides a venue in which information from distant brain systems can interact through relatively local circuitry. During learning, reward-related signals could act on the PFC to strengthen pathways—the associative links—between the neurons that processed the information that led to a reward. As a result, the PFC rapidly constructs a pattern of activity that represents goals and the means to achieve them. This, in essence, is a representation of the logic of the task, a task model that reflects the constellation of relevant information and their interrelations. In neural terms, this could amount to a 'map' of the neural pathways in the brain that is needed to solve the task. Cognitive control results from the PFC sending excitatory signals from this representation back to the brain structures that provide the PFC with input. These signals arise from the ability of many PFC neurons to sustain their activity. These chronic signals reflecting the task demands can enhance the activity of neurons that process task-relevant information (that match the model) in other brain systems and thereby select the forebrain neural pathways that are needed to solve the task at hand.

To understand how this selection takes place, consider visual attention. In the visual system, neurons processing different aspects of the visual scene compete with each other for activation, by mutually inhibiting one another. This is thought to be important for enhancing contrast

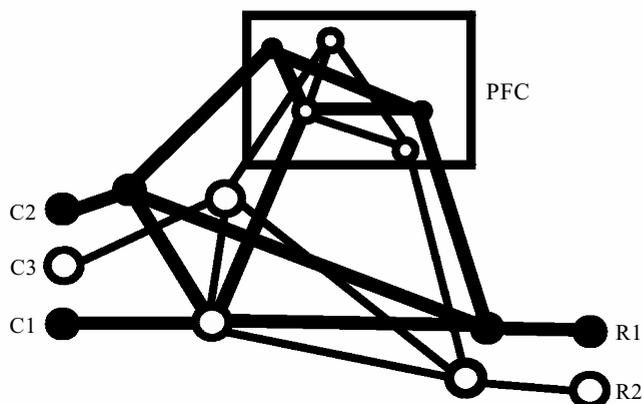


Figure 12. Schematic diagram of a posited role for the PFC in cognitive control. Information on sensory inputs, current motivational state, memories, etc. (e.g. 'cues' such as C1, C2 and C3) as well as information about behaviour (e.g. 'responses' such as R1 and R2) is indicated. Reward signals foster the formation of a task model, a neural representation that reflects the learned associations between task-relevant information. A subset of the information (e.g. C1 and C2) can then evoke the entire model, including information about the appropriate response (e.g. R1). Excitatory signals from the PFC feed back to other brain systems to enable task-relevant neural pathways. Thick lines indicate activated pathways, thin lines indicate inactive pathways.

and separating a figure from its background. The neurons that 'win' the competition and remain active are those that incur a higher level of activity. The biased competition model posits that visual attention exploits this circuitry (Desimone & Duncan 1995). In voluntary shifts of attention, a competitive advantage comes from excitatory signals (thought to originate from the PFC) that represent the expected stimulus. These excitatory signals enhance the activity of neurons in the visual cortex that process that stimulus and, by virtue of the mutual inhibition, suppress activity of neurons processing other stimuli. This notion of excitatory bias signals that resolve local competition can be extended from visual attention to cognitive control in general (Miller 1999, 2000). By enhancing the activity of neurons representing task-relevant information, those representing irrelevant information are simultaneously suppressed and neural activity is steered down the pathways needed to solve the task at hand.

For an illustrative example of how this might work, consider the cartoon shown in figure 12. Processing units are shown that correspond to cues (C1, C2, C3). They can be thought of as neural representations of sensory events, internal states, stored memories, etc. in corresponding neural systems. Also shown are processing units that correspond to the motor circuits mediating two responses (R1 and R2). We have set up the sort of flexible situation for which the PFC is thought to be important. Namely, one cue (C1) can lead to one of two responses (R1 or R2), depending on the situation (C2 or C3). Imagine that you suddenly decide that you want a beer (and let us consider that to be cue C1). If you are at home (C2), then you get up and get one (R1). But if you are in a pub (C3), you ask for one instead (R2). These conditional associations form the 'if-then' rules that are fundamental building blocks of voluntary behaviour (Passingham 1993). How does the PFC construct these representations?

In an unfamiliar situation, information flows into the PFC relatively unchecked. But then reward-related signals from successful (rewarded) experiences foster the formation of associations between the PFC neurons that had processed the information immediately preceding reward. This signal may be an influx of dopamine from the mid-brain VTA neurons that are sensitive to reward and, through the basal ganglia, influence the PFC (Passingham 1993; Schultz & Dickinson 2000). As this neural ensemble becomes established, it becomes self-reinforcing. It sends signals back to other brain systems, biasing their processing towards matching information and thus refining the inputs to the PFC. As learning proceeds, reward-related signals from VTA neurons appear progressively earlier as they become evoked by the events that first predict reward (Schultz & Montague 1997). Through repeated iterations of this process, more and more task-relevant information is linked into the PFC representation; it 'bootstraps' from direct associations with reward to a multivariate network of associations that can describe a complex, goal-directed task.

Once the PFC representation is established, a subset of the information (such as the cues) can activate the remaining elements (such as the correct response). So, if we want a beer (C1) and we are at home (C2), the corresponding PFC representation containing the correct response (R1) is activated and sustained until the response is executed. The resulting excitatory bias signals from the PFC then feed back to other brain regions, selecting the appropriate pathway needed for the task (e.g. C1-R1). A different pattern of cues (e.g. cues 1 and 3 in figure 12) evokes a different PFC model and a different pattern of bias signals selects other neural pathways (C1-R2). With repeated selection of these pathways, they can become established independently of the PFC. As this happens, the PFC becomes less involved and the behaviour becomes habitual or automatic.

This particular view of PFC function is not without peer or precedent. Fuster first proposed that PFC neurons encode task-relevant contingencies between stimuli and/or responses, particularly when they are separated by gaps in time (as so often happens with extended, goal-directed behaviours) (Fuster 1985). Neurons that explicitly encode task contingencies and rules are used in neural network models of cognitive control by Changeux and Dehaene (Changeux & Dehaene 1993; Dehaene *et al.* 1998). The models of Cohen and colleagues use a layer (thought to correspond to the PFC) that represents task demands or 'context' (Cohen & Servan-Schreiber 1992). Wise *et al.* (1996) proposed that a cardinal PFC function is the acquisition of behaviour-guiding rules. Shimamura independently proposed a role for the PFC directly analogous to the Miller and Cohen model (Shimamura 2000).

Central to our model, and indeed all physiologically inspired models of PFC function, is the ability of PFC neurons to sustain their activity for several seconds in the absence of further stimulation. This is crucial for several reasons. As previously mentioned, gaps in time are an inevitable consequence of extended goal-directed behaviours. Thus, sustained activity allows PFC neurons to learn relationships (associations) between stimuli and/or responses that are separated in time (Fuster 1985). It also allows task rules to be maintained until the task is com-

pleted. But conveying information by sustained activation affords more than the ability to bridge temporal gaps. O'Reilly and Munakata have pointed out that it is an ideal format for transmitting the knowledge needed for cognitive control to other brain systems (O'Reilly & Munakata 2000).

A tenet of modern neuroscience is that long-term storage in the brain depends on strengthening some neural connections and weakening others, that is, by changing synaptic weights. Encoding information in structural changes has obvious advantages for information storage; it allows for very long-term memories. But the resulting memories are relatively inflexible; once a neural circuit is established, it will tend to fire in the same way every time it is triggered. Also, changing the strength of a synapse only affects the neurons that share the synapse. Thus, the information encoded in a pattern of synaptic weights only affects the firing of the neurons that form that particular circuit and is only expressed when that circuit is fired. Cognitive control, however, requires that a given pattern of information (the task demands) affect many brain circuits; it is used to orchestrate processing in many different brain systems. It is apparent, therefore, that the information needs to be encoded in a different format. Sustained activity is such a format. Because information is encoded in a pattern of sustained activity (rather than only in a pattern of synaptic weights), it can be propagated across the brain. Thus, the ability of sustained activity to tonically influence other brain systems is probably important for coordinating diverse processing around a specific goal. It also affords flexibility; if cognitive control stems from a pattern of information maintained in the PFC, changing behaviour is as easy as changing the pattern (O'Reilly & Munakata 2000; Miller & Cohen 2001).

Finally, the central role of sustained activity might explain the severely limited capacity of controlled processes. While we can carry out a number of automatic processes simultaneously, our ability to carry out controlled processes is limited by the low capacity of our attention. If the information for cognitive control is expressed in a unique pattern of ongoing activity distributed across many simultaneously active neurons—a population code—then there will be a natural capacity limitation. Trying to represent more than just a few items at the same time would degrade information because the unique patterns impinging on a given set of neurons would overwrite and interfere with one another.

6. SUMMARY AND CONCLUSIONS

The ability to take charge of one's actions and direct them towards future, unseen goals is called cognitive control. Virtually all theories of cognition posit that cognition depends on functions specialized for the acquisition of information about goals and the means to achieve them. These functions exert a top-down influence on the lower-level automatic processes that mediate sensory analysis, memory storage and motor outputs, orchestrating and directing them toward a given goal.

The PFC, a brain structure that reaches its greatest complexity in the primate brain, seems to have a central role in cognitive control. It has access to, and the means to influence processing in, all major forebrain systems and

can provide a means to synthesize the diverse information related to a given goal. As we have established, PFC neurons seem to have a crucial ability for cognitive control; they convey the knowledge that animals acquire about a given goal-directed task. Their ability to develop abstracted representations frees the organism from specific associations and endows it with the ability to generalize and develop overarching concepts and principles. This ability is consistent with observations of a loss of flexibility after PFC damage and may form a foundation for the complex, intelligent behaviour that is often seen in primates.

This work has been supported by the National Institutes of Health, the RIKEN-MIT Neuroscience Research Center, the Pew Foundation, The McKnight Foundation, The Sloan Foundation and the Whitehall Foundation. The authors thank Marlene Wicherski for her valuable comments on the manuscript.

REFERENCES

- Asaad, W. F., Rainer, G. & Miller, E. K. 1998 Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407.
- Asaad, W. F., Rainer, G. & Miller, E. K. 2000 Task-specific activity in the primate prefrontal cortex. *J. Neurophysiol.* **84**, 451–459.
- Baddeley, A. & Della Sala, S. 1996 Working memory and executive control. *Phil. Trans. R. Soc. Lond. B* **351**, 1397–1403.
- Barbas, H. & Pandya, D. 1991 Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In *Frontal lobe function and dysfunction* (ed. H. S. Levin, H. M. Eisenberg & A. L. Benton), pp. 35–58. New York: Oxford University Press.
- Beymer, D. & Poggio, T. 1996 Image representations for visual learning. *Science* **272**, 1905–1909.
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F. & Knauss, K. S. 1988 Conceptual behavior in pigeons: categorization of both familiar and novel examples from four classes of natural categories. *J. Exp. Psych: Anim. Behav. Process.* **14**, 219–234.
- Bichot, N. P. & Schall, J. D. 1999 Effects of similarity and history on neural mechanisms of visual selection. *Nat. Neurosci.* **2**, 549–554.
- Bichot, N. P., Schall, J. D. & Thompson, K. G. 1996 Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* **381**, 697–699.
- Changeux, J. P. & Dehaene, S. 1993 Formal models for cognitive functions associated with the prefrontal cortex. In *Exploring brain functions: models in neuroscience* (ed. T. A. Poggio & D. A. Glaser). Chichester: Wiley.
- Cohen, J. D. & Servan-Schreiber, D. 1992 Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol. Rev.* **99**, 45–77.
- Dehaene, S., Kerszeberg, M. & Changeux, J. P. 1998 A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl Acad. Sci. USA* **95**, 14 529–14 534.
- Desimone, R. & Duncan, J. 1995 Neural mechanisms of selective visual attention. *A. Rev. Neurosci.* **18**, 193–222.
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. 1984 Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062.
- Dias, R., Robbins, T. W. & Roberts, A. C. 1996 Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behav. Neurosci.* **110**, 872–886.

- Dickinson, A. 1980 *Contemporary animal learning theory*. Cambridge University Press.
- Duncan, J., Emslie, H., Williams, P., Johnson, R. & Freer, C. 1996 Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cogn. Psychol.* **30**, 257–303.
- Eacott, M. J. & Gaffan, D. 1992 Inferotemporal–frontal disconnection—the uncinete fascicle and visual associative learning in monkeys. *Eur. J. Neurosci.* **4**, 1320–1332.
- Fabre-Thorpe, M., Richard, G. & Thorpe, S. J. 1998 Rapid categorization of natural images by rhesus monkeys. *NeuroReport* **9**, 303–308.
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. 2001 Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316.
- Fuster, J. M. 1985 The prefrontal cortex, mediator of cross-temporal contingencies. *Hum. Neurobiol.* **4**, 169–179.
- Fuster, J. M. 1989 *The prefrontal cortex*. New York: Raven.
- Fuster, J. M. 1995 *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.
- Fuster, J. M., Bodner, M. & Kroger, J. K. 2000 Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* **405**, 347–351.
- Gaffan, D. & Harrison, S. 1988 Inferotemporal–frontal disconnection and fornix transection in visuomotor conditional learning by monkeys. *Behav. Brain Res.* **31**, 149–163.
- Gaffan, D. & Harrison, S. 1991 Auditory–visual associations, hemispheric specialization and temporal–frontal interaction in the rhesus monkey. *Brain* **114**, 2133–2144.
- Gainotti, G. 2000 What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex* **36**, 539–559.
- Goldman-Rakic, P. S. 1987 Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In *Handbook of physiology: the nervous system* (ed. F. Plum), pp. 373–417. Bethesda, MD: American Physiology Society.
- Gross, C. G. 1973 Visual functions of inferotemporal cortex. In *Handbook of sensory physiology* (ed. R. Jung). Berlin: Springer.
- Hoshi, E., Shima, K. & Tanji, J. 1998 Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *J. Neurophysiol.* **80**, 3392–3397.
- Lisker, L. & Abramson, A. 1970 *The voicing dimension: some experiments in comparing phonetics*. Prague: Academia.
- Miller, E. K. 1999 The prefrontal cortex: complex neural properties for complex behavior. *Neuron* **22**, 15–17.
- Miller, E. K. 2000 The neural basis of top-down control of visual attention in the prefrontal cortex. In *Attention and performance 18* (ed. S. Monsell & J. Driver). Cambridge, MA: MIT Press.
- Miller, E. K. & Cohen, J. D. 2001 An integrative theory of prefrontal function. *A. Rev. Neurosci.* **24**, 167–202.
- Milner, B. 1963 Effects of different brain lesions on card sorting. *Arch. Neurol.* **9**, 100–110.
- Mishkin, M. 1982 A memory system in the monkey. *Phil. Trans. R. Soc. Lond. B* **298**, 83–95.
- Murray, E. A., Bussey, T. J. & Wise, S. P. 2000 Role of prefrontal cortex in a network for arbitrary visuomotor mapping. *Exp. Brain Res.* **133**, 114–129.
- Murray, E. A. & Wise, S. P. 1997 Role of the orbitoventral prefrontal cortex in conditional motor learning. *Soc. Neurosci. Abstr.* **27**, 12.
- Nauta, W. J. H. 1971 The problem of the frontal lobe: a reinterpretation. *J. Psychiatr. Res.* **8**, 167–187.
- O'Reilly, R. C. & Munakata, Y. 2000 *Computational explorations in cognitive neuroscience: understanding the mind*. Cambridge, MA: MIT Press.
- Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. 2000 Macaque monkeys categorize images by their ordinal number. *Nature* **404**, 77–80.
- Pandya, D. N. & Barnes, C. L. 1987 Architecture and connections of the frontal lobe. In *The frontal lobes revisited* (ed. E. Perecman), pp. 41–72. New York: The IRBN Press.
- Parker, A. & Gaffan, D. 1998a Memory after frontal/temporal disconnection in monkeys: conditional and non-conditional tasks, unilateral and bilateral frontal lesions. *Neuropsychologia* **36**, 259–271.
- Parker, A. & Gaffan, D. 1998b Memory after frontal/temporal disconnection in monkeys: conditional and non-conditional tasks, unilateral and bilateral frontal lesions. *Neuropsychologia* **36**, 259–271.
- Passingham, R. 1993 *The frontal lobes and voluntary action*. Oxford University Press.
- Petrides, M. 1985a Deficits in non-spatial conditional associative learning after periarculate lesions in the monkey. *Behav. Brain Res.* **16**, 95–101.
- Petrides, M. 1985b Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* **23**, 601–614.
- Petrides, M. 1990 Nonspatial conditional learning impaired in patients with unilateral frontal but not unilateral temporal lobe excisions. *Neuropsychologia* **28**, 137–149.
- Rainer, G., Asaad, W. F. & Miller, E. K. 1998a Memory fields of neurons in the primate prefrontal cortex. *Proc. Natl Acad. Sci. USA* **95** (15), 15 008–15 013.
- Rainer, G., Asaad, W. F. & Miller, E. K. 1998b Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393**, 577–579.
- Rainer, G., Rao, S. C. & Miller, E. K. 1999 Prospective coding for objects in the primate prefrontal cortex. *J. Neurosci.* **19**, 5493–5505.
- Rao, S. C., Rainer, G. & Miller, E. K. 1997 Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824.
- Roberts, W. A. & Mazmanian, D. S. 1988 Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J. Exp. Psychol. Anim. Behav. Proc.* **14**, 247–260.
- Schultz, W. & Dickinson, A. 2000 Neuronal coding of prediction errors. *A. Rev. Neurosci.* **23**, 473–500.
- Schultz, W. & Montague, P. R. 1997 A neural substrate of prediction and reward. *Science* **275**, 1593–1599.
- Shallice, T. & Burgess, P. W. 1991 Deficits in strategy application following frontal lobe damage in man. *Brain* **114**, 727–741.
- Shelton, C. 2000 Morphable surface models. *Int. J. Comp. Vis.* **38**, 75–91.
- Shimamura, A. P. 2000 The role of the prefrontal cortex in dynamic filtering. *Psychobiology* **28**, 207–218.
- Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. 1991 Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* **66**, 170–189.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. & Miyashita, Y. 1999 Top-down signal from prefrontal cortex in executive control of memory retrieval (see comments). *Nature* **401**, 699–703.
- Vaadia, E., Benson, D. A., Hienz, R. D. & Goldstein Jr, M. H. 1986 Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. *J. Neurophysiol.* **56**, 934–952.
- Vogels, R. 1999a Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur. J. Neurosci.* **11**, 1223–1238.
- Vogels, R. 1999b Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur. J. Neurosci.* **11**, 1239–1255.
- Wallis, J. D., Anderson, K. C. & Miller, E. K. 2001 Single neu-

- rons in the prefrontal cortex encode abstract rules. *Nature* **411**, 953–956.
- Watanabe, M. 1990 Prefrontal unit activity during associative learning in the monkey. *Exp. Brain Res.* **80**, 296–309.
- Watanabe, M. 1992 Frontal units of the monkey coding the associative significance of visual and auditory stimuli. *Exp. Brain Res.* **89**, 233–247.
- White, I. M. & Wise, S. P. 1999 Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* **126**, 315–335.
- Wise, S. P., Murray, E. A. & Gerfen, C. R. 1996 The frontal–basal ganglia system in primates. *Crit. Rev. Neurobiol.* **10**, 317–356.
- Wytenbach, R. A., May, M. L. & Hoy, R. R. 1996 Categorical perception of sound frequency by crickets. *Science* **273**, 1542–1544.
- Young, M. E. & Wasserman, E. A. 1997 Entropy detection by pigeons: response to mixed visual displays after same-different discrimination training. *J. Exp. Psychol. Anim. Behav. Process.* **23**, 157–170.

GLOSSARY

- DMC: delayed match-to-category
FEF: frontal eye field
ITC: inferior temporal cortex
ITI: inter-trial interval
PFC: prefrontal cortex
VTA: ventral tegmental area