Categorical Representation of Visual Stimuli in the Primate Prefrontal and Inferior Temporal Cortices

David J. Freedman

B.A., Brain and Cognitive Sciences
University of Rochester, Rochester, NY 1997

Submitted to the Department of Brain & Cognitive Sciences in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Systems Neuroscience

at the

Massachusetts Institute of Technology

February, 2002

Signature of Author: ____________________________

Department of Brain & Cognitive Sciences
October 19, 2001

Certified by: ____________________________

Earl K. Miller
Associate Professor of Neuroscience
Thesis Supervisor

Accepted by: ____________________________

Earl K. Miller
Associate Professor of Neuroscience
Chairman, Department Graduate Committee

Categorical Representation of Visual Stimuli in the Primate Prefrontal and Inferior Temporal Cortices

by

David J. Freedman

Submitted to the Department of Brain and Cognitive Sciences
on September 20, 2001 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Systems Neuroscience

ABSTRACT

The ability to group stimuli into meaningful categories is a fundamental cognitive process though little is known its neuronal basis. To address this issue, we trained monkeys to perform a categorization task in which they classified visual stimuli into well defined categories that were separated by a "category-boundary". We recorded from neurons in the prefrontal (PFC) and inferior temporal (ITC) cortices during task performance. This allowed the neuronal representation of category membership and stimulus shape to be independently examined.

In the first experiment, monkeys were trained to classify the set of morphed stimuli into two categories, "cats" and "dogs". Recordings from the PFC of two monkeys revealed a large population of categorically tuned neurons. Their activity made sharp distinctions between categories, even for stimuli that were visually similar but from different classes. Likewise, these neurons responded similarly to stimuli from the same category even if they were visually dissimilar from one another.

In the second experiment, one of the monkeys used in the first experiment was retrained to classify the same stimuli into three new categories. PFC recordings collected after the monkeys were retrained revealed that the population of neurons reflected the three new categories but not the previous (now irrelevant) two categories.

In the third experiment, we recorded from neurons in the ITC while a monkey performed the two-category "cat" vs. "dog" task. There were several differences between ITC and PFC neuronal properties. Firstly, a greater proportion of ITC neurons were only stimulus selective but not category tuned. Secondly, while many PFC neurons displayed category tuning that persisted into the memory delay, such tuning in the ITC was primarily observed during stimulus presentation. Thirdly, whereas many PFC neurons reflected the monkeys' decisions about whether a stimulus indicated a behavioral response, most ITC neurons conveyed information about the visual stimuli only, and not about the monkey's task-related decisions.

In conclusion, our results suggest that neurons in the PFC and ITC can convey information about the category of visual stimuli. The differences in neuronal responses between the ITC and PFC support the hypothesis that the ITC plays an important role in object recognition and visual learning while the PFC is more involved in cognitive functions related to executive control.

Thesis Supervisor: Earl K. Miller
Title: Associate Professor of Neuroscience

Contents

Experiments:

Vita

1975	Born New Haven, CT
1997	B.A. Brain and Cognitive Sciences, University of Rochester
2001	Ph.D. Systems Neuroscience, Department of Brain and Cognitive Sciences, M.I.T.
2001-	Post-doctoral Associate, Center for Learning and Memory, M.I.T.

Acknowledgements

These past four years at M.I.T. have far exceeded my previous expectations of graduate school. Earl Miller has provided an exceptional environment for research in his laboratory and I have been, time and time again, so impressed by his generosity and scientific integrity and always appreciative of his friendship. It is a great tribute to Earl that most of his students (myself included), both undergraduate and graduate, can not seem to bring themselves to leave his laboratory after graduation. It seems that we've had a hard time finding somewhere else that compares.

I never would have found myself here at M.I.T. without the influence of my mentors during my undergraduate years at the University of Rochester, particularly Walt Makous and David Williams. Through their enthusiasm and excellent teaching, they sparked my interest in the fields of visual perception and neuroscience and encouraged me to get involved in research and apply to graduate school. For this, I am forever grateful.

I would especially like to thank Max Riesenhuber and Tommy Poggio, my collaborators on this project since it began in 1998. We have worked together closely for the past three years, and their enthusiasm and approach to brain research have been powerful driving forces both for me and this project.

When I first joined the lab, Wael Asaad and Gregor Rainer taught me many experimental techniques and set an example that I still strive to emulate. I also thank the other members of Earl's lab (Kathy Anderson, Wael Asaad, Mark Histed, Cindy Kiddoo, Emily Le, Pamela Meadors, Michelle Machon, Andreas Nieder, Anitha Pasupathy, Gregor Rainer, Jon Wallis, Melissa Warden, and Richard Wehby) for both their camaraderie and constant criticisms that have made the many late nights, weekends and holidays spent here in the lab so enjoyable and intellectually stimulating.

I have benefited greatly from the guidance and wisdom of the many faculty members here at M.I.T., particularly those on my thesis committee (Tommy Poggio, Pawan Sinha, Earl Miller and Nancy Kanwisher). I thank them for their time and effort.

It is a special privilege to have Charlie Gross on my thesis committee as well. He was the first to shed light on the neuronal representations of complex visual stimuli and his research to this day continues in that pioneering spirit.

Of course, none of this would have been possible without the love, support and encouragement from my family and friends. Thanks especially to my girlfriend, Constance Orlins, for her love and support and to my roommate and longtime friend Chris "Cricket" Collier for being like the brother I never had. Most of all, I am forever grateful to my first teachers, my parents, for their love, support and encouragement. To them, this thesis is dedicated. However, I will certainly understand if they do not feel compelled to read the whole thing.

INTRODUCTION

As we interact with the world around us, the brain is faced with the challenge of making sense of a continuous stream of sensory inputs. After constructing useful perceptions of the physical features of our environment, the brain has to ascribe meaning to these percepts in order to plan successful behavior. To make matters worse, many stimuli represent objects or situations that require immediate attention such as an approaching predator or a friend's call for help. Hence, categorical decisions about the meaning of stimuli must be made quickly, reliably and in the context of the current behavioral situation. Though parsing stimuli into meaningful categories is a fundamental cognitive process, very little is known about its neuronal basis.

Categorization is not an ability limited to humans. In fact, categorical perception has been demonstrated in a variety of species including: insects, frogs, birds and monkeys (Ehret, 1987). In its simplest form, perceptual categorization occurs when a continuous set of stimuli is divided into two discrete subsets, separated by a sharp "category boundary"; stimuli from different subsets would elicit different perceptual or behavioral responses while stimuli from the same subset would evoke similar responses. For example, Wyttenbach et al., (1996) observed that crickets are attracted to sounds with a frequency less than 16kHz but are repelled by frequencies greater than 16kHz. A sharp "category boundary" was seen at 16kHz; the crickets were attracted or repelled by stimuli that were slightly less or greater than 16kHz, respectively. All crickets, regardless of their past experiences, will exhibit this behavior which suggests that this categorical phenomenon has genetic rather than experience-dependent origins.

In more advanced animals, perceptual categories can be shaped by learning. We are not born with knowledge of categories like "chair", "fruit" and "vehicle". Rather, the ability to recognize these classes is acquired through experience. Furthermore, once we are familiar with a perceptual category, we can correctly classify novel stimuli from that class based on their physical features and similarity to other category members. At the other end of the spectrum are conceptual categories usually attributed only to humans. Abstract categories such as "right" and "wrong" are separated by "fuzzy" category boundaries which are not tied to specific stimuli but rather, can be applied in many situations and contexts (Lakoff, 1987). Abstract categories are shaped by learning. For example, one's concept of "right" and "wrong" develops through one's personal experiences, education and culture.

Logical places to look for the neuronal mechanisms of category learning are brain areas that are involved in object recognition and perceptual learning. Hence, the prefrontal (PFC) and inferior temporal cortices (ITC) are likely candidates. Damage to the inferior temporal cortex in both humans and monkeys causes profound deficits in visual discrimination, object recognition, perceptual leaning (Kluver and Bucy, 1938, 1939; Blum et al., 1950; Mishkin, 1954; Mishkin and Pribram, 1954; Mishkin, 1966) and even category-specific agnosias, disorders in which the ability to recognize stimuli from a particular category (most notably for faces) is selectively impaired (Damasio et al., 1982; De Renzi, 2000). In addition, neurons in the ITC respond to complex visual stimuli (Gross, 1972; Desimone et al., 1984; Perret et al., 1992; Tanaka et al., 1991; Kobatake and Tanaka 1994) and their activity can reflect stimulus associations acquired though visual learning (Miyashita et al., 1998) . By contrast with the ITC, prefrontal cortex damage does not typically cause deficits in visual discrimination or object recognition but impairs more cognitive functions such as attention and working memory. PFC lesions or reversible inactivations cause deficits in working memory, attention and response inhibition (Mishkin, 1957; Gross and Weiskrantz, 1962; Mishkin et al., 1969; Goldman and Rosvold, 1970; Goldman et al., 1971; Passingham, 1975; Mishkin and Manning, 1978; Funahashi et al., 1993;

Dias et al. 1996). Furthermore, the activity of prefrontal neurons during complex behavioral paradigms often reflects the relevant information that must be kept "in mind" to successfully solve the task at hand (Miller, 1999). The aim of the experiments described in this thesis is to determine the respective roles of the PFC and ITC by recording the activity of single neurons while monkeys perform a visual categorization task.

The dorsal and ventral visual streams

The results of anatomical, neuropsychological and neurophysiological studies over the past several decades have led to the prevailing view that different types of visual information are processed in parallel by two cortical streams (Felleman and Van Essen, 1991). A "dorsal stream" projects from cortical layers 4Cα and 4β of the primary visual cortex through the middle temporal (MT) and middle superior temporal (MST) areas and into the posterior parietal cortex (LIP, VIP and 7a). A "ventral stream" proceeds from the 4Cβ layer of V1 onwards to V4 and visual areas in the posterior and anterior inferior temporal cortex. Damage to structures in the dorsal pathway, also referred to as the "where" stream, causes pronounced perceptual deficits in the spatial, motion and high-frequency flicker domains. However, damage to the dorsal pathway does not markedly affect the perception of color or form nor does it cause impairments of fine visual acuity. By contrast to the dorsal stream, damage to the ventral "what" stream impairs visual acuity, form and color vision but spares spatial and motion perception (Maunsell, 1992). The results of neuronal recordings have, for the most part, supported the dissociation of function between the two streams. Dorsal stream neurons show tuning to linear motion in area MT, radial and optic flow motion in MST and spatial location in the posterior parietal lobe (Colby and Goldberg, 1999). Ventral stream recordings reveal selectivity to increasingly complex visual features; optimal stimuli range from simple geometric shapes in V4 and posterior IT to complex shapes such as faces in anterior IT (Tanaka, 1992; Logothetis and Sheinberg, 1996).

Though the concept of two independent parallel streams for visual processing is a simple and attractive model, the dissociation between spatial and form processing in the dorsal and ventral streams does not fully describe their importance for visual perception and visually guided behavior. Lateral connections between areas in the dorsal and ventral streams suggest that their processing of visual information may not be entirely independent of one-another. In fact, this hypothesis has been supported by neuronal recordings in the parietal cortex that revealed a population of neurons that carried information about stimulus shape (Sereno and Maunsell, 1998). Likewise, theoretical studies of inferior temporal neurons' receptive fields have suggested that, though they do not individually carry fine spatial information, detailed information about objects' spatial locations could be extracted from a population of neurons that, individually, show only modest spatial tuning (as is the case for most IT neurons) (Logothetis et al., 1995). In addition to its importance for spatial vision, the parietal cortex seems to be critically involved in planning motor actions (Andersen, 1989). These shortcomings of the form vs. space explanations of the visual streams have provoked re-evaluation of the data resulting in new models of the visuomotor system. As such, Goodale and Milner (1992) have proposed that, instead of form and space, the ventral and dorsal streams may in fact underlie "vision for perception" and "vision for action" respectively.

The prefrontal cortex is directly interconnected with both dorsal and ventral stream areas. The dorsolateral prefrontal cortex (area 46) receives direct projections from the parietal cortex (Barbas and Mesulam, 1981) while the ventrolateral prefrontal cortex (areas 45 and 12) receives inputs from the inferior temporal cortex (Webster et al., 1992). This pattern of connectivity suggested to some that the separate representations of form and space in the ventral and dorsal visual system were mirrored in the ventral and dorsal prefrontal cortex as well. In fact, suggestive evidence for such a dissociation resulted both from neuropsychological studies and neuronal recordings that pointed towards such an organization (Funahashi et al., 1993;Wilson et al., 1993). However, it was subsequently shown that the majority of neurons in the dorsal and ventral prefrontal cortex were selective for both stimulus shape and location (Rao et al., 1997; Rainer et al., 1998) when monkeys were required to use both spatial and form information to solve a task. These results suggest that the function of intrinsic connections within the prefrontal cortex (Barbas and Pandya, 1989) may serve to bring together disparate types of information when such integration is required by the cognitive demands of a behavioral task (Miller, 1999).

The ventral stream and object recognition

The earliest evidence that temporal lobe structures are important for visual recognition came from the work of Brown and Schafer in the late nineteenth century and Kluver and Bucy in the 1930s. They found that large bilateral lesions of the temporal lobes of monkeys (including both cortical and subcortical structures such as the amygdala and hippocampus) produced a "psychic blindness" or visual "agnosia" as well as the loss of emotional responses. Monkeys were able to see and react to stimuli around them though they showed little evidence of recognizing their meaning. Blum, Chow and Pribram in the late 1940s and early 1950s found that the visual and emotional deficits of the Kluver-Bucy syndrome could be fractionated by restricted lesions to the temporal cortex and medial temporal structures, respectively, suggesting that the inferior cortical regions of the temporal lobe were important for visual recognition while medial temporal structures such as the amygdala and hippocampus were more involved in emotional processing (Gross, 1998). It was subsequently discovered by Iwai and Mishkin (1969) that the inferior temporal cortex could be further subdivided into the posterior (PIT or TEO) and anterior regions (AIT or TE); posterior lesions caused deficits that were more perceptual in nature while damage to anterior IT caused more mnemonic deficits such as visual agnosia.

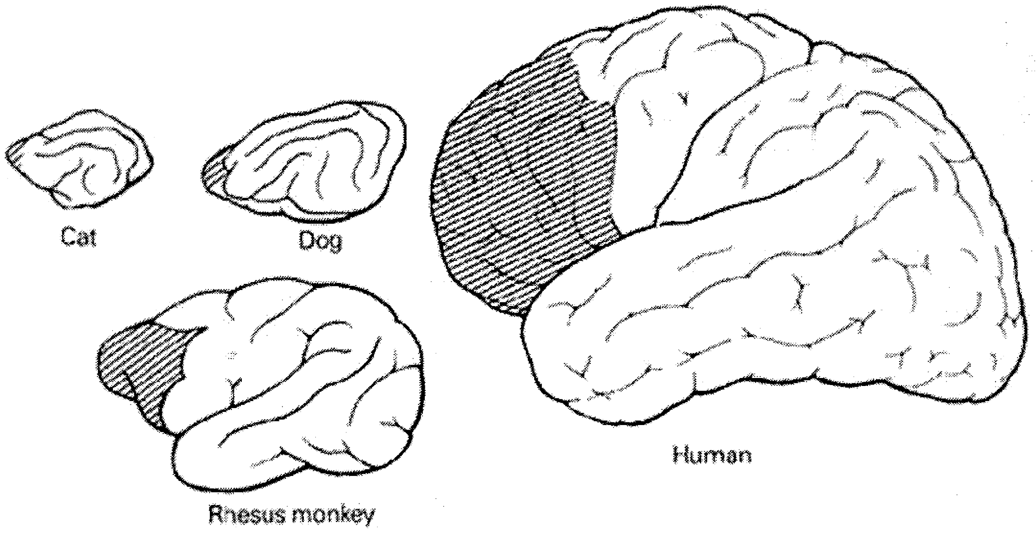
Electrophysiological recordings have further supported the concept of a hierarchy of increasingly complex visual functions throughout the ventral stream. Extrastriate area V4, in the prelunate gyrus, receives direct projections from V1 and V2 and is the primary source of visual input to the inferior temporal cortex. Neurons in V4 have larger receptive fields (~3°) than those in V1 and V2 (< 1°) (Desimone and Ungerleider, 1989) and are preferentially activated by stimuli that are much more complex than the oriented bars that elicit maximal responses from V1 and V2 (Hubel and Wiesel, 1977). Kobatake and Tanaka (1994) found V4 neurons that responded preferentially to complex geometric shapes compared to simple oriented bars and gratings. Pasupathy and Connor (1999) found robust tuning of V4 neurons to contour features, and Gallant et al., (1993, 1996) demonstrated V4 selectivity to curvilinear gratings.

Recordings from the inferior temporal cortex were first conducted by Gross and colleagues in the early 1970s (Gross, 1973). They found that, unlike neurons in the striate and extrastriate cortices, receptive fields of TE neurons were very large (6° PIT, 25° AIT) and usually included the fovea. Furthermore, robust neuronal activity was frequently elicited by complex shapes such as faces, brushes and hands and was often invariant to changes in stimulus size, orientation and position. Careful studies by Tanaka and colleagues (Tanaka, 1996) set out to determine the critical stimulus features that were necessary to evoke a neuronal response from IT neurons. Once a stimulus was found that elicited a strong response from a TE neuron, that stimulus was systematically simplified in search of simpler geometric figures that would still elicit strong neuronal responses. The typical simplified figures that elicited a maximal neuronal response were often complex in shape and were not easily described by a linear combination of the responses to their geometric components. This suggests that the generation of such complex stimulus selectivity is likely a more complicated process than that postulated for the transformation of V1 orientation selectivity into direction or motion selectivity. Indeed, the problem of understanding the mechanisms through which complex shape tuning arises remains as one of the fundamental challenges facing visual neuroscience.

More recent studies have shown that, in addition to the representation of complex stimuli such as faces (Perrett et al., 1992), neurons in anterior IT are involved in visual short term memory (Miller and Desimone, 1994), visual recognition and perceptual awareness (Sheinberg and Logothetis, 1997) and the process of identification learning (described later in this chapter). Hence, there is a fundamental agreement between the results of neuropsychological and neuronal recordings, implying that the ventral stream brain areas are important for a wide range of visual behaviors from simple discrimination and recognition in V4 and PIT to perceptual learning and mnemonic representations in AIT.

The prefrontal cortex: working memory and executive control of behavior

The prefrontal cortex of the rhesus monkey is, phylogenetically, the newest addition to the mammalian neocortex and is thought to be critical for executive cognitive functions such as planning, reasoning and problem solving. Moving up the mammalian evolutionary hierarchy, one notices that the PFC has expanded at an exaggerated pace as compared to the rest of the neocortex (see figure 1), reaching its maximum relative size in humans where it accounts for almost one third of the entire cortical area. This observation alone suggests that the functions of the PFC may underlie the enhanced cognitive abilities that so dramatically differentiate humans from other mammals (Fuster, 1995).



More concrete evidence in support of the PFC's role in higher cognitive functions comes from studies of human and animal subjects with frontal lobe damage. One of the earliest, and most famous cases of a patient with damage to the frontal lobes is that of Phineas Gage in the mid nineteenth century, a foreman for a railroad construction company in Vermont. During a construction accident, the detonation of an explosive charge propelled a pointed iron rod through his jaw and destroyed much of his orbital and medial prefrontal cortices before exiting though the top of his skull. Amazingly, Gage not only survived the accident – he never lost consciousness. He recovered and went on to live for more than a decade after the accident, though he did so with a host of cognitive deficits and personality changes. Having been a respectable and responsible member of society before the accident, he spent the years following his accident as an impulsive, childish, profane man who could not hold a job or successfully plan his actions (Fuster, 1997).

Patterns of deficits similar to Gage's have been observed in a number of studies of frontal lobe damage in both monkeys and humans. Thus, the prefrontal cortex is thought to underlie the supra-sensory cognitive functions that control goal-directed voluntary behavior: attention, inhibition of inappropriate actions, short-term "working" memory and the ability to flexibly modify familiar stimulus-action associations, to name a few. Subjects with PFC damage show profound impairments on behavioral tasks that rely on these cognitive functions. For example, monkeys with prefrontal cortex damage can easily learn that picture 'A' means 'look rightwards to receive a reward' while picture 'B' means 'look leftwards to receive a reward.' After learning this scheme, if the stimulus-reward contingencies are reversed such that 'A' now means 'look left' and 'B' means 'look right,' normal monkeys can quickly learn the new associations while PFC damaged monkeys will continue to perseverate on the old rules as if they had become hard wired to do so (Mishkin et al., 1969; Fuster, 1997).

Insights into the neuronal mechanisms underlying cognitive functions have been achieved by recording the activity of single neurons in the prefrontal cortices of awake monkeys while they performed complex behavioral tasks. During working memory tasks in which a stimulus and response are separated by a delay, many PFC neurons often exhibit stimulus-selective sustained activity during the delay period, often referred to as "delay activity" (Fuster and Alexander, 1971; Fuster et al., 1982). Working memory is much more than just passive maintenance of a sample stimulus; the term "working" memory implies that some active process or transformation is applied to that memory trace. Accordingly, PFC neurons can represent much more than retrospective information about passing stimuli: it has also been demonstrated that PFC neurons' activity can be modulated by shifts in attention (Rainer et al., 1998) and can carry information about an upcoming stimulus or response (Asaad et al., 1998; Rainer et al., 1998; Fuster et al., 2000), the currently relevant task or rule (White and Wise, 1999; Asaad et al., 2000;

Wallis et al., 2001), or the motivational state of the monkey (Hasagawa et al., 2000). The strong correspondence between the cognitive impairments caused by PFC damage and the kinds of information encoded by its neurons provides strong evidence that the PFC does indeed play a critical role in the executive control of behavior and is a likely candidate for involvement in visual categorization and categorical learning.

Neuronal mechanisms for perceptual learning

Visual categorization and identification are closely related perceptual abilities. Often, determining the category of a stimulus relies on first recognizing its identity. Conversely, a prevailing view of many cognitive scientists is that the cognitive mechanisms for categorization are separate from those of the "simpler" process of visual identification (Logothetis and Sheinberg, 1996). Thus, it is unclear whether categorization and other types of perceptual learning such as identification or discrimination are served by the same or different neuronal mechanisms. Though the neuronal processes that underlie visual identification have not yet been fully elucidated, the results of several studies provide a useful context for which to interpret our own findings.

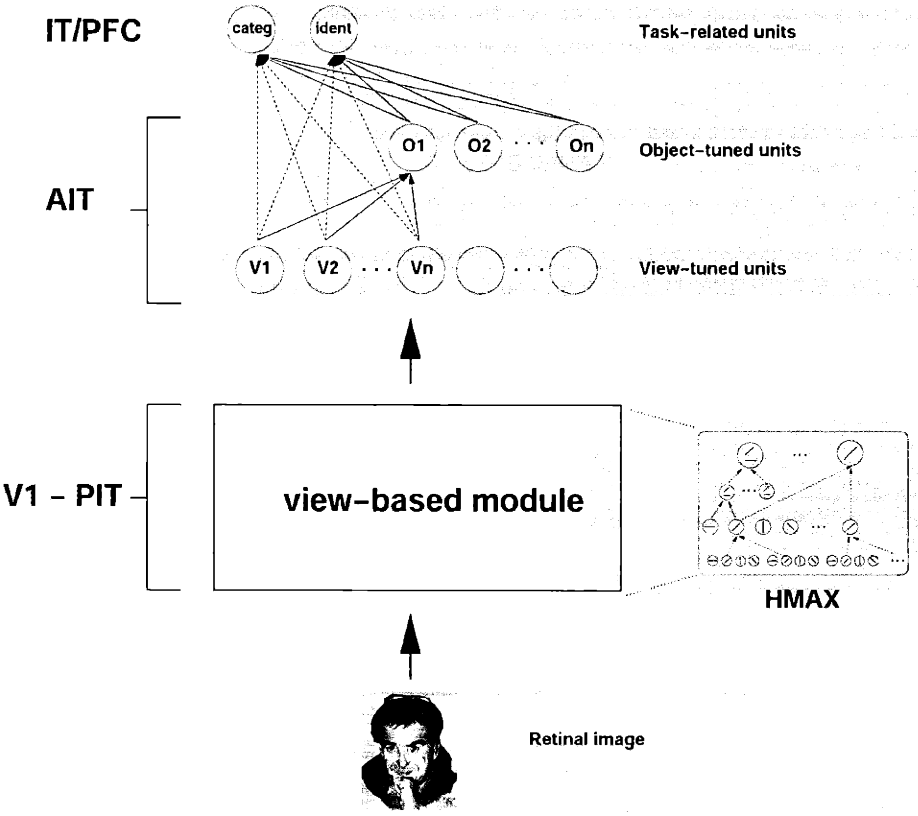
Most previous studies have explored one of two types of perceptual learning: visual identification or visual discrimination learning. In visual identification tasks, subjects are required to search for or identify familiar objects despite stimulus transformations such as rotation, illumination or occlusion. In visual discrimination learning, subjects are extensively trained to make fine perceptual judgments between visually similar stimuli in tasks such as orientation discrimination. Discrimination and identification tasks share a common important feature: task performance improves markedly with practice and experience. The primary goal of these studies was to illuminate the changes in neuronal activity that formed the basis for these perceptual improvements.

A study by Logothetis et al., (1995) trained monkeys to recognize familiar 3D computer generated wireframe objects from multiple viewpoints. Once the monkeys were able to correctly identify objects from all viewpoints around the vertical axis, recordings from the inferior temporal cortex revealed a population of neurons that had become tuned to specific views of the stimuli. Furthermore, no neurons were encountered that were tuned to views of objects that the monkey did not recognize. In addition, a small number of neurons showed a high degree of rotational invariance in that they fired selectively for a specific object and their firing rate was consistent across all viewpoints. The authors concluded that the view and object-tuning observed in IT arose as a result of experience and suggests that the brain may compute the 3D structure of familiar objects by interpolating across neurons tuned to particular static views. View invariant object-tuned neurons carry much more abstract information about stimuli than most other neurons in the visual system. Their response does not carry much information about the actual stimulus that is in view, as visually dissimilar views of the same object would elicit the same response. Rather, they seem better suited to identifying the familiar object in view despite its exact physical features.

Rainer and Miller (2000) explored the role of the prefrontal cortex in object identification by training monkeys to perform a delayed match-to-sample task using both familiar and novel sample stimuli. Varying degrees of random visual noise was parametrically added to the sample stimuli, making the monkeys' recognition task more difficult. Psychophysical analysis of the monkeys' behavior showed that recognition performance decreased more sharply for the novel than familiar stimuli as they were degraded. Degradation of the novel and familiar stimuli had a differential effect on their neuronal representation which was consistent with the monkeys' behavioral responses. There was a progressive waning of neuronal selectivity as the stimuli were degraded that agreed well with the drop in behavioral performance. Furthermore, novel stimuli activated a greater proportion of neurons than familiar stimuli though tuning to familiar stimuli was sharper and less affected by degradation. These results suggest that, while novel stimuli may selectively activate many neurons to some degree, familiar stimuli are encoded by fewer neurons that are more robustly tuned. It is possible that though the process of visual learning, these neurons can carry reliable information about the identity of familiar stimuli despite reductions of stimulus quality or stimulus transformations such as rotation or occlusion.

A recent study has shown that the effects of perceptual learning are not limited to "higher" visual and cognitive brain areas such as the inferior temporal cortex and the prefrontal cortex. Schoups et al., (2001) have demonstrated that neurons in primary visual cortex show a robust enhancement of orientation selectivity after monkeys are extensively trained on an orientation discrimination task. After viewing an oriented grating (always in the same retinotopic location), monkeys had to indicate, with an eye movement to one target or another, whether its orientation was greater-than or less-than a reference orientation. Their results demonstrated that V1 neurons with receptive fields in the region where the trained stimuli were presented showed sharper orientation tuning that did those neurons with receptive fields in other parts of the visual field suggesting that, somewhat surprisingly, even the primary sensory cortical areas are involved in visual learning.

These and other studies have provided substantial evidence that neurons throughout the striate, extrastriate and prefrontal cortices show changes in their tuning properties concomitant with improvements in perceptual performance. The results of the Logothetis et al., (1995) study of object recognition as well as the theoretical model of object recognition and categorization of Riensenhuber and Poggio (2000), illustrated below, are of particular relevance to our studies of categorical learning. The invariance to size and rotation-in-depth observed for some IT neurons suggests that learning to perform an identification task can cause IT neurons to respond similarly to visually dissimilar views of the same object and agrees well with the predictions of the Riesenhuber and Poggio model. Their model also suggests that categorization and identification may share some of the same computational mechanisms that allow for invariance to object transformation such as changes in orientation, size or illumination. A similar neuronal process could then, in principal, serve visual categorization since visually dissimilar stimuli can belong to the same category. Therefore, category learning could produce category-tuned neurons that would be selective for a specific category but would not make distinctions between visually dissimilar members of the same class.



Previous studies of the neuronal representation of categories

In the only published study that investigated the neuronal representation of categories in the monkey, Vogels (1999) trained monkeys to perform two visual categorization tasks and recorded from neurons in the inferior temporal cortex. In this study, monkeys were trained to perform a tree vs. non-tree and a fish vs. non-fish task. Vogels reported finding a subset of IT neurons that were 'category-specific' in that they responded to many of the stimuli from the trained class (photographs of trees or fish) but not to the distracter objects (photos of household objects or scenes containing neither trees nor fish). The responses of these single neurons were not entirely invariant in their response to all members of a category, however. Even the best example of a neuron that preferred stimuli from the category 'tree' responded more strongly to several non-'tree' distrator stimuli than to many 'tree' stimuli. In addition, the firing rates of that neuron varied over a six-fold range depending on which stimulus from the category (trees) was presented.

The results from this study could be explained by the physical similarity of the category members as compared to the distracters. The trees used in this experiment were visually similar and had many features in common (i.e. branches, same orientation, etc…) and more dissimilar from the set of non-tree distracters. The same was true for the set of fish images. Therefore, the apparent category tuning of these IT neurons could be explained by their being tuned to stimulus features that were present in many of the tree stimuli but not the distracters. Hence, the degree to which categorical learning influenced neuronal tuning in this study is not easy to determine.

A recent study by Kreiman et al., (2000) investigated the responses of human medial temporal lobe neurons to images from nine categories. Epileptic patients were implanted with electrodes for the purpose of localizing seizure foci. Neuronal recordings were obtained while they performed a face detection task on images from each of the nine categories. It was determined that the majority of visually responsive neurons were category-specific in that they responded more similarly to stimuli from one category than to those from different categories. This study suffered from a similar problem as the Vogels study. The visual similarity of the photographic stimuli from each of the categories was not controlled in a careful way. Therefore, it is likely that stimuli from the same class (cars) would be visually more similar to members of their class (cars) than to members of other classes (i.e. trees or faces). Again, it is hard to determine whether neurons that responded selectively for one category reflected subjects' category recognition or merely reflected the physical similarity of stimuli within each category and larger differences in stimulus features between categories.

The purpose of experiments reported here

The experiments presented in this thesis were designed to determine the role of PFC and ITC neurons in representing visual categories and, more generally, to better understand their respective roles underlying visually guided behavior. In contrast to previous studies of categorization, our monkeys were trained to categorize computer generated stimuli for which the visual similarity of stimuli within a category and between-categories was parametrically controlled and quantified. By so doing, we ensured that any signals related to the category of stimuli were indeed a product of learning and not due to inherent properties of the stimulus set. Recordings from the PFC revealed many neurons that were categorically tuned: their activity was similar for members of the same category but made sharp distinctions between stimuli from different classes.

In the second experiment, we tested the hypothesis that the apparent category tuning of PFC neurons was a result of experience. We did so by training one of the monkeys from the first experiment to "unlearn" the old categories and to reclassify the same stimuli into three new categories. Following retraining, recordings from the PFC revealed that the neuronal tuning properties had changed: the (now irrelevant) two categories were no longer represented. Rather, it seemed that training the monkey to learn the three new categories had caused them to be selectively represented. Hence, we established that categorical tuning in the PFC was a product of learning.

Past studies have established that the PFC receives direct inputs from the ITC (Webster et al., 1994). Therefore, it is possible that the observed category tuning in the PFC could have been due to categorical inputs from the ITC. Alternatively, the ITC may have conveyed information only about stimulus shape to the PFC, which could then have generated categorical representations from these inputs. In an attempt to distinguish between these possibilities, we conducted ITC recordings in one monkey using the same behavioral paradigm. Our results from this study demonstrate that many ITC neurons did reflect the category of visual stimuli and that the PFC and ITC likely played different roles in solving the categorization task; the ITC seemed primarily involved in rapid visual analysis while PFC responses appeared more suited for using that highly processed visual information to guide behavior.

References

Andersen RA. (1989) Visual and eye movement functions of the posterior parietal cortex. Annu Rev Neurosci 1989;12:377-403.

Asaad WF, Rainer G and Miller EK. (1998) Neural activity in the primate prefrontal cortex during associative learning. N*euron* 21, 1399-407.

Asaad WF, Rainer G and Miller EK. (2000) Task-specific neural activity in the primate prefrontal cortex. J Neurophysiol Jul;84(1):451-9.

Barbas H and Mesulam MM. (1981) Organization of afferent input to subdivisions of area 8 in the rhesus monkey. J Comp Neurol Aug 10;200(3):407-31.

Barbas H and Pandya DN. (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. Comp Neurol Aug 15;286(3):353-75.

Blum JS, Chow KL, and Pribram KH. (1950) A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *J Comp Neurol* 93: 53-100.

Colby CL and Goldberg ME. (1999) Space and attention in parietal cortex. Annu Rev Neurosci;22:319-49.

Damasio AR, Damasio H, and Van Hoesen GW. (1982) Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology* 32: 331-341.

De Renzi E. (2000) Disorders of visual recognition. Semin Neurol;20(4):479-85.

Desimone R, Albright TD, Gross CG, and Bruce C. (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051-2062.

Desimone R and Ungerleider LG. (1989) Neural mechanisms of visual processing in monkeys. in Handbook of Neuropsychology, vol 2. New york: elsevier

Dias R, Robbins TW, and Roberts AC. (1996) Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 380: 69-72.

Ehret, G. (1987) Categorical perception of sound signals: Facts and hypotheses from animal studies. In: *Categorical Perception*. S. Harnad, Ed. Cambridge University Press, United Kingdom, pp 301-331.

Felleman DJ and Van Essen DC. (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex Jan-Feb;1(1):1-47.

Freedman DJ, Riesenhuber M, Poggio T and Miller EK. (2001) Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science,* Jan 12;291:312-316.

Funahashi S, Bruce CJ and Goldman-Rakic PS. (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". J Neurosci. Apr;13(4):1479-97.

Fuster JM (1995) Memory in the cerebral cortex. Cambridge, MA: MIT Press.

Fuster JM, Bodner M and Kroger JK. (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. Nature May 18;405(6784):347-51.

Fuster JM and Alexander GE. (1971) Neuron activity related to short-term memory. Science Aug 13;173(997):652-4.

Fuster JM, Bauer RH and Jervey JP. (1982) Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. Exp Neurol Sep;77(3):679-94.

Gallant JL, Braun J and Van Essen DC. (1993) Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex . Science 259:100-103.

Gallant JL, Connor CE, Rakshit S, Lewis JW, and Van Essen DC. (1996) Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey. J. Neurophysiol. 76: 2718-2739.

Goldman PS and Rosvold HE. (1970) Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Exp Neurol* 27: 291-304.

Goldman PS, Rosvold HE, Vest B, and Galkin TW. (1971) Analysis of the delayed-alternation deficit produced by dorsolateral prefrontal lesions in the rhesus monkey. *J Comp Physiol Psychol* 77: 212-220.

Goodale MA and Milner AD. (1992) Separate visual pathways for perception and action. Trends Neurosci. Jan;15(1):20-5.

Gross, CG (1973) Visual functions of the inferotemporal cortex. In: *Handbook of Sensory Physiology,* Vol. 7/3B. R. Jung, Ed. Springer-Verlag, Berlin, pp. 451-482.

Gross CG (1998) Brain, vision, memory. Cambridge, MA: MIT Press.

Gross CG, Rocha-Miranda CE and Bender DB. (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophsiol 35: 96-111.*

Gross CG and Wieskrantz L. (1962) Evidence for dissociation of impairment on auditory discrimination and delayed response following lateral frontal lesions in monkeys. *Exp Neurol* 5: 453-476.

Hasegawa RP, Blitz AM, Geller NL and Goldberg ME. (2000) Neurons in monkey prefrontal cortex that track past or predict future performance. Science Dec 1;290(5497):1786-9.

Hubel DH and Wiesel TN. (1977) Functional architecture of macaque monkey visual cortex. Proc R Soc Lond B Biol Sci. Jul 28;198(1130):1-59.

Iwai E and Mishkin M. (1969) Further evidence on the locus of the visual area in the temporal lobe of the monkey. Exp Neurol Dec;25(4):585-94.

Kluver H and Bucy L. (1938) An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness." *J Psychol* 5: 33-54.

Kluver H and Bucy L. (1939) Preliminary analysis of functions of the temporal lobes in monkeys. *Arch Neurol Psychiatry* 42: 979-1000.

Kobatake E and Tanaka K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. Neurophysiol Mar;71(3):856-67.

Kreiman G, Koch C, and Fried I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci.* Sep;3(9):946-53.

Lakoff, G. (1987) Women, Fire, and Dangerous Things. University of Chicago Press, Chicago.

Logothetis NK, Pauls J and Poggio T. (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol.* May 1;5(5):552-63.

Logothetis NK and Sheinberg DL. (1996) Visual object recognition. Annu Rev Neurosci; 19:577-621.

Marr D and Ullman S. (1981) Directional selectivity and its use in early visual processing. Proc R Soc Lond B Biol Sci Mar 6;211(1183):151-80.

Maunsell JH. (1992) Functional visual streams. Curr Opin Neurobiol Aug;2(4):506-10

Miller EK. (1999) The prefrontal cortex: complex neural properties for complex behavior. Neuron Jan;22(1):15-7.

Miller EK and Desimone R. (1994) Parallel neuronal mechanisms for short-term memory. Science Jan 28;263(5146):520-2

Mishkin M. (1954) Visual discrimination performance following partial ablations of the temporal lobe. II. Ventral surface vs. hippocampus. *J Comp Physiol Psychol* 47: 187-193.

Mishkin M. (1957) Effects of small frontal lesions on delayed alternation in monkeys. *J Neurophysiol* 20: 615-622.

Mishkin M. (1966) Visual mechanisms beyond the striate cortex. In: *Frontiers in Physiological Psychology.* R.W. Russell, ed. New York: Academic Press.

Mishkin M. and Manning FJ. (1978) Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res* 143: 313-323.

Mishkin M. and Pribram K. (1954) Visual discrimination performance following partial ablations of the temporal lobe. I. Ventral vs. lateral. *J Comp Physiol Psychol* 47: 14-20.

Mishkin M, Vest B, Waxler M, and Rosvold HE. (1969) A re-examination of the effects of frontal lesions on object alternation in monkeys. *Neuropsychologia* 7: 357-364.

Miyashita Y, Kameyama M, Hasegawa I and Fukushima T. (1998) Consolidation of Visual Associative Long-Term Memory in the Temporal Cortex of Primates. Neurobiol Learn Mem Jul;70(1/2):197-211.

Passingham RE. (1975) Delayed matching after selective prefrontal lesions in monkeys (*Macacca Mulatta*). *Brain Res* 92: 89-102.

Pasupathy A and Connor CE. (1999) Responses to contour features in macaque area V4. J Neurophysiol Nov;82(5):2490-502.

Perrett DI, Hietanen JK, Oram MW, and Benson PJ. (1992) Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans Roy Soc Lond B Biol Sci* 335: 23-30.

Rainer G, Asaad WF and Miller EK. (1998) Memory fields of neurons in the primate prefrontal cortex. Proc Natl Acad Sci U S A Dec 8;95(25):15008-13.

Rainer G, Asaad WF and Miller EK. (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature Jun 11;393(6685):577-9.

Rainer G, Rao SC and Miller EK. (1999) Prospective coding for objects in primate prefrontal cortex. *J.Neurosci.* 19, 5493-505.

Rainer G. and Miller EK. (2000) Visual Experience and Prefrontal Object Representation. *Neuron* **27**, 179-189.

Rao SC, Rainer G and Miller EK. (1997) Integration of what and where in the primate prefrontal cortex. Science May 2;276(5313):821-4.

Riesenhuber M and Poggio T. (2000) Models of object recognition. Nat Neurosci 2000 Nov;3 Suppl:1199-204.

Schoups A, Vogels R, Qian N and Orban G. (2001) Practising orientation identification improves orientation coding in V1 neurons. Nature Aug 2;412(6846):549-53.

Sereno AB and Maunsell JH. (1998) Shape selectivity in primate lateral intraparietal cortex. Nature. Oct 1;395(6701):500-3.

Sheinberg DL and Logothetis NK. (1997) The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci U S A Apr 1;94(7):3408-13.

Tanaka K. (1992) Inferotemporal cortex and higher visual functions. Curr Opin Neurobiol. Aug;2(4):502-5.

Tanaka K. (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci;19:109-39.

Tanaka K, Saito H, Fukada Y and Moriya M. (1991) Coding visual images of objects in the inferotemporal cortex of monkeys. *J Neurophysiol* 66: 170-189.

Vogels R. (1999) Categorization of complex visual images by rhesus monkeys. Eur J Neurosci Apr;11(4):1239-55.

Wallis JD, Anderson KC and Miller EK. (2001) Single neurons in prefrontal cortex encode abstract rules. Nature Jun 21;411(6840):953-6.

Webster MJ, Bachevalier J and Ungerleider LG. (1994) Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. Cereb Cortex. Sep-Oct;4(5):470-83.

White IM and Wise SP. (1999) Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res Jun;126(3):315-35.

Wilson FA, Scalaidhe SP, and Goldman-Rakic PS. (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science. Jun 25;260(5116):1955-8.

Wyttenbach RA, May ML, and Hoy RR. (1996) Categorical perception of sound frequency by crickets. Science. Sep 13;273(5281):1542-4.

Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex

David J. Freedman [1,2,5], Maximilian Riesenhuber [3,4,5], Tomaso Poggio [3,4,5],
and Earl K. Miller * [1,2,5]

[1]Center for Learning and Memory, [2]RIKEN-MIT Neuroscience Research Center, [3]Center for Biological and Computational Learning, [4]McGovern Institute for Brain Research, [5]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA 02139, U.S.A.

Science, 2001, Jan 12;Vol. 291, pp. 312-316.

ABSTRACT

The ability to group stimuli into meaningful categories is a fundamental cognitive process. To explore its neural basis, we trained monkeys to categorize computer-generated stimuli as "cats" and "dogs." A morphing system was used to systematically vary stimulus shape and precisely define the category boundary. Neural activity in the lateral prefrontal cortex reflected the category of visual stimuli even when a monkey was retrained with the stimuli assigned to new categories.

Categorization refers to the ability to react similarly to stimuli when they are physically distinct and vice-versa (1). We consider an apple and a billiard ball to be different even though they are similar in shape and sometimes color. Categorization is fundamental; our raw perceptions would be useless without our classification of items as furniture or food. While a great deal is known about the neural analysis of visual features, little is known about the neural basis of the categorical information that gives them meaning.

In advanced animals, most categories are learned. Monkeys can learn to categorize stimuli as animal or non-animal (2), food or non-food (3), tree or non-tree, fish or non-fish (4), and by ordinal number (5). The neural correlate of such perceptual categories might be found in brain areas that process visual form. The inferior temporal (IT) and prefrontal (PF) cortices are likely candidates; their neurons are sensitive to form (6 7 8 9) and they are important for a wide range of visual behaviors (10 11 12).

The hallmark of perceptual categorization is a sharp "boundary" (13). That is, stimuli from different categories that are similar in appearance (e.g., apple/billiard ball) are treated as different while distinct stimuli within the same category (e.g., apple/banana) are treated alike. Presumably, there are neurons that also represent such sharp distinctions. This is difficult to assess with a small subset of a large, amorphous category (e.g., food, human, etc). Because the category boundary is unknown, it is unclear whether neural activity reflects category membership or physical similarity.

We employed a 3D morphing system to generate stimuli that spanned two categories, "cats" and "dogs", by using three species of cats and three breeds of dogs as prototypes (14 15 16). The morphed images were linear combinations of all possible arrangements between them (Fig. 1). By blending different amounts of "cat" and "dog" we could continuously vary shape and precisely define the category boundary (17). Thus, stimuli that were close to but on opposite sides of the boundary could be similar, while stimuli that belong to the same category could be dissimilar (e.g. the "cheetah" and "housecat") (18).

Two monkeys performed a delayed match-to-category task (DMC, Fig. 2a) that required judging whether a sample and test stimulus were from the same category (19). Fig. 2b shows the monkeys' behavior. Performance was high (about 90% correct), even when the samples were close to the category boundary; the monkeys classified dog-like cats (60:40 cat:dog) correctly about 90% of the time, and misclassified them as dogs only 10% of the time, and vice-versa.

We recorded from 395 neurons from the lateral prefrontal cortices of two monkeys (20)(fig. 3a). The majority of neurons were activated during the sample and/or delay interval (253/395, or 64%) (21). They often reflected the sample's category. Nearly one third of responsive neurons (82/253) were category-selective in that they exhibited an overall difference in activity during the sample and/or the delay interval to cats versus dogs. Similar numbers preferred cats (sample interval: 35/65, delay interval: 21/44) as dogs (sample: 30/65, delay: 23/44).

Fig. 3b shows a single neuron. It exhibited greater activity to dogs than cats and responded similarly to samples from the same category regardless of their degree of dogness or catness. Its activity was different to stimuli near the category boundary, the cat-like dogs (60:40 dog:cat) versus the dog-like cats (60:40 cat:dog) (22), but there was no difference in activity between these stimuli and their respective prototypes (the 100% cat or dog) (23). The inset in fig. 3b shows its activity to each of the 54 samples. It exhibited overall greater activity to dogs than cats but small differences within categories. Just a few stimuli elicited activity that was similar to that from the other category. These stimuli were not consistent across different neurons, however. Across the population of neurons, category activity appeared at the start of neural responses to the sample, about 100 msec after sample onset (24).

We examined all stimulus-selective neurons, irrespective of whether or not they were category-selective per se (25). For each neuron, we computed the difference in activity between pairs of samples at different positions along each between-category morph line (fig. 1a). In fig. 4a,b, each neuron's average difference to pairs of samples from the same category (within-category difference, WCD) is plotted against its difference to samples from different categories (between-category difference, BCD). If neurons were not sensitive to categories, these measures should be similar (i.e., BCD/WCD ratios should equal 1 and cluster around the diagonal). Instead, the BCD values are significantly higher than WCD values indicating greater activity differences to samples from different categories, especially during the delay (26).

Fig. 4c,d illustrates the average activity of all stimulus-selective neurons at different morph levels (27). There was a significant difference in activity between the categories (28), but activity was similar at the different morph levels within each category (29), indicating greater sensitivity to stimulus category than identity. In fact, few category-selective neurons conveyed significant identity information (sample interval: 20/65, or 31%, delay interval: 10/44 or 23%)(30). Also, PF neural responses to the test stimulus seemed to reflect category evaluation. Many PF neurons showed enhanced or suppressed activity when the test stimulus matched the category of the sample (112/395, or 28%)(31). Similar effects were reported for identity matches in the PF and inferior temporal cortex (32).

As our monkeys had no experience with cats or dogs prior to training, it seemed likely that the categories were learned. We thus retrained one monkey on the DMC task after defining two new category boundaries that were orthogonal to the original boundary (Fig. 1a). This created three new classes, each containing morphs centered around one cat prototype and one dog prototype (e.g., the cheetah and the "doberman"). Following training, the monkey was able to perform the new 3-category DMC task at >85% correct. We then recorded from 103 PF neurons from the same depths and locations in the PF cortex and using the same samples as the original 2-category task.

Neural responsiveness (58% or 60/103)(33) and stimulus-selectivity (35%, or 21/60)(34) during the 3-category task was similar to that during the 2-category task (64%, or 253/395 and 28%, or 73/253, respectively), but the original categories were no longer reflected in activity (35). Instead, the three new categories were evident in delay activity (36). As during the 2-category task, category information was stronger during the delay (37). This may be because it is relevant for the judgment after the delay. "Prospective activity" is stronger nearer the relevant event (38 39) and appears earlier within a trial as task proficiency increases (40). The monkey was not as proficient at the 3-category task and its reaction times were significantly longer (41).

Categorization of sensory inputs is the nexus between perception and cognition; thoughts and behaviors depend on knowledge of the types of things around us. The sharp transition in neural activity we observed is consistent with a "classical," perceptual category boundary. More conceptual categories can have "fuzzy" boundaries and are unlikely to exhibit such properties (42). Perceptual categorization relies on extraction of the combinations of features defining a category. They were not explicitly instructed, were acquired by training, and were necessarily multivariate abstractions; the categories differed by more than a few simple features. PF activity could have reflected, and/or resulted in, a shifting of attention to those features (43). These results fit well with studies suggesting that PF neural circuitry is malleable. Experience has been shown to induce and modify the sensitivity of PF neurons to specific stimuli (44 45), and PF activity reflects learned associations and rules (40 46 47).

Of course, the PF cortex is not likely to be the only brain area involved in categorization. The PF cortex is interconnected with temporal lobe structures important for long-term memory (48), including the inferior temporal cortex whose neurons have stimulus specificities that could contribute to categorization (49 50). Interactions between the PF and IT cortices underlie the storage and/or recall of visual memories and associations (51 52 53), but not necessarily visual short-term memory (54). The storage and recall of categories may also require such collaboration.

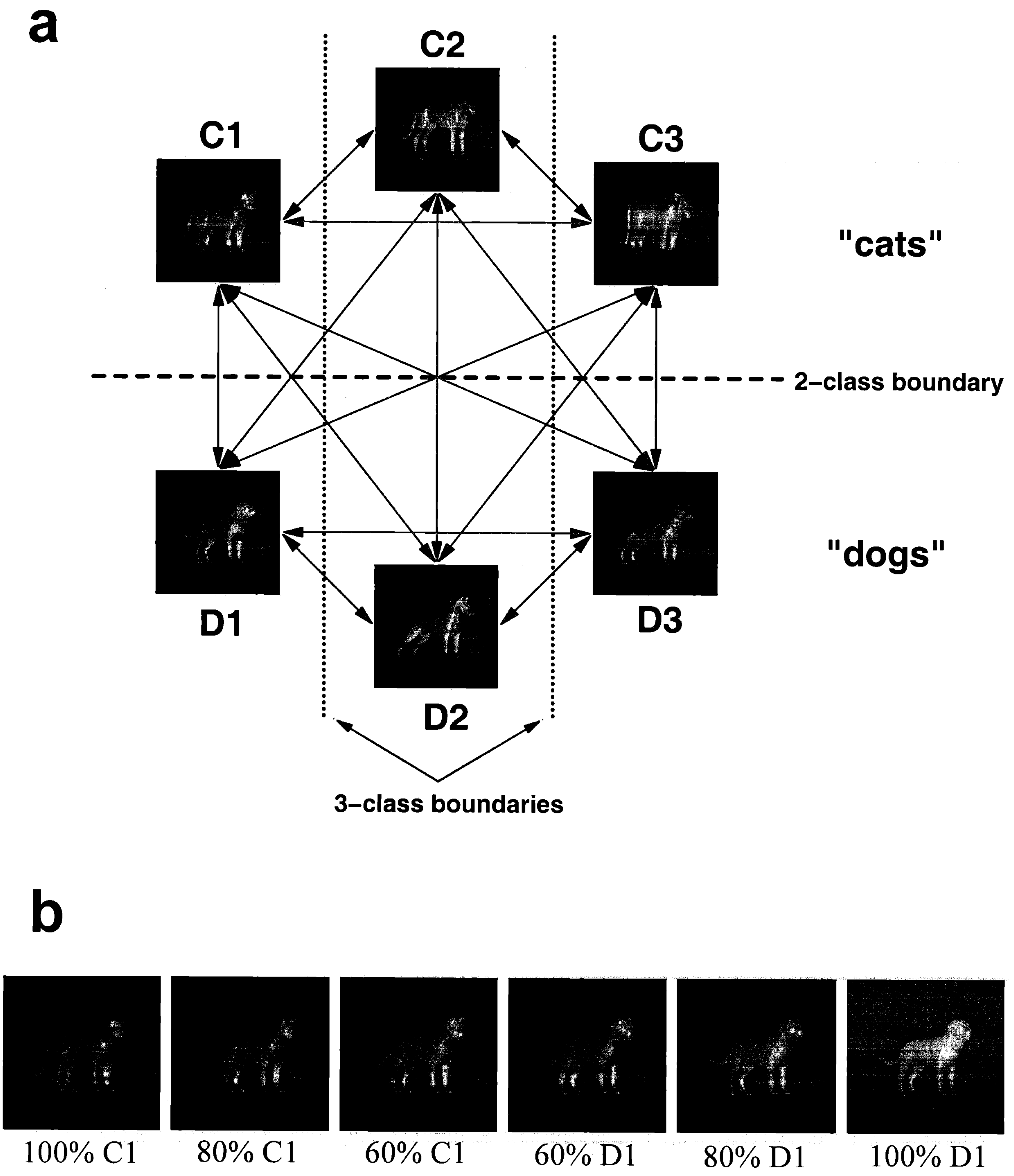
Figure Legends

Fig, 1 The stimuli. 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. The placement of the prototypes on this figure does not reflect their similarity. B. Morphs along the C1-D1 line.

Fig. 2 Task design and behavior. A. 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. B. Average performance of both monkeys. Red bars indicate the percent of samples classified as "cat" and blue bars the percent classified as "dog".

Fig. 3 Recording locations and single neuron example. A. Recording locations in both monkeys. A-anterior, P-Posterior, D-dorsal, V-ventral. There was no obvious topography to task-related neurons. B. The average activity of a single neuron to stimuli at the six morph blends. The vertical lines correspond (from left to right) to sample onset, offset and test stimulus onset. The inset shows the neuron's delay activity to stimuli along each of the 9 between-class morph lines (see fig. 1). The prototypes (C1, C2, C3, D1, D2, D3) are represented in the outermost columns; each appears in three morph lines. A color scale indicates the activity level.

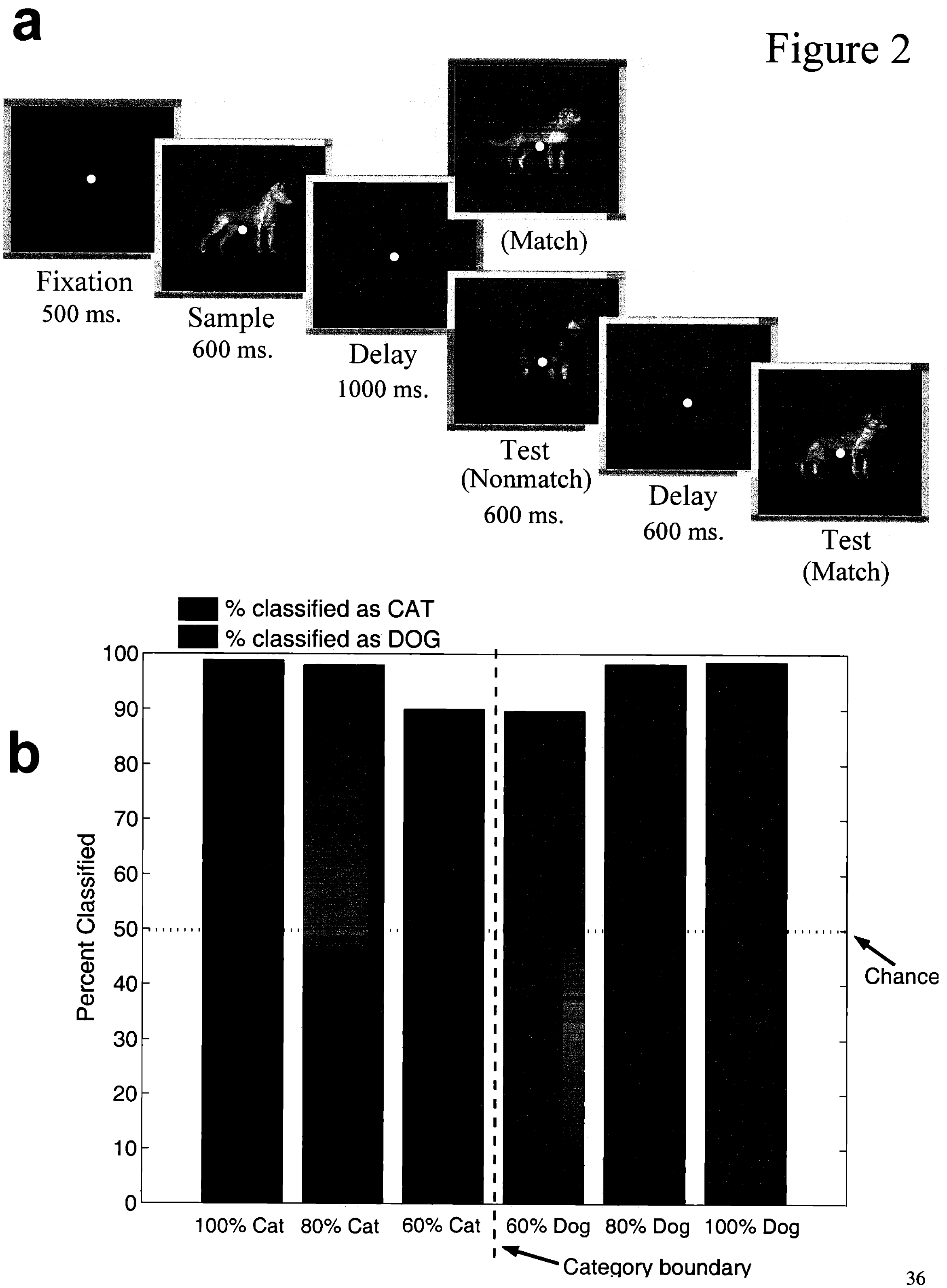
Figure 4 Category effects in a neural population. A and B show the average differences in activity to samples from the same (WCD) and different (BCD) categories for the sample (A) and delay interval (B). Each point represents one neuron. The dotted line indicates equal differences irrespective of category. The solid line indicates the regression line. C and D show average activity of the population (and standard error) to stimuli at different morph levels of their preferred and non-preferred categories for the sample (C) and delay (D) intervals.



100% C1 80% C1 60% C1 60% D1 80% D1 100% D1

Figure 1

Figure 2


a
Fixation
500 ms.
Sample
600 ms.
Delay
1000 ms.
(Match)
Test
(Nonmatch)
600 ms.
Delay
600 ms.
Test
(Match)
b
% classified as CAT
% classified as DOG
Percent Classified
100
90
80
70
60
50
40
30
20
10
0
Chance
100% Cat
80% Cat
60% Cat
60% Dog
80% Dog
100% Dog
Category boundary

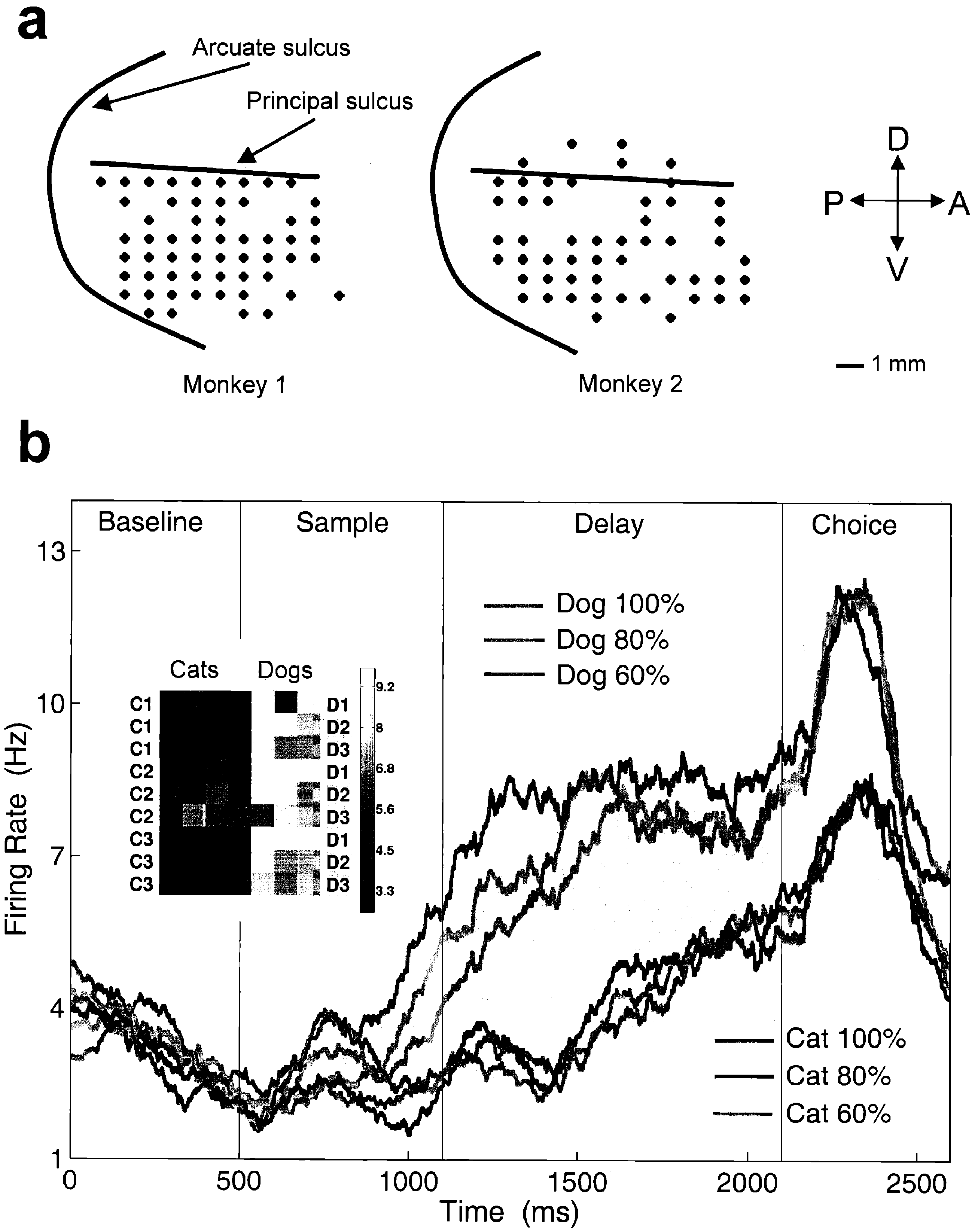

a
Arcuate sulcus
Principal sulcus
D
P
A
V
1 mm
Monkey 1
Monkey 2
b
Baseline
Sample
Delay
Choice
Dog 100%
Dog 80%
Dog 60%
Cats
Dogs
C1
C1
C1
C2
C2
C2
C3
C3
C3
D1
D2
D3
D1
D2
D3
D1
D2
D3
9.2
8
6.8
5.6
4.5
3.3
Cat 100%
Cat 80%
Cat 60%
Firing Rate (Hz)
1
4
7
10
13
0
500
1000
1500
2000
2500
Time (ms)

Figure 3

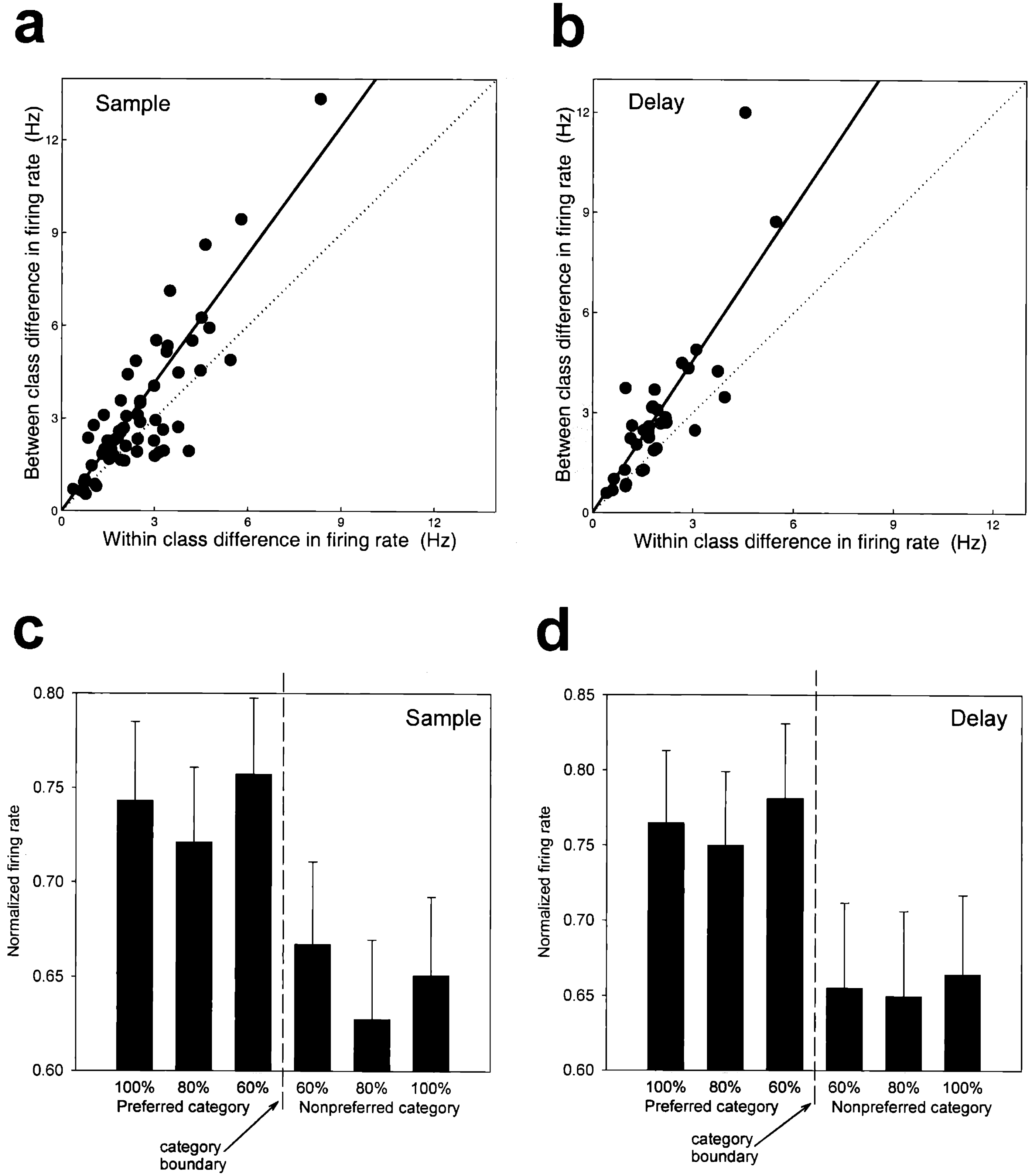

a
b
Sample
Delay
Between class difference in firing rate (Hz)
Within class difference in firing rate (Hz)
0
3
6
9
12
c
d
Normalized firing rate
0.60
0.65
0.70
0.75
0.80
0.85
100%
80%
60%
Preferred category
Nonpreferred category
category boundary

Figure 4

References and Notes

---

[1] L. W. Barsalou, *Cognitive Psychology: An Overview for Cognitive Scientists* (Lawrence Erlbaum, Hillsdale, 1992).

[2] W. A. Roberts, D. S. Mazmanian, *J. Exp. Psychol. Anim. Behav. Proc.* **14**, 247 (1988).

[3] M. Fabre-Thorpe, G. Richard, S. J. Thorpe. Rapid categorization of natural images by rhesus monkeys. *Neuroreport* **9**, 303 (1998).

[4] R. Vogels. Categorization of complex visual images by rhesus monkeys.*E ur. J. Neurosci*, **11**, 1223 (1999).

[5] T. Orlov, V. Yakovlev, S. Hochstein, E. Zohary. Macaque monkeys categorize images by their ordinal number. *Nature* **404**, 77 (2000).

[6] R. Desimone, T. D. Albright, C. G. Gross, C. Bruce. Stimulus-selective properties of inferior temporal neurons in the macaque. *J.Neurosci.* **4**, 2051 (1984).

[7] E. Kobatake, K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. of Neurophysiol.* **71**, 856 (1994).

[8] N. K. Logothetis, J. Pauls, T. Poggio. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552 (1995).

[9] D. I. Perrett, J. K. Hietanen, M. W. Oram, P. J. Benson. Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. Roy. Soc. B* **335**, 23 (1992).

[10] C. G. Gross, in *Handbook of Sensory Physiology* R. Jung, Ed. (Springer-Verlag, Berlin, 1973).

[11] K. Tanaka. Inferotemporal cortex and object vision. *Ann. Rev. of Neurosci.* **19**, 109 (1996).

[12] E. K. Miller. The prefrontal cortex and cognitive control. *Nat. Rev. Neurosc.* **1**, 59 (2000).

[13] R. A. Wyttenbach, M. L. May, R. R. Hoy. Categorical perception of sound frequency by crickets. *Science* **273**, 1542 (1996).

[14] D. Beymer, T. Poggio. Image representations for visual learning. *Science* **272**, 1905 (1996).

[15] C. Shelton. Morphable Surface Models. *International Journal of Computer Vision* **38**, 75-91 (2000).

[16] We used an algorithm that found corresponding points between one of the prototypes and the others and computed their differences as vectors. Morphs were linear combinations of these vectors added to the prototype. For more information see

---

http://www.ai.mit.edu/people/cshelton/corr/. The cat and dog stimuli differed along multiple features and were smoothly morphed, i.e., without sudden appearance of any feature. Images were 4.2 degrees in diameter, had identical color, shading and orientation and scale and were presented at the center of gaze.

[17] For neurophysiological recording, morphs were six levels of blends of cat and dog (100:0, 80:20, 60:40, 40:60, 20:80, 0:100) and two levels within categories (60:40, 40:60).

[18] "Similarity" as defined by the morphing technique and confirmed by an image correlation analysis.

[19] Monkeys maintained gaze within two degrees of a fixation point throughout the trial. Eye movements were monitored using an eye tracking system (ISCAN Cambridge, MA). We excluded stimuli that were less than 60% of a given category, as they carried little or no category information. To prevent memorization of sample–test pairs, the test stimuli were a set of 200 randomly generated morphs that were at least 70% of a category. All main effects were observed in both monkeys. For brevity, we summarize their data.

[20] 130 neurons from one monkey, 265 from the other. Sample interval activity was summed over 800 ms beginning 100 ms after stimulus onset. The delay interval activity was summed from 300 ms after sample offset to 100 ms after the end of the delay. Baseline activity was from the 500 ms of fixation before sample onset.

[21] T-Test versus baseline activity, $P < 0.01$. Parametric statistics such as T-Tests assume normal distributions. Because neuronal activity is sometimes not normally distributed, we also computed non-parametric statistics for all main effects. They yielded a virtually identical pattern of results.

[22] T-Tests on activity from the sample and delay intervals, both $P < 0.001$.

[23] $P > 0.6$

[24] Paired T-Tests between activity of all stimulus-selective neurons to the two categories computed in successive 100 msec time bins. A significant difference ($P < 0.01$) began 100-200 msec after sample onset, when the earliest PF neurons began responding. The immediate appearance of category information was also evident in average histograms across the neuron population.

[25] One-way ANOVA on the 54 sample stimuli. Sample interval: 62 neurons, delay interval: 33 neurons, $P < 0.01$.

[26] T-Test that BCD/WCD ratios were significantly different from 1. Sample interval: BCD/WCD mean = 1.30; delay interval: BCD/WCD mean = 1.49, both $P < 0.001$. Category information was significantly stronger during the delay, one-tailed T-Test, $P = 0.04$. An index of (BCD – WCD)/(BCD+WCD) yielded similar results.

---

[27] Excluding neurons with firing rates below 2 Hz (which produce spurious values when normalized) yielded 55 and 29 neurons with selectivity in the sample and delay intervals, respectively. We normalized each neuron's activity as a proportion of its activity to the most effective single stimulus. Preferred and non-preferred category was determined by only this stimulus to ensure analyses were not biased toward a category effect.

[28] Two-way ANOVA of category membership and level of category (60%, 80%, 100%), test of the category factor. $P < 0.01$ for both intervals.

[29] Two-way ANOVA, test of the level factor, $P>0.6$ for both intervals.

[30] ANOVAs on the 27 samples from the preferred or non-preferred category, either $P < 0.01$.

[31] T-Test on all match vs. all non-match test stimuli, $P < 0.01$

[32] E. K. Miller, C. A. Erickson, R. Desimone. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J.Neurosci.* **16**, 5154 (1996).

[33] T-Test vs baseline for the sample and/or delay intervals, $P<0.01$.

[34] One-way ANOVA on all 54 samples for the sample and/or delay intervals, $P < 0.01$.

[35] Sample interval: mean 2-category BCD/WCD = 1.13, T-Test $P = 0.22$; delay interval: 2-category BCD/WCD mean = 0.96, T-Test $p = 0.58$. This analysis was limited to morphs between corresponding cat and dog prototypes (i.e., C1-D1, C2-D2, C3-D3, the vertical morph lines in figure 1a) because the other morph lines crossed both the 2-category and 3-category boundaries. We confirmed that this test could detect 2-category information by applying it to the data from the 2-category task. The results were virtually identical to the 2-category test described above (sample interval: BCD/WCD ratio = 1.33, T-Test, $P < 0.001$; delay interval: BCD/WCD ratio = 1.57, T-Test, $P < 0.001$).

[36] 3-category BCD/WCD mean = 1.51, T-Test $P<0.01$. As for the 2-category test, we compared samples at equivalent distances along between-category morph lines but now using morph lines that crossed the 3-category boundaries, but not the 2-category boundary. The early appearance of category information in PF activity also suggested that training had altered neuronal selectivity.

[37] It was not evident during the sample interval, 3-category BCD/WCD mean = 1.03, T-Test, $p = 0.79$.

[38] J. Quintana, J. M. Fuster. Mnemonic and predictive functions of cortical neurons in a memory task. *Neuroreport.* **3**, 721 (1992).

[39] G. Rainer, S. C. Rao, E. K. Miller. Prospective coding for objects in primate prefrontal cortex. *J.Neurosci.* **19**, 5493 (1999).

---

[40] W. F. Asaad, G. Rainer, E. K. Miller. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399 (1998).

[41] Performance was lower on the 3-category than 2-category task (87% vs 96%, T-Test $P<0.001$) and reaction times were longer (average 307 ms vs. 264 ms, T-Test $P<0.001$).

[42] G. Lakoff, *Women, fire, and dangerous things* (University of Chicago Press, Chicago, 1987).

[43] R. Dias, T. W. Robbins, A. C. Roberts, Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* **380**, 69 (1996).

[44] N. P. Bichot, J. D. Schall, K. G. Thompson. Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* **381**, 697 (1996).

[45] G. Rainer, E. K. Miller. Visual Experience and Prefrontal Object Representation. *Neuron* **27**, 179 (2000).

[46] I. M. White, S. P. Wise. Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* **126**, 315 (1999).

[47] W. F. Asaad, G. Rainer, E. K. Miller. Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* **84**, 451 (2000).

[48] M. J. Webster, J. Bachevalier, L. G. Ungerleider. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex* **4**, 470 (1994).

[49] R. Vogels. Categorization of complex visual images by rhesus monkeys. *Eur J Neurosci* **11**, 1239 (1999).

[50] E. Kobatake, K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. of Neurophysiol.* **71**, 856 (1994).

[51] S. A. Gutnikov, Y. Y. Ma, D. Gaffan. Temporo-frontal disconnection impairs visual-visual paired association learning but not configural learning in Macaca monkeys. *Europ. J. Neurosci.* **9**, 1524 (1997).

[52] G. Rainer, S. C. Rao, E. K. Miller. Prospective coding for objects in primate prefrontal cortex. *J.Neurosci.* **19**, 5493 (1999).

[53] H. Tomita, M. Ohbayashi, K. Nakahara, I. Hasegawa, Y. Miyashita. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* **401**, 699 (1999).

[54] D. Gaffan, M. J. Eacott. Uncinate fascicle section leaves delayed matching-to-sample intact, with both large and small stimulus sets. *Exp Brain Res* **105**, 175 (1995).

---

[55] We thank C. Shelton for the morphing software, and K. Anderson, W. Asaad, M. Histed, M. Mehta, J. Wallis, R. Wehby, and M. Wicherski for valuable comments, help, and discussions. This work was supported by a NINDS grant, a NSF-KDI grant, the RIKEN-MIT Neuroscience Research Center, a Merck/MIT Fellowship (M.R.), the Whitaker Chair (T.P.) and the Class of 1956 Chair (E.K.M.)

Visual Categorization and the Primate Prefrontal Cortex: Neurophysiology and Behavior

David J. Freedman, Maximilian Riesenhuber, Tomaso Poggio, Earl K. Miller*
Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, Center for Biological and Computational Learning and The Department of Brain and Cognitive Sciences
Cambridge, MA, USA, 02139

Submitted for publication

ABSTRACT

The ability to group stimuli into meaningful categories is a fundamental cognitive process. To explore its neural basis, we trained monkeys to categorize computer-generated stimuli as "cats" and "dogs". A morphing system was used to systematically vary stimulus shape and precisely define a category boundary. Psychophysical testing and analysis of eye movements suggested that the monkeys categorized the stimuli by attending to multiple stimulus features. Neuronal activity in the lateral prefrontal cortex reflected the category of visual stimuli and changed with learning when a monkey was retrained with the same stimuli assigned to new categories. Further, many neurons showed activity that appeared to reflect the monkey's decision about whether two stimuli were from the same category or not. These results suggest that the lateral PFC is an important part of the neuronal circuitry underlying category learning and category-based behaviors.

INTRODUCTION

Our perception of the environment is not a faithful registration of its physical attributes. Instead, we carve the world into meaningful groupings, or categories. This process of abstracting and storing the commonalities among like-themed individuals is fundamental to cognitive processing because it imparts knowledge. For example, knowing that a new gadget is a "camera" instantly and effortlessly provides a great deal of information about its relevant parts and functions and spares us from having to learn anew each time we encounter a new individual. The usefulness of storing categories is likely to be related to their ubiquity. They are evident in all sensory modalities and range from relatively simple (e.g., color perception) to the most abstract human concepts.

Because perceptual categories often group together very different-looking things, their representation must involve something beyond the sort of neuronal tuning that typifies encoding of physical appearance: gradual changes in neuronal activity as features gradually change (e.g., shape, orientation, direction). In fact, evidence that a human or animal has stored a category is that behavior does not track smoothly with changes in physical appearance: categories have sharp boundaries (not gradual transitions) between them and members of the same category are treated as equivalent even though their physical appearances may vary widely. A simple example is crickets sharply dividing a continuum of pure tones into “mate” versus “bat” (a predator); their degree of approach or avoidance is virtually identical across a wide range of frequencies that are to one side or the other of the boundary, but flips suddenly at 16 kHz (Wyttenbach et al., 1996). It is advantageous to represent this information categorically because it optimizes reproductive behavior while minimizing fatal mistakes. Similar effects are evident in humans’ perception of the phonemes “b” versus “p” (Lieberman et al., 1967) and the facial expressions of emotion (Beale and Keil, 1995).

The elaborate behavioral repertoire of advanced animals naturally depends on more elaborate categorization abilities. Their mental lexicon includes categories that are characterized along multiple dimensions and are often difficult to precisely define. In addition, advanced animals have an enormous capacity to learn and adapt. Monkeys, for example, have been taught categories such as animal versus non-animal (Roberts and Mazmanian, 1988), food versus non-food (Fabre-Thorpe et al., 1998), tree versus non-tree, fish versus non-fish (Vogels, 1999) and ordinal numbers (Orlov et al., 2000).

Where such categories are encoded in the brain has not been clear. In primates, they could be represented and stored in the same areas of the visual cortex involved in form analysis and object recognition, such as the inferior temporal cortex (ITC) (Gross, 1973; Desimone et al, 1984; Tanaka, 1996; Logothetis and Sheinberg, 1996). They might also be evident in the brain regions, such as the prefrontal cortex (PFC) that receive direct projections from the ITC (Ungerleider et al., 1989; Webster et al., 1994) and are involved in a variety of higher cognitive functions (Goldman-Rakic, 1987; Fuster, 1997; Miller, 2000; Miller and Cohen, 2001). Both the ITC and PFC contain neurons that often exhibit highly specific responses to complex stimuli such as trees, fishes, faces, brushes, etc. (Gross et al., 1972; Bruce et al., 1981; Perrett et al., 1982; Desimone et al., 1984; Tanaka et al., 1991; Miller et al., 1996; Scalaidhe et al., 1999) and are influenced by experience (Logothetis et al., 1995; Booth and Rolls, 1998; Kobatake et al., 1998; Miyashita, 1988; Rainer and Miller, 2000). Whether or not their activity reflects stimulus categories has not been clear. These neurons have not been tested for the diagnostic characteristics of categories, (e.g., sharp boundaries and within-category generalization); their specificity might reflect similarities and differences in physical appearance of the stimuli, not necessarily their category membership.

To test for neuronal correlates of visual categories, we trained monkeys to categorize computer-generated stimuli into two categories, cats and dogs. A novel 3D 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 and Poggio, 1996; Shelton, 2000). This allowed us to establish a sharp category boundary between stimuli that were physically similar, yet include in the same category stimuli that were visually dissimilar. A brief report of these results appeared previously (Freedman et al., 2001).

MATERIALS AND METHODS

Subjects: Two female adult rhesus monkeys (*Macacca mulatta*) weighing 6.0 and 7.5 kg were used in this study. Using previously described methods (Miller et al., 1993), they were implanted with a head bolt to immobilize the head during recording and with recording chambers. Eye movements were monitored and stored using an infrared eye tracking system (Iscan, Cambridge, MA). All surgeries were performed under sterile conditions while the animals were anesthetized with isoflurane. The animals received postoperative antibiotics and analgesics and were handled in accord with National Institutes of Health guidelines and the recommendations of the Massachusetts Institute of Technology Animal Care and Use Committee.

Recording Techniques: Electrode penetration sites were determined using magnetic resonance imaging scans obtained prior to surgery. The recording chambers were positioned stereotactically over the lateral prefrontal cortex such that the principal sulcus and ventrolateral prefrontal cortex were readily accessible. Neuronal activity was isolated using arrays of four to eight independently moveable tungsten microelectrodes (FHC Instruments, Bowdoinham, ME). The electrodes were advanced using custom made screw-driven mini-microdrives (Nichols et al., 1998) mounted on a plastic grid (Crist Instruments, Damascus, MD). Neuronal activity was amplified, filtered, and stored for off-line sorting into individual neuron records (Plexon Systems, Dallas, TX). This allowed us to isolate, on average, nearly two neurons per electrode. We did not prescreen neurons for task-related activity such as visual responsiveness or stimulus selectivity. Rather, we randomly selected neurons for study by advancing each electrode until the activity of one or more neurons was well isolated and then began data collection. This procedure was used to ensure an unbiased estimate of prefrontal activity.

Stimuli: A large continuous set of images was generated from three prototype cats and three prototype dogs (Figure 1) with a novel algorithm (Shelton, 2000). It found corresponding points between one of the prototypes and the others and then computed their differences as vectors. Morphs were created by linear combinations of these vectors added to that prototype. For more information see http://www.ai.mit.edu/people/cshelton/corr/. By morphing different amounts of the prototypes we could generate thousands of unique images, continuously vary shape and precisely define one or more arbitrary category boundaries. For most of the experiments, the images were divided into two groups, cats and dogs, with the boundary at an equal blend of cat and dog. Thus category membership was defined by whichever category contributed more (>50%) to a given morph. As a result, stimuli that were close to but on opposite sides of the boundary were visually similar, while stimuli that belonged to the same category could be visually dissimilar (e.g. the "housecat" (C1) and "cheetah"(C2) prototypes). The stimuli differed along multiple features and were smoothly morphed, i.e., without sudden appearance or disappearance of any feature. They were 4.2 degrees in diameter, had identical color, shading and orientation and scale and were presented at the center of gaze.

We confirmed that the morphs did indeed vary smoothly and at a constant rate over different parts of morph space by using an image correlation analysis. A 2D correlation coefficient was calculated for neighboring images at six levels of blends of cat and dog (cat:dog: 100:0, 80:20, 60:40, 40:60, 20:80, 0:100) along each of the nine between-category morph lines. The correlation was calculated by computing the 2D correlation coefficient separately for each color plane and then averaging across planes. The correlation coefficient between neighbors remained constant and high (about 0.9) across the entire morph space. The coefficients between stimuli directly across the cat/dog boundary did not differ from the other coefficients from equal distance in morph space that did not cross the boundary (one-way ANOVA, $P = 0.44$).

Behavioral Tasks: The monkeys performed a delayed match-to-category task that required them to judge whether two successive stimuli were from the same category (Figure 2). The trial began when the monkey grasped a metal bar and fixated a small (0.3 deg) white spot at the center of a CRT screen. They were required to maintain gaze within a ±2 deg window around the fixation spot for the entire trial. Following the initial 500 ms of fixation, a sample image was presented at the center of the screen for 600 ms, followed by a 1000 ms delay. Then, a choice image appeared. If the sample and choice stimuli were from the same category (a category match), the monkeys were required to release the lever before the stimulus disappeared 600 ms after its onset, to receive a juice reward. If the choice image was from a different category (a category non-match), there was an additional brief delay (600 ms) followed by another image that was always a match and thus required a lever release. As a result, a category judgment was only required for the first choice image. The second delay and match image were used so that a behavioral response would be required on every trial. This ensures that the monkeys were always paying attention. Because a decision was only required for the first choice image and the forthcoming behavioral response was predictable from the second delay onwards, that delay and subsequent match image will not be considered further. Note that with this design, the behavioral response (lever release) is not uniquely associated with a category (it was used to signal "match", not cat or dog) and, further, the monkeys could not predict whether the first choice stimulus would require a response. Thus, any differential activity to the sample categories could not be related to the behavioral response. A 2000 ms to 3000 ms inter-trial interval followed correct trials. An error was defined as a lever release to a non-match or failure to release to a match; breaks of fixation were not counted among the error rates in behavioral analyses. An additional 3000 ms "timeout" was added to the inter-trial interval following an error. Monkeys typically performed over 700 correct trials per day.

The monkeys were gradually trained to categorize the images as cats and dogs by beginning with a delayed match-to-sample task in which the prototypes were used as samples, the match was always identical to it, and the non-matches were a prototype from the other category. We then gradually included more and more morphs as samples by choosing images at increasing distances from the prototypes. In parallel, matches were chosen from an increasingly greater distance of morph space around the sample while respecting the category boundary. Non-matches were always from the other category.

During the course of training, over 1000 sample stimuli were used from all over the morph space. This prevented monkeys from solving the task by simply remembering specific stimulus–response contingencies. Neurophysiological recording, however, requires that a limited number of stimuli be used so that each can be repeated multiple times and neuronal variability can be assessed. Thus, for recording experiments we limited the samples to 54 images. This included the six prototype images and morphs from equally spaced intervals across each of the nine morph lines that connected each cat prototype to each dog prototype (Figure 1a). There were six levels of blends of cat and dog (cat:dog) (100:0, 80:20, 60:40, 40:60, 20:80, 0:100) along the nine morph lines that crossed the two-category boundary (the red lines in Figure 1a) and two levels along the six within-category morph lines (60:40, 40:60) (the blue lines in Figure 1a). To prevent monkeys from learning to memorize specific stimulus-response contingencies during the recording experiments, the choice stimuli were 100 randomly generated morphs from each category that were randomly paired with sample stimuli of the appropriate category. To ensure that category judgment errors were due to confusion over the sample category, the choice stimuli unambiguously belonged to a given category: they were always chosen to be at a distance of at least 20% from the boundary.

The monkeys' categorization abilities were further examined with separate psychophysical tests employing an additional 14 morphs that were equally and tightly spaced (6.67% intra-stimulus distance) along each of the morph lines that crossed a category boundary. This allowed for a more precise description of the monkeys' ability to categorize stimuli near the category boundary. This task was identical in all timing and behavioral events except that the monkeys were randomly rewarded on trials in which the sample stimulus was very close to the category boundary (less than 10% difference). This was done to ensure that reward history did not influence a monkey's choices on trials in which the samples were near the boundary and more likely to elicit guessing.

To test the effects of learning on neuronal activity, we trained a monkey to re-categorize the cat and dog images into three new categories. Two new category boundaries were defined that were orthogonal to the original two-category boundary (figure 1a). This resulted in three new classes that each contained morphs centered around one cat prototype and one dog prototype. The same 54 sample stimuli were used for neurophysiological recording under the two and three-category schemes. As in the two-category experiment, the choice stimulus set consisted of 100 randomly generated morphs from each category that had a maximum component of 20% from each of the other two categories.

Data Analysis: Neuronal activity level was calculated in four time epochs: baseline, sample presentation, first delay and first choice stimulus presentation. Baseline neuronal activity was averaged over the 500 ms of fixation preceding sample presentation. Sample period activity was averaged over an 800 ms epoch beginning 100 ms after sample onset to account for the latency of PFC neuronal responses and included the first 300 ms following sample offset to include any activity related to that event. Delay activity was assessed over an 800 ms epoch beginning 300 ms after sample offset and ending 100 ms after first choice stimulus onset. Activity to that choice stimulus was averaged over an epoch that began 100 ms after its onset and ended two standard deviations before the monkeys' average reaction time during each recording session to exclude any effects related to the execution of the behavioral response.

Category information in neuronal activity was assessed using several methods. We computed an index of category tuning by calculating each neuron's average firing rate difference to pairs of sample morphs from the same category (within-category difference, WCD) and its average difference to samples from different categories (between-category difference, BCD) using images from the morph lines that crossed the category boundary. The WCD was defined by computing the absolute difference between the 100% and 80% morphs and between the 80% and 60% morphs for both categories and averaging these values. The BCD was computed by averaging the across-boundary differences between the 60% cats and 60% dogs. As a result, the distance between stimuli in morph space was identical (20%) for the BCD and WCD comparisons. A standard index was computed for each neuron by dividing the difference between its BCD and WCD values by their sum. This index can have values ranging from -1 to 1. Positive values indicate a larger difference between categories while negative values reflect larger differences within a category than between categories. BCD and WCD values were computed for neurons recorded during the three-category task in a similar fashion, by determining differences in activity to samples that differed by 20% along the morph lines that crossed the three-category boundaries (Figure 1). To ensure that the previously learned two-category scheme did not contribute to the values obtained when calculating category effects in the three category scheme, we excluded from this analysis the morph lines that crossed both the two and three-category boundaries (e.g., the morph line connecting cat prototype #1 and dog prototype #2).

In addition to computing an index, we also compared between and within category differences by using a receiver-operating characteristics (ROC) analysis (Green and Swets 1966; Tolhurst et al., 1983; Vogels and Orban, 1990). The ROC analysis measures the degree of overlap between two distributions of values. It has several advantages. First, it makes no assumptions about the two distributions, A and B, and thus returns an unbiased estimation of overlap. Second, it can be interpreted as the performance of an ideal observer in a two-way forced choice task; values of 0.5 indicate 50% correct classification (guessing) while values of 0 or 1 indicate error-free classification. Third, it is independent of neuronal firing rate and number of observations. While the category index described above explicitly tests for sharp tuning across the category boundary, the ROC value gives a general measure of the degree of category selectivity.

To determine the time-course of category information in neuronal activity, we computed the ROC area within a time bin of 200 ms that was slid in 10 ms steps. We began 500 ms prior to sample stimulus onset and ended 100 ms following the first choice stimulus onset. This was computed for all neurons that were category selective (according to a t-test evaluated at $P < 0.01$) during the sample and/or delay epochs.

The latency for neuronal activation (irrespective of category information) was determined by compiling the average histogram of firing rate values for all responsive neurons (i.e. neurons that showed significantly different activity during the sample and/or delay periods compared to baseline, evaluated by two-tailed t-test at $P < 0.01$.) This average histogram was smoothed with a 30 ms gaussian window and the latency was defined as the point of maximum inflection (determined by computing the second derivative at all points along the histogram) of this curve following sample onset.

To account for the wide range of firing rates across different neurons, individual neurons' firing rates were normalized when computing population average histograms of neural activity. For each neuron, the mean firing rates at each of the six steps from the cat to dog prototypes were computed. Then, the range of firing rates for each neuron were rescaled according to the minimum and maximum values across those six groups such that each neuron's minimum and maximum rate was 0.0 and 1.0, respectively. This allowed each neuron's range of modulation to contribute equally to the population average.

RESULTS

Behavioral Data

Category judgments

The monkeys were very accurate at the two-category judgments. During the recording sessions, performance was high (about 90% correct), even when the samples were close to the category boundary; the monkeys classified dog-like cats (60:40 cat:dog) correctly about 90% of the time, and misclassified them as dogs only 10% of the time, and vice-versa (Figure 3). The results of psychophysical tests with more closely spaced morphs are shown in Figure 4a. Even very near the boundary, when stimuli were very similar to (only 3% different from) the other category (i.e., a 53.3% cat or dog), the monkeys still performed significantly above chance (~65%, chance = 50%). Thus even with closely spaced morph images, the sudden change in behavior characteristic of category representations were evident in behavior.

Figure 4b shows the performance of monkey A after it had been trained to re-categorize the same images under the three-category scheme. Performance here was somewhat lower than during the two-category task. This is presumably because there were two boundaries in the three-category task and thus a higher percentage of stimuli were close to the border (the data in the figure are collapsed across the boundaries). Still, the sharp drop-off in performance indicative of a category representation was evident; monkeys continued to perform above chance for morphs that were only about 3% different from the boundary. The greater difficulty of the three-category task was also evident in the monkey's behavioral reaction times. They were significantly longer during the three-category task (average = 307 ms) than the two-category task (264 ms, t-test at $P < 0.01$).

Stimulus features used for categorization

To explore which features the monkeys were using to categorize the images into cats and dogs, we conducted further psychophysical testing. In one set of experiments, we removed the requirement to maintain central fixation (by removing the fixation point) and allowed the monkeys to freely gaze at the images. Given the close link between attention and gaze during unconstrained viewing, the assumption was that monkeys would spend more time gazing at the features that they were using to define the categories.

It was apparent that the monkeys were not focusing on a single feature to categorize the images. Even though the sample presentation was brief (600 ms), they typically made several saccades while viewing the stimulus. One monkey made an average of 3.45 saccades and the other monkey averaged 2.25 saccades during sample presentation (defined as the number eye movements exceeding 50 degrees/second, equivalent to 0.5 degrees of movement in adjacent 10 ms time bins). Interestingly, the two monkeys seemed to use different combinations of features to categorize the images. One monkey tended to look toward the tail of the sample images; its gaze was on average 1.46 deg to the left and 0.60 deg below the center of the screen. The other monkey tended to direct its gaze toward the head region; on average, its gaze was 0.57 deg to the right and 1.16 deg above central fixation. Figure 5a shows representative traces from one trial for each monkey. The gaze patterns for the two monkeys were significantly different from one another (along both the horizontal and vertical axes, t-tests, $P < 0.01$).

We also tested the monkeys' ability to categorize the images after removing the 'heads' or 'tails' of the morph stimuli and then interleaving them with non-degraded samples. The assumption here was that if the monkey was relying on a single feature at the front or back of the image, its removal should decrease categorization performance to chance. This was not the case. As shown in Figure 5b, we found that the monkeys' performance remained high (~80% correct) when either the head or tail was absent. This pattern of results suggests that monkeys each used a unique combination of features to categorize the images.

Neuronal Data

Basic properties

A total of 395 lateral prefrontal cortex neurons were recorded from three hemispheres of two monkeys during performance of the two-category DMC task (130 from monkey A, 265 from monkey B, figure 6). Visual responsiveness was evaluated by comparing the activity in the sample and delay intervals to baseline activity using paired t-tests (evaluated at $P < 0.01$). Based on this criterion, 259/395 (66%, 113 from monkey A, 146 from monkey B) of neurons were activated during one or more task intervals. The onset of neuronal responsiveness across the population of responsive PFC neurons occurred at approximately 100 ms following sample stimulus onset (Methods).

An initial assessment of neuronal category selectivity was made with a t-test of the activity to all cat stimuli versus all dog stimuli (evaluated at $P < 0.01$). This revealed that nearly a quarter of all (randomly selected) neurons (96/395, 24%, 60 and 36 in monkeys A and B, respectively) showed a significant difference in their overall activity to cats versus dogs in the sample and/or delay intervals (74 sample, 51 delay). Many neurons (78/395, or 20%; 67 sample, 32 delay) also showed a significant effect of the individual samples (i.e., were stimulus-selective) according to a one-way ANOVA (with the 54 sample stimuli as the factor; evaluated at $P < 0.01$). A majority of these stimulus-selective neurons also showed an overall effect of category (56/78, or 72%; 46 sample, 21 delay, t-test, $P < 0.01$). Similar numbers of category selective neurons preferred cats (39/74 sample, 27/51 delay) as dogs (35/74 sample, 24/51 delay). In both monkeys, there was a greater incidence of category selective neurons in the sample than the delay interval (monkey A: 48 sample, 31 delay; monkey B: 26 sample, 20 delay).

The activity of many neurons showed a sharp differentiation between the two categories that mirrored the monkeys' behavior. That is, they showed relatively large differences in activity to samples from different categories and relatively similar activity to samples from the same category. Two examples of single neurons are shown in figure 7a and 7b. They show each neuron's average activity to all samples at different blends of cats and dogs. Both seem to encode category membership rather than relative physical similarity. Note that their activity was significantly different to dog-like (60%) cats and cat-like (60%) dogs (t-test, $P < 0.001$) but there was no difference in activity between these stimuli and their respective prototypes ($P > 0.1$).

These effects were also evident in the average activity across the population of all stimulus-selective neurons. For this analysis, we chose neurons that were stimulus-selective, not category-selective per se (ANOVA with the individual samples as a factor, $P < 0.01$, $N = 55$ for the sample interval, 29 for the delay, excluding neurons with mean firing rates below 2 Hz, as they can produce spurious results when normalized). Figure 8 shows the mean normalized firing rates for the six levels of morphs. Each neuron's preferred category was determined by the category of the single sample stimulus (of 54) that evoked the maximal firing rate, computed separately for the sample and delay intervals. By determining the preferred category by a single stimulus instead of the average across all category members, we ensured that this test was not biased towards finding a category effect. During both time epochs, there was a significant difference between the categories ($P < 0.01$) but no differences between the different morph levels within each category ($P > 0.6$, two-way ANOVA with category and distance from the category boundary as factors).

Quantification of category effect

To quantify the effect of category membership on the neuronal population, we computed a category index that reflected each neuron's average difference in activity to samples across the category boundary versus its difference to samples that were from the same category (see Methods). Positive values indicate greater differences across the category boundary than within each category and negative index values indicate the opposite.

We examined all stimulus-selective neurons, irrespective of whether they were category-selective per se ($N = 78$, 67 sample, 32 delay). The distributions of category index values for the sample and delay periods are shown in figure 9. During both epochs, mean category index values were significantly greater than zero, i.e., the distribution was shifted toward category tuning (sample: 0.08, delay: 0.14, one-tailed t-test, $P < 0.001$). The indices were significantly larger (more shifted toward category tuning) during the delay than the sample interval (one tailed t-test, $P = 0.04$). Similar comparisons were also made by computing ROC values, which reflect how well an ideal observer would do at categorization using each neuron's firing rate (see Methods).

Across the population of stimulus selective neurons, the average ROC value was 0.59 (range = 0.50 to 0.75) in the sample interval and 0.59 in the delay (range: 0.50 to 0.82).

These analyses demonstrate that a significant degree of category information is evident even across the entire population of stimulus-selective neurons. The average index or ROC values obtained were somewhat modest because activity was averaged across an entire trial epoch and across all stimulus-selective neurons. As will be shown next, the strength of category signals varied widely with individual neurons and with time; individual neurons could convey very strong category signals at particular points in the trial.

Temporal characteristics of category information

To examine the temporal dynamics of the representation of category information in PFC activity, we used a sliding ROC analysis (see Methods). For this analysis, we included neurons whose average activity in the sample and/or delay intervals was significantly category-selective (t-test on activity to all cats versus all dogs, evaluated at $P < 0.01$, $N = 96$ neurons).

Figure 10a shows the ROC values for each neuron in 10 millisecond time steps. The ROC values are sorted by their magnitude separately for each time bin in order to better illustrate the number of neurons exhibiting ROC values above 0.5 (chance) at each moment in time. This revealed that, in general, more neurons conveyed category signals late in the sample epoch than during the delay interval, but that the strongest category signals occurred in the late delay and early choice presentation epoch. Figure 10 indicates that there were a greater number of neurons with moderate or small ROC values for the time bins during the sample epoch (i.e., there are more "foothills" leading up the "peaks") but that the highest ROC values occurred during the late delay/choice presentation (the "peaks" are highest then).

Effects of learning on category representations

As our monkeys had no prior experience with cats and dogs, it seemed likely that the category information in the PFC was acquired through learning. To test the effects of learning on category representations, we retrained one monkey with the samples reassigned to three new categories (see Figure 1 and Methods). We then recorded from 103 neurons at similar depths and locations as those recorded during the two-category task. The incidence of neuronal responsiveness and stimulus selectivity during the three-category task was similar to that during the two-category task: about 63% (65/103) of neurons were visually responsive (t-test vs. baseline, as above, $P < 0.01$) and about 23% (24/103, 14 sample, 14 delay) were stimulus selective (ANOVA with stimulus as factor, $P < 0.01$).

An example of a neuron recorded during the three-category task is shown in Figure 11. It showed a significant effect of category during the delay period when the data was sorted according to the new, currently relevant, three-category scheme (ANOVA, $P < 0.001$); it distinguished one of the categories from the other two. By contrast, when the data was sorted according the old (now irrelevant) cats and dogs category scheme, there was no significant effect (ANOVA, $P = 0.74$).

To test for these effects in this population of neurons, we first examined all those that were stimulus-selective ($N = 13$, ANOVA, $P < 0.01$). When the category index was computed using the old (now irrelevant) cat and dog categories, there was no evidence of category effects; the two-category index was not significantly greater than zero for the sample interval (two-category index = 0.01, one-tailed t-test P = 0.5) nor the delay (two-category index = -0.10, one-tailed t-test P = 0.9). However, when the category index was computed using the new (relevant) three category boundaries, a significant category effect was observed in the delay (three-category index = 0.16, one-tailed t-test P = 0.008). As we found for the two-category task, three-category tuning was stronger during the delay than the sample interval (one-tailed t-test, P = 0.02). In fact, we did not detect significant category tuning across the population of stimulus selective neurons during the sample interval (three-category index = -0.01, P = 0.5), although it was detected when we computed the index for all neurons recorded during the three-category task (N = 103, see below).

The same pattern of effects was observed across the entire population of neurons. Figure 12 shows the distribution of the category indices for all 103 (randomly sampled) cells recorded during the three-category task. The indices computed using the three-category scheme revealed significant category information (i.e., the distribution was shifted to the right) for both the sample interval (Figure 12a, three-category index = 0.065, one tailed T-test P = 0.0007) and for the delay (Figure 12b, three-category index = 0.08, one tailed t-test, P = 0.0005). By contrast, when the indices were computed using the two-category scheme, there were no significant category effects during the sample (Figure 12c, two-category index = -0.02, one tailed t-test, P = 0.83) nor the delay interval (Figure 12d, two-category index = -0.03, one tailed t-test, P= 0.82). Thus, while information about the three-category scheme was evident in the population of PFC neurons, we could no longer detect information about the previously-learned, now-irrelevant, cat and dog categories.

Category match/non-match effects

When the choice stimulus was presented, the monkeys needed to categorize it and then decide whether or not its category matched that of the sample. Both signals were present in neuronal responses to the choice stimulus. We evaluated activity in this interval with a two-way ANOVA (factor one: choice stimulus category, factor two: match vs. non-match, evaluated at $P < 0.01$). Just over 9% (37/395) of the entire population of PFC neurons reflected the category of the choice stimulus while 11% (43/395) reflected its match/non-match status. Over two thirds of the latter neurons (29/43) showed an effect of matching/non-matching that was similar regardless of whether the choice stimulus was a cat or dog (main effect of match/non-match, no interaction with choice stimulus category). An example of a neuron that exhibited greater activity to matches is shown in Figure 13a and an example of a neuron showing greater activity to non-matches is shown in Figure 13b. The remaining third of these neurons (14/43) showed an interaction between the match/non-match status and the category of the choice stimulus ($P < 0.01$). In other words, they showed match/non-match effects that were limited, or much stronger, to one of the categories. An example of a "cat match" neuron is shown in Figure 13c. For match/non-match selective neurons, a similar number preferred matches (22/43 or 51%) as non-matches (21/43 or 49%).

Analysis of Error Trials

For insight into neuronal correlates of the monkey's errors, we compared category effects and match/non-match effects on correctly performed trials versus those in which monkeys made errors in category judgments. For these analyses, we included neurons that showed significant effects on correct trials. Figure 14 shows the results of these comparisons. Category information was evident during the sample interval on both correct and error trials; the average activity to the preferred versus non-preferred category was significantly different for both types of trials (t-test, $P<0.001$, Figure 14a). But category information seemed to be lost in the delay. A significant difference between the average activity to the two categories was evident on correct trials ($P < 0.001$), but not on error trials ($P = 0.79$, Figure 14b). Match/non-match effects also depended on whether the trial was correctly performed or not. For these analyses, the choice stimulus status (match or non-match) that elicited the greater activity on correct trials was termed the "preferred condition." For all neurons that showed pure match vs. non-match effects ($N = 25$, i.e. match vs. non-match factor: $P < 0.01$, choice-category and interaction factors: $P > 0.01$), there was a significant difference ($P < 0.001$) in average activity to the preferred and non-preferred conditions on correct trials. On error trials, however, the pattern reversed; there was an increase in activity to non-preferred over preferred conditions that reached significance at $P < 0.05$ (Figure 14c). This is presumably because the monkeys mistakenly responded to non-matches as if they were matches.

DISCUSSION

We report that neurons in the PFC, a brain region central to many visual behaviors, exhibited properties that mirrored the behavioral characteristics of perceptual categories. They made sharper distinctions between stimuli from different categories than between stimuli from the same category, irrespective of their relative physical similarity. This explicit encoding of category membership in the activity of single neurons did not have to be the case. In principle, categories might have only been reflected on the ensemble level, as an emergent property of neurons encoding different defining features. Our results illustrate instead that familiar categories are reflected on the single-neuron level, much as physical attributes of stimuli are.

This ability to carve category membership into the tuning of single neurons might allow for the quick and effortless classification of familiar objects. We also observed neuronal correlates of category match/non-match effects, suggesting a role for the PFC in these judgments, and that neuronal correlates of categories and category judgments waned or changed on error trials, suggesting that PFC activity was directly related to task performance.

Category information in the PFC is likely to depend on inputs to the PFC from the ITC; the traditional roles of the PFC and ITC are in cognitive functions versus object vision and recognition, respectively (Robbins, 1998; Baddely, 1998; Gross, 1973; Miyashita 1993). PFC lesions or reversible inactivations in monkeys cause deficits in attention, working memory and response inhibition (Mishkin, 1957; Gross and Weiskrantz, 1962; Mishkin et al., 1969; Goldman and Rosvold, 1970; Goldman et al., 1971; Passingham, 1975; Mishkin and Manning, 1978; Funahashi et al., 1993; Dias et al. 1996) but usually spares object recognition, long term memory and "high level" visual analysis. By contrast, ITC damage causes deficits in visual discrimination, recognition and learning (Kluver and Bucy, 1938, 1939; Blum et al., 1950; Mishkin, 1954; Mishkin and Pribram, 1954; Mishkin, 1966) and category-specific agnosias (e.g. for faces) in humans (Damasio et al., 1982).

Indeed, studies of ITC and PFC neuronal properties support their respective roles in object recognition and executive functions (Logothetis and Sheinberg, 1996; Miller and Cohen, 2001). Since the seminal work of Gross and coworkers in ITC, who reported a small population of ITC “face cells”, numerous studies have shown that ITC neurons show selectivity for objects that cannot be explained by sensitivity to low-level features, such as orientation or color (Gross, 1972; Desimone et al., 1984; Perret et al., 1992; Tanaka et al., 1991; Kobatake and Tanaka

1994). There has been some recent evidence that suggests that these neurons play a direct role in categorization. Vogels (1999) recorded from ITC in monkeys trained to categorize stimuli as tree vs. non-tree or fish vs. non-fish and found that many neurons were selectively activated by the trained class (photographs of trees or fish) but not by distracter objects (photos of household objects or scenes containing neither trees nor fish). Kreiman et al. (2000) recorded from medial temporal lobe neurons in epileptic human patients while they classified stimuli into nine categories (e.g. faces, cars, food) and found neurons that selectively responded to stimuli from one of the categories.

However, it has not been clear whether these neurons encode the category membership of stimuli or, rather, their physical appearance. With a large, amorphous set of stimuli (such as trees or food), the category boundaries are unknown and the sharp transitions that are diagnostic of categories cannot be evaluated independently of stimulus similarity. Hence, neuronal selectivity for, say, trees could reflect the fact that trees look more like one another than other stimuli. Our results indicate that PFC neurons can convey the category of stimuli irrespective of their physical appearance.

Our results might reflect a relative specialization of the PFC in encoding category membership; because categories are defined by their functional relevance, they might be preferentially represented in the PFC, a cortical region specialized for behaviorally-related functions. Indeed, a recent theory of object recognition, suggests that category tuning in the PFC could arise from converging inputs from ITC neurons that are stimulus, but not category, tuned (Riesenhuber and Poggio, 2000). On the other hand, a recent study by Tomita et al. (1999) suggested that recall of long-term visual memories involved top-down signals from the PFC that activate representations stored in the ITC. Similar mechanisms might underlie the retrieval of category information stored in the ITC. This may be discerned by directly comparing the relative timing and the tuning properties of PFC and ITC neurons.

From a computational point of view, the categorization task learned by the monkeys, while subjectively not easy, is also not intrinsically difficult. Artificial classifiers, including a biologically plausible model of recognition in cortex (Riesenhuber & Poggio, 1999), can easily be trained to do the task using the same training and choice images employed in our experiments. An interesting question is to study the categorization performance of a monkey and of its PFC neurons if only a limited set of training examples were used. Examining their ability to generalize to the new images in comparison with model performance may give us clues about underlying mechanisms.

In sum, our results have provided insight into how perceptual categories are encoded in the PFC, a brain area that receives the outputs of sensory cortex and uses them to guide goal-directed behaviors. How and whether category membership is encoded in sensory systems and the respective roles of the PFC and visual areas like the ITC in representing and storing category information remains to be determined.

ACKNOWLEDGEMENTS

We thank C. Shelton for the morphing software and K. Anderson, W. Asaad, M. Mehta, A. Nieder, A. Pasupathy, J. Wallis and M. Wicherski for valuable comments, help and discussions. This work was supported by a NINDS grant, a NSF-KDI grant, the RIKEN-MIT Neuroscience Research Center, a McDonnell Pew Fellowship (M.R.), the Whitaker Chair (T.P.), and the Class of 1956 Chair (E.K.M.).

FIGURE LEGENDS

Figure 1 Organization of stimulus set. A. The 6 prototype images and 12 morph lines. The sample stimulus set was composed of 54 unique images: 6 prototypes (as shown), 4 images evenly placed (20%, 40%, 60%, 80%) along the 9 lines (in red) connecting each cat to each dog prototype, and 2 images (at 40% and 60%) along each of the 6 lines (in blue) between prototypes of the same category (with respect to the 2-class boundary). B. An example of the morphs generated between the C1 and D1 prototypes.

Figure 2 Task design and behavioral performance. The trial began with central fixation (500 ms) after which a sample stimulus appeared at the center of gaze for 600 ms. This was followed by a one second delay and then by a choice stimulus (600 ms). If the category of the choice matched that of the sample, monkeys had to release a lever to the choice stimulus within 600 ms of its presentation to obtain a fruit juice reward. If the choice was a non-match, there was another delay interval (600 ms) followed by a presentation of a match, which required a lever release for a reward. There were an equal number of match and non-match trials and they were randomly interleaved.

Figure 3 Average performance of both monkeys during neurophysiological recordings for the two-category task. Red bars indicate the percent of samples classified as cat and blue bars the percent classified as dog.

Figure 4 Psychophysical performance for the two and three-category tasks. For both the two-category (A) and three-category (B) tasks, the monkeys' error rates did not increase linearly as stimuli approached the category boundary, but changed more sharply at the category boundary. The sigmoidal shape for classification in both tasks indicates that the monkeys responded to the stimuli categorically.

Figure 5 Eye movements and degraded stimuli during behavioral testing. A. An example of monkey A's (left) and monkey B's (right) eye movements during the sample period of a single trial superimposed upon the sample stimulus shown on that trial. Monkeys were allowed to gaze freely at the stimuli during behavioral testing as fixation was not required. During neurophysiological recordings, monkeys were always required to maintain fixation within a ±2° window. B. Behavioral performance to degraded stimuli.

Figure 6 Anatomical location of recording sites and category selective neurons in both monkeys. A-anterior, P-Posterior, D-dorsal, V-ventral. The recording sites at which cat or dog selective neurons were found are indicated with an "X" or "O", respectively. Each location at which no category selective neurons were encountered is marked by a black dot. There was no obvious topography to task-related neurons.

Figure 7 Single neuron examples (two-category task). A. The average activity of a single neuron that showed greater activity to dogs during the memory delay. Each histogram trace represents the neuron's average activity to stimuli at each of the six morph levels. B. The average activity of a single neuron that showed greater activity to cats during the late sample and early delay period.

Figure 8 Average neuronal response to preferred and non-preferred categories during the sample (A) and delay (B). Each bar represents the population's average normalized response to stimuli at each of the six morph levels. The error bars represent the standard error of the mean.

Figure 9 Distribution of two-category index values across the population of 67 and 32 stimulus selective neurons during the sample (A) and delay (B) epochs, respectively. Positive values indicate larger differences in neuronal firing to samples across the category boundary than within a category. Negative values indicate larger differences within category than between categories.

Figure 10 Time course of category selectivity. A. Category selectivity across the population of 96 category selective neurons was computed using a sliding ROC analysis (see Methods). The ROC values for all 96 neurons were sorted from minimum to maximum for each time bin independently. Higher ROC values indicating stronger category tuning. Time is aligned to the end of the 200 ms sliding window (i.e. the values at time = 0 indicates the ROC values during the –200 to 0 time epoch relative to sample onset).

Figure 11 An example of a single neuron recorded during the three-category task that showed selectivity for the newly learned three categories (A) but not the old (now irrelevant) two categories (B).

Figure 12 Distribution of three-category (A: sample, B: delay) and two-category (C: sample, D: delay) index values across the entire population of 103 neurons recorded during the three-category task. The index can range from -1 to 1. Positive values indicate larger differences in neuronal firing between categories than within categories. Negative values indicate larger differences within categories than between categories. The two asterisks superimposed on figures A and B indicate significantly positive values of the category index (one tailed t-test, $P < 0.001$).

Figure 13 Single neuron examples of match/non-match effects. Neuronal activity is grouped according to the category of the choice stimulus and whether or not the category of the choice stimulus matched the category of the sample. A. An example of a neuron that showed enhanced activity to category matches. B. Enhanced activity to category non-matches. C. A neuron that showed enhanced activity to a cat choice stimulus when it was a match and a suppressed response to a cat non-match. It did not differentiate between match and non-match trials when the choice stimulus was a dog.

Figure 14 Comparison of neuronal selectivity on correct and error trials. Two asterisks indicate significance (evaluated by a t-test) at $P < 0.001$. One asterisk indicates significance at $P < 0.05$. The average response of all category selective neurons during the sample ($N = 74$) (A) and delay ($N = 51$) (B) phases is shown for correct and incorrect trials. Each neuron's preferred category was determined by the category that evoked greater average firing rates during correct trials. (c) The average activity to choice stimuli for all match/non-match selective neurons ($N = 25$) for correct and incorrect trials. Each neuron's preferred response (match or non-match) was determined according to the group that elicited greater average firing rates during correctly executed trials.

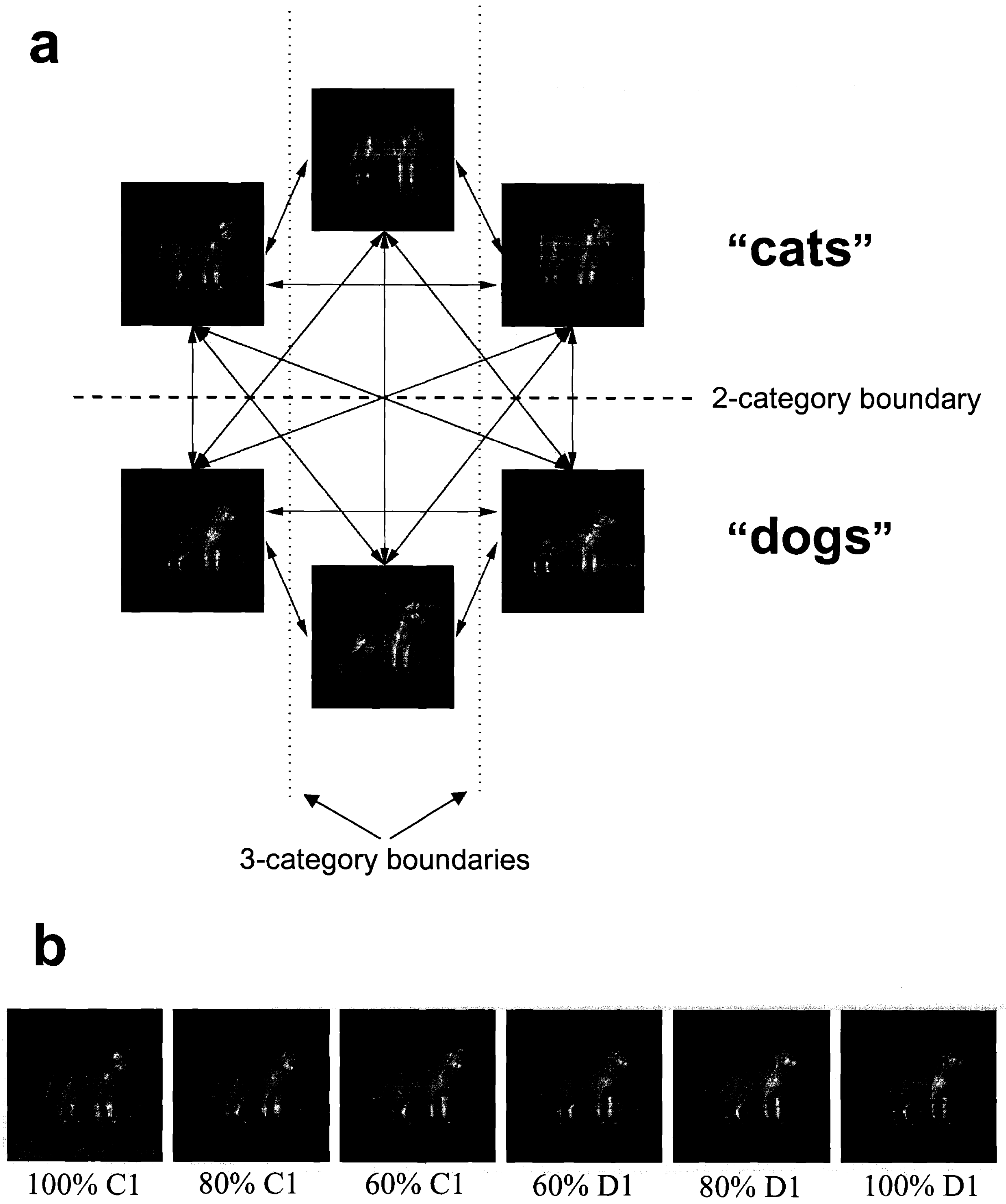

a
"cats"
2-category boundary
"dogs"
3-category boundaries
b
100% C1
80% C1
60% C1
60% D1
80% D1
100% D1

Figure 1

Figure 2

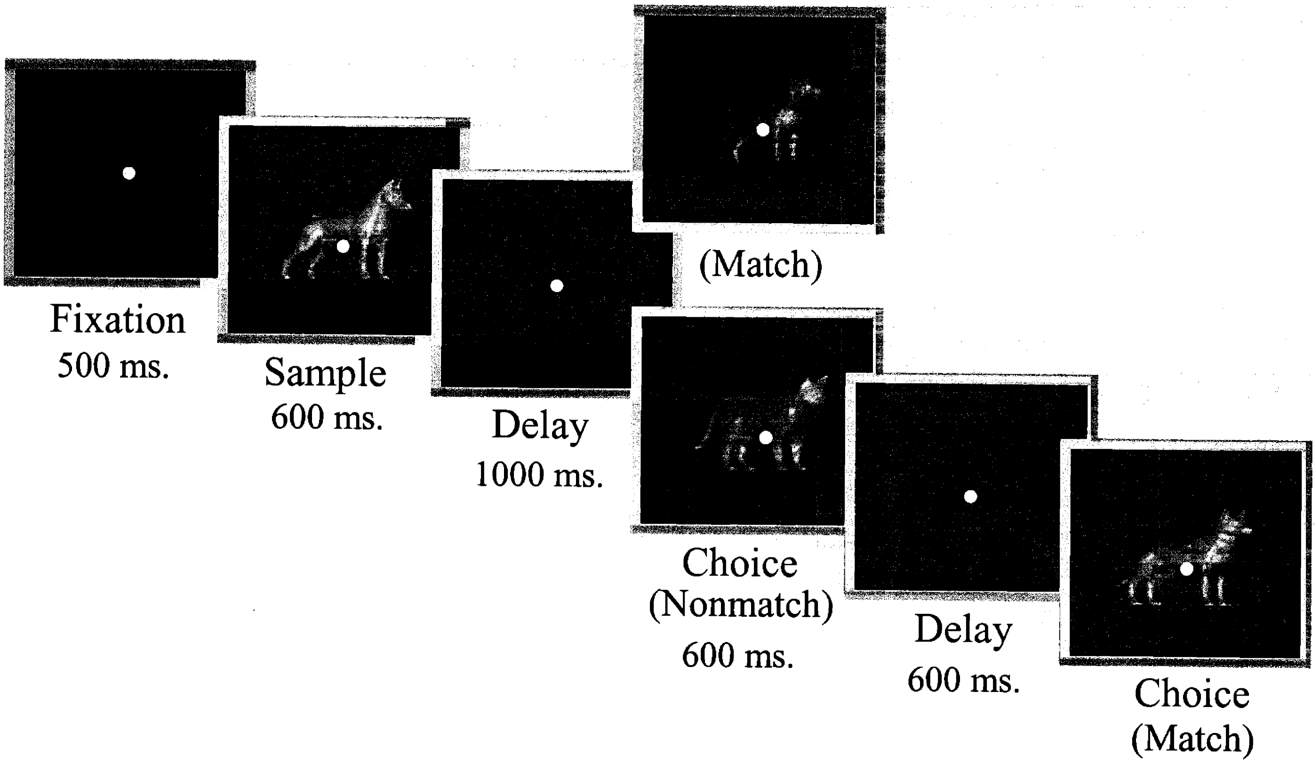

Fixation
500 ms.
Sample
600 ms.
Delay
1000 ms.
(Match)
Choice
(Nonmatch)
600 ms.
Delay
600 ms.
Choice
(Match)

Figure 3

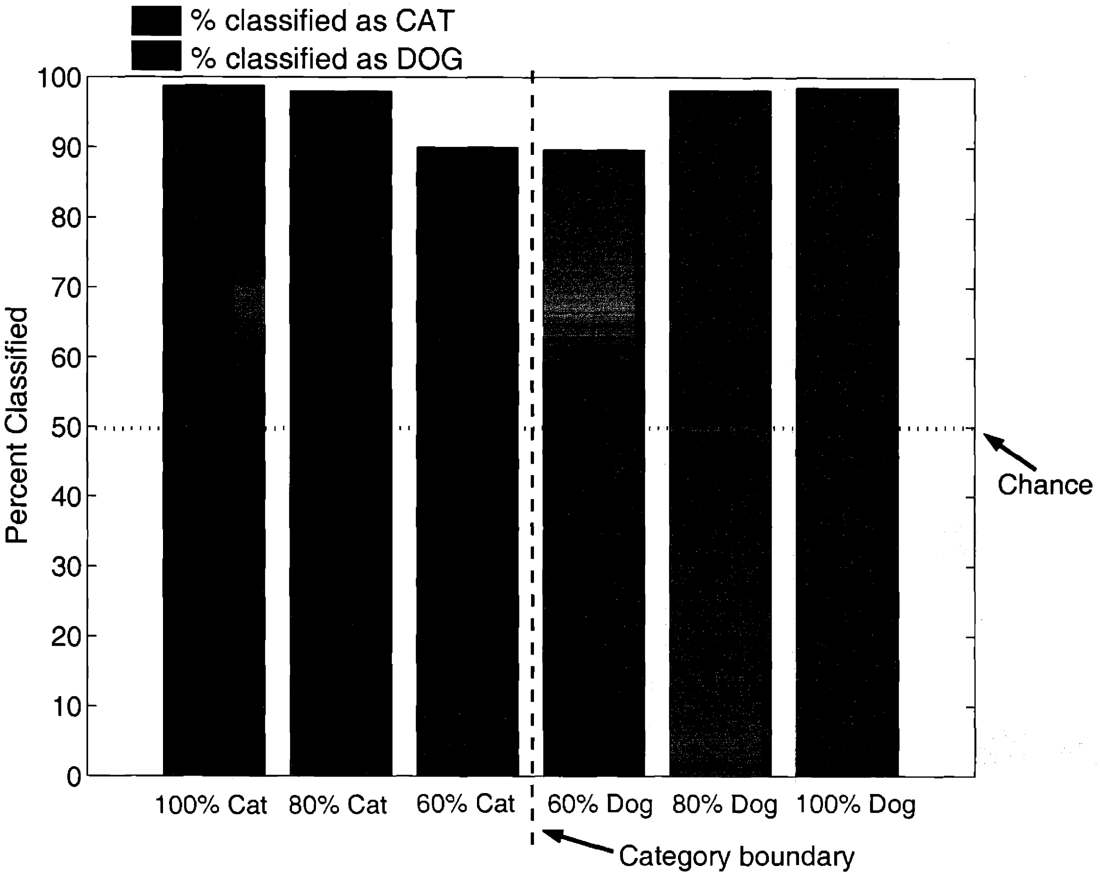

% classified as CAT
% classified as DOG
Percent Classified
100
90
80
70
60
50
40
30
20
10
0
100% Cat
80% Cat
60% Cat
60% Dog
80% Dog
100% Dog
Chance
Category boundary

Figure 4

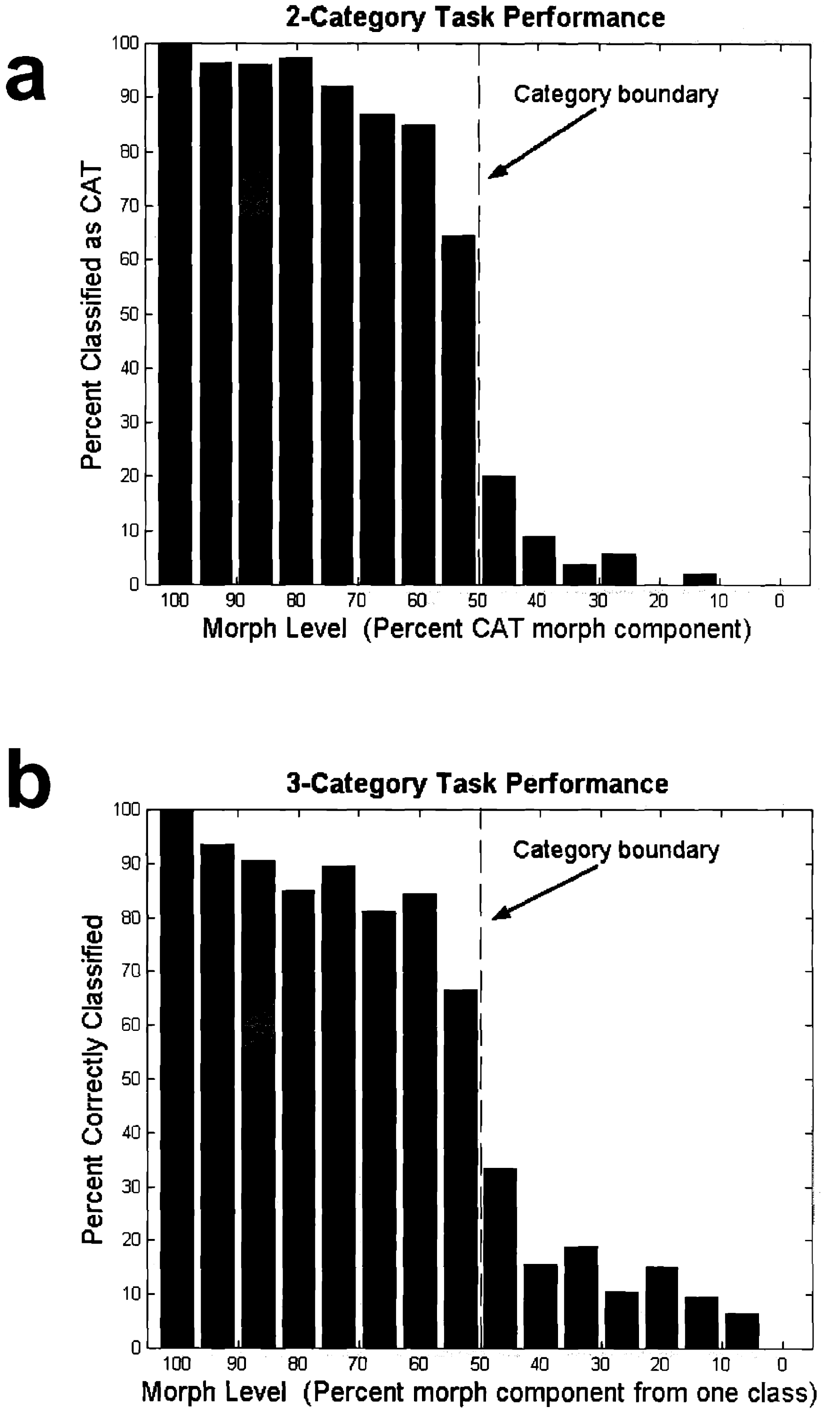
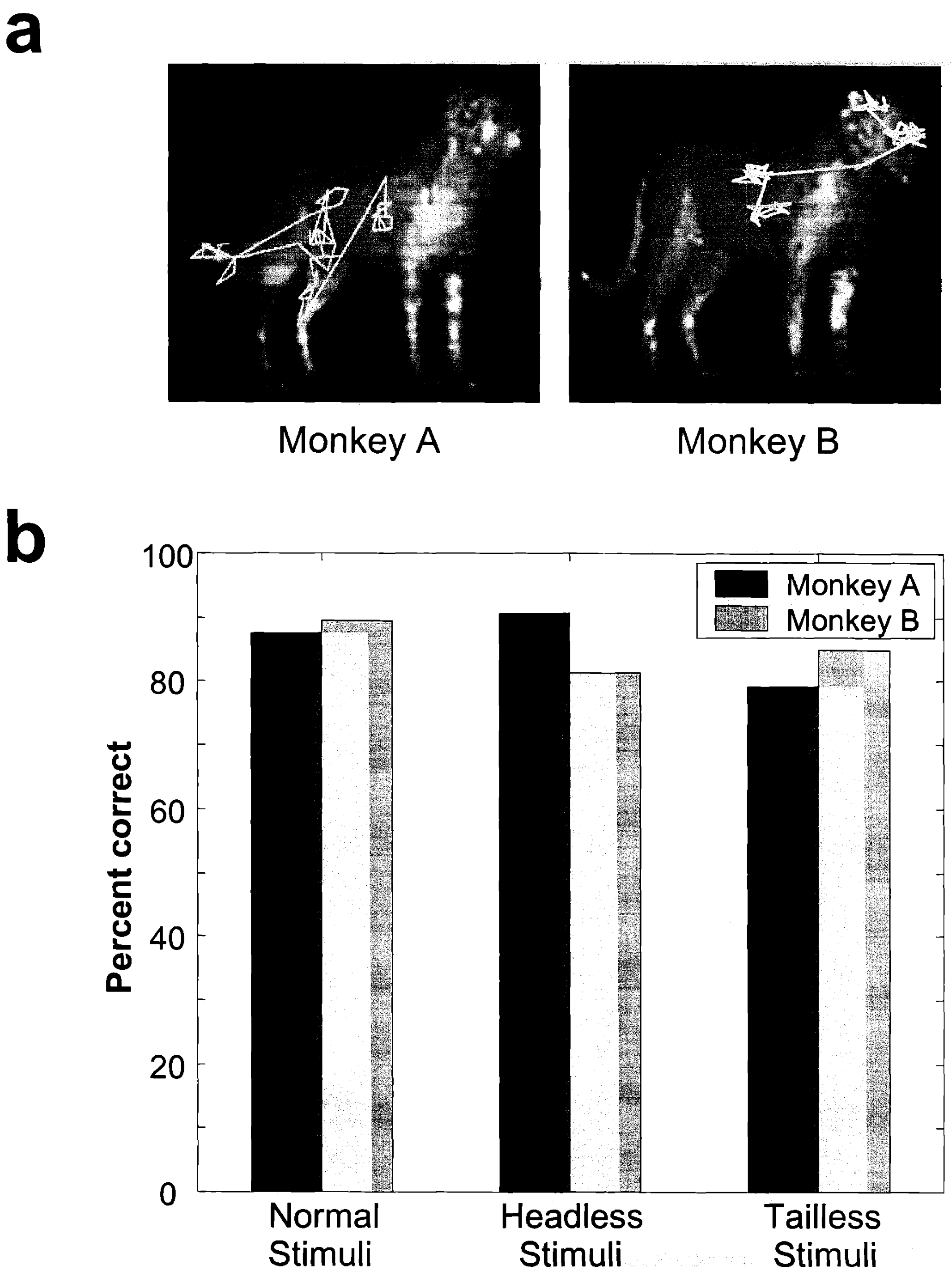

a
2-Category Task Performance
Percent Classified as CAT
100
90
80
70
60
50
40
30
20
10
0
Category boundary
100
90
80
70
60
50
40
30
20
10
0
Morph Level (Percent CAT morph component)
b
3-Category Task Performance
Percent Correctly Classified
100
90
80
70
60
50
40
30
20
10
0
Category boundary
100
90
80
70
60
50
40
30
20
10
0
Morph Level (Percent morph component from one class)

Figure 5


Monkey A

Monkey B

Figure 6

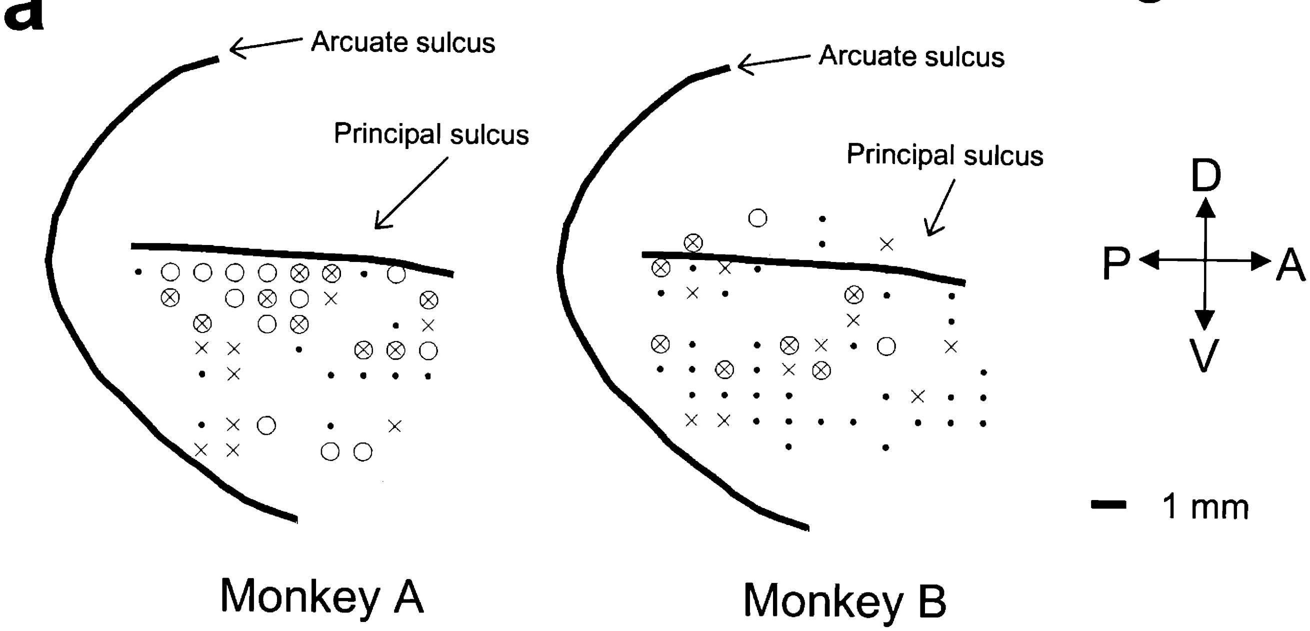

a
Arcuate sulcus
Principal sulcus
Monkey A
Arcuate sulcus
Principal sulcus
Monkey B
D
P
A
V
1 mm

Figure 7

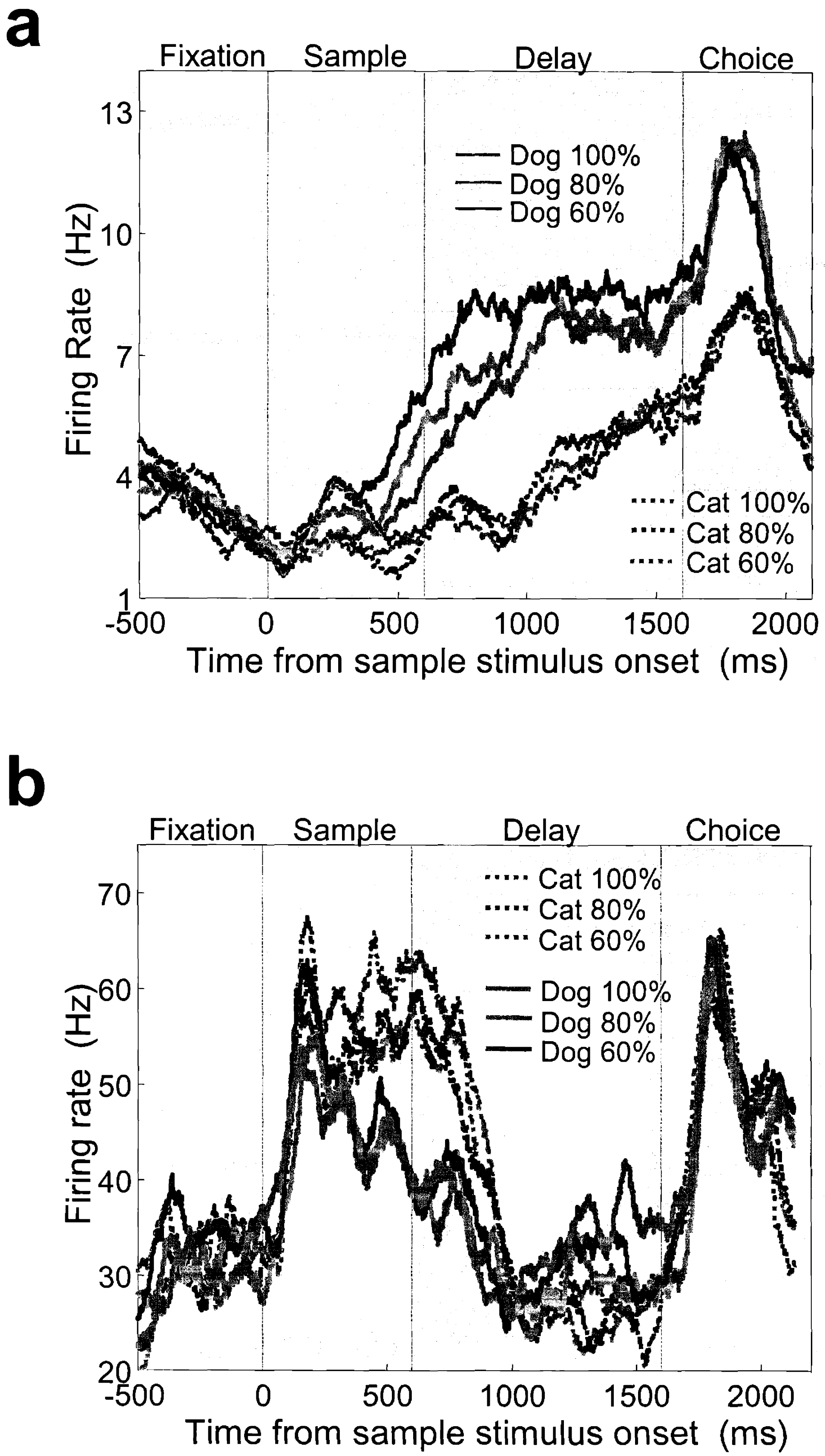

a
Fixation
Sample
Delay
Choice
Dog 100%
Dog 80%
Dog 60%
Cat 100%
Cat 80%
Cat 60%
Firing Rate (Hz)
1
4
7
10
13
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
b
Fixation
Sample
Delay
Choice
Cat 100%
Cat 80%
Cat 60%
Dog 100%
Dog 80%
Dog 60%
Firing rate (Hz)
20
30
40
50
60
70
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)

Figure 8

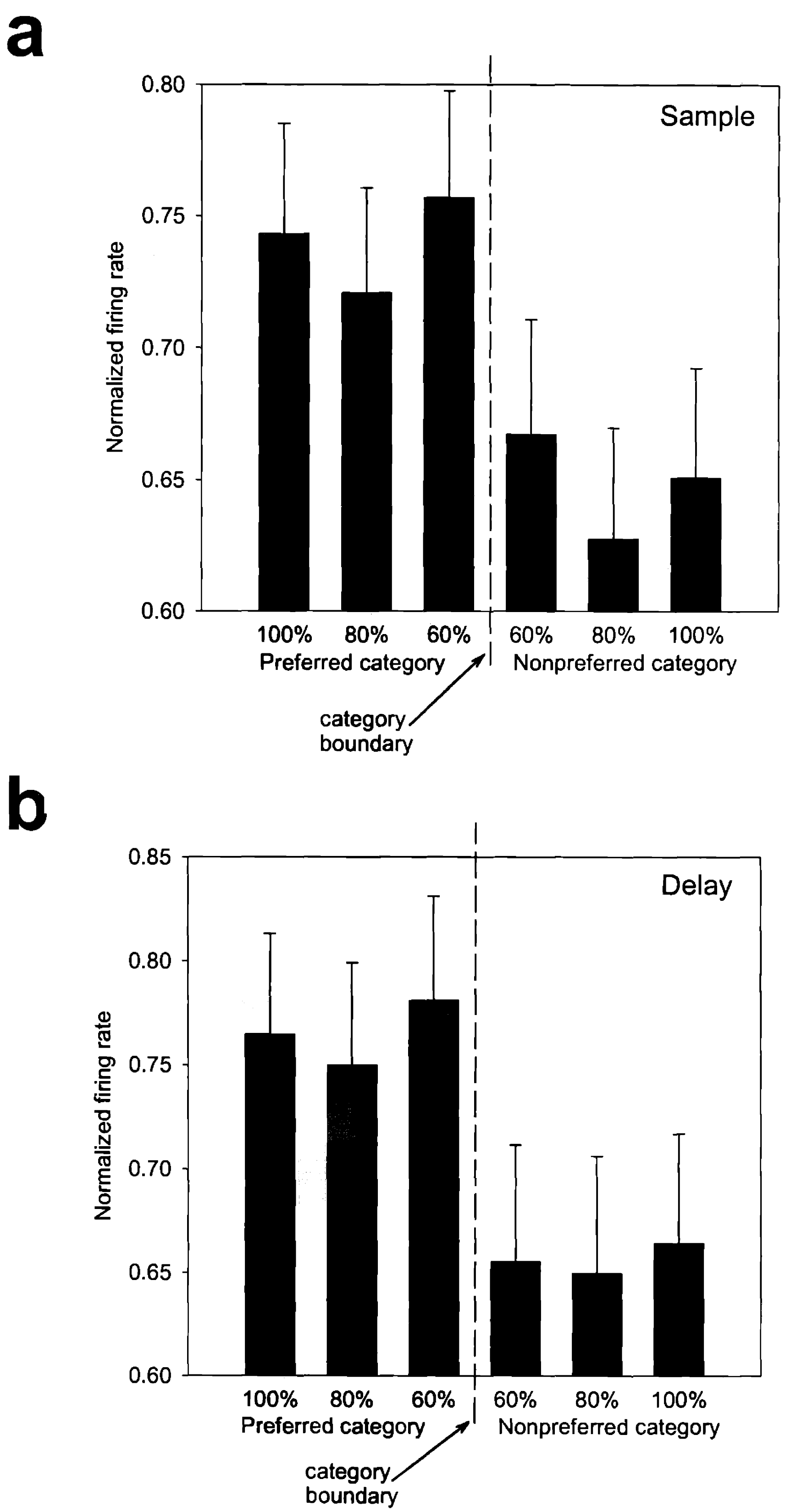

a
0.80
0.75
0.70
0.65
0.60
Normalized firing rate
Sample
100%
80%
60%
Preferred category
60%
80%
100%
Nonpreferred category
category boundary
b
0.85
0.80
0.75
0.70
0.65
0.60
Normalized firing rate
Delay
100%
80%
60%
Preferred category
60%
80%
100%
Nonpreferred category
category boundary

Figure 9

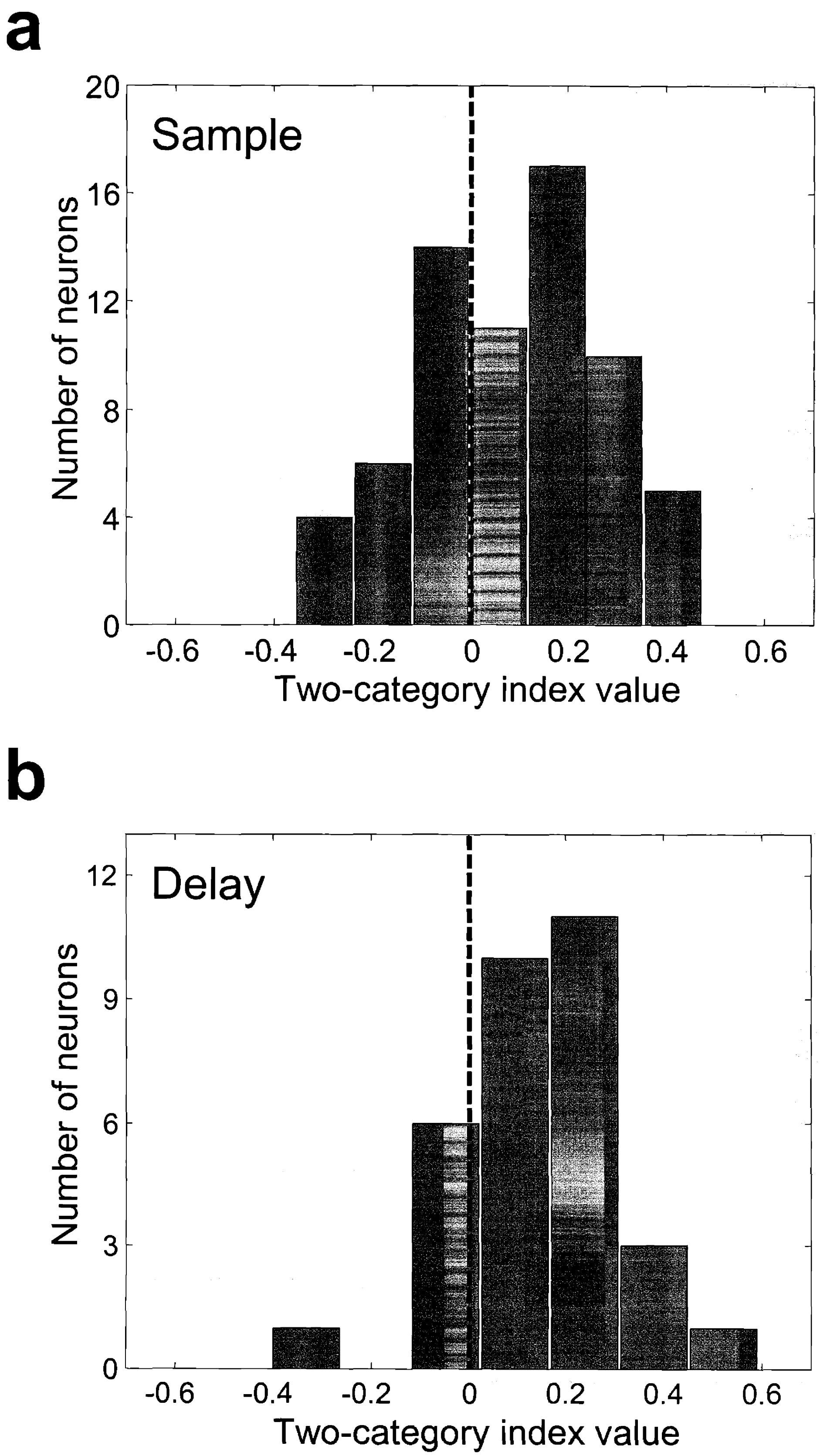

a
Sample
Number of neurons
20
16
12
8
4
0
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Two-category index value
b
Delay
Number of neurons
12
9
6
3
0
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Two-category index value

Figure 10

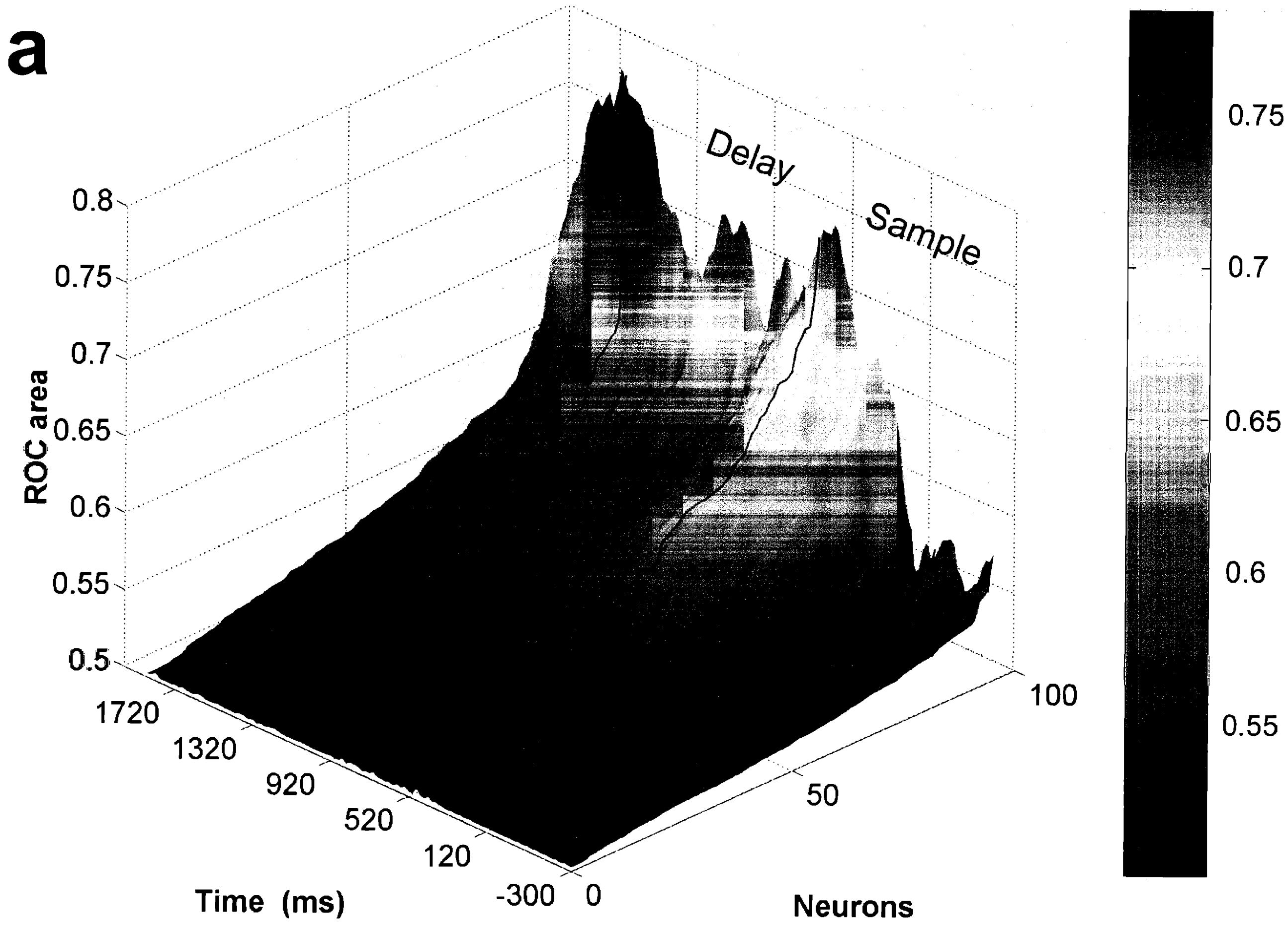
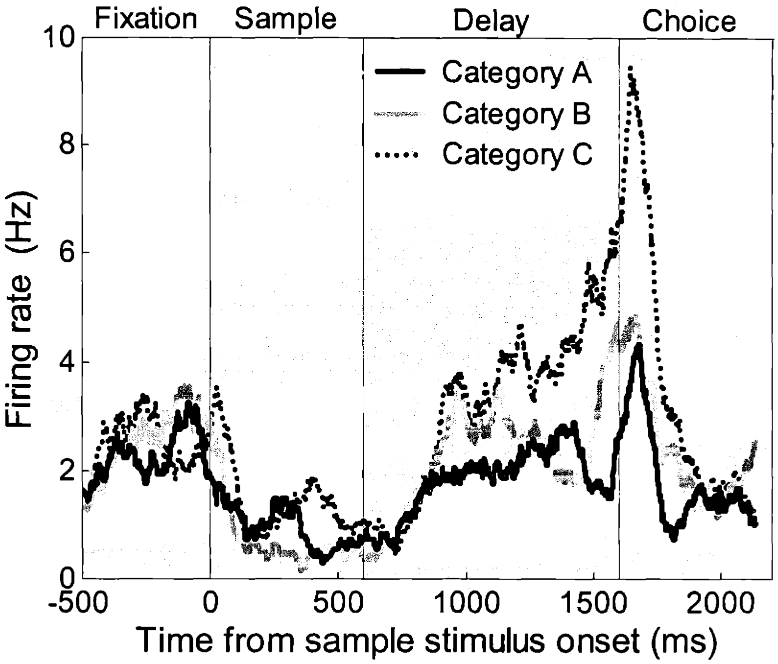

a
Delay
Sample
ROC area
0.8
0.75
0.7
0.65
0.6
0.55
0.5
1720
1320
920
520
120
-300
Time (ms)
0
50
100
Neurons
0.75
0.7
0.65
0.6
0.55

Figure 11

a


Fixation
Sample
Delay
Choice
Category A
Category B
Category C
Firing rate (Hz)
0
2
4
6
8
10
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)

b

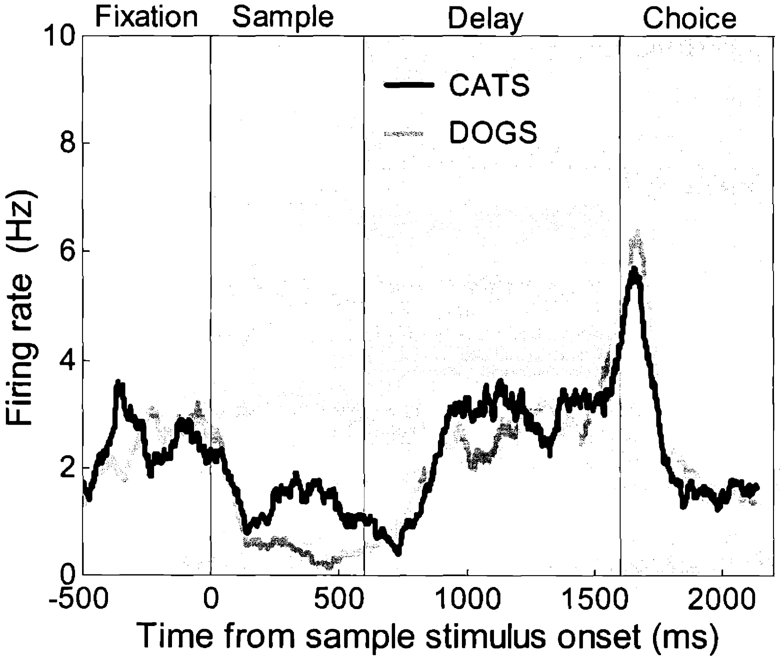
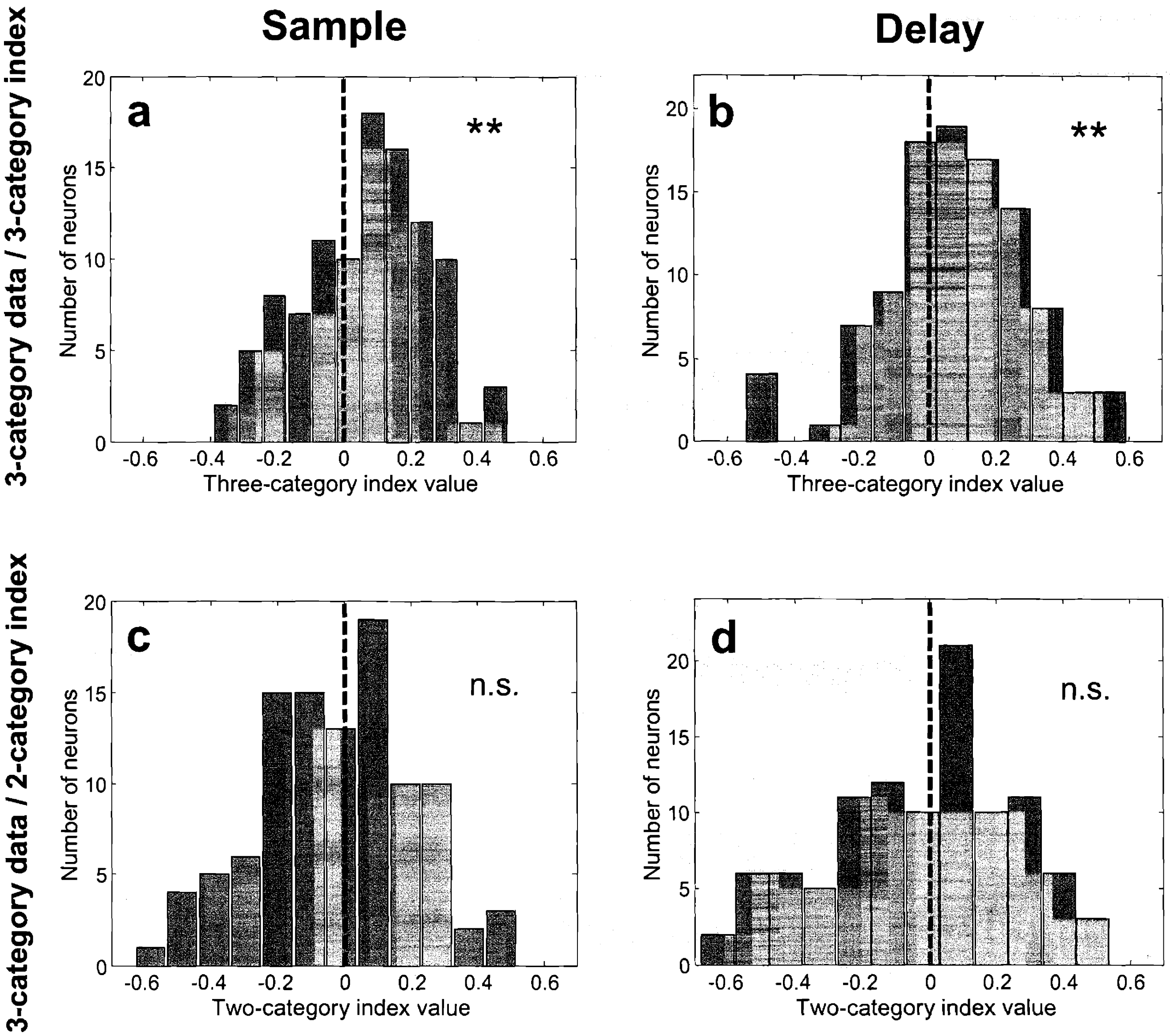

Fixation
Sample
Delay
Choice
CATS
DOGS
Firing rate (Hz)
0
2
4
6
8
10
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)

Figure 12



Sample
Delay
3-category data / 3-category index
3-category data / 2-category index
a
b
c
d
**
**
n.s.
n.s.
Number of neurons
0
5
10
15
20
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Three-category index value
Two-category index value

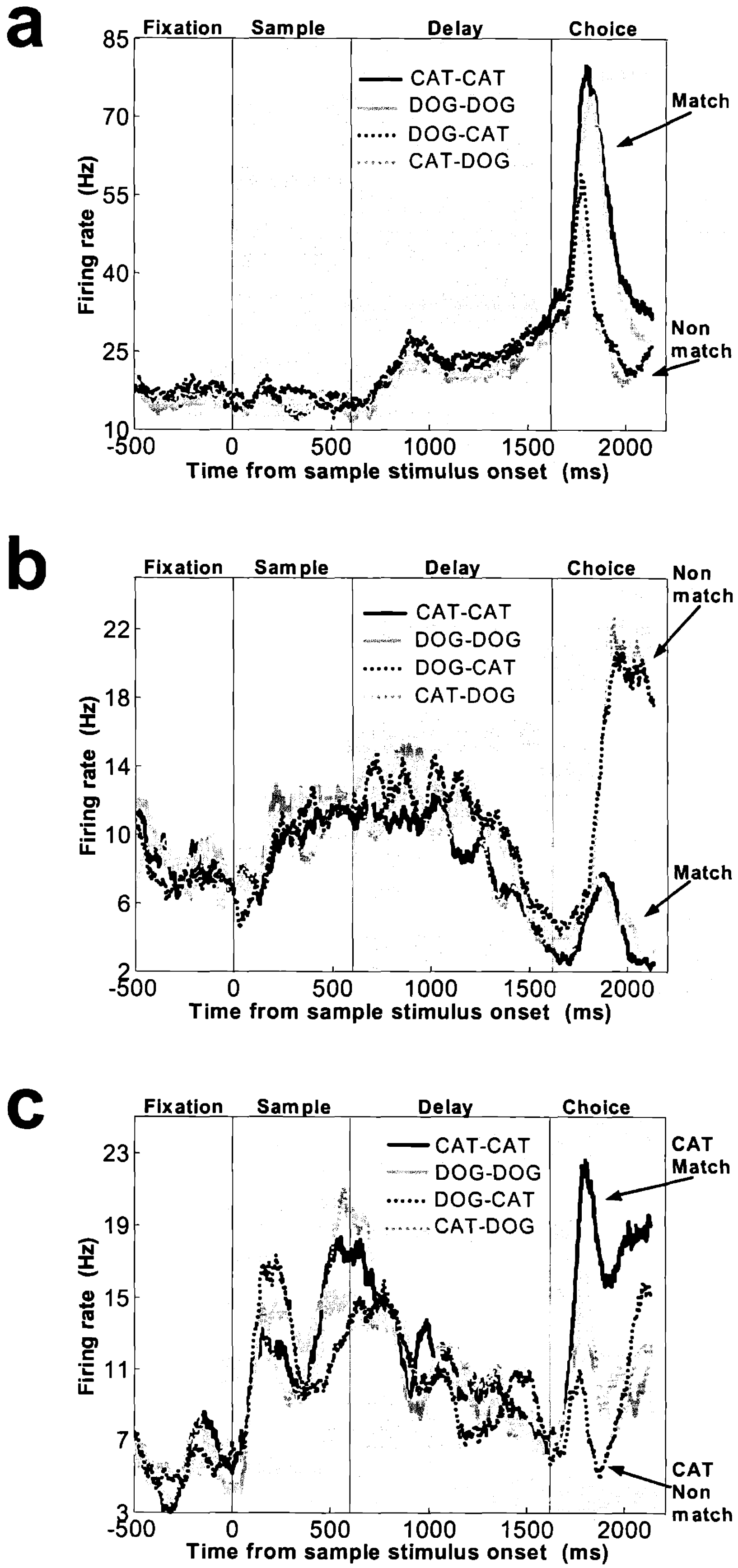

a
Fixation
Sample
Delay
Choice
CAT-CAT
DOG-DOG
DOG-CAT
CAT-DOG
Match
Non match
Firing rate (Hz)
85
70
55
40
25
10
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
b
Fixation
Sample
Delay
Choice
CAT-CAT
DOG-DOG
DOG-CAT
CAT-DOG
Non match
Match
Firing rate (Hz)
22
18
14
10
6
2
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
c
Fixation
Sample
Delay
Choice
CAT-CAT
DOG-DOG
DOG-CAT
CAT-DOG
CAT Match
CAT Non match
Firing rate (Hz)
23
19
15
11
7
3
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)

Figure 13

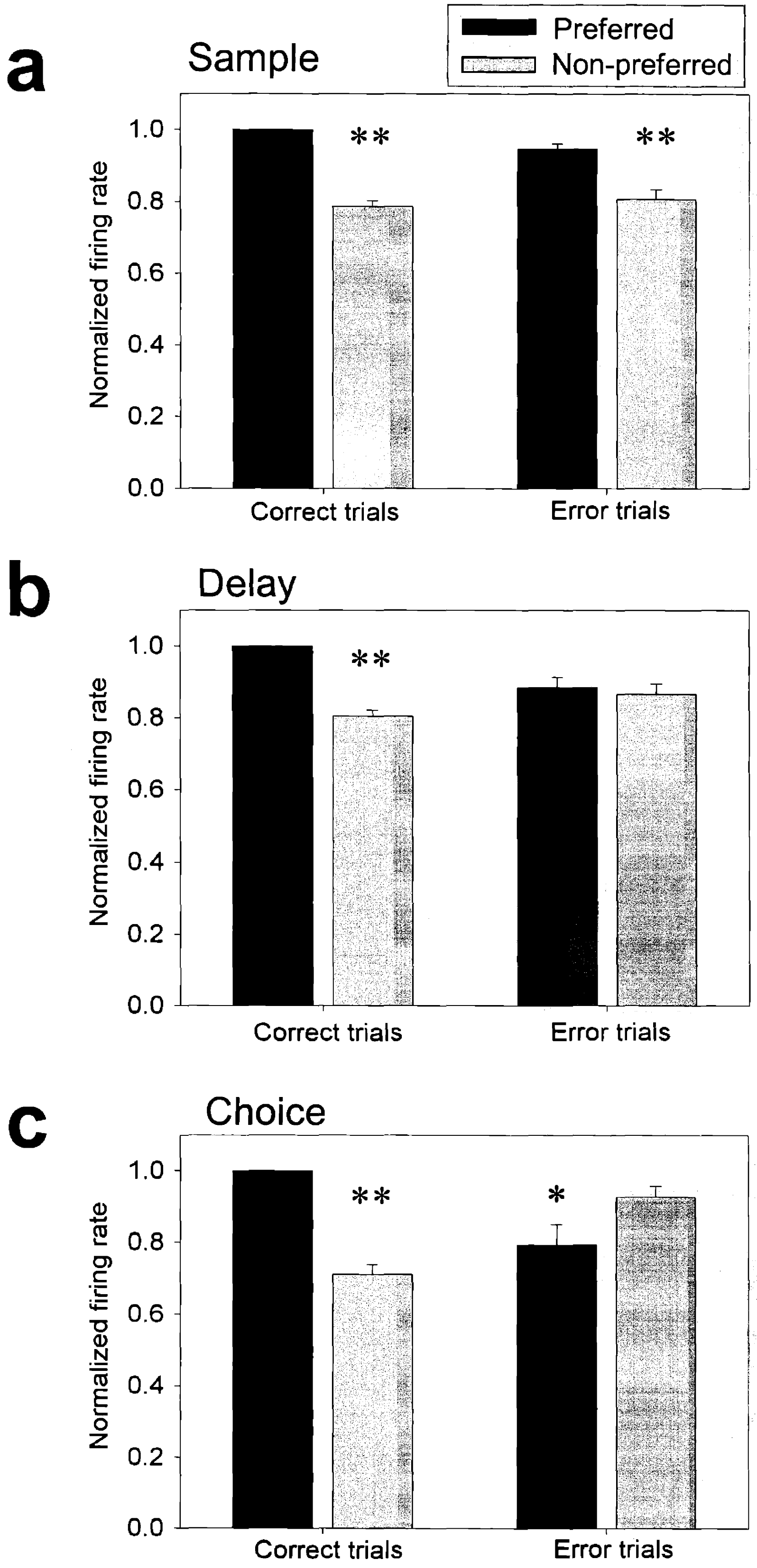

a
Sample
Preferred
Non-preferred
Normalized firing rate
1.0
0.8
0.6
0.4
0.2
0.0
**
**
Correct trials
Error trials
b
Delay
Normalized firing rate
1.0
0.8
0.6
0.4
0.2
0.0
**
Correct trials
Error trials
c
Choice
Normalized firing rate
1.0
0.8
0.6
0.4
0.2
0.0
**
*
Correct trials
Error trials

Figure 14

REFERENCES

Baddley A and Della Sala S. Working memory and executive control. In: *The Prefrontal Cortex* (Roberts AC, Robbins TW, Weiskrantz L., ed.). Oxford: Oxford University Press, 1998.

Beale JM and Keil FC. Categorical effects in the perception of faces. *Cognition* 57: 217-239, 1995.

Beymer D and Poggio T. Image representations for visual learning. *Science* 272: 1905-1909, 1996.

Booth MC and Rolls ET. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8: 510-523, 1998.

Bruce C, Desimone R, and Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus in the macaque. *J Neurophysiol* 46: 369-384, 1981.

Blum JS, Chow KL, and Pribram KH. A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *J Comp Neurol* 93: 53-100, 1950.

Damasio AR, Damasio H, and Van Hoesen GW. Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology* 32: 331-341, 1982.

Desimone R, Albright TD, Gross CG, and Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051-2062, 1984.

Dias R, Robbins TW, and Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 380: 69-72, 1996.

Fabre-Thorpe M, Richard G, and Thorpe SJ. Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9: 303-308, 1998.

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK. Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science* 291: 312-316, 2001.

Funahashi S, Bruce CJ, and Goldman-Rakic. Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". *J Neurosci* 13: 1479-1497, 1993.

Fuster JM. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe.* Philadelphia: Lippincott-Raven Publishers, 1997, p. 66-101.

Goldman-Rakic PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of Physiology: The Nervous System.* (Plum F, ed),. Bethesda: Am. Physio. Soc., 1987.

Goldman PS and Rosvold HE. Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Exp Neurol* 27: 291-304, 1970.

Goldman PS, Rosvold HE, Vest B, and Galkin TW. Analysis of the delayed-alternation deficit produced by dorsolateral prefrontal lesions in the rhesus monkey. *J Comp Physiol Psychol* 77: 212-220, 1971.

Green DM and Swets JA. *Signal Detection Theory and Psychophysics.* New York: J. Wiley and Sons, 1966.

Gross CG, Rocha-Miranda CE and Bender DB. Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophsiol 35: 96-111, 1972.*

Gross CG. Visual functions of inferotemporal cortex. In: *Handbook of Sensory Physiology,* Vol. VII/3B. H. Autrum, R. Jung, W. Lowenstein, D. Mckay, and H.-L. Teuber, eds. Berlin: Springer-Verlag, 1973.

Gross CG and Wieskrantz L. Evidence for dissociation of impairment on auditory discrimination and delayed response following lateral frontal lesions in monkeys. *Exp Neurol* 5: 453-476, 1962.

Kluver H and Bucy L. An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness." *J Psychol* 5: 33-54, 1938.

Kluver H and Bucy L. Preliminary analysis of functions of the temporal lobes in monkeys. *Arch Neurol Psychiatry* 42: 979-1000, 1939.

Kobatake E and Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856-867, 1994.

Kobatake E, Wang G, and Tanaka K. Effects of shape discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80: 324-330, 1998.

Kreiman G, Koch C, and Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci* 3: 946-953, 2000.

Lieberman AM, Cooper FS, Shankweiler DP, and Studdert-Kennedy M. Perception of the speech code. *Psychol Rev*, 74: 431-461, 1967.

Logothetis NK, Pauls J, and Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552-563, 1995.

Logothetis NK and Sheinberg. Visual Object Recognition. *Annu Rev Neurosci* 19: 577-621, 1996.

Miller EK. The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 1: 59-65, 2000.

Miller EK and Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24: 167-202, 2001.

Miller, EK, Erikson, CA, and Desimone, R. Activity of neurons in anterior inferior temporal cortex during a short term memory task. *J Neurosci* 13: 1460-1478, 1993.

Miller EK, Erickson CA, and Desimone R. Neuronal mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16: 5154-5167, 1996.

Mishkin M. Visual discrimination performance following partial ablations of the temporal lobe. II. Ventral surface vs. hippocampus. *J Comp Physiol Psychol* 47: 187-193, 1954.

Mishkin M. Effects of small frontal lesions on delayed alternation in monkeys. *J Neurophysiol* 20: 615-622, 1957.

Mishkin M. Visual mechanisms beyond the striate cortex. In: *Frontiers in Physiological Psychology*. R.W. Russell, ed. New York: Academic Press, 1966.

Mishkin M. and Manning FJ. Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res* 143: 313-323, 1978.

Mishkin M. and Pribram K. Visual discrimination performance following partial ablations of the temporal lobe. I. Ventral vs. lateral. *J Comp Physiol Psychol* 47: 14-20, 1954.

Mishkin M, Vest B, Waxler M, and Rosvold HE. A re-examination of the effects of frontal lesions on object alternation in monkeys. *Neuropsychologia* 7: 357-364, 1969.

Miyashita Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335: 817-820, 1988.

Miyashita Y. Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci* 16: 245-263, 1993.

Nichols AM, Ruffner TW, Sommer MA, and Wurtz RH. A Screw microdrive for adjustable chronic unit recording in monkeys. *J Neurosci Methods* 81: 185-188, 1998.

Orlov T, Yakovlev V, Hochstein S, and Zohary E. Macaque monkeys categorize images by their ordinal number. *Nature* 404: 77-80, 2000.

Passingham RE. Delayed matching after selective prefrontal lesions in monkeys (*Macacca Mulatta*). *Brain Res* 92: 89-102, 1975.

Perrett DI, Hietanen JK, Oram MW, and Benson PJ. Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans Roy Soc Lond B Biol Sci* 335: 23-30, 1992.

Perrett DI, Rolls ET, and Caan W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47: 329-342, 1982.

Riesenhuber M and Poggio T. A note on object class representation and categorical perception. *AI Memo 1679, CBCL Paper 183,* Artificial Intelligence Lab and Center for Biological & Computational Learning, Massachusetts Institute of Technology, 1999.

Riesenhuber M and Poggio T. Models of Object Recognition. *Nat Neurosci 3 supp.*: 1199-1204, 2000.

Robbins TW. Dissociating executive functions of the prefrontal cortex. In: *The Prefrontal Cortex* (Roberts AC, Robbins TW, Weiskrantz L., ed.). Oxford: Oxford University Press, 1998.

Roberts WA and Mazmanian DS. Concept at different levels of abstraction by pigeons, monkeys and people. *J Exp Psychol Anim Behav Process* 14: 247-260, 1988.

Scalaidhe SP, Wilson FA, and Goldman-Rakic PS. Face selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cereb Cortex* 9: 459-475, 1999.

Shelton C. Morphable Surface Models. *Int J Comp Vis* 38: 75-91, 2000.

Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109-139, 1996.

Tanaka K, Saito H, Fukada Y, and Moriya M. Coding visual images of objects in the inferotemporal cortex of monkeys. *J Neurophysiol* 66: 170-189, 1991.

Tolhurst DJ, Movshon JA, and Dean, AF. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res* 23: 775-785, 1983.

Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, and Miyashita Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401: 699-703, 1999.

Ungerleider LG, Gaffan D, and Pelak VS. Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp Brain Res* 76: 473-484, 1989.

Vogels R. Categorization of complex visual images by rhesus monkeys. *Eur J Neurosci* 11: 1223-1238, 1999.

Vogels R and Orban GA. How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. *J Neurosci* 10: 3542-3558, 1990.

Webster MJ, Bachevalier J, and Ungerleider LG. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb Cortex* 4: 470-483, 1994.

White IM and Wise SP. Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126: 315-335, 1999.

Wyttenbach RA, May ML, and Hoy RR. Categorical perception of sound frequency by crickets. *Science* 273: 1542-1544, 1996.

Responses of Inferior Temporal Cortex Neurons During Visual Categorization

David J. Freedman, Maximilian Riesenhuber, Tomaso Poggio, Earl K. Miller
Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, Center for Biological and Computational Learning and The Department of Brain and Cognitive Sciences
Cambridge, MA, USA, 02139

ABSTRACT

The ability to group stimuli into meaningful categories is a fundamental cognitive process. However, very little is known about how the brain solves this computationally difficult problem. To explore the neuronal basis of categorical perception, we trained monkeys to categorize computer-generated stimuli as "cats" and "dogs". A morphing system was used to systematically vary stimulus shape and precisely define a category boundary. We have previously shown that neuronal activity in the prefrontal cortex (PFC) reflected the category of visual stimuli. In this study, we recorded from 286 neurons in the inferior temporal cortex (ITC) and found many that were categorically tuned. A comparison of the pattern of results from the PFC and ITC suggest that the ITC is primarily involved in high-level visual analysis while the PFC may be more involved in using categorical information to guide behavior.

INTRODUCTION

The results of many physiological and neuropsychological studies have demonstrated that the inferior temporal cortex (IT) plays a critical role in high-level visual analysis and awareness (Gross, 1994; Tanaka, 1996; Sheinberg and Logothetis, 1997; Perrett et al., 1998) Damage to these cortical areas in monkeys and humans can result in impaired visual recognition and learning (Gross, 1973) and even category-specific agnosia in which a subject has a selective impairment for recognizing stimuli from a specific category (i.e. faces) (Nachson, 1995). Electrophysiological recordings from the temporal lobe have revealed neurons that are selectively activated by specific stimuli with invariance for scale, position and rotation (Logothetis, 1998) and have demonstrated that their responses are affected by visual learning and experience (Erickson et al., 2000; Booth and Rolls, 1998; Logothetis et al., 1995). The goal of this study is to determine whether the activity of IT neurons is strictly stimulus-specific or whether it can convey more abstract information about stimuli such as their category membership.

To address this question, we trained monkeys to categorize stimuli into two categories, “cats” and “dogs”. A large continuous set of stimuli was parametrically generated from six prototype images (three species of cats and three breeds of dogs) using a 3D morphing system (Beymer and Poggio, 1996; Shelton, 2000). Through training, an arbitrary category boundary was introduced that separated the two categories such that stimuli from the same category could be visually dissimilar (i.e. cheetah and tiger prototypes) while stimuli from different categories could be physically similar (i.e. morphs that are very close to, but on opposite sides of the category boundary).

In a previous study (Freedman et. al, 2001) using the same behavioral paradigm, we found that neurons in the lateral prefrontal cortex (PFC) reflected stimulus category. It is well established that the PFC receives direct inputs from IT cortex. Therefore, it is possible that the observed category tuning in the PFC could be due to categorical inputs from IT. To explore this possibility, we have conducted IT recordings in one of the monkeys used in our PFC study using the same 'cat' vs. 'dog' task. There were two primary goals of this study. The first was to determine whether IT neurons can become categorically tuned as a result of experience. Secondly, we wanted to compare the neuronal tuning properties between IT and PFC in order to elucidate their respective roles in the context of this behavioral paradigm. Our results from this study demonstrate that many IT neurons do reflect the category of visual stimuli and that the PFC and IT cortex likely play different roles in solving the categorization task: IT seems primarily involved in the rapid analysis of visual stimuli while PFC responses appear more suited for using this highly processed visual information to guide behavior.

MATERIALS AND METHODS

Subjects: One female adult rhesus monkey (Macacca mulatta) weighing 7.5 kg, was used in this study. This monkey was one of two monkeys used in our previous study of prefrontal cortex (Freedman et al., 2001). Using previously described methods (Miller et al., 1993), the monkey was implanted with a head bolt to immobilize the head during recording and a recording chamber. Eye movements were monitored and stored at rates of 60-120 frames per second using an infrared eye tracking system (IScan, Cambridge, MA). Electrode penetration sites were determined using structural magnetic resonance imaging scans obtained prior to surgery. The recording chamber was positioned stereotactically over the inferior temporal cortex such that the superior temporal sulcus and lateral convexity were readily accessible. All surgeries were performed under sterile conditions while the animal was anesthetized with isoflurane. The animal received postoperative antibiotics and analgesics and was always handled in accord with National Institutes of Health guidelines and the recommendations of the Massachusetts Institute of Technology Animal Care and Use Committee.

Recording Techniques: A stainless steel recording chamber (Crist Instruments. Damascus, MD.) was stereotaxically placed over the anterior inferior temporal cortex using coordinates obtained from structural MRI scans. The location of recording sites was estimated from the MRI scans and from the characteristic white/gray matter transitions encountered while lowering electrodes. The 20 mm. diameter recording chamber was centered over the right anterior medial temporal sulcus 15 mm. anterior to the ear canal. Neuronal recordings were conducted in the region between 14-19 mm lateral to the midline and 14-19 mm anterior from the intra-aural line. According to electrode track estimation using structural MRI scans, recordings were concentrated in lower bank of the superior temporal sulcus (areas TEa and TEm) and the ventrolateral surface of the temporal lobe between the superior temporal and anterior medial temporal sulci (areas TE1 and TE2). We utilized a hydraulic multielectrode manifold (FHC Engineering. Bowdoin, ME) that enabled independent delivery of four insulated tungsten microelectrodes into IT cortex through a single 23 gauge guide tube. This allowed us to isolate, on average, nearly two neurons per electrode. Neurons were not prescreened for task-related activity such as visual responsiveness or stimulus selectivity. Rather, we advanced each electrode until the activity of one or more neurons was well isolated and then began data collection. This procedure was used to ensure an unbiased estimate of neuronal activity.

Behavioral Task: The monkey was trained to perform a delayed match-to-category task in which it released a lever when the category of a test stimulus matched the category ("cat" or "dog") of a previously presented sample stimulus (FIGURE 1). This task was identical in all respects to the task used in our previously reported study of prefrontal cortex (Freedman et al., 2001).

The trial began when the monkey grasped a metal bar and fixated a small (0.3°) white spot at the center of a CRT screen. The monkey needed to maintain its gaze within a ±2° window around the fixation spot, which was presented for the entire duration of the trial. Following 500 ms of fixation, a sample stimulus was presented at the center of the screen for 600 ms, followed by a 1000 ms delay. Following the delay, a test stimulus was presented for 600 ms that was either a categorical match or non-match. In the event that the test stimulus was a non-match to the sample, a second delay (600 ms) followed, after which a second test stimulus (always matching the category of the sample) was presented. To receive a juice reward, the monkeys had to release the lever during the presentation of a test stimulus that matched the category of the previously presented sample stimulus. A 2000 ms to 3000 ms inter-trial interval followed correct trials. In the event of an incorrect trial due to an erroneous bar release or lapse in fixation, the monkey received no reward and waited an additional 3000 ms before beginning the next trial. Match and non-match trials were randomly interleaved. Monkeys often performed over 1000 correct trials per day. This balanced task design allows a decoupling of information about stimulus category from that related to the monkey's motor response; following onset of the sample stimulus, the monkey could do no better than guess at whether the upcoming test stimulus would be a categorical match or non-match.

Stimuli: Stimuli in the experiments were generated using a 3D object morphing system. A large continuous set of images was generated from six prototypes objects that spanned two natural categories, "cats" and "dogs" (FIGURE 2). From these three "cat" and three "dog" prototypes, novel images or "morphs" were generated whose identity was determined by the relative contribution of each of the six prototypes. The category boundary during the 2-category task was arbitrarily defined at the 50% cat/dog level so that the category membership of a morph was determined by which prototype class ("cat" or "dog") contributed most. Images that belonged to different categories were visually similar if they were close to but on opposite sides of the category boundary. Likewise, images that belonged to the same category could be visually dissimilar (i.e. the cheetah and housecat prototypes). Corresponding features on each of the prototypes were defined to ensure that images changed smoothly across a morph line without a sudden appearance, disappearance or distortion of features. Images were approximately 4 X 3 degrees of visual angle in size, had identical color, shading, orientation and were appropriately centered and scaled.

The sample stimulus set consisted of 54 unique images generated by the morphing system. In addition to the 6 prototype images, 4 morphs were generated at equally spaced intervals across each of the 9 morph lines that connect each "cat" prototype to each "dog" prototype. Along each of the within-class class lines that connect the "cat" prototypes to one another and likewise for the "dogs", one stimulus was generated at both the 40% and 60% positions. The set of choice stimuli consisted of 100 randomly generated morphs from each category. To ensure that errors during task performance were due to misclassification of the sample stimulus rather than the test stimulus, each choice stimulus was a minimum distance of 15% from the category boundary. The choice stimuli were randomly paired with sample stimuli of the appropriate categories to ensure that the monkeys could not solve the categorization tasks using a paired associate strategy.

Data Analysis: IT neurons' firing rates were calculated in four time epochs: baseline (fixation), sample, delay and test. These time epochs are identical to those used for the analysis of data collected from the PFC during our previous studies (Freedman et al., 2001). Hence, the results from the PFC and IT cortex can be directly compared. Neuronal activity during the baseline period was calculated over the 500 ms of fixation preceding sample presentation. Sample period activity was calculated over an 800 ms epoch beginning 100 ms after sample onset to account for the latency of IT neuronal responses. We included the first 300 ms following sample offset in the calculation of sample evoked activity as some IT neuronal responses occurred several hundred ms after stimulus onset and persisted several hundred ms after sample offset but not for the remainder of the delay. For the analysis of delay period activity, we calculated the average firing rate over an 800 ms epoch beginning 300 ms after sample offset. Analysis of neuronal responses to the first choice stimulus (when the monkey must decide whether or not to make a behavioral response) was restricted to a time window beginning 100 ms after its onset and ending 2 standard deviations before the monkeys average reaction time during each recording session. This was done in order to exclude any effects related to the execution of the motor response. We used data from correctly executed trials only. There were, on average, 12 correct trials per sample stimulus during neuronal recordings. As the monkeys performed at such a high rate of accuracy during recordings (>90% correct) there were not sufficient numbers of error trials to allow a complete analysis of neuronal activity during erroneous categorization.

To determine the reliability of category tuning, we employed a receiver-operator characteristics (ROC) analysis (Green and Swets 1966; Tolhurst et al., 1983; Vogels and Orban, 1990). The ROC analysis measures the degree of overlap between two distributions of values. Given two distributions of values A and B, we start by plotting for each possible firing rate the proportion of distribution A that exceeded this criterion versus the proportion of distribution B that exceeded it. Integration across the curve returns the a single number for that comparison, the ROC area (with a range of 0 to 1). ROC area values near 0.5 indicate large overlap between A and B. Values of 0 or 1 indication that the two distributions are entirely non-overlapping. The ROC analysis has several advantageous over mean and variance based statistics such as T-Tests or ANOVAs. First, it makes no assumptions about the two distributions, A and B, and thus returns an unbiased estimation of overlap. Second, it can be interpreted as the performance of an ideal observer in a two-way forced choice task; values of 0.5 indicate 50% correct classification (guessing) while values of 0 or 1 indicate error-free classification. Third, it is independent of neuronal firing rate and number of observations.

For some comparisons of category selectivity, we computed an index of categorical tuning. We computed the average within category difference (WCD) and between category difference (BCD) at equal distances across all morph lines that crossed the category boundary. The WCD was defined by computing the absolute difference between the 100%-80% morphs and 80%-60% morphs for both categories and averaging these four values. The BCD was computed simply by averaging the across-boundary differences between the 60% "cats" and 60% "dogs". By so doing, the WCD and BCD were computed over an equal morph distance of 20% and could be directly compared. A category index was computed according to the following equation:

$$C = (BCD - WCD) / (BCD + WCD)$$

The category index can have values ranging from –1 to 1. Positive values indicate a larger difference between categories while negative values reflect larger differences within a category than between categories. As the identical set of stimuli was used for the two and three category tasks, both the two and three-category tuning indices could be applied to the neuronal data obtained during both the two and three-category tasks.

Neuronal selectivity following onset of the test stimulus (during which the monkey must decide whether or not to release the lever) was evaluated using a two-way ANOVA (factor 1: trial type (match or non-match), factor 2: category of test stimulus).

RESULTS

The results from our study of the PFC have been previously presented in great detail (Freedman et al., 2001). Therefore, only the results from new analyses of that data and those that are crucial for comparisons with the IT cortex data will be described here.

Single neuron analysis

A total of 286 IT cortex neurons were recorded from the right hemisphere of one monkey during performance of the two-category DMC task. Visual responsiveness was evaluated by independently comparing sample and delay period spike-rates for all trials to that during baseline fixation using two paired-sample T-Tests (evaluated at $p < 0.01$). Based on this criterion, 200/286 (70%, 178 during the sample, 99 delay) of neurons responded significantly to the sample stimuli.

Category selectivity during the sample and delay epochs was determined by performing one-way ANOVAs (independently for each epoch) with one level for each category (evaluated at

$p < 0.01$) comparing the firing rate for "dog" trials to those for "cat" trials. This revealed that nearly a quarter of all neurons (69/286, 24%, 60 and 15 during the sample and delay, respectively) responded selectively to the two categories. Similar numbers of categorically tuned neurons preferred "cats" (27/60 sample, 7/15 delay) as "dogs" (33/60 sample, 8/15 delay). Two examples of category selective neurons are shown in figure 6a and 6b, respectively, both of which are category selective at $p<0.001$.

Stimulus selectivity was determined for all neurons with a one-way ANOVA with 54 levels (evaluated at $p < 0.01$), one level for each unique sample stimulus. Based on this criterion, 113/286 (40%) neurons were stimulus selective during either the sample or delay (110 sample, 11 delay). Approximately half of stimulus selective neurons were also category selective (n=57/113, 52 sample, 3 delay).

To further quantify the degree of neuronal category tuning, we employed a receiver-operator characteristics (ROC) analysis that computed the degree to which the values from two distributions were overlapping. This analysis returned a value indicating the accuracy with which an ideal observer could classify values from the two distributions. Completely overlapped distributions would yield ROC values of 0.5 (values from the two distributions cannot be reliably discriminated from one another) while completely non-overlapping distributions would give ROC values of 1.0 (values can be discriminated from one another with 100% accuracy). Across the population of stimulus selective neurons, the mean delay period "cat" vs. "dog" ROC value was 0.54 (range: 0.50 to 0.65). The mean sample period ROC value was 0.57 (range: 0.50 to 0.72). Across the population of category selective neurons, the mean delay period ROC value was 0.58 (range: 0.56 to 0.65). The mean sample period ROC value was 0.60 (range: 0.54 to 0.72). The same analysis applied to all neurons that were neither stimulus nor category selective (p>0.1) reveals a mean ROC value of 0.52 (range: 0.50 to 0.55) during the sample and 0.52 (range: 0.50 to 0.55) during the delay.

Population analysis

Master Histogram

Figures 3a and 3b illustrate the average activity across all stimulus selective neurons from IT and PFC, respectively (113 from ITC, 78 from PFC). Note that a neuronal response occurred approximately 100 ms following stimulus onset across both the IT and PFC populations. The level of delay period activity in IT returns to near that prior to stimulus onset. By contrast, persistent elevated delay activity is evident in the PFC.

Category Index

As a strict measure of tuning across the category boundary, we computed a category index that precisely measured each neuron's firing rate difference both within and between categories. The within class difference (WCD) was defined as a neuron's firing rate difference evoked by adjacent stimuli of the same category along a morph line (i.e. 80% and 60% "cats") averaged over all 9 morph lines that cross the category boundary. The between class difference (BCD) was defined as a neuron's firing rate difference between adjacent stimuli from different categories (i.e. 60% "cat" and 60% "dog"), again averaged over all 9 morph lined that cross the category boundary. (see methods) Category tuned neurons should exhibit large differences in firing rate to stimuli from distinct categories (BCD) and small differences in firing rate for stimuli from the same class (WCD). The index was computed for each neuron by dividing the difference between BCD and WCD by their sum. Thus, the index can take values with a range of –1 to 1. Positive values indicate larger differences between categories than within (i.e. category tuning) while negative values indicate smaller differences between category than within (i.e. no tuning to the category boundary).

The mean category index across the population of 60 category selective neurons during the sample period was 0.05, and significantly greater than 0 (one tailed T-Test, $p<0.001$). Though there were few category selective neurons during the delay period ($n=15$), we observed a mean category index of 0.07 (one tailed T-Test, $p<0.05$). By contrast with our results from the PFC, we did not observe significant category tuning across the population of stimulus selective neurons (IT, sample: 0.015 delay: 0.00). For comparison, across the population of PFC stimulus selective neurons, category index values were significantly greater than zero during both time epochs (PFC, sample: 0.08, delay: 0.14) as determined by a one-tailed T-Test ($p<0.001$). This suggests that most PFC neurons that responded differentially between stimuli tended to convey information about stimulus category. In contrast, the finding that the population of stimulus selective IT neurons did not show significant category tuning suggests that, while some stimulus selective neurons carried categorical information, many others responded to specific stimuli independent of their category-membership. In the PFC, category index values were significantly greater across the population of stimulus selective neurons during the delay than during the sample period as shown by a one tailed T-Test at $p<0.05$. We observed a similar, yet non-significant ($p=0.35$), trend across IT category selective neurons; delay period category index values were greater than during the sample (0.07 for the delay vs. 0.05 in the sample).

To compare the category tuning between IT and the PFC, we computed the breadth of category tuning. This index gave a measure of the degree of overlap between the neuronal responses to two sets of stimuli by computing the proportion of sample stimuli from the preferred category (the category which, on average, evoked the greater response) that evoked a greater response than the best stimulus from the non-preferred category. Index values of 0 would indicate that no stimulus from the preferred category evoked a larger response than the best stimulus from the non-preferred category. An index of 1 would indicate that all sample stimuli from the preferred class evoked larger responses than the best stimulus from the non-preferred category. Hence, a neuron that responds strongly to every dog stimulus and weakly to every cat stimulus will give a breadth of tuning index of 1.0. This analysis revealed that the population of PFC sample period category selective neurons had a significantly greater breadth of tuning (PFC mean = 0.23; range: 0.05 to 0.76) than did those from IT (IT mean = 0.16; range: 0.05 to 0.57) at $p < 0.01$ according to a two-tailed Wilcoxon rank-sum test. A similar pattern of results was found during the delay where PFC index values (PFC mean = 0.24; range: 0.0 to 1.0) were significantly greater than those from IT (IT mean = 0.15; range: 0.05 to 0.33) according to a one-tailed T-Test at $p<0.05$.

Latency of Selectivity

The latency of category selectivity for both the IT cortex and PFC was determined using a sliding version of the category index described above. Category index values were computed using a 250 ms time window that was incremented in 25 ms steps from 150 ms before sample onset to 50 ms before sample stimulus offset (29 steps). This analysis was applied to all neurons that were found to be category selective at $p < 0.001$ during a 600 ms epoch beginning 100 ms after sample onset. The normalized results from IT and PFC (shown individually for each monkey in figure 7) clearly show that the onset of category selectivity in IT preceded that in the PFC. The time at which the maximum value of the category index occurred was computed for these groups of neurons in both IT and PFC. These values, shown in figured 8a and 8b, and demonstrate a statistically significant difference in selectivity latency between the two populations of neurons (Wilcoxon rank-sum test, $p<0.01$).

Decision-related effects

During the presentation of the first test stimulus following the memory delay, the monkeys decided whether or not that stimulus matched the category of the previously presented sample stimulus and whether or not to make a motor response (lever-release). During this phase of the task, we found that the activity from 21% of IT neurons (61/286) differentiated between the four types of trials ("cat"-match, "cat"-non-match, "dog"-match and "dog"-non-match) as evaluated by a one-way ANOVA at $p<0.01$. We then performed a two-way ANOVA (factor 1: match or non-match trial type, factor 2: category of test stimulus) on these 61 neurons evaluated at $p<0.01$ to better determine their tuning properties. According to this criterion, the vast majority of neurons (85% or 52/61) reflected the category of the test-stimulus (that was currently in view). An example of such a neuron is shown in figure 9. Very few neurons (5% or 3/61) reflected whether the category of the test stimulus matched that of the sample. The activity of approximately 21% (13/61) of neurons was memory-related in that it reflected the category of the previously presented sample stimulus.

These results contrast sharply with our previous results from the PFC and suggest that IT and PFC play complementary roles in solving the behavioral task. While only 5% of test-period selective IT neurons reflected whether or not the category of the test stimulus matched the category of the sample, this distinction was observed in nearly half of PFC neurons (47%, 37/78). Furthermore, while most test-period selective IT neurons (85%) reflected the category of the test stimulus, this was observed in only 32% (25/78) of test-period selective PFC cells. Comparatively more neurons in the PFC (35/78, 45%) had memory-related test-period activity that reflected the category of the previously presented sample stimulus than did those in the ITC (21%). This pattern of results suggests that IT neurons are more involved in representing information about the stimulus currently in view while PFC may be more involved in maintaining relevant information about previously presented stimuli in short term memory and using that representation in the context of the behavioral task to make appropriate responses.

DISCUSSION

The purpose of this study was to examine the properties of IT neurons during visual categorization and to compare these findings with those of our previous study of the PFC. We report that, as in the PFC, the activity of visual neurons in IT can reflect the category of visual stimuli during the performance of a category-matching task. There were several notable differences in the pattern of results from IT as compared with those from PFC. Firstly, whereas robust category tuning was detected across the entire population of stimulus selective PFC neurons, this was not the case in the IT; a subset of stimulus selective IT neurons was category-tuned while others responded selectively to individual stimuli but were not explicitly tuned across the category boundary. This suggests that while some IT neurons carried signals about stimulus category, others may have been tuned to the physical shape of the stimuli and could have represented information useful for stimulus-specific identification. A second difference between IT and PFC was found in the time-course of category tuning. Across the population of category-tuned neurons, category-tuning following sample stimulus onset occurred earlier in IT than PFC. Furthermore, IT category tuning was mostly restricted to the sample period and rarely persisted into the delay. By contrast, comparatively more PFC neurons exhibited category selective activity that persisted throughout the memory delay. A third difference was found during the decision phase of the task during which the monkey decided whether a test stimulus matched a previously presented sample. During this task phase, the vast majority of selective IT neurons reflected the category of the test stimulus (that was currently visible to the monkey) and very few neurons reflected the monkey's decision about whether that stimulus matched the prior sample stimulus. In the PFC however, many neurons reflected the monkey's decision about whether the test stimulus matched the sample while fewer neurons reflected the category of the test stimulus currently in view. This suggests that while IT neurons' activity primarily represented the stimuli that were currently in view, activity in the PFC reflected the conjunctions of signals about the sample and test stimuli that were necessary for successfully solving the task.

Previous studies that have examined the properties of neurons during categorization did not use parametrically controlled stimuli and utilized behavioral tasks that required subjects to report whether a given stimulus (one of many photographs of objects or scenes) was a member of the target class or not. Using this design, stimuli from the same category (i.e. trees) were, on average, more similar to members of their own category (other trees) than to the distrator stimuli (i.e. animals, faces, cars… etc). As a result, it is unclear whether the observed category tuning in those studies was a result of categorical learning or rather that stimulus similarity was greater within a category than between categories. By training our subjects to learn two distinct well defined categories for which the visual similarity of stimuli was parametrically controlled, the effect of category learning on neuronal stimulus representations could be more precisely evaluated.

These results fit in well with those of other neurophysiological studies which show that the shape tuning of IT and PFC neurons is modified by visual experience and may underlie improvements in perceptual tasks. In a study by Logothetis et al., (1995), monkeys were trained to recognize static images of complex 3D wireframe stimuli from multiple viewpoints. Once the monkeys had learned to recognize objects from a wide range of viewpoints, neuronal recordings from the IT cortex revealed a population of neurons that had become view-tuned; these neurons were tuned to one or more viewpoints of an object that typically corresponded with the viewpoints for which the monkeys had been extensively trained. A smaller subset of neurons displayed view-invariant object-tuning in that they responded robustly to all viewpoints of a preferred object but not to other stimuli. In a similar study by Booth and Rolls (1998), monkeys were familiarized with a number of real plastic objects. Recordings from the ITC revealed a small population of view-invariant neurons that responded to static views of these familiar objects from multiple viewpoints suggesting, again, that visual experience could result in an enhanced stimulus representation. In a study of object representations in the PFC, Rainer et al., (2000) trained monkeys to recognize familiar and novel stimuli that had been degraded by the addition of varying amounts of noise. It was found that, though novel stimuli activated a larger proportion of neurons, the neuronal responses to familiar stimuli were more tightly tuned and robust to stimulus degradation.

Taken together, the results of these previous studies from other researchers suggest that the neuronal basis of learning to recognize complex stimuli may be in the development of neurons robustly tuned to specific objects and with enhanced response invariance to image transformations such as rotation and degradation. Our results are compatible with these results and with theoretical models of object recognition and categorization (Riesenhuber and Poggio, 2000). Learning to recognize a stimulus despite changes in viewpoint, or that a group of visually dissimilar stimuli are of the same category, may be accomplished by a neuronal representation that is invariant across the relevant dimension: viewpoint in the case of rotated stimuli or category-member in the case of category recognition. Hence, category learning may, in fact, cause neurons to become category-tuned, responding with invariance—not to transformations of individual stimuli—but, rather, to the members of a given category.

Figure Legends

Figure 1 Task design and behavioral performance. The trial began with central fixation (500 ms.) followed by a one second delay and then by a test stimulus (600 ms.). Monkeys had to release a lever to the match with the 600 ms. of its presentation to obtain a fruit juice reward. If the test was a non-match, there was another delay interval (600 ms) followed by a presentation of a match, which required a lever release. Monkeys did not need to remember anything over the second delay interval, it was included to insure a behavioral response on each trial. There were an equal number of match and non-match trials and they were randomly interleaved.

Figure 2 Organization of stimulus set. a, The 6 prototype images and 12 morph lines. The sample stimulus set was composed of 54 unique images: 6 prototypes (as shown), 4 images evenly placed (20%, 40%, 60%, 80%) along the 9 lines connecting each 'cat' to each 'dog' prototype, and 2 images (at 40% and 60%) along each of the 6 lines between prototypes of the same category (with respect to the 2-class boundary). b, An example of the morphs generated between the C1 and D1 prototypes.

Figure 3 Average histogram of ITC and PFC neuronal activity. a. The average activity is shown for the population of 113 stimulus selective ITC neurons. b. The average histogram across 78 stimulus selective PFC neurons.

Figure 4 Single neuron examples (two-category task). a. The average activity of a single category selective neuron shows greater activity to DOG stimuli at all morph levels during the sample period. Each histogram trace represents the neuron's average activity to each of the six morph levels. b. The average activity of a single category selective neuron that shows greater activity to CAT stimuli at all morph levels during the sample period.

Figure 5 Distribution of ITC category index values. The index can take values ranging from –1 to 1. Positive values indicate larger differences in neuronal firing between categories than within category. Negative values indicate larger differences within category than between categories. a. The distribution of sample-period category index values across the population of 110 neurons that showed stimulus selective sample stimulus evoked activity does not show a significant shift towards positive values (that would have indicated a categorically tuned population). b. Significant category tuning is observed across the populations of 60 and 15 category selective neurons in the sample (b) and delay (c), respectively.

Figure 6 Breadth of category tuning in the ITC and PFC. The breadth, or magnitude, of category tuning was computed across the population of category selective neurons for both the ITC (a) and PFC (b) during the sample period. A significantly larger breadth of tuning is observed in the PFC, indicating an enhanced representation of stimulus category there than in the ITC.

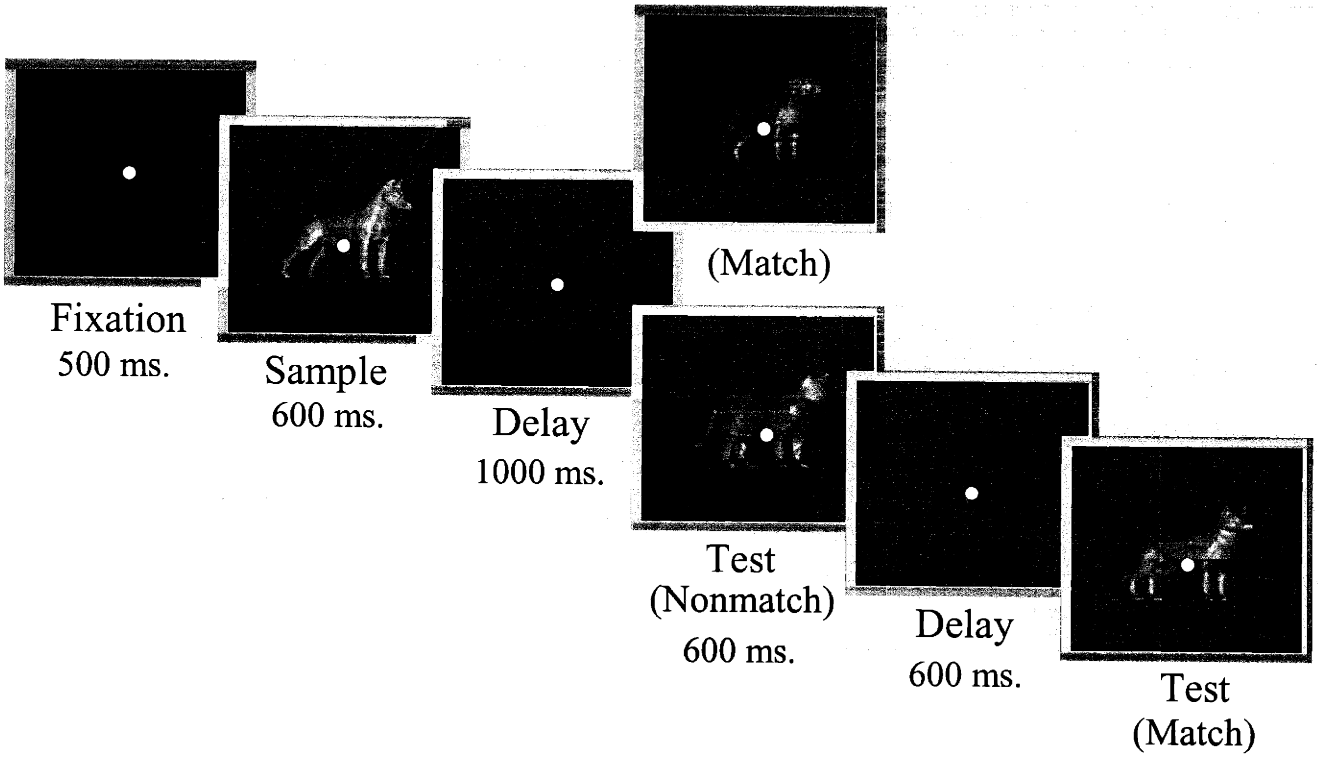
Figure 7 Time course of category selectivity in the ITC and PFC. CAT vs. DOG category selectivity across the population of category selective ($p<0.001$) neurons was computed for the ITC and PFC. a. Normalized values of the sliding category index are shown for 45 ITC neurons from monkey A (blue), 10 PFC neurons from monkey A, and 31 PFC neurons from monkey B. Increased category index values in the ITC occurred with a significantly shorter latency following sample onset than in the PFC.

Figure 8 Histograms of PFC and ITC category selectivity latency. Figures a and b show histograms of the time at which each ITC (a) and PFC (b) neuron reaches its maximum category index value.

Figure 9 Single neuron example of decision-phase selectivity. In this peri-stimulus time histograms, neuronal activity is grouped according to the category of the test stimulus and whether or not the category of the test stimulus matched the category of the sample stimulus (four groups: CAT-match, DOG-match, CAT-non-match, and DOG-non-match). The single neuron in figure 8a differentiates between the two categories during both the sample and test period. Following onset of either the sample or test stimuli, this neuron responds preferentially to stimuli from the 'cat' category.

Table 1 Incidence of test-phase selectivity in the PFC and ITC. Note that in the ITC, the majority of selective neurons reflected the category of the test stimulus. The population of PFC test period selective neurons was more heterogeneous; substantial proportions of the test-period selective neurons represented the category of the sample stimulus, test stimulus or whether or not the test stimulus was the same category as the sample.

Figure 1


Fixation
500 ms.
Sample
600 ms.
Delay
1000 ms.
(Match)
Test
(Nonmatch)
600 ms.
Delay
600 ms.
Test
(Match)

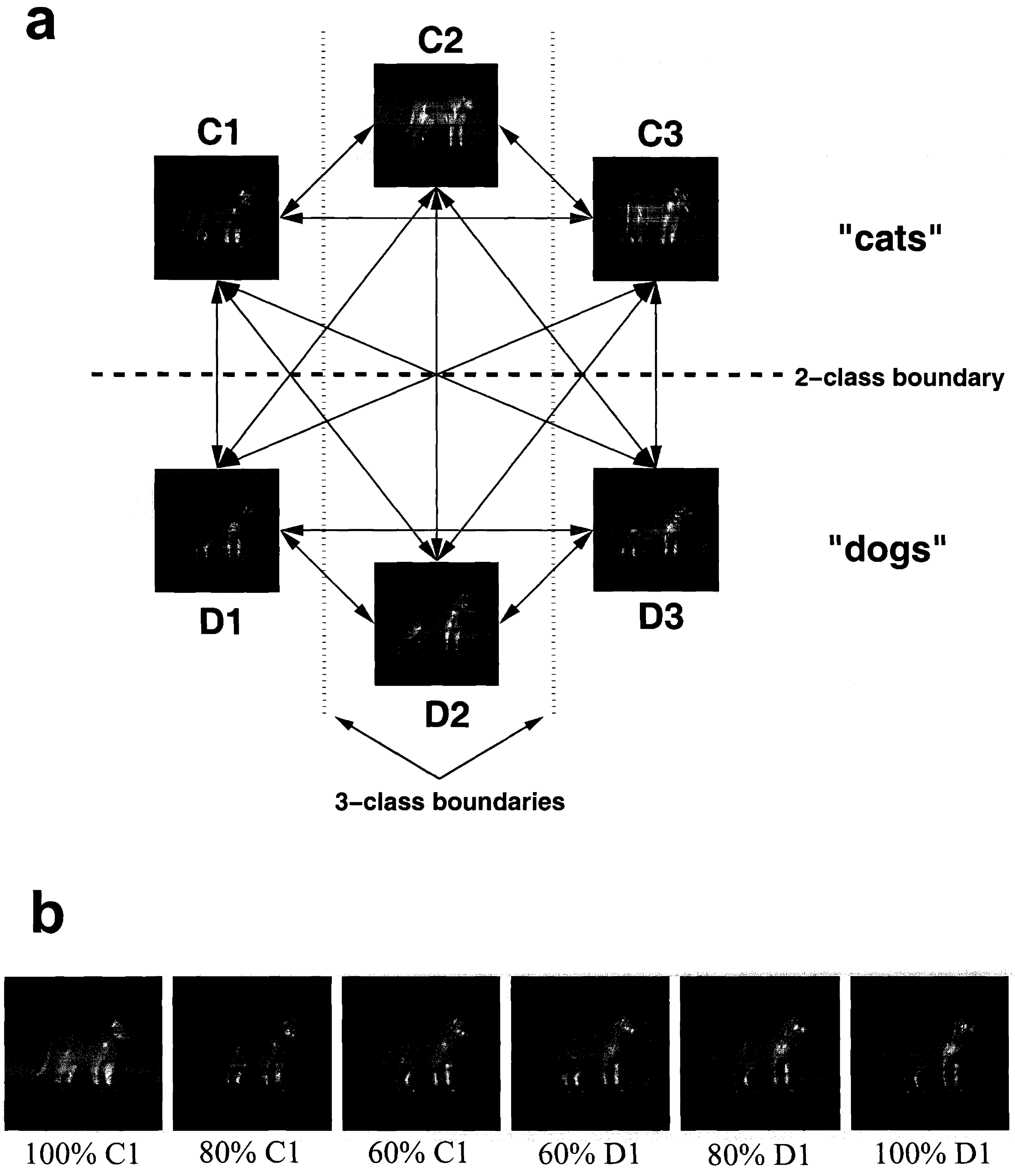

a
C2
C1
C3
"cats"
2-class boundary
"dogs"
D1
D3
D2
3-class boundaries
b
100% C1
80% C1
60% C1
60% D1
80% D1
100% D1

Figure 2

Figure 3

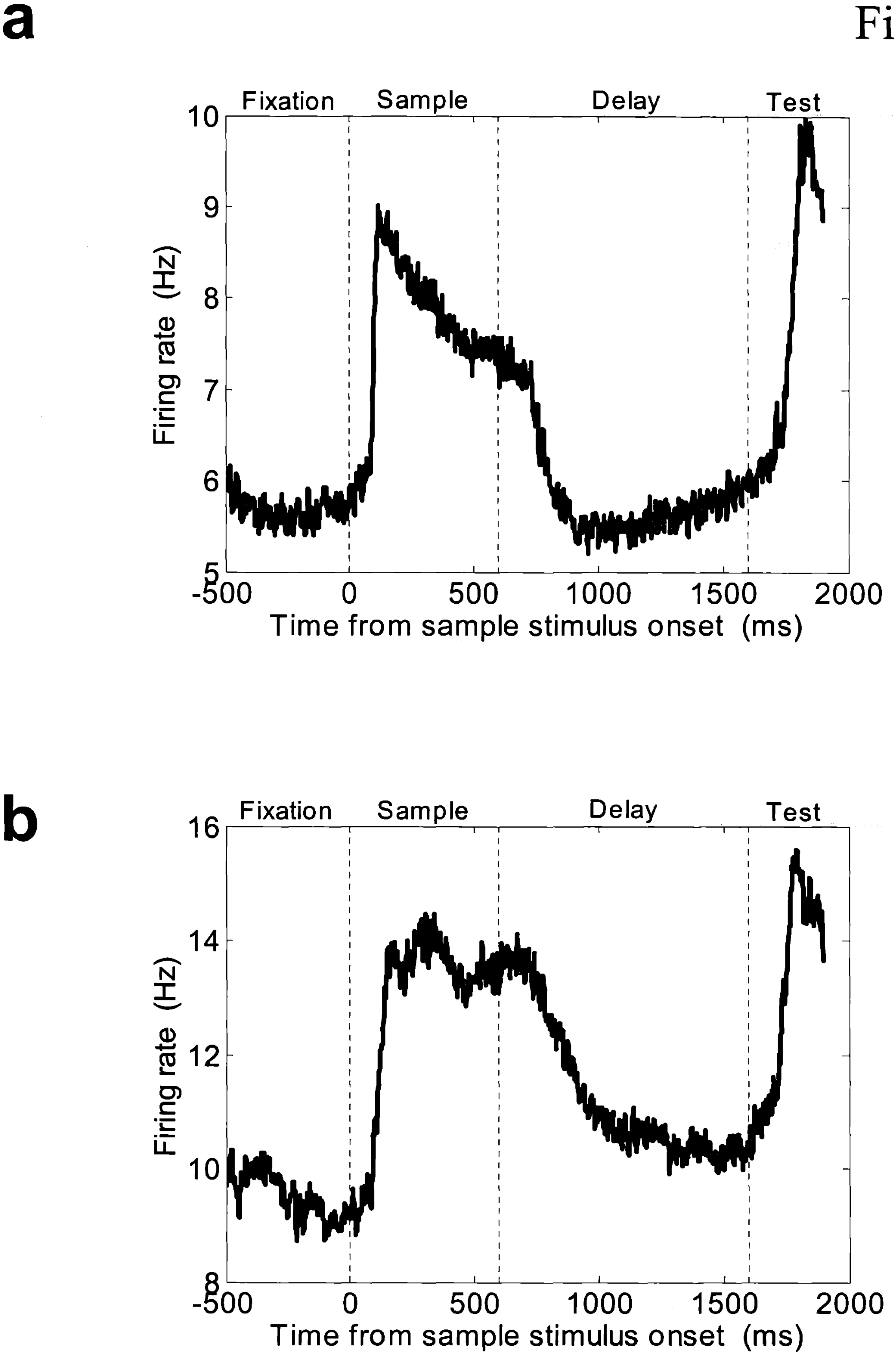

a
Fixation
Sample
Delay
Test
10
9
8
7
6
5
Firing rate (Hz)
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
b
Fixation
Sample
Delay
Test
16
14
12
10
8
Firing rate (Hz)
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)

Figure 4

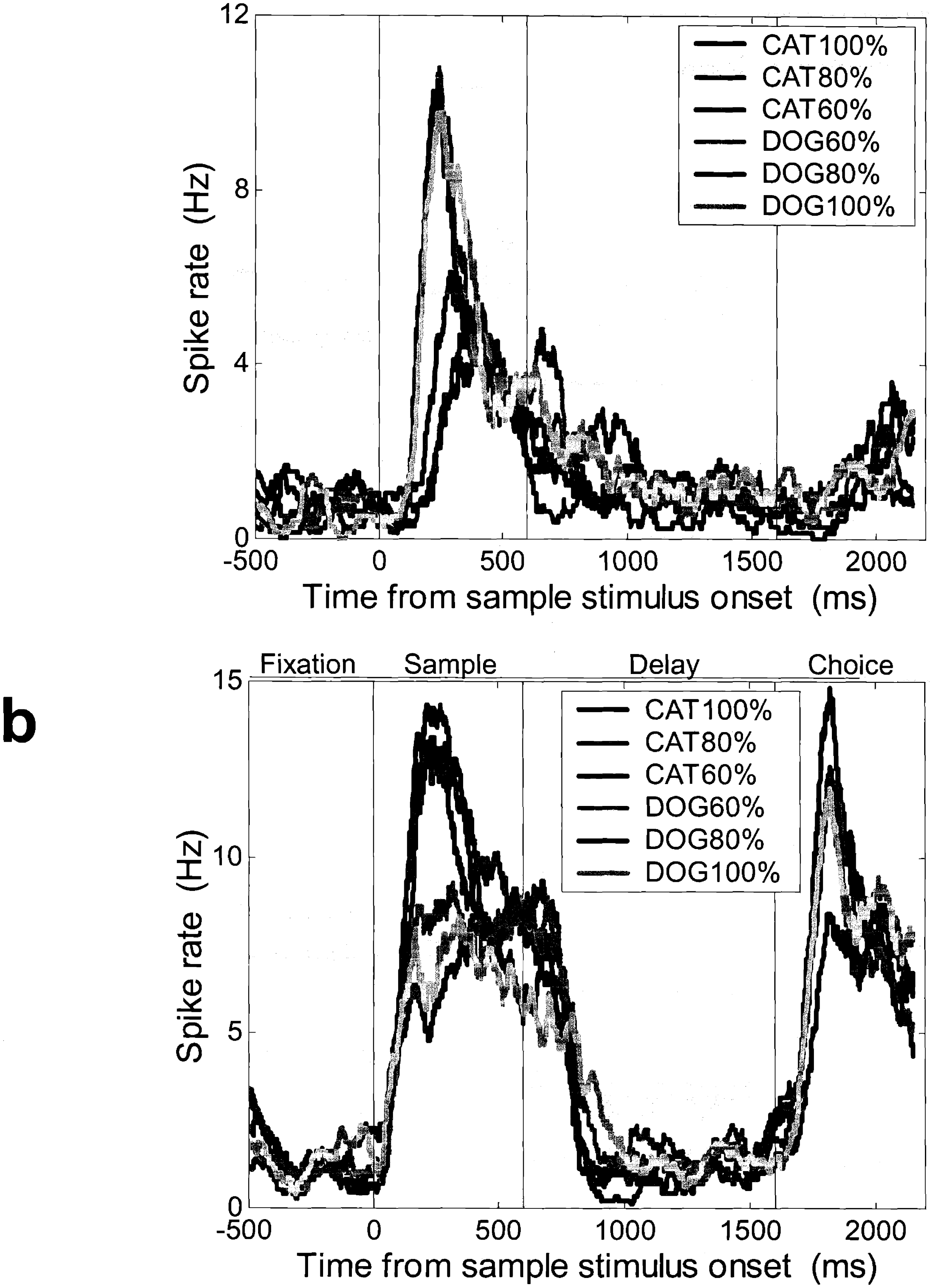
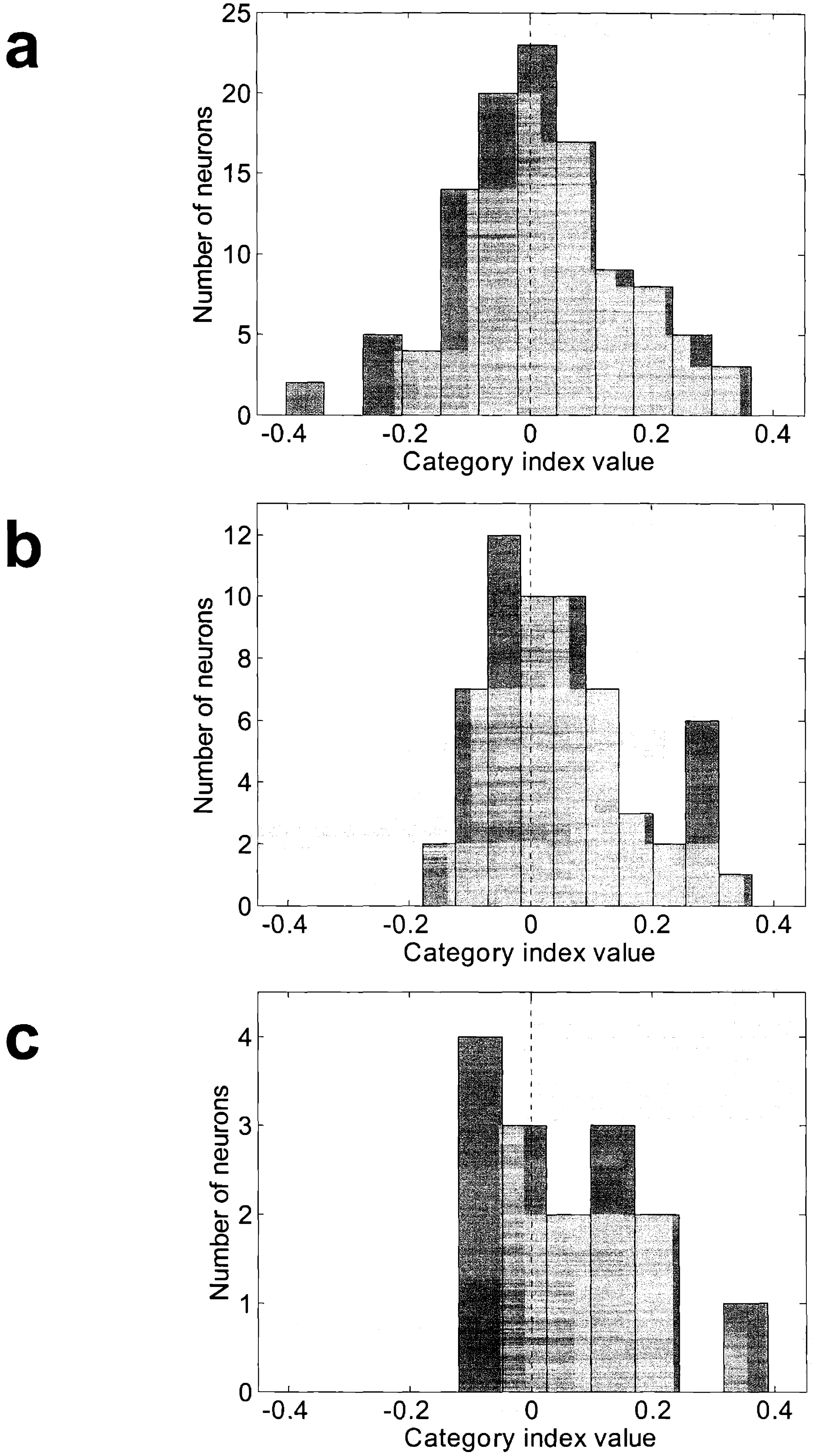

a
12
8
4
0
Spike rate (Hz)
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
CAT100%
CAT80%
CAT60%
DOG60%
DOG80%
DOG100%
b
Fixation
Sample
Delay
Choice
15
10
5
0
Spike rate (Hz)
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
CAT100%
CAT80%
CAT60%
DOG60%
DOG80%
DOG100%

Figure 5

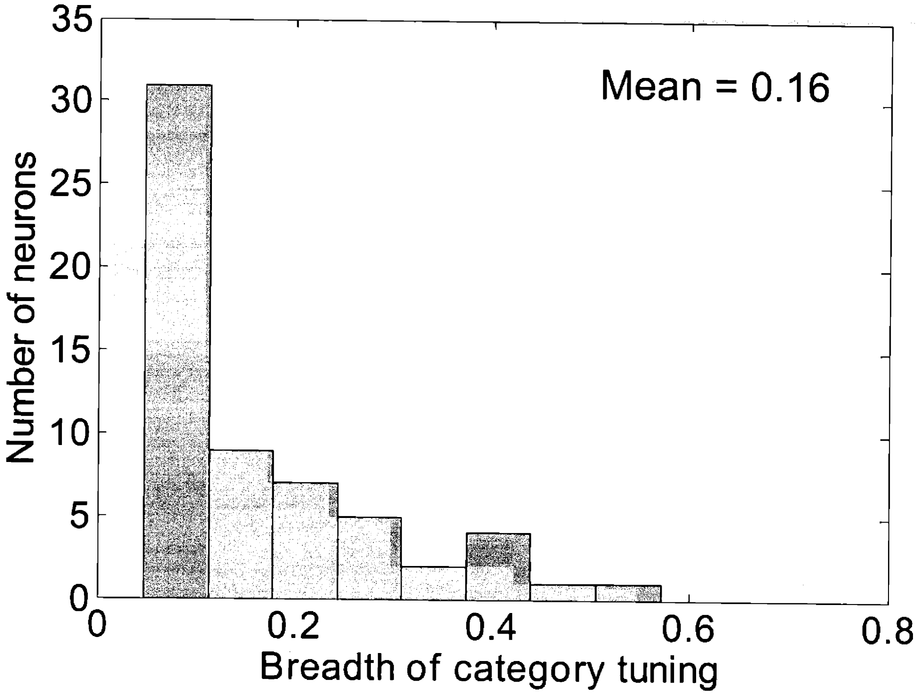


110 stimsel neurons
Index=0.015
P=0.13 (1tailed)

60 categsel neurons
Index=0.053
P=0.0008 (1tailed)
800 ms window

15 categsel neurons
Index=0.067
P=0.039 (1tailed)

a


35
30
25
20
15
10
5
0
Number of neurons
Mean = 0.16
0
0.2
0.4
0.6
0.8
Breadth of category tuning

b

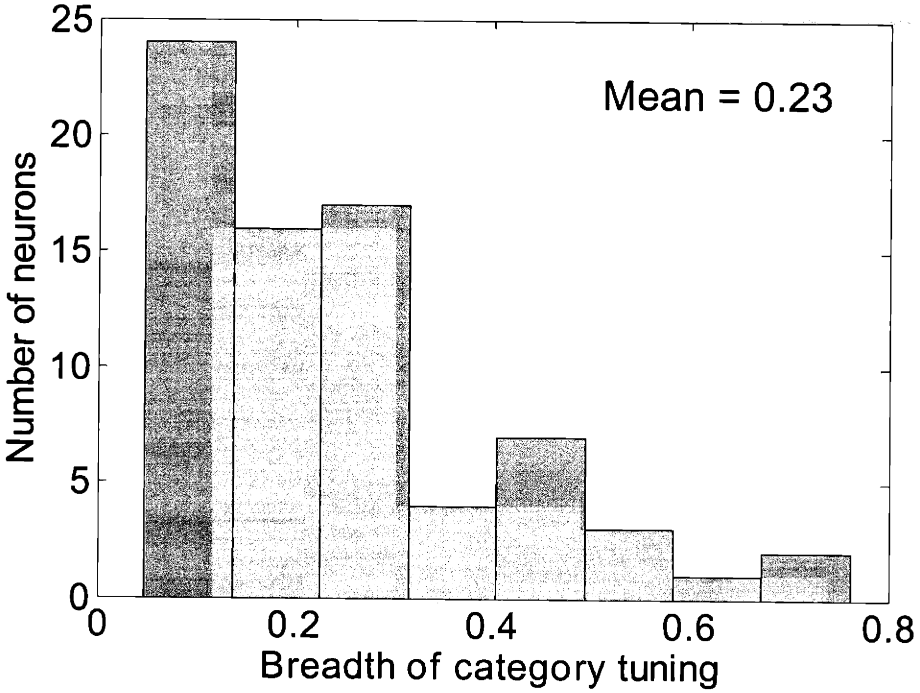
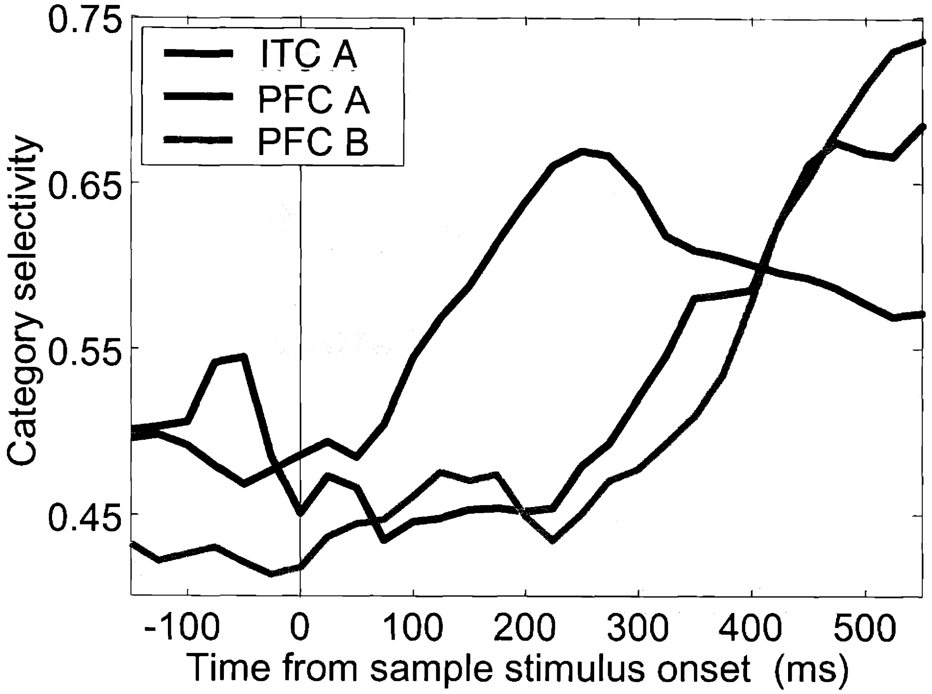

25
20
15
10
5
0
Number of neurons
Mean = 0.23
0
0.2
0.4
0.6
0.8
Breadth of category tuning

Figure 7



0.75
0.65
0.55
0.45
Category selectivity
ITC A
PFC A
PFC B
-100
0
100
200
300
400
500
Time from sample stimulus onset (ms)

Figure 8

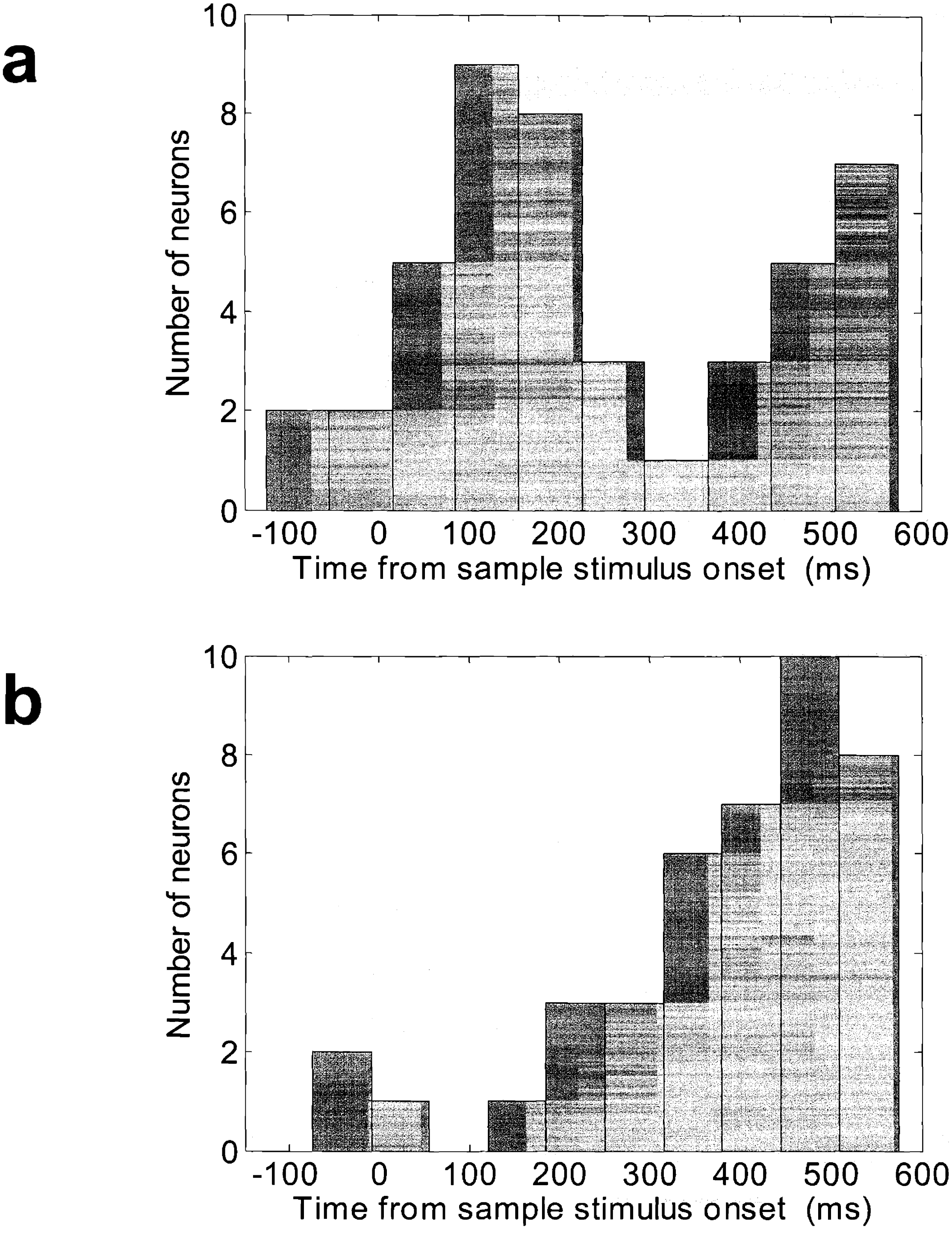

a
10
8
6
4
2
0
Number of neurons
-100
0
100
200
300
400
500
600
Time from sample stimulus onset (ms)
b
10
8
6
4
2
0
Number of neurons
-100
0
100
200
300
400
500
600
Time from sample stimulus onset (ms)

Figure 9

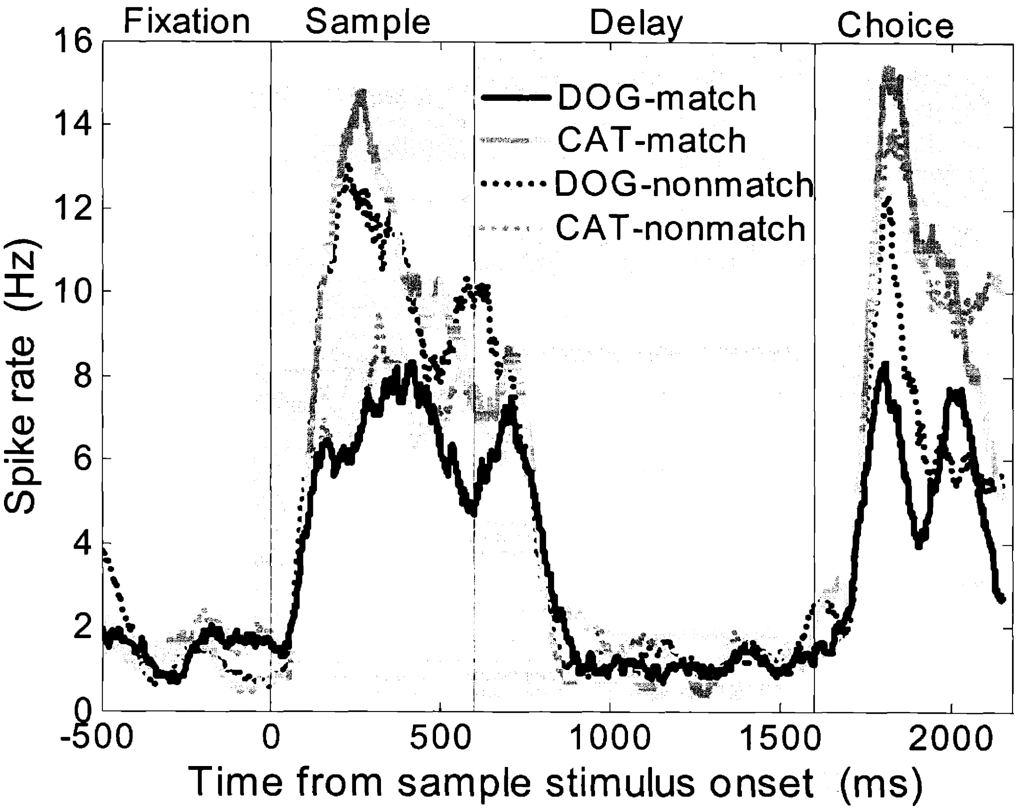

Fixation
Sample
Delay
Choice
16
14
12
10
8
6
4
2
0
Spike rate (Hz)
-500
0
500
1000
1500
2000
Time from sample stimulus onset (ms)
DOG-match
CAT-match
DOG-nonmatch
CAT-nonmatch

Table 1

ITC

	# selective neurons
Total selective	61
Match/Nmatch	3/61 (5%)
Test category	52/61 (85%)
Sample category	13/61 (21%)
Main effect + interaction	7/61 (11%)

PFC

	# selective neurons
Total selective	78
Match/Nmatch	37/78 (47%)
Test category	25/78 (32%)
Sample category	35/78 (45%)
Main effect + interaction	16/78 (21%)

References

Beymer D and Poggio T. (1996) Image representations for visual learning. *Science* 272, 1905-1909.

Booth MC and Rolls ET. (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* Sep;8(6):510-23

Erickson CA, Jagadeesh B and Desimone R. (2000) Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nat Neurosci.* Nov;3(11):1143-8.

Freedman DJ, Riesenhuber M, Poggio T and Miller EK. (2001) Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science,* Jan 12;291:312-316.

Green DM and Swets JA. (1966). Signal Detection Theory and Psychophysics (New York: J. Wiley and Sons).

Gross CG. (1973) Visual functions of the inferotemporal cortex. In: *Handbook of Sensory Physiology,* Vol. 7/3B. R. Jung, Ed. Springer-Verlag, Berlin, pp. 451-482.

Kluver H and Bucy L. (1938) An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to “psychic blindness”. *J. Psychol.* 5:33-54.

Kreiman G, Koch C and Fried I. (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci.* Sep;3(9):946-53.

Logothetis NK, Pauls J and Poggio T. (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol.* May 1;5(5):552-63.

Logothetis N. (1998) *Curr. Opinion in Neurobiol.* **8**:536-544.

Miller EK, Erikson CA and Desimone R. (1993) Activity of neurons in anterior inferior temporal cortex during a short term memory task. J. Neurosci. 13:1460-1478.

Perrett DI, Oram MW and Ashbridge E. (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* Jul;67(1-2):111-145.

Nachson I. (1995) On the modularity of face recognition: the riddle of domain specificity. *J Clin Exp Neuropsychol,* Apr;17(2):256-75.

Rainer G and Miller EK. (2000) Visual Experience and Prefrontal Object Representation, *Neuron* **27**, 179.

Riesenhuber M and Poggio T. (2000) Models of object recognition. 2000 Nov;3 Suppl:1199-204.

Sheinberg DL and Logothetis NK. (1997) The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci U S A. Apr 1;94(7):3408-13.

Shelton C. (2000) Morphable Surface Models. *International Journal of Computer Vision* 38, 75-91.

Tanaka K. (1996) Inferotemporal cortex and object vision. *Annual Review of Neuroscience* **19**, 109-139.

Tolhurst DJ, Movshon JA and Dean AF. (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res. 23, 775-785.

Vogels R and Orban GA. (1990) How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. J. Neurosci. 10, 3542-3558.

Vogels R. (1999) Categorization of complex visual images by rhesus monkeys. *Eur J Neurosci.* **11**, 1223-38.

Webster MJ, Bachevalier J and Ungerleider LG. (1994) Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. Cereb Cortex. Sep-Oct;4(5):470-83.

DISCUSSION AND CONCLUSIONS

Results summary

In the first experiment, examining the responses of prefrontal cortex neurons during visual categorization, we found two primary results. Firstly, many prefrontal neurons reflected the category membership of stimuli—neuronal activity made sharp distinctions between visually similar stimuli from different categories and, likewise, responded similarly to physically different stimuli that belonged to the same category. Tuning to the two categories was evident both while the sample stimulus was in view and while the monkey maintained information about the sample stimulus in short term memory. Secondly, many prefrontal cortex neurons reflected the monkeys' decisions about whether a test stimulus matched the category of a previously presented sample stimulus.

The second experiment was designed to further test that the categorical tuning observed in the first experiment was a product of learning. After a monkey was trained to reclassify the same stimuli from the first experiment into three new categories, we no longer found neurons that were tuned to the old two-categories. Rather, the population showed tuning to the newly-learned three-categories during both the sample and memory-delay periods suggesting that the category tuning observed in the first experiment was most likely a result of learning and could not be explained by the inherent visual properties of the stimuli.

The third experiment examined the properties of inferior temporal cortex neurons during the same categorization paradigm used for the first experiment. This allowed for a direct comparison of neuronal properties in the prefrontal and inferior temporal cortices, yielding three primary results. Firstly, as in the prefrontal cortex, many temporal lobe neurons reflected stimulus category. However, there was a greater proportion of temporal lobe neurons that were only stimulus selective—their activity conveyed information about stimulus-identity and not category. Secondly, following the onset of a stimulus, category tuning was apparent with a shorter latency in the temporal lobe as compared with the prefrontal cortex. Furthermore, comparatively fewer temporal lobe neurons exhibited category tuning that persisted into the memory delay. Thirdly, whereas many prefrontal cortex neurons reflected the monkeys' decisions about whether a test stimulus matched a previously presented sample stimulus, the vast majority of temporal lobe neurons reflected the category of the test-stimulus that was currently in view and did not carry a representation of the monkeys' upcoming motor response.

Taken together, these results suggest that, through the process of learning, neurons in the prefrontal and inferior temporal cortex can convey information about the category of visual stimuli. The differences in neuronal responses between the ITC and PFC support the hypothesis that the temporal lobe cortex is primarily involved in the rapid processing of visual information while the prefrontal cortex underlies high-level cognitive functions related to the executive control of behavior.

Open questions for further research

The results from these experiments represent only the first step in understanding the neuronal mechanisms that form the basis of categorical perception and learning. As is true for most promising lines of research, these studies have provided several new insights but at least as many new questions for further investigation. To address these questions, we have already begun to think about the next generation of experiments to further investigate the mechanisms of category learning.

Neuronal representations of multiple category schemes

An obvious question is raised by the second (three-category) experiment. Namely, did training the monkey to reclassify the stimuli (that had been grouped into two categories) into three new categories cause the population of two-category neurons to become tuned to the three new categories or were the three new categories represented by a separate population of neurons? The finding that very few (if any) neurons reflected the old two-categories while the monkey performed the three-category task can be taken as weak evidence suggesting that, in the prefrontal cortex, multiple category schemes may be encoded by the same population of neurons. Subsequent behavioral testing of that monkey revealed that it was quickly able to relearn the two-category task with a small amount of practice after the three-category recordings were complete. Therefore, the monkey still maintained some long-term knowledge about the old categories though they were not reflected in the prefrontal cortex when the three new categories were currently relevant. Where, then, was this information about the old two-categories represented?

One possibility is that different category schemes may be stored by separate populations of inferior temporal cortex neurons and retrieved, when needed, by the prefrontal cortex. In this model, separate populations of inferior temporal neurons, each dedicated to a specific category scheme, would arise through learning the two and three-category tasks. The prefrontal cortex is known to be critical for the ability to switch between one task and another and its neurons show task related modulation concomitant with this capacity. Hence, a single population of task-related prefrontal cortex neurons might select their inputs from either the two or three-category populations of temporal lobe neurons depending on the task that was currently in effect. In fact, our results provide preliminary support for this model; Categorical information was expressed earlier in the inferior temporal than in prefrontal cortex. Furthermore, a greater proportion of stimulus selective prefrontal neurons were category-selective and showed sharper category-tuning than did those in the inferior temporal cortex; while some inferior temporal neurons carried categorical information, others carried stimulus specific information independent of category. That the majority of stimulus selective prefrontal cortex neurons were category selective suggests that they may have been driven primarily by inputs from category selective inferior temporal neurons.

In theory, this hypothesis could be investigated by recording simultaneously from a large group of neurons in both the prefrontal and inferior temporal cortices for numerous consecutive days while a monkey learned a new category scheme. This would reveal whether individual neurons changed their stimulus tuning properties during learning or whether new groups of neurons came to encode new categories. Furthermore, neuronal correlations with the monkeys' behavioral improvements would give great insights into the precise changes in electrophysiological activity that accompany perceptual learning. Unfortunately, such an experiment is not currently feasible due to the technical limitations of chronic extra-cellular recordings. Once it is possible to reliably isolate and hold individual neurons for extended periods of time, great strides will be made in understanding the changes in neuronal activity that accompany long-term learning.

As a technically feasible alternative, we propose to teach monkeys both the two and three-category tasks simultaneously and, furthermore, to flexibly switch between the two tasks within a single recording session. Recordings from the prefrontal and inferior temporal cortices while monkeys switch between one categorization task and another could test the above hypotheses: Firstly, whether different category schemes were indeed encoded by separate populations of inferior temporal cortex neurons and, secondly, whether the tuning of individual prefrontal cortex neurons changed flexibly to reflect the categories that were currently relevant for solving the task.

Common Neuronal Mechanisms for Categorization and Identification

Neurons that are perfectly category-tuned, that respond identically to all members of a category and not to stimuli from other classes, make the task of classifying stimuli extremely easy—they convey all the information need to make a categorical judgment. However, this seems to be a computationally expensive solution as these neurons would be good for nothing else besides categorizing their preferred stimuli. If a task required specific identification of individual category members, these neurons would carry no useful information. Conversely, neurons tuned exclusively to single stimuli, or specific views of single stimuli, make simple the task of stimulus specific identification but give rise to the 'grandmother cell' dilemma: the brain is not big enough to contain a neuron specifically tuned to every stimulus that we can recognize. Furthermore, neurons with such specific tuning make the task of recognizing stimuli despite various transformations (i.e. changes in rotation, illumination or scale) very difficult. How then does the brain arrive at such a fine balance between specificity and invariance? The category tuning that we have observed in the prefrontal and inferior temporal cortices was, save for a few outstanding examples, far from perfect. Though there was a sharp distinction between categories in many neurons, some information was often conveyed about individual stimuli within a category as well. In principal then, the same neurons could be used for both categorization and identification; reliable information about the category and identity of a stimulus could be conveyed by a population of imperfectly tuned category selective neurons.

The neuronal mechanisms underlying categorization and identification could be simultaneously investigated by training monkeys to alternately categorize and identify the same set of stimuli by during a single recording session. Neuronal recordings while monkeys performed the two distinct tasks would indicate whether the same population of neurons was active during both tasks and could reveal whether their stimulus tuning properties changed according to the perceptual demands of the task. One possibility, commensurate with the above hypothesis, is that the same neurons would be active during both tasks and show similar tuning properties. This would suggest that categorization and identification can be served by the same neuronal mechanisms. Alternatively, different populations of neurons, each with tuning properties specifically suited to their respective tasks, could alternately become active as the monkey switched from one task to the other. This would suggest that learning to categorize and identify stimuli is served by distinct neuronal populations. A third possibility is that the same neurons may be active during both categorization and identification but that their response properties would change, for example, from broad category-tuning to sharper stimulus-specific tuning when the monkey switched from categorization to identification, respectively. This result would suggest that neuronal stimulus tuning, though acquired through learning, can be modified rapidly and dynamically according to the current perceptual demands and, furthermore, that single stimulus-selective neurons are likely involved in encoding multiple tasks or different category-groupings of stimuli. Any differences observed in the pattern results from different brain areas would, once again, further dissociate the respective roles that they play in solving complex perceptual tasks.

Concluding remarks

The results of the studies presented here suggest that learning new perceptual categories can cause neurons in the inferior temporal and prefrontal cortices to become categorically tuned. While monkeys categorized visual stimuli into previously learned classes, neurons responded more similarly to stimuli from the same category and make sharp distinctions between stimuli from different categories. Furthermore, our results suggest that the prefrontal and inferior temporal cortex may play different, though complementary roles, in visually guided behavior. While the inferior temporal cortex seemed to be primarily involved in the rapid analysis of visual stimuli currently in view, responses of neurons in the prefrontal cortex appeared more suited for using that highly processed visual information to guide behavior in the context of the currently relevant task.