Scuola Internazionale Superiore di Studi Avanzati - Trieste


# A multidisciplinary approach to the study of shape and motion processing and representation in rats

Candidate:

Federica Rosselli

Supervisor:

Davide Zoccolan

Thesis submitted for the degree of Doctor of Philosophy in Neuroscience Area

Trieste, 2013

SISSA - Via Bonomea 265 - 34136 Trieste, Italy.

*To the loving memory*

*of Prof. Ferdinando Rossi.*

*'What else?'*

# Acknowledgements

It is the fine brains and mighty hearts that one has the privilege to meet along the way that make this job worth doing, and this life worth living. These are: Davide Zoccolan, the best Supervisor anyone could find, and one of the best persons I have ever met.

All my labmates, for their support and patience in sharing this ride with me. Special thanks to: Alireza Alemi-Neissi, a great collaborator and friend; Francesca Pulecchi, who taught me everything I know about microsurgery; Houman Safaai and Alessio Ansuini for their enormous contribution and help in data analysis; Liviu Soltuzu, for his kind help with the editing part.

All the Master Students (Dario Campagner, Matilde Fiorini, Gioia De Franceschi, Silvia Rossi, Angela Morriello, Elettra Grdina) who passed by the lab and decided to spend some time there with us, sharing their knowledge, skills and friendship.

Mathew Diamond's Lab members, for their invaluable help and support. Special thanks to Fabrizio Manzino and Marco Gigante, for helping me out when I was in need, even when they were not supposed to.

All lab technicians, the backbone of whatever enterprise one might think of in this field, and SISSA administration members, for making complicated things as simple as possible. Special thanks to: Tullio Bigiarini, Andrea Tomicich, Renato Tomizza, Alessandra Janousek, Gabriella Rossi.

My family: were it not for them, I wouldn't be here.

All my friends and teachers; the old, the new and the lost ones.

Finally, all my love goes to Francesco Caracciolo. Who is probably jamming with Jimi Hendrix right now.

# Note

Part of the data reported in the present thesis has been published in the following article:

Alemi-Neissi, A., Rosselli, F.B., and Zoccolan, D. (2013). Multifeatural shape processing in rats engaged in invariant visual object recognition. J. Neurosci. Off. J. Soc. Neurosci. 33, 5939–5956.

Federica Rosselli

# Contents

# Abstract

During my PhD I investigated how shape and motion information are processed by the rat visual system, so as to establish how advanced is the representation of higher-order visual information in this species and, ultimately, to understand to what extent rats can present a valuable alternative to monkeys, as experimental models, in vision studies. Specifically, in my thesis work, I have investigated:

1) The possible visual strategies underlying shape recognition.
2) The ability of rat visual cortical areas to represent motion and shape information.

My work contemplated two different, but complementary experimental approaches: psychophysical measurements of the rat's recognition ability and strategy, and *in vivo* extracellular recordings in anaesthetized animals passively exposed to various (static and moving) visual stimulation.

The first approach implied training the rats to an invariant object recognition task, i.e. to tolerate different ranges of transformations in the object's appearance, and the application of an mage classification technique known as The Bubbles to reveal the visual strategy the animals were able, under different conditions of stimulus discriminability, to adopt in order to perform the task.

The second approach involved electrophysiological exploration of different visual areas in the rat's cortex, in order to investigate putative functional hierarchies (or streams of processing) in the computation of motion and shape information.

Results show, on one hand, that rats are able, under conditions of highly stimulus discriminability, to adopt a shape-based, view-invariant, multi-featural recognition strategy; on the other hand, the functional properties of neurons recorded from different visual areas suggest the presence of a putative shape-based, ventral-like stream of processing in the rat's visual cortex.

The general purpose of my work is and has been the unveiling the neural mechanisms that make object recognition happen, with the goal of eventually 1) be able to relate my findings on rats to those on more visually-advanced species, such as human and non-human primates; and 2) collect enough biological data to support the artificial simulation of visual recognition processes, which still presents an important scientific challenge.

# Chapter I | Introduction

Each environment-exploring living system must answer accurately to one, perhaps *the* biologically relevant question: 'what is that, and where is it going?'.

In other words, survival-driven behavior implies the unambiguous assignment of values of *identity*, *velocity* and *direction* to objects in the world.

The issues concerning 'what' and 'where' are solved with apparent simplicity by our visual system; research has shown that human and non-human primates are able to effortlessly identify static and moving objects from among a huge number of alternatives and within fractions of seconds (Thorpe et al., 1996; Fabre-Thorpe et al., 1998; Intraub, 1980; Keysers et al., 2001; Rousselet et al., 2002).

But in spite of the apparent effortlessness with which the brain performs it, the astonishing computational complexity of visual object recognition strikes when we consider that its achievement is by nature underconstrained: we cannot access the physical world directly, but only by means of retinal patterns. Hence, the sole source of input our neural system affords is inherently ambiguous, since any individual object can produce an infinite set of different images on the retina (reviewed by Logothetis and Sheinberg, 1996; Purves et al., 2011).

When getting acquainted to the theory behind the study of object and motion processing, we find ourselves facing two orders of problems: the Invariance Problem and the Aperture Problem (explained below).

Again, the purpose of visual percepts is to generate successful behavior based on the information conveyed by retinal stimuli. In order to support recognition, visual object representations must be endowed with the property of *invariance*, i.e. they must tolerate the changes produced, at the level of the retinal image, by the many transformations each object may undergo (e.g., position, scale, pose and illumination conditions, and the presence of visual clutter/background). At the same time, visual object representation must meet the requirement of *selectivity*, i.e. they must support the ability to uniquely identify individual objects.

In other words, the visual system must produce representations that selectively allow

identification of visual objects in a manner that is largely tolerant to variations in their appearance.

Researchers have named this issue the 'Invariance Problem', which anatomically appears to find a particularly effective solution in the inferotemporal (IT) cortex of humans and monkeys. In fact, IT is the latest stage of the ventral visual stream, the neural pathway devoted to the processing of visual information supporting complex object identification (for a review, see: Logothetis and Sheinberg, 1996; Rolls, 2000; Tanaka, 1996).

On the other hand, a critical step in the interpretation of the visual world is the integration of the various local motion signals generated by moving objects. This process is complicated by the fact that local velocity measurements can differ depending on contour orientation and spatial position. Specifically, when an object larger than its receptive field moves across the visual field, any local motion detector can only measure the component of motion perpendicular to a contour that extends beyond its field of view. This 'Aperture Problem' (reviewed by Chagas et al., 1986) is particularly relevant to direction-selective neurons early in the visual pathways, where small receptive fields allow only a limited view of a moving object. Neurons in the middle temporal visual area (known as MT or V5), which marks the next-to-last stage of the dorsal visual stream (the neural pathway devoted to the processing of visual motion information, further projecting to the parietal lobe), appear to offer a dynamic solution to this problem, by integrating information from upstream units and producing a 'global' motion percept (Chagas et al., 1986; Hildreth and Koch, 1987; Rust et al., 2006). However, recent research has shown that such problem might be already solved at the first stages of motion detection by particular groups of neurons in the primary visual cortex, known as 'end-stopped cells' (Pack and Born, 2001; Pack et al., 2003; Tinsley et al., 2003; van Wezel and van der Smagt, 2003). End-stopped cells were firstly described by Hubel and Wiesel (1968), who referred to them as 'hyper-complex cells'.

In sum, areas IT and MT appear to perform a fast and thorough integration of inputs coming from the early visual stages, leading to accurate and coherent representations of the external world; nevertheless, understanding nature of the visual input at each step of the processing stream, i.e. the kind of information that is conveyed from one stage to the next and the way such information is parametrized, still remains a major challenge for the vision science community.

# 1. The Ventral and Dorsal Streams of Processing

Anatomical and physiological studies have identified over 50 distinct visual processing areas in the primate cortex (Felleman and Van Essen, 1991). About thirty years ago, Mishkin and Ungerleider (reviewed by Mishkin and Ungerleider, 1983) proposed that these multiple areas are organized into two major processing streams, known as the ventral (or temporal) pathway and the dorsal (or parietal) pathway, both originating in the primary visual cortex (Fig. 1).

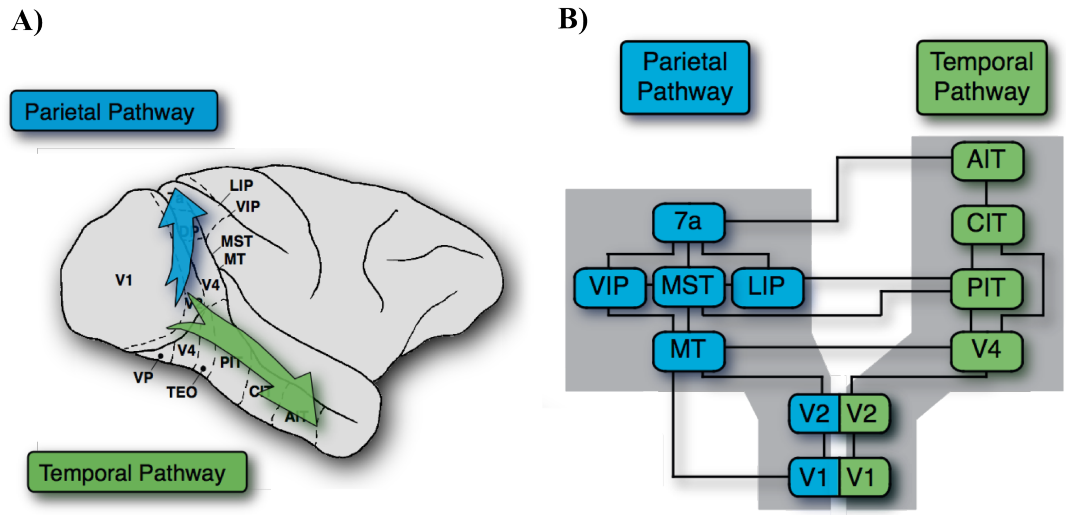


**Figure 1. The ventral and dorsal streams of processing in the monkey brain.**
**A)** Cortical localization of the pathways. **B)** Cortical areas subdivision. Both streams originate in V1; the parietal stream projects to the parietal areas VIP (ventral intraparietal sulcus), LIP (lateral intraparietal sulcus) and 7a, including areas V5 or MT (middle-temporal cortex) and MST (medial superior temporal cortex). The temporal streams projects to the temporal areas PIT (posterior inferotemporal cortex, aka TEO), CIT (central inferotemporal cortex) and AIT (anterior inferotemporal cortex, aka TE), including areas V2 and V4.
Adapted from Mishkin and Ungerleider, 1983.

The ventral stream, or pathway, travels into the temporal lobe, including cortical areas V2, V4, and the IT (Inferotemporal) cortex, and is thought to be crucial for the visual recognition of objects (also known as the "what" stream). The dorsal stream, or pathway, projects to the posterior parietal cortex, and includes areas V2, the MT (middle-temporal) cortex, and the MST (medial superior temporal) cortex, and is thought to be crucial for motion integration, for encoding spatial relationships between objects and for visual guidance

toward objects (also known as the “where” stream) (Baizer et al., 1991; Mishkin et al., 1983; Morel and Bullier, 1990).

For almost three decades, this theory has been established as one of the most dominant theories of primate visual system organization.

As a matter of fact, single-unit recording studies are consistent with the two streams hypothesis. For example, neurons in the ventral stream show selectivity for orientation, shape and texture while those in the dorsal stream show selectivity for the direction and speed of visual motion (Maunsell and Newsome, 1987; Ungerleider and Pasternak, 2004).

Hence, motion processing would involve first computation by direction-selective sensors in V1, followed by integration up to MT to solve the aperture problem; form processing would involve filtering by orientation-selective units in V1, up to size-selective and other features-selective units, followed by integration up to IT to encode object shape and solve the invariance problem.

Modularity in the processing of different visual attributes has the advantage of allowing each stream to optimize its computations for the relevant visual attribute, ensuring that limitations or errors in processing output remain confined, rather than be propagated across attributes.

However, recent works have started to question the established principle of parallel, modular processing streams, being more in favor of a constant and active interplay at different stages between form and motion processing systems.

Such interplay is already evident when analyzing the properties of neurons in the early stages of processing: the apparent direction of a moving line depends on how the local orientation and direction of motion combine to match the receptive field properties of motion-selective neurons in the early visual areas. Since the pioneering works by Hubel and Wiesel (1968), there has been extensive physiological evidence that the receptive fields of direction selective neurons in V1 extract the motion component orthogonal to local orientation, so their directional response is ambiguous (the ‘aperture problem’). Such ambiguity is thought to be solved, as already mentioned and as we will see later in detail, by area MT, which appears to integrate the responses of different V1 cells (Simoncelli and Heeger, 1998), though it might be solved already at earlier stages of processing by the so-called end-stopped cells (Pack and Born, 2001; Pack et al., 2003; Tinsley et al., 2003; van Wezel and van der Smagt, 2003). However, although such cells in V1 respond to the motion of line-terminators independently of line orientation, the orthogonal motion component has been shown to affect the response of motion-selective neurons at later stages in MT (Pack and Born, 2001). The interaction

between orientation and motion direction at the very early levels of cortical analysis show that processing of these two attributes is inextricably linked (reviewed by Mather et al., 2013).

Moreover, behavioral studies have shown that motion information *per se* is also important for object recognition. Intuitively, motion provides visual cues for object boundaries, local part structures and view-specific features. For example, the direction of rotation in depth, which preserves 3D shape and view information, affects observers' performance across a range of stimuli and recognition tasks. Several studies have shown that reversing learned motion patterns impairs recognition performance (Vuong and Tarr, 2004; Wallis and Bülthoff, 2001), and that such 'motion reversal effect' can be influenced by similarities in the 3D structure of objects (Vuong and Tarr, 2006).

Interestingly, several human fMRI studies investigating structure-from-motion (SFM) processing have identified multiple sites along the ventral and dorsal pathways that are concurrently involved in deriving 3D shape from the motion of random-dots displays.

Deliver an invisible form composed of randomly arranged dots against a dotted background. As soon as the form moves, it becomes visible, by virtue of what the Gestalt psychologists have named 'the common fate' of its components, which all move together with a common speed and direction (Ledgeway and Hess, 2002; Uttal et al., 2000). This phenomenon clearly shows that form perception can emerge from motion processing, in the absence of any other cue, suggesting that motion information represented in the dorsal stream can feed object recognition performed in the ventral stream.

Kriegeskorte et al. (2003) investigated the motion flowfield component of object recognition with functional magnetic resonance imaging. SFM stimuli consisted of face surfaces, random three-dimensional shapes with matched curvature properties. Control stimuli consisted of static and moving random dots patterns. The dots in the moving conditions were SFM-matched. The authors showed that face surfaces elicited a stronger response in the whole ventral stream, particularly the fusiform face area (FFA) as compared to random shapes. Interestingly, the human motion complex (+hMT, homologue of primate MT/MST) increased its activity in response to the random moving dot stimuli, but also to motion-defined object (shape or face), showing to play a central role in SFM object recognition.

Finally, a couple of other fMRI study have reported object-selective responses in motion-associated areas (Kourtzi and Kanwisher, 2000; Kourtzi et al., 2002). Object-selective responses were observed in +hMT for moving objects as well as for static objects implying motion (e.g., a still photograph displaying a motion scene), but not for static 2D objects. Also, +hMT displayed higher responses for intact than for scrambled images of objects.

Evidently, the influence of form signals on motion processing, and vice-versa, is more extensive than previously thought.

For the sake of simplicity, I will first describe separately the two 'relay stations' of object and motion neural processing, IT cortex and MT cortex. Similarly, I will separately explore the functional mechanisms of object recognition and motion processing in each section.

## 1.1. IT Cortex and Object Recognition

Our capacity to identify visual objects relies on a large neural network, from the retina to the prefrontal cortex, which includes the so-called 'ventral pathway', a set of posterior cortical areas extending from the primary visual cortex (V1) to the infero-temporal (IT) cortex. The inferotemporal cortex (IT) of the monkey brain has been divided into subregions in several different manners. Posterior IT corresponds to TEO, and anterior IT to TE.

IT projects to various brain sites outside the visual cortex, including the perirhinal cortex (areas 35 and 36), the prefrontal cortex, the amygdala, and the striatum of the basal ganglia. The IT cortex, is homologue of human LOC (Lateral Occipital Cortex), whose role in object recognition has been well established (Grill-Spector et al., 1999; James et al., 2003; Kim et al., 2009).

### 1.1.1. Functional Properties of IT Neurons

The whole ventral pathway can be characterized by a hierarchical architecture in which neurons in higher areas code for progressively more complex representations by pooling information from lower areas (reviewed by Rolls, 2012; Tanaka, 1996).

In other words, the fundamental object processing steps appear to be reflected in the functional architecture, meaning that the complexity of the critical features and the receptive field sizes increase in a bottom-up fashion through the stream. Cells in IT respond to complex stimulus features, dealing mostly with object's attributes like shape, texture and color (Gross, 1992; Kobatake and Tanaka, 1994; Tanaka, 1996).

Particularly, these neurons are activated by moderately complex combinations of visual features (Brincat and Connor, 2004; Kobatake and Tanaka, 1994; Rust and Dicarlo, 2010; Yamane et al., 2006) and they are often able to maintain object selectivity over changes in object position and size (Ito et al., 1995; Li et al., 2009; Rust and Dicarlo, 2010; Tovee et al.,

1994), pose (Logothetis et al., 1994), illumination (Vogels and Biederman, 2002) and clutter (Li et al., 2009; Missal et al., 1999; Zoccolan et al., 2005).

Thus, even though IT cells are just "a few synaptic connections away" from primary visual cortex, the receptive field size and the complexity of the critical stimuli (i.e. the stimuli that elicit optimal responses by the neuron) substantially increase from V1 to IT. In primate V1, neurons have small RFs and thus provide precise information about the position of the stimulus in the visual field. However, because TE neurons receive massive convergent inputs from lower areas, they integrate information from larger areas of the visual field and consequently have large RFs. Hence the capability of these cells to respond to complex stimulation and a more structured visual information, as compared to the earlier visual stages, and their ability to convey information about their preferred shapes in spite of variation in 2D retinal position, 2D rotation in the image plane, 3D rotation in depth, illumination, and so forth (Fig. 2). Amongst these multiple degrees of freedom, different subspaces of parameters represent the effective stimuli of IT cells.

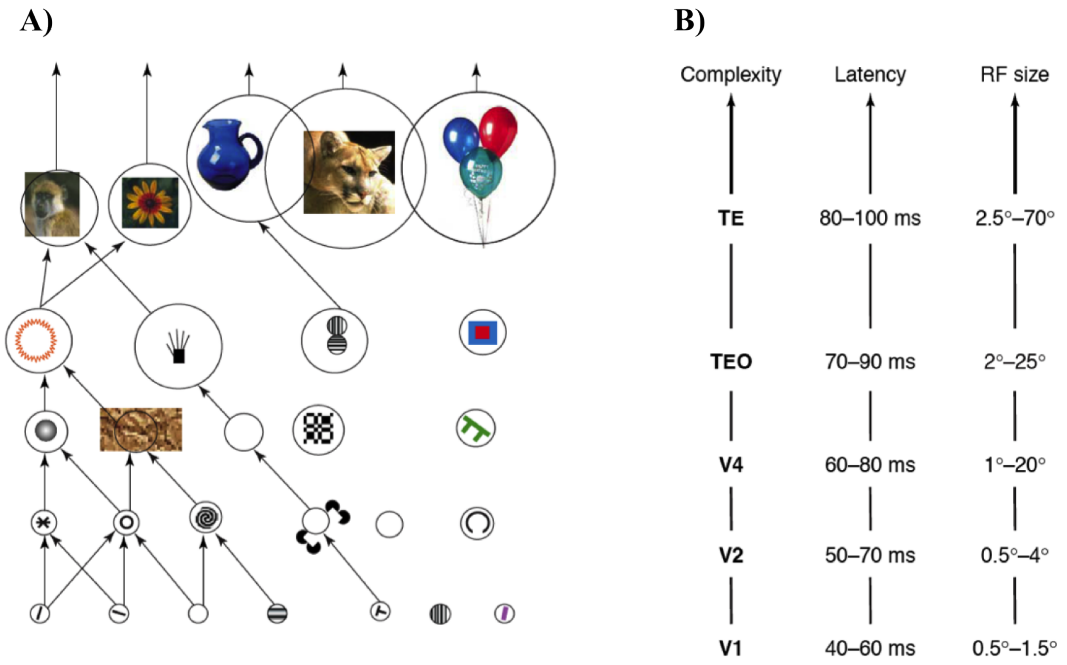


**Figure 2. A Schematic representation of the ventral visual pathway.**

**A)** Through a hierarchy of cortical areas, complex object representations are progressively built by integrating convergent inputs from lower levels. Examples of elements for which neurons respond selectively are represented inside receptive fields (RFs; represented by circles) of different sizes.

**B)** Schematics of the progressive increase in the 'complexity' of the neuronal representations from V1 through temporal-occipital cortex (TEO/LIP) to temporal cortex (TE/AIP). The rightmost column displays estimates of the smallest and largest RF sizes reported in the literature. As RF

size increases (presumably to allow translation and size invariance), neurons at higher levels typically receive inputs from more than one object at a time during natural scene perception. The central column displays an estimate of the minimum and the average response latencies reported in the literature.
Adapted from Rousselet et al., 2004.

Single-cell and optical imaging recordings have shown that IT cells responsive to similar object features cluster into vertical columns across the cortical layers (Fujita et al., 1992; Tsunoda et al., 2001; Wang et al., 1998). Moreover, specialized "modules" in human and monkey IT cortex have been described, in which specific regions (e.g., the fusiform face area (FFA) and the parahippocampal place area (PPA)) are selectively activated by discrete object categories (i.e., faces and places, respectively) (Desimone, 1991; Desimone and Gross, 1979; Epstein et al., 1999, 2003; Grill-Spector et al., 2004; Kanwisher et al., 1997; Zangenehpour and Chaudhuri, 2005).

Although early studies have described RFs in area TE as being large, it is only recently that a fully systematic study has been performed, by Op De Beeck and Vogels (2000). This study found a high degree of variability in RF size, where some are actually rather small (~ 3˚). Also, the authors found a positive effect of stimulus size on the size of the RFs.

This would seem to go against the classic description of TE neurons as being insensitive to changes in stimulus size. However, although true for many neurons, some actually modulate their response as a function of object size and others respond selectively to a stimulus at a particular size (Lueschow et al., 1994), suggesting a task-related adjustment of RF size.

### 1.1.2. How Does IT Provide a Solution to the Invariance Problem?

The study of object recognition must provide an account of how observers compensate for a wide variety of changes in the object's appearance.

As we mentioned in the beginning, object recognition is computationally difficult for many reasons, but the most fundamental is that any individual object can produce an infinite set of different images on the retina, due to variation in object position, scale, pose and illumination, and the presence of visual clutter.

The primate ventral stream produces a particularly effective solution to such 'Invariane problem' in the inferotemporal (IT) cortex.

As we have seen, IT cortex represents an important step in the transformation from low-

level shape signals to complex object representation, by producing an integration of elementary information processed by upstream areas, through both linear and non-linear mechanisms (Brincat and Connor, 2004; DiCarlo and Cox, 2007; DiCarlo et al., 2012; Wallis and Rolls, 1997).

Along the stream neurons process features that are more and more complex, until IT where the objects seem to be processed in a complex and complete way. But how is the flow of information processed and conveyed at each step along the stream?

In this regard, representational geometry has proven to allow an effective formalization, and a possible solution, to the invariance problem, as to many other issues concerning sensory processing.

We can think of a sensory representation in a certain brain region in terms of a multidimensional space. The dimensions of the space correspond to the neuronal responses, and a point corresponds to an activity pattern (i.e., each neuron's activity provides the coordinate value for one of the dimensions). A visually perceived object, for example, will correspond to a point in the representational space of a given visual area. The set of all possible objects corresponds to a vast set of points in the space. It is the geometry of these points that defines the nature of the representation (reviewed by Kriegeskorte and Kievit, 2013; Fig. 3).

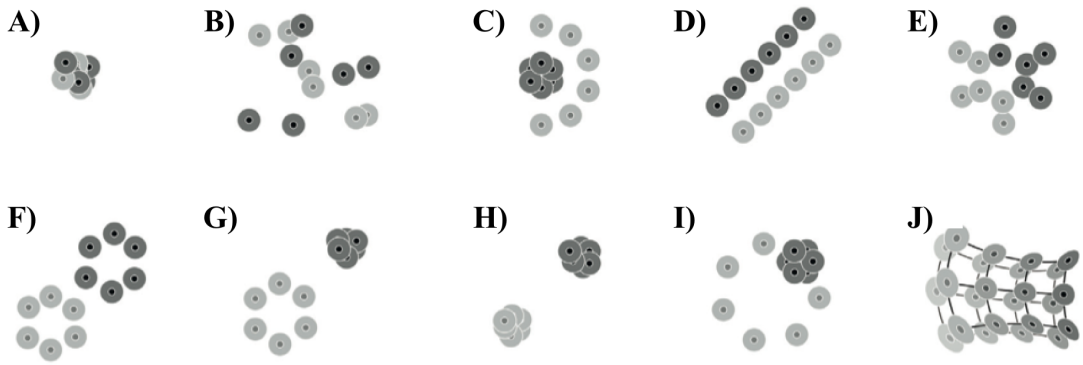


**Figure 3. Illustration of ten hypothetical representational geometries for ten different brain regions.**

Each dot corresponds to the representation of a particular piece of content (e.g., a visual object). The space in which the dots are placed is the space of representational patterns (illustrated as two-dimensional, but high-dimensional in reality). The halo regions around the dots indicate the margin of error; dots with overlapping error halos are indistinguishable. The items fall into two categories (dark or light), or in the case of geometry 10, on a continuous manifold (shades of gray). **A)** No item is distinct from any other item.

**B)** Most items are distinctly represented but the categories cannot be separated by any simple boundary.

**C)** Only the light items are distinctly represented and they are separable from the dark items by a quadratic boundary.

**D)** Dark and light items are linearly separable and arranged along parallel lines with pairs of dark and light dots matched up across the boundary.

**E)** The items form a single cluster but the categories are linearly separable.

**F)** The items form two category clusters that are linearly separable and within which all items are distinct.

**G)** Like the previous case, but the items in the dark category are indistinguishable. H) Like the previous case, but only the category distinction is represented; items within each category are indistinguishable from each other.

**I)** The dark items are indistinguishable and located among the distinctly represented light items on a circle.

**J)** Items fall on two manifolds that closely follow each other, with pairs of items matched up across them.

Adapted from Kriegeskorte and Kievit, 2013.

A particularly effective geometrical description of the invariance problem draws from the notion that the ventral visual pathway would gradually 'untangle' information about object identity (DiCarlo and Cox, 2007; DiCarlo et al., 2012; Fig. 4).

When an object undergoes a transformation, such as a shift in position or a change in pose, it produces a different pattern of population activity, which corresponds to a different response vector. The ensemble of response vectors corresponding to all possible identity preserving transformations defines a low-dimensional surface in this high-dimensional space: an object manifold. For neurons with small RFs that are activated by simple light patterns, such as retinal ganglion cells or simple cells in V1 cortex, each object manifold will be highly curved, and the manifolds corresponding to different objects will be entangled. At higher stages of visual processing, neurons tend to maintain their selectivity for objects across changes in view; this translates to discrete and more flat object manifolds.

Thus, object manifolds are thought to be gradually untangled through selectivity and invariance computations applied at each stage of the ventral pathway.

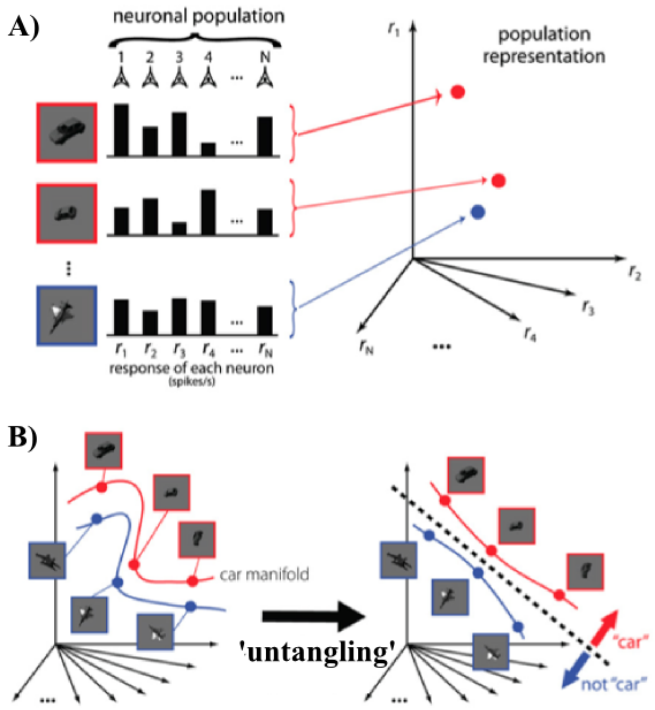


**Figure 4. A geometrical description of the invariance problem.**

**A)** The response pattern of a population of visual neurons to each image (three images shown) is a point in a high-dimensional space where each axis is the response level of each neuron.

**B)** All possible transformations of an object will form a low-dimensional manifold of points in the population vector space, i.e., a continuous surface (represented here, for simplicity, as a one-dimensional trajectory; see red and blue lines). Neuronal populations in early visual areas (V1) contain object identity manifolds that are highly curved and tangled together (see red and blue manifolds in left panel). The solution to the recognition problem is conceptualized as a series of successive re-representations along the ventral stream (black arrow) to a new population representation (IT) that allows easy separation of one object's manifold (e.g., a car; see red manifold) from all other object identity manifolds (of which the blue manifold is just one example). Geometrically, this amounts to remapping the visual images so that the resulting object manifolds can be separated by a simple weighted summation rule (i.e., a hyperplane, see black dashed line). Adapted from Di Carlo et al., 2012.

In short, object recognition is the ability to segregate images that contain one particular object from images that do not. In this geometrical perspective, this means the setting of a decision boundary, like a hyperplane, to separate all the views of one object from all the views of the other objects.

In principle, one can think of such a decision boundary as approximating a higher-order neuron, or a population of them, pooling from upstream neurons and computing object

identity via a simple weighted sum of each neuron's responses. Such operation would be impossible in the early stages of shape processing: only at later stages, manifolds discretize, and a simple hyperplane would be sufficient to separate them.

Another way to see this phenomenon is considering the properties of selectivity for the object identity and tolerance for its identity-preserving transformations as signals that propagate through the ventral visual stream. As already mentioned, object recognition requires both selectivity among different objects and tolerance to vastly different retinal images of the same object, resulting from natural variation in its appearance. Thus, identifying neuronal responses that have object selectivity and tolerance to identity-preserving transformations is fundamental to understanding object recognition.

Zoccolan et al. (2007) showed that IT neurons that attained the highest object selectivity were, on average, the least tolerant to identity-preserving image transformations and vice versa. That is, selectivity and tolerance trade off within IT.

On the other hand, Rust and Dicarlo (2010) found that neurons at lower levels of visual processing encode more local structure whereas neurons at higher stages of the visual system become more sensitive to specific conjunctions of those local features, affording the possibility of complex object selectivity.

Moreover, IT's encoding strategy is more tolerant (more invariant) to identity-preserving object transformations than the V4 population. From a geometrical point of view, in IT the population activity tends to cluster in the same manifolds, in spite the transformations it may undergo. The parallel increase in selectivity and tolerance of IT cells can be seen as the gradual increase of the ability to untangle the object manifolds along the ventral stream.

## 1.2. MT Cortex and Motion Processing

The middle temporal area (MT or V5) receives direct and primary input from layer 4B of V1, which represents 90% of the total input of MT (Born and Bradley, 2005), and provides a major output to the dorsal stream of processing, particularly to MST and VIP (ventral intraparietal) cortices. As we have seen, the dorsal stream represents the neural network traditionally associated with the processing of motion information. In humans, brain imaging studies have revealed a network of brain areas responsive to motion; among these is the hMT+ complex, which is a homologue of macaque MT, MST and VIP (Tootell et al., 1995; Zeki et al., 1991).

### 1.2.1. Motion Sensitivity in V1 and MT Neurons

In primates, the first place in which we can find motion-sensitive neurons is the primary visual cortex (V1), where about 25% of the neurons display direction selectivity (Hubel and Wiesel, 1968), meaning, they respond to one direction of motion and show little or no responses to the opposite direction.

In primary visual cortex (V1) of macaque monkeys, direction selectivity appears first in layer 4Cα (Gur and Snodderly, 2007; Hawken et al., 1988; Snodderly and Gur, 1995), and it initiates motion selectivity in the dorsal cortical stream through its projections to layer 4B and layer 6, both of which have many direction selective cells (Gur and Snodderly, 2007; Hawken et al., 1988; Livingstone and Hubel, 1984; Orban et al., 1986). Layers 4B and 6 project monosynaptically to MT (Movshon and Newsome, 1996; Nassi and Callaway, 2009), where almost all cells are direction selective.

Directionality in MT is organized in a columnar pattern, such that neurons across the cortical layers within a column prefer the same direction of motion while direction preference systematically changes across columns. These changes can either be gradual or abrupt with preferred direction changing by 180 degrees (Albright, 1984; Dubner and Zeki, 1971; Geesaman et al., 1997).

A columnar organization is also present in terms of binocular disparity tuning (DeAngelis and Newsome, 1999) and speed preference (Liu and Newsome, 2003).

The fairly ordered structure of MT can be also appreciated by its complete retinotopic representation of the contralateral visual hemifield. Finally, neurons in MT are heavily myelinated, and display a latency of 50-60 milliseconds (Van Essen et al., 1981; Maunsell and Van Essen, 1987). Area MT is also known to be involved in pursuit eye movements (Dürsteler et al., 1987; Movshon et al., 1990; Newsome et al., 1985).

Since sensitivity to motion stimuli is already present at the very early stages of the dorsal stream, one might question in what guise, i.e. by which functional properties, V1 neurons are to be considered different from MT neurons.

A number of studies in the macaque have specifically tackled the issue by investigating the response of areas MT and V1 to a set of moving and static stimuli.

One of the first of these studies was conducted by Albright (1984), who found, as mentioned before, a dramatically higher neural population in MT selective for motion, measured as the response to random dot fields, a moving spot and moving slits, as compared to V1: approximately 1/3rd of V1 cells are directionally-selective; this is true for ~90% of MT

cells.

Also, the direction tuning bandwidth for moving stimuli is slightly broader for MT neurons as compared to V1 neurons. Orientation tuning bandwidth, measured as the response to stationary slits, is slightly narrower than direction tuning bandwidth, but still broader than the one of V1 cells.

Finally, the response magnitude (spikes/s) of MT neurons is higher when processing moving stimuli as compared to stationary ones, while V1 cells display responses of equal magnitude to moving stimuli and stationary stimuli.

Noticeably, in area MST which adjoins the inner border of MT, neurons selective for spiral and contraction/expansion motion, as tested by means of moving dot patterns, have been reported (Graziano et al., 1994; Tanaka and Saito, 1989).

Importantly, Albright was one the first to report two sub-populations of MT cells: one (~60%) which displays an orientation preference nearly perpendicular to the direction of motion (as V1 cells), another showing an orientation bias nearly parallel to the preferred direction, suggesting specific neural mechanisms underlying pattern-motion sensitivity in area MT (see 1.2.2. Pattern and component responses in MT neurons).

Another important difference between these two populations concerns the size and properties of the receptive fields (RFs). Mikami et al. (1986) have shown that MT cells detect directional differences of moving spots, presented in both stroboscopic and smooth motion, over spatial intervals that are three times larger than those detected by V1 cells. This capability is related to RF size, which is one order of magnitude larger that V1 cells at equivalent eccentricities (Albright and Desimone, 1987; Maunsell and Van Essen, 1987).

In contrast, no significant difference in the maximal temporal interval for direction selectivity was found. According to the authors, these findings might imply that the role of MT is to extend direction selectivity at higher stimulus speeds by increasing the spatial width of the RF subunits that detect directional differences, rather than increasing the temporal properties.

A more recent work by Snowden et al. (1992) characterized the tuning properties of V1 and MT neurons in response to random dot patterns stimulation. Although nearly all V1 cells tested were driven by the random dot patterns, many cells did not respond in a uniform manner during the stimulus presentation time. Instead these cells tended to fire at a certain time (and not at other times) during stimulus presentation. Other V1 cells gave responses which were much more consistent over the time course of the stimulus. Noticeably, nearly all MT cells gave responses that were similar to this continuous, uniform response.

In other words, according to the authors, MT cells exhibit a uniform (or 'field', in the terminology of Gulyás et al., 1987, who had reported the same property in cat's V1) response, while V1 cells show a not uniform ('grainy', Gulyás et al., 1987) response, and dot density cannot be accounted for this effect. This would suggest that V1 cells fire to a specific feature or phase relationship within the dot pattern, rather than its motion properties.

Hence, MT appears to add some relevant value in the kind of information the early stages of the dorsal stream would be able to process alone. One question that arises at this point is: what is the nature of such added value? As a corollary, does the properties of MT neurons simply produce an integration of the properties of its inputs from V1, or additional processing are needed to occur to significantly modify their responses?

### 1.2.2. Patterns, Components and the Aperture Problem

As we have seen for object recognition, motion processing also requires characteristics of *invariance* and *selectivity*.

For MT neurons to support motion perception their direction tuning properties must be invariant with respect to the details of the stimulus. Moreover, the process of integration of local information about the stimulus must be selective, meaning it must unambiguously convey information about the direction of motion.

The mechanisms of integration and the possible strategies by which selectivity is obtained is typically addressed by the use of a particular type of visual stimulation, the so-called 'plaid patterns', formed by the superimposition of two or more drifting gratings (Adelson and Movshon, 1982; Chagas et al., 1986; Jazayeri et al., 2012; Rust et al., 2006; Tinsley et al., 2003; Fig. 5).

Such a kind of stimulation was first used by Movshon and colleagues (in Chagas et al., 1986). The authors proposed a model according to which motion processing is performed through different steps, initiating with a mere analysis of local orientation and culminating in the integration of low-level features to produce a 'global' motion percept. The authors hypothesized a 'component' and a 'pattern' response to be found at the different levels of the process: the first step units, or 1D analyzers, would respond to the direction of motion of their components, and not to the resulting moving direction of the combination of the two; 2D analyzers, would, instead respond to this latter parameter.

Clearly, this kind of stimulation happens to be particularly suitable when addressing what we have mentioned before as 'The Aperture Problem': a critical step in the interpretation

of the visual world is the integration of the various local motion signals generated by moving objects. Local measurements are a prerogative of V1, where neurons selective for orientation and direction first compute the motion of oriented edges in visual scenes, by filtering the image in both space and time. Neurons in this area typically have small RF sizes; therefore, the object's contours will extend beyond the neuron's region of excitability, making a 'partition' of the motion information necessary. However, the motion of a single contour does not allow by itself to determine the motion of the surface containing that contour. This would be solved by postulating a hierarchical model in the visual cortex in which a velocity value could be unambiguously assigned to a 2D pattern, given knowledge of the motion of its 1D components. The speed and direction of motion of a plaid pattern can be calculated from the motion of its components using the 'intersection of constraints' (IOC) computation (Adelson and Movshon, 1982): if each component's motion is represented as a vector, the only velocity vector that is compatible with both component velocities is the one that ends at the intersection of the constraint lines of the two components (Fig. 5).

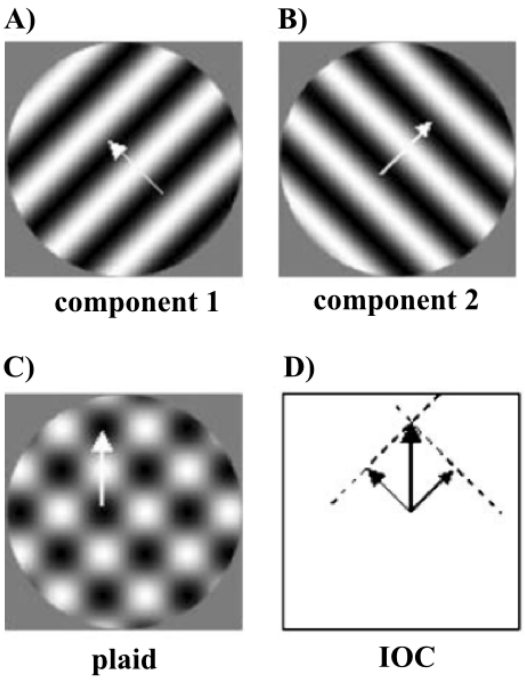


**Figure 5. Plaid pattern, its constituent gratings and intersection of constraints (IOC) computation.**

Gratings in **A)** and **B)** add together to form the plaid in C). Each grating has direction and speed of motion indicated by white arrows. The plaid has direction and speed of motion that can be calculated from intersection of constraints (IOC)(D).

Adapted from Tinsley et al., 2003.

Consistent with the results of Albright (1984), the majority of cells in V1 displayed a 'component direction selectivity', since their tuning curves revealed exclusive preferences for the components' direction of motion. On the other hand, MT neural population displayed both component direction selectivity (about 40% of the recorded cells), and pattern direction selectivity (25%), as measured to their response to the plaid motion direction.

Such findings led to the suggestion that the intersection of constraints is calculated in area MT, and naturally suggest that direction-selective neurons in V1 and MT are the neural correlates of the hypothesized first and second stages of motion processing. The issue is: in what way are V1 neurons that project to MT different from other randomly sampled V1 neurons?

Movshon and Newsome (1996) addressed the question by asking, i) whether the projection neurons are directionally selective at all; and ii) whether the projection neurons respond to plaid patterns in a manner characteristic of the first ('component') or second ('pattern') stage of motion processing.

To answer the questions, the authors analyzed a population of V1 neurons that were antidromically activated by electrical stimulation of MT. The large majority of these neurons were strongly directional and responded selectively to the motion of the component gratings of plaid patterns. The MT projection neurons were typically responding to a broad range of spatial and temporal frequencies (namely 'special complex', aka end-stopped, aka hypercomplex cells). The projection neurons thus comprise a homogeneous and highly specialized subset of V1 neurons, consistent with the notion that V1 performs the basic visual measurements, and then properly distributes information to higher cortical areas for specialized analysis.

It would seem that the aperture problem is handled by the latest stages of the processing stream, in area MT, which would perform a dynamic integration of downstream local information, and such integration appears to be prerogative of specialized pattern-selective neurons.

At this point one could question: what is the influence of shape, or 'component' information, as conveyed by V1 neurons in motion processing? In other words, how do pattern-selective neurons integrate information from component-selective neurons?

According to Pack and Born (2001), the temporal tuning dynamics of MT neurons in response to plaids show that their motion sensitivity is largely independent by stimulus

orientation, but such independence takes some time to become fully developed, with a stereotyped time-course. The authors reported 60 MT cells in the macaque brain displaying an initial orientation-dependent motion sensitivity, which started decreasing during the course of prolonged stimulation, after about 60ms. Within 150ms after the onset of stimulus motion, MT cells primarily encoded the actual stimulus direction, irrespective of orientation.

This would suggest that MT cells need some time to get independent from stimulus shape, and that the primate visual system derives an initial estimate of motion direction by integrating local motion and refines this estimate over time. The authors report a behavioral correlate of these neural responses: the initial velocity of pursuit eye movements deviates in a direction which is perpendicular to local contour orientation, indicating the influence of the earliest neural computations on the oculomotor reaction.

Priebe et al. (2003) investigated whether speed in MT is coded in a way that is invariant to the shape of the moving stimulus, and if so, how. The authors pointed out that the idea of investigating motion by characterizing the visual scene according to its spatial and temporal sine-wave components (in Fourier space) might be misleading. For example, sine-wave gratings are characterized by a spatial frequency, defined in cycles per degree as the inverse of the width of a single cycle of the grating, and a temporal frequency, defined in cycles per second as the inverse of the time required for the intensity of a single pixel to go through a full cycle of sinusoidal modulation. The speed of a moving grating is the ratio of the temporal frequency and the spatial frequency. Although sine-wave gratings are commonly used in the laboratory setting to assess the response properties of neurons, moving real-world objects contain multiple spatial and temporal frequencies. Therefore, motion processing of natural objects might be quite different from what happens in laboratory conditions, and whatever model for motion tuning must take this into account.

When tested with single sine-wave gratings of different spatial and temporal frequencies, MT neurons show a continuum in the degree to which preferred speed depends on spatial frequency: 75% of MT neurons displayed dependence, while the other 25% maintain speed tuning despite changes in spatial frequency. However, when tested with stimuli constructed by adding two superimposed sine-wave gratings, the preferred speed of MT neurons became less dependent on spatial frequency. Analysis of these responses revealed a speed-tuning nonlinearity that selectively enhances the responses of the neuron when multiple spatial frequencies are present and moving at the same speed. Consistent with the presence of the nonlinearity, MT neurons show speed tuning that is close to be form-invariant when the moving stimuli comprise square-wave gratings, which contain multiple spatial frequencies

moving at the same speed. The authors conclude that MT neurons derive form-invariant speed tuning in a way that takes advantage of the multiple spatial frequencies that comprise moving objects in natural scenes.

In line with this report, Jazayeri et al. (2012) show that pattern-selective neurons integrate component information non-linearly, i.e. they do not respond to the entire pattern as it would to the individual components. The authors addressed the issue of invariance in MT neurons by measuring their response to a particular type of dynamic stimulus, namely a 'triplaid', which contains 3 component and 3 pattern motions simultaneously.

The actual visual stimulation consisted of 3 successive epochs going from a single grating, to a plaid, to a triplaid. The authors used the responses during the first 2 epochs to characterize cells as 'component-direction selective' (CDS) or 'patter-direction selective' (PDS), and then used responses during the triplaid epoch to assess whether these characterizations were robust when the stimulus contained other competing motion signals. The direction tuning of the CDS cells during the triplaid epoch had 3 peaks that corresponded to the 3 constituent gratings, compatible with the behavior of an ideal CDS cell that maintains its selectivity for component motion even in the presence of other motion signals. However, the triplaid tuning function for PDS cells did not have clear peaks for the 3 pattern motions; instead, tuning functions became nearly flat with no strong preference for either the component or pattern motion signals. These findings indicate that pattern selectivity, unlike component selectivity, pattern selectivity is highly susceptible to the presence of other competing motion signals.

The authors explained this reduced selectivity by an MT model that explains pattern direction selectivity as gained by pooling V1 afferents with broad excitation and strong opponent inhibition (Rust et al., 2006): the reason why PDS cells do not respond to the preferred plaid in the triplaid stimulus might be due to the activation of this opponent mechanism by the third grating that moves opposite to the preferred plaid.

### 1.2.3. An Early Solution to the Aperture Problem: V1 End-Stopped Cells

Interestingly, some of the properties of MT neurons, such as independence from stimulus shape features such as orientation and time-dependent learning dynamics can already be found in a particular sub-population of V1 neurons, the so-called 'end-stopped' (or hyper-complex) cells.

V1 'simple' cells respond to lines and edges at a very specific orientation, and at a very

specific position within the field of view; V1 ‘complex’ cells fire, like simple cells, to lines with a specific orientation, but the position of the line within the field of view is not critical. Thus, the essential difference between the two types consists in, for a simple cell, the importantly narrow range of positions over which an optimally oriented line evokes a response; for a complex cell, the responses to a properly oriented line irrespective of where it is placed in the receptive field. This behavior is related to the explicit on and off regions of a simple cell, and to the lack of such regions in a complex cell.

A sub-type of the complex cells (hyper-complex cells) responds well to short contours or endings of contours, but the response is suppressed by longer contours. This characteristic of these neurons is referred to as ‘end-stopping’, since these cells respond best to oriented edges which are ‘stopped’, i.e., their end does not extend beyond a specific part of their receptive field.

In fact, an ordinary simple or complex cell usually shows length summation: the longer the stimulus line, the better is the response, as far as the line is as long as the receptive field; making the line still longer has no effect. For an end-stopped cell, lengthening the line improves the response up to some limit, but exceeding that limit in one or both directions results in a weakening or suppression of the response (Hubel, 1995; Hubel and Wiesel, 1968).

A number of recent studies have suggested that neurons with end-stopping behavior are optimally suited to solving the aperture problem.

For instance, an end-stopped neuron responds to the contour-ending of the wing of a plane in accordance with the perceived motion direction (Fig. 6). Non-end-stopped cells, however, respond to the long contour of the wing but signal the wrong direction.

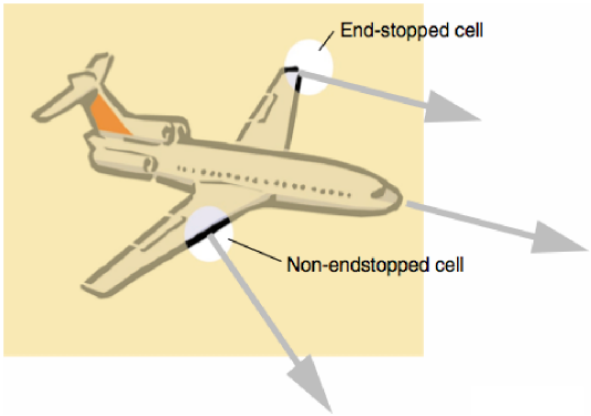


**Figure 6. The aperture problem of motion vision could be solved by end-stopped cells.**
A non-end-stopped V1 neuron can signal an incorrect motion direction of a contour of a moving object. Through an aperture the contour seems to move perpendicular to the orientation of the

contour, and this does not always coincide with the direction of the whole object. An end-stopped neuron signals the correct motion direction for a contour end.
Adapted from van Wezel and van der Smagt, 2003.

In contrast to these considerations, neurophysiological evidence has tended, as we have seen, to support the idea that the aperture problem is solved at a later stage of motion processing, where information from V1 neurons is combined. The work by Pack et al. (2003) seems to refute this generally accepted hierarchical model, showing that this subtype of V1 cells responds to the overall direction of motion, and not to the components that create it.

The authors trained macaque monkeys to maintain visual fixation while a stimulus of white and black bars, moving in different directions in combination with different orientations, was presented on a grey background at the receptive field location of a recorded cell. End-stopped cells responded best when the bar endpoints were in the receptive field, and poorly when the bar was centered on, and extended beyond, the receptive field, and displayed a preferred motion direction which was independent of the bar's orientation. It is this independence of stimulus orientation that makes these cells suitable candidates for solving the aperture problem, without the need for further integration in area MT.

To sum up, end-stopped cells are strongly direction-selective, and such selectivity is held consistent throughout different stimulus orientations: rotations of the bar stimulus of 45° clockwise and counter-clockwise led to similar profiles of response. As for the time course of the response, the authors found that, just like pattern-selective responses in MT, end-stopping features develop with a slight delay: the end-stopped response profile takes 20-30 milliseconds to show up.

We have seen that another well-established view holds that gratings are processed in V1, and that those signals are combined in a subsequent stage of visual motion processing. Evidence for this integration stage has been found in area MT, which contains neurons responding specifically to moving plaids.

However, Tinsley et al. (2003) have shown that the response strength to patterns or components is significantly correlated with the interrelated properties of direction tuning width, and receptive-field subunit aspect ratio of neurons: neurons with broad direction tuning and short, wide receptive-field subunits gave their greatest response when the plaid moved in their preferred direction. Conversely, neurons with narrow direction tuning and long, narrow receptive-field subunits gave their greatest responses when the plaid moved in a direction such that one of its components moved in the preferred direction.

The authors explain such finding by referring to a previous work by (Derrington and Badcock, 1992), who had suggested that cells which are sensitive to the motion of local elements in the plaid, such as edges (i.e. already in V1), could also contribute to extract pattern-motion signals without any IOC.

Taken together, these findings point out that although MT can be accounted as a motion-specialized area, V1 contains already a fair extent of correct and detailed information on motion patterns.

In the light of such tight interplay between MT and the early stages, and such possible independency of the early stages in motion processing, it would be interesting to investigate how fundamental is one area as compared to the other at the purpose of the creation of motion representations, particularly in a neural system where no functional subdivision has been reported yet.

## 2. Using Rats in Vision Research

One important question in the study of object recognition concerns the scarce experimental manageability of its optimal test cases: the primates. To date, nonhuman primates have been the model-of-choice in the study of the mechanisms underlying object vision, since their visual system mirrors the human one. However, experiments with monkeys are typically slow and labor-intensive, generally affording only a limited number of subjects per study; needless to say, genetic, molecular, and highly invasive manipulations are often unpractical. Hence, an increasing number of visual neuroscientists has recently started to look into rodents as potential models of visual functions, implying both behavioral (Brooks et al., 2013; Douglas et al., 2006; Forwood et al., 2007; Hupfeld and Hoffmann, 2006; Minini and Jeffery, 2006; Petruno et al., 2013; Tafazoli et al., 2012; Zoccolan et al., 2009) developmental (Sia and Bourne, 2008; Sun et al., 2009) and electrophysiological/ imaging approaches (Andermann et al., 2011; Girman et al., 1999; Greenberg et al., 2008; Marshel et al., 2011; Niell and Stryker, 2008).

Although many rodent species (like rats, as we will see) display a fully functioning visual neural machinery, their ability to engage in high-level visual processing has been long underestimated. Such underestimation is based, in part, on the observation that rodents have, as we will see in details in the following sections, a poor visual acuity (approximately one cycle/degree in pigmented rats). Furthermore, it is commonly accepted that these animals

make extensive and preferential use of their whiskers (Diamond et al., 2008; von Heimendahl et al., 2007) and sense of smell (Rubin and Katz, 2001; Uchida and Mainen, 2003) when exploring their environment.

The following sections will explore the rat's visual machinery and behavioral potential, introducing this animal as a possible advantageous model for vision research.

## 2.1. The Rat's Visual System

### 2.1.1. The Rat's Eye

The rat has a rather sophisticated and effectively functioning visual system (reviewed by Sefton et al., 2004). Its laterally placed eyes provide it with a panoramic view but there is a binocular overlap, estimated to be 40–60 degrees in front of the animal. The visual field of the rat spans from -40° to more than 60° with respect to the horizontal plane of the animal if the pupil is pointing 20° in the elevation plane, and can reach more than 100° temporally (Adams and Forrester, 1968; Fig. 8A).

The rat's lens cannot perform accommodation, preventing light from being focused on a discrete point on the retina. However, with its pupil constricted the rat's eye shows a considerable depth of focus, and its lens power appears to increase with beams of light entering the eye obliquely with respect of the optic axis, displaying a point of least negative aberration (Hughes, 1977, 1979).

Being non-diurnal animals, rods are the predominant photoreceptors; however cones, though relatively rare (about 0.85% of photoreceptors), are also present and the retina is therefore capable of functioning in both scotopic and photopic conditions (Cicerone, 1976; Green and Powers, 1982; Jacobs et al., 2001; Muntz, 1967; Szél and Röhlich, 1992).

Of the two types of cones identified, the majority (93%) contain a photopigment with a peak sensitivity at about 500–520 nm (Deegan and Jacobs, 1993). The second, albeit rare, cone type (7% of cones, i.e., about 0.05% of all photoreceptors; Szél and Röhlich, 1992) contains a photopigment with a peak sensitivity at about 370 nm, in the ultraviolet range (Akula et al., 2003; Deegan and Jacobs, 1993, 1993; Szél and Röhlich, 1992).

The rat's eye is capable of optically resolving 12 min of visual angle (Hughes and Wässle, 1979). The maximal spatial resolution of dark- and light-adapted retinal ganglion cells recorded from optic tract of the hooded rat is, however, only about 1.2 cycles/degree or about 25 min of visual angle (Friedman and Green, 1982). Similarly, visual acuity estimated

on the basis of the sampling theorem from the peak density of ganglion cells is about 1.3 cycles/degree (Pettigrew et al., 1988).

Consistent with this, the upper limit of spatial resolution of single neurons recorded from the primary visual cortices of Lister hooded rats is 1.2 cycles/degree (Girman et al., 1999). Behaviorally measured visual acuity or acuity extrapolated on the basis of evoked potentials in the visual cortex of awake rats is also about 1.2 cycles per degree (Birch and Jacobs, 1979; Prusky et al., 2000a, 2002; Seymoure and Juraska, 1997). Interestingly, it has recently been shown that visual acuity of hooded rats is prone to experience-dependent plasticity during the critical period (Prusky et al., 2000b).

As we said, rat acuity can also be measured by examining the density of ganglion cells in the retina. Ganglion cells are relatively evenly distributed across the rat retina, with the variation from the highest to the lowest density being 5:1 (from 3000 to 600 cells per square millimeters). Dendritic trees and therefore the sizes of the receptive field centers of ganglion cells located in the area of highest density (area centralis) are not significantly different from their counterparts located peripherally, in the areas of lowest density (Huxlin and Goodchild, 1997; Perry and Cowey, 1979). Therefore, exploratory fixation (i.e. eye movement bringing the image of an object of interest to a region of high resolution) is not likely to occur.

However, studies on the dynamics of the rat eye movements report the presence of stereotypic, spontaneous saccades, directed primarily along the horizontal axis, with small vertical components (Chelazzi et al., 1989, 1990; Zoccolan et al., 2010).

It is worth mentioning a recent report by (Wallace et al., 2013), which has shown that rats move their eyes according to the precise head movement they need to perform while exploring the environment: when the head points downward, the eyes move back, away from the tip of the nose. When the head is lifted, the eyes look forward and the visual fields cross. If the animal puts its head on one side, the eye on the lower side moves up and the other eye moves down. This implies that line of vision varies by as much as 40 degrees in the horizontal plane and up to 60 degrees in the vertical plane. This would allow the rat to gain maximum surveillance of the environment, but at the expenses of a fine and detailed binocular vision.

### 2.1.2. The Rat's Visual Cortex: V1

In a series of electrophysiological studies, Montero and co-workers (reviewed in Montero, 1981) distinguished seven visuotopically organized areas in the occipital cortex of the rat. The largest of these areas is area 17 (V1, the striate area) distinguished by Krieg

(1946a, 1946b) on the basis of qualitative cytoarchitectonic criteria.

Primary visual cortex of rat occupies the most posterior surface of both hemispheres. Anteroposteriorly (AP) it spans an area of about 5mm (from bregma -4.4 mm to -9.36 mm, i.e. 4.4 mm and 9.36 mm posterior from bregma). Mediolaterally (ML) it reaches its maximal width around bregma -7.92 mm, where the most medial side is ~1.4 mm far from the central suture and the most lateral ~5.8 mm (Fig. 7B). At bregma -4.44 mm, only its monocular part (V1m) is visible, and its width is less than 1 mm and its location is ~4 mm far from the central suture (Fig. 7A). At bregma -9.36 mm it becomes small again and is located between ~2 mm and ~4.4 mm far from the central suture (Fig. 7C) (Paxinos and Watson, 2005).

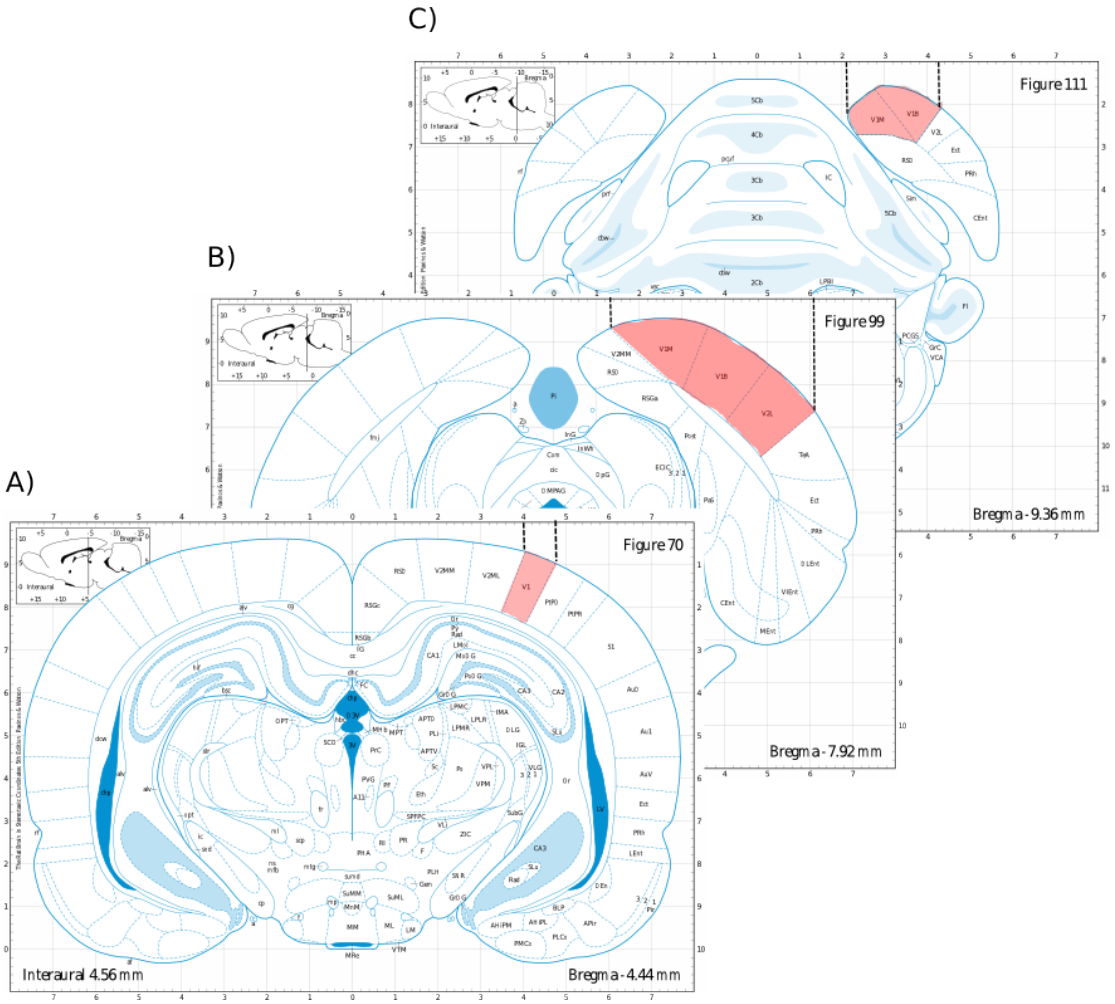


**Figure 7. The rat's primary visual cortex at three different bregmas.**

The visual cortex is highlighted by red patches, and its margins marked by black dashed lines.

**A)** At bregma -4.4 V1M is less than 1 mm wide.

**B)** At bregma -7.92, the whole V1 (binocular and monocular) has its maximum width.

**C)** At bregma -9.32, V1m and V1b cover less than 2 mm.

Adapted from Paxinos and Watson, 2005.

V1 neurons can be clustered into two different populations according to whether they integrate visual inputs coming from only one eye (monocular neurons) or both eyes (binocular neurons). These two populations of cells are organized into two topographic distinct cortical areas: one more medial (monocular V1, V1M) and one more lateral (binocular V1, V1B). The percentage of binocular cells in the rat primary visual cortex is about 80% (Sefton et al, 2004). When the pupil of the animal is pointing 65° with respect to the sagittal plane passing through the nose of the animal, the ipsilateral binocular hemifield occupies 40° starting from the vertical meridian (Fig. 8A, green angle). Panel B of Figure 8 shows the primary visual cortex of the animal. The green surface in the figure represents the population of neurons which collects the information from the binocular part of the visual field.

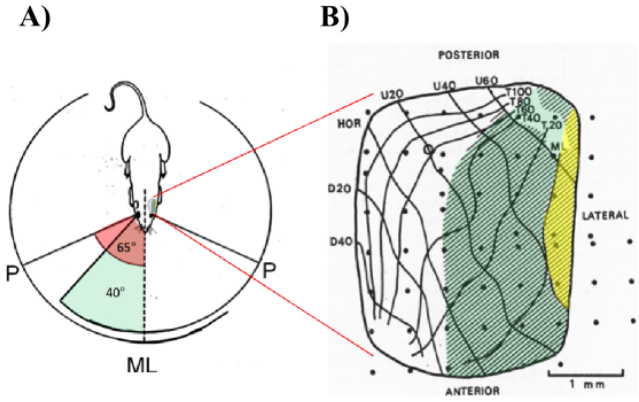


**Figure 8. Cortical projections of the rat's visual field.**

**A)** The presumed fields of view of the rat's eyes in the horizontal plane. ML is the meridian which lies in the sagittal plane of the rat. The red angle represents the angle between the sagittal plane of the animal and the fixation point of the pupil of the right eye P. The green angle represent the binocular region of the visual field of the right eye. The left primary visual cortex of the animal is shown.

**B)** The projection of the visual field on the left primary visual cortex. Abbreviations: HOR: the horizon; U20: a parallel 20° above the horizon; D20: a parallel 200°below the horizon; white circle: position of the optic disc (which is not itself represented). The binocular area, to which the left eye projects as well as the right, is green.

Adapted from Adams and Forrester, 1968.

Monocular and binocular V1 have a retinotopic organization (Fig. 9). RF size decreases at the border of the vertical meridian (lower than 3°). In the periphery, it can reach 20°. The mean RF size in V1 has been reported to be about 13° (Espinoza and Thomas, 1983; Girman

et al., 1999).

As documented in other mammals, rat V1 neurons are tuned by a set of properties of visual stimuli like spatial frequency, temporal frequency, velocity, contrast, orientation, direction and so forth (Girman et al., 1999).

The authors document that about half of the population of cells responds well to flashing uniform stationary on–off stimuli. The other half responds best to optimally oriented moving gratings of an appropriate spatial frequency. The spatial frequency tuning covers a range of 0–1.2 cycle/degree, peaking at 0.08 cycle/degree. Around 75–80% of cells in the primary visual cortex are orientation selective, although no evidence of iso-orientation columns has been reported (Ohki et al., 2005).

Almost 63% of the orientation-selective cells display an orientation tuning width at half-height of 60º or less. Furthermore, an additional 16% of cortical neurons are orientation biased. Thus almost 95% of cells in the primary visual cortex appear to exhibit some orientation selectivity.

About 35% of orientation selective or orientation biased cells show preference for bars and/or gratings oriented horizontally. Cells in layers II and III tend to exhibit little background activity, tend to be sharply orientation selective, and respond to stimuli of relatively low temporal frequencies (low velocities) and relatively high spatial frequencies. Cells in layer IV tend to exhibit high temporal resolution and respond well to relatively high stimulus velocities (over 500º/s).

Analysis of the response modulation of V1 cells in the rat evidences a bimodal distribution consistent with the simple/complex distinction: simple cells are characterized by spatially separate ON and OFF region in the RF, summation within region, antagonism between ON and OFF subregion, linear behavior and relative modulation of response to gratings major than 1.0; complex cells are characterized by superimposed ON-OFF region, highly non-linear behavior and relative modulation minor than 1.0.

Parnavelas et al. (1981) documented a similar simple/complex segregation in rat's V1.

The receptive field properties of neurons were examined in the visual cortex, area 17, of Long-Evans pigmented rats. Visually responsive cells comprised 90% of cells recorded in area 17. Cells were classified as i) simple (27% of the population), displaying small RFs, low baseline rate, preference to slowly moving stimuli, and strong orientation selectivity; ii) complex (44%), displaying stronger baseline firing, preferring fast movements and being direction selective; iii) hypercomplex (13%), which, as mentioned before, are selective for edges ending in their RF; iv) non-oriented (16%), which responded to motion regardless of

the orientation of the stimulus.

According to Girman et al. (1999), almost 60% of visual cortical cells recorded in rats exhibit either clear directional selectivity or directional bias. For many cortical neurons the magnitude of responses is reduced substantially when stimuli extend beyond the RF. In some cases, the reduction in the magnitude of responses is independent of the orientation of the stimuli. In most cases, however, the level of suppression of the responses is dependent on the orientation of the stimuli outside the receptive field. Only pyramidal cells have complex or end-stopped RFs, with complex cells being located in layers 2 to 4 and end-stopped cells in layers 2, 3, and 5. In contrast, simple cells may have either pyramidal or nonpyramidal morphology and are found in layers 2, 3, and 4. Of those cells that do not exhibit orientation selectivity, some have on- or off-discharge centers with antagonistic surrounds, whereas others have on- or off-discharge centers with silent suppressive surrounds; some do not respond to stationary flashing stimuli, although their responses to moving stimuli are quite vigorous.

### 2.1.3. The Rat's Visual Cortex: Extrastriate Areas

Rat V1 is surrounded by a number of satellite areas whose functions are almost unknown.

Electrophysiological and anatomical reports individuated a number (7±3) areas around V1 (Coogan and Burkhalter, 1993; Espinoza and Thomas, 1983; Montero, 1993; Sanderson et al., 1991). I will, in the following, adopt Espinoza and Thomas (1983) classification, who individuated 7 visual areas by means of electrophysiological assessment. Two of them are located medial to V1, anteromedial (AM) and posteromedial (PM) and the others are located laterally, lateromedial (LM), anterolateral (AL), laterointermediate (LI), laterolateral (LL).

These areas are retinotopically organized, displaying RF reversal at each border, and each of these areas encode a complete representation of the visual field (Fig. 9).

In LM, the upper VF is represented caudally and the nasal VF medially, being thus a mirror image of V1. In AL (third visual area) the upper VF is represented rostrally and the nasal VF, medially, being thus a mirror image of LM. In LI, the upper VF is medial and the nasal VF, lateral, being thus a mirror image of LM, or a reduced copy of V1. More medially, there are two representations of the temporal VF, labeled anteromedial (AM) and posteromedial (PM). In AM, the upper temporal VF is medial and the lower temporal VF is lateral, the extreme temporal field being rostral. The 30˚ azimuth provides the boundary

between AM and PM. Thus, AM is organized as a counter-clockwise rotation by 90° of the V1 representation. In PM, the upper lower VF topography is like in AM, but the extreme temporal VF is caudal, being thus a mirror image of AM.

The analysis of the cortico-cortical connections, mainly by means of tracer injections, between these satellite areas (Montero, 1993; Sanderson et al., 1991) revealed that each of them receives direct connections from the V1, and in turn, each area sends feedback connections to the striate cortex. Moreover, Coogan and Burkhalter (1993) identified asymmetrical connectivity patterns between reciprocally connected areas, a marker of hierarchical ranking (Van Essen et al., 1986), and concluded that V1 is at the first level in the hierarchy, LM ranks higher and is likely at the second level, AL ranks above LM, and AM is at a still higher level.

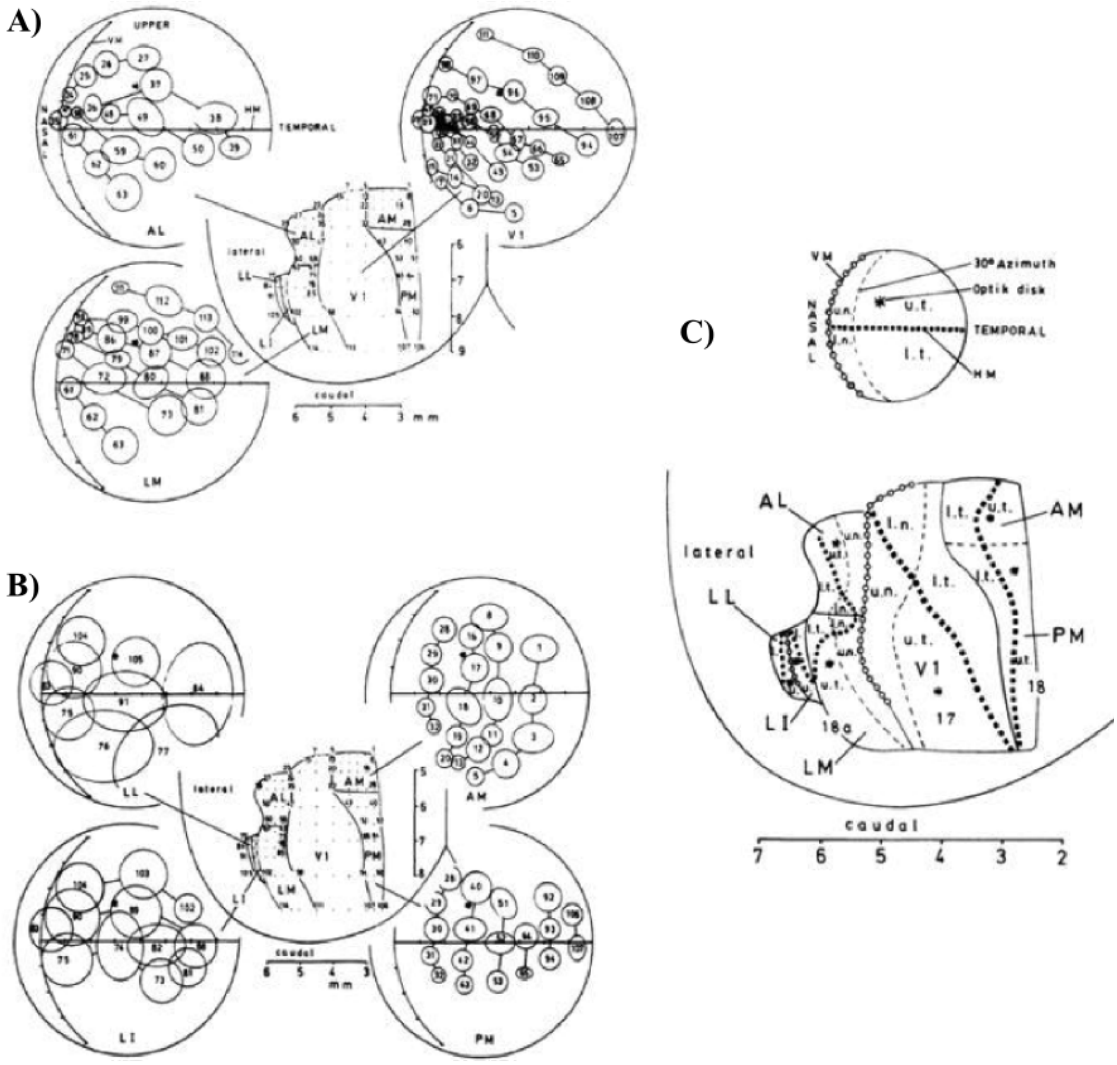


**Figure 9. Retinotopic organization in rat's striate and extrastriate areas.**

**A)** dorsal view of the left posterior cortex , illustrating the recording sites (numbered from right to

left and from top to bottom) and the locations of the areas containing representations of the right eye visual field; Bregma taken as zero for the reference axes; scale in ram. Receptive fields identically numbered and corresponding to the recording sites in V1, AL and LM are plotted in the perimeter charts. VM and indicate vertical and horizontal (equator) meridians, respectively; asterisk represents the projection of the optic disk; each scale subdivision represents 20 °.

**B)** Receptive field organization of LL, LI, AM and PM (same conventions as in (A)).

**C)** Relationship between the retinotopic map and the cytoarchitectonic subdivisions. A schematics of the retinotopic map of visual cortex and of the left cerebral hemisphere and of the right eye visual field is shown. Abbreviations: u. upper , l. lower, n. nasal, t. temporal .

Adapted from Espinoza and Thomas 1983.

## 2.2. Is There a Ventral/Dorsal Stream in Rats?

The studies on the cortico-cortical connections we just reviewed suggest the presence of a hierarchical organization in the visual cortex of the rat. Such suggestion is made stronger by the observation that the more one moves away from V1, the bigger the RF sizes get. As shown by Espinoza and Thomas (1983), the RF size gets particularly wide in extra-striate areas LM, LL, LI and AL (see Fig. 9). The increase in RF size is one hallmark of progressively higher-level neural representations, the ones found travelling upstream in both ventral and dorsal pathways in the primate and human brain. It is then reasonable to wonder: is there an IT and MT counterparts in the rat's brain?

Recent studies approached the question in mice (Andermann et al., 2011; Marshel et al., 2011; Wang et al., 2011). Interestingly, all these studies identified a putative subdivision between ventral and dorsal streams of processing in the mouse brain (reviewed by Niell, 2011; Fig.10). Importantly, although maybe not surprisingly, the rat and mouse visual system are extremely similar, in terms of anatomical structure, receptive field organization and area subdivision (and nomenclature).

Andermann et al. (2011) largely concentrated on two areas: AL, which was proposed to be the 'gateway' into the dorsal stream (Wang et al., 2011), and PM, which also receives a strong direct input from V1 and was also a candidate dorsal region, although this assignment is less clear. Similarly to Marshel et al. (2011), the authors made use of drifting sinusoidal gratings, they found a striking dichotomy between these two areas: AL was responsive to low spatial frequencies and high temporal frequencies (large features moving fast) while PM was responsive to high spatial frequencies and low temporal frequencies (fine detail moving

slowly). The first property is suggestive of optic flow, the movement of objects and landmarks across the visual field as one moves through the environment, and the authors note that the very high speeds these neurons responded to could correspond to the stimuli seen by a running mouse. The responses of the second area, PM, are more indicative of an object recognition area, except that their analysis revealed a further specialization for motion processing: as spatial frequency was varied, the preferred temporal frequency changed in a manner to keep the preferred speed constant. This form of speed tuning was relatively uncommon in V1, suggesting that it is a new feature being computed in PM, perhaps specifically for tracking objects in motion. Results on putative ventral stream areas were less conclusive.

Both Marshel and Andermann's groups studied LM, the proposed gateway to the ventral stream (Wang et al., 2011): LM-projecting V1 neurons are specialized for the processing of high spatial frequency/ high contrast information, whereas AL-projecting V1 neurons are optimally tuned to low spatial frequency/low contrast stimuli.

However, they either found it similar to V1 or more like the dorsal areas. The other putative ventral region studied by Marshel et al., LI, showed high spatial frequency preference, but no other specialization for processing shape or form. It is clear that further studies of these areas will be needed to make any definitive statement about their homology to the primate ventral areas.

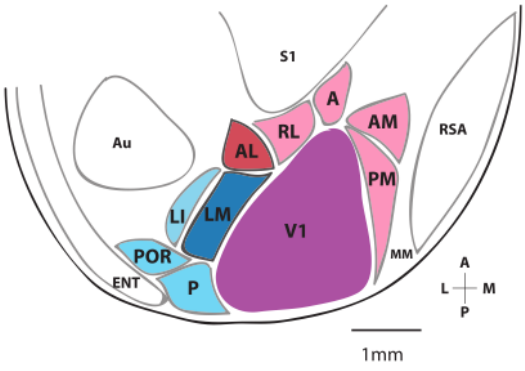


**Figure 10. Putative ventral and dorsal pathways in the mouse visual cortex.**
Map of the visual areas in mouse cortex, showing nine extrastriate areas circumscribing primary visual cortex (V1). Proposed dorsal stream ventral stream areas are shown in red and blue, respectively, with emphasis on putative gateway areas LM and AL.
Adapted from Niell, 2011.

Inconclusive evidence has been reported for a rat's dorsal stream; a couple of studies

have reported impairment of visuospatial perception after lesioning the posterior parietal cortex (Kolb and Walkey, 1987; Kolb et al., 1982). Interestingly, a developmental study (Sun et al., 2009), investigated cortical changes in response to the vision of moving stimuli in dark-reared P9 rats. When exposed to visual stimulation consisting in moving images, the animals displayed substantial conformational changes (in terms of spine density) in area V2M (corresponding to areas AM/PM), suggesting a role of dorsal areas in the processing of motion information.

On the other hand, damaging temporal and perirhinal cortex, downstream of the main outflow of LM, disrupted object recognition but spared spatial memory (Aggleton et al., 2010; Bussey et al., 1999; Davies et al., 2007; Prusky et al., 2004).

In conclusion, the presence of a ventral or dorsal stream of processing in the rat brain has not been reported yet.

Importantly, the few studies which have proposed a ventral/dorsal subdivision in the mouse or in the rat's brain do so, at best, by quantifying the categories 'shape' and 'motion' in terms of spatial and temporal frequency processing and integration. This can only give a small hint on the nature of the problem, since a primal and robust way to investigate the presence (or lack thereof) a function requires properly-designed experiments (and close-to-natural visual stimulation) for the behavioral assessment of that particular function.

## 2.3. Behavioral Studies: Object Recognition in Rats

Are rats able of solving object recognition tasks?

An interesting study, reporting that rats could distinguish amongst different images and remember them, is the one by Forwood et al. (2007). The authors found that rats could discriminate and keep in memory pictures of objects. Importantly, exploration of 3D objects and images yielded comparable performances. Also, when rats are placed in front of 3 images, 2 of which are identical, the animals display an exploratory tendency towards the odd image, demonstrating that they recognize it as different.

Recently, Meier et al. (2011) reported that rats can be easily trained to perform visual tasks that involve distracters, and that their vision is sensitive to the spatial arrangement of features: rats were trained to detect an oriented visual target grating located between two flanking stimuli ("flankers"). Flankers varied in contrast, orientation, angular position, and sign. Results show that rats are impaired at detecting visual targets with collinear flankers,

compared to configurations where flankers differ from the target in orientation or angular position. In particular, rats are more likely to miss the target when flankers are collinear. The same impairment is found even when the flanker luminance was sign-reversed relative to the target. These findings suggest that contour alignment alters visual processing in rats, despite their lack of orientation columns in the visual cortex. This was the first report that the arrangement of visual features relative to each other affects visual behavior in rats.

More recently, Brooks et al. (2013) trained eight rats on two visual categorization tasks using photographs of eight objects from each of four basic-level categories: chairs, flowers, cars, and humans. In Experiment 1, rats learned to categorize chairs versus flowers; in Experiment 2, rats learned to categorize cars versus humans. After rats learned each categorization, they were tested with eight novel pictures from each of the categories. The rats performed at reliably above-chance levels during these generalization tests. To determine which dimension(s) of the stimuli controlled the rats' behavior, the authors conducted regression analyses using several image dimensions. The chair versus flower discrimination was mainly controlled by the convexity of the stimuli, whereas the car versus human discrimination was mainly controlled by the aspect ratio of the stimuli. These results demonstrate that rats can categorize complex visual objects using shape-based properties, or features.

Are rats capable of *invariant* object recognition tasks, i.e. adopt a strategy that is both object-selective and transformation-tolerant?

A study by Minini and Jeffery (2006) reports evidence against this hypothesis: the performance of the animals at a two geometrical shapes discrimination task dropped at chance level when the stimuli underwent changes in appearance. The authors conclude that rats lack advanced shape-processing abilities and rely, instead, on low-level image cues, i.e. overall brightness, as local discriminant features. By contrast, two recent studies have shown that rats can recognize objects despite remarkable variation in their appearance (e.g., changes in size, position, lighting, in-depth and in-plane rotation), thus arguing in favor of a sophisticated recognition strategy in this species (Zoccolan et al., 2009; Tafazoli et al., 2012).

Interestingly, a lesion study conducted by Tees (1999) investigated the functional consequences of posterior temporal cortical lesions in rats (what we would call areas LI/LL) using behavioral paradigms. Lesioned animals did not recognize changes in the visual characteristics of the objects, suggesting a role for this area for visual object recognition process.

Taken together, these studies add to the growing list of visual tasks demonstrated in rats;

however, giving the both in favor and contrary evidence, it is not clear if rats can be a suitable model for the investigation of high-level functions.

## 2.4. Behavioral Studies: Motion Processing in Rats

With the purpose of finding the most suitable animal model for the study of motion processing, Douglas et al. (2006) compared the thresholds of visual motion coherence in rats and mice. The authors used a water-maze-like task to induce the animals to discriminate different levels of coherence of a moving dot pattern. They found a comparable threshold of 25% of coherence needed by the animals to perform the task with confidence. However, rats and mice were not so similar in practice: rats displayed, as compared to mice, faster learning, higher behavioral flexibility and lesser adjustments in the experimental parameters, proving to be better suited for motion studies with respect to mice.

Hupfeld and Hoffmann (2006) investigated which strain of rats had better motion discrimination performance, comparing albino Wistar and pigmented Long-Evans rats. First, it has to be known that albino rats suffer from a defective development of the retina: they show "a reduced number and a reduced peak density of rods, a reduced ipsilateral projection of ganglion cell axons, as well as physiological deficits in cortical visual centers" (Hupfeld and Hoffmann (2006)). This generates lower visual acuity, depth perception and a reduced monocular visual field. These parameters *per se* suggest that albino rats are not suited to study vision. However, the authors found that all rats could discriminate between a noise pattern stimulus (0% coherence) and a blank screen, and between a 0% coherence and a 100% coherence dot pattern. Investigating the coherence thresholds, they found that pigmented Long-Evans rats had a significantly lower coherence threshold with respect to albinos. Long-Evans, in fact, could discriminate a pattern of dots with about 12% coherence from a noise pattern; albino Wistar, on the other hand, reach their threshold at 30% coherence. A quick analysis of other species' threshold revealed that Long-Evans rats have a lower coherence threshold with respect to cats, ferrets and pigeons; only primates show even lower motion coherence thresholds. This study suggests, then, that Long-Evans rats are a good strain to investigate motion perception.

It is interesting to mention a lesion study by Petruno et al. (2013), which investigated separately the contribution of V1 and of extrastriate areas to this skill. The authors compared a direction of motion recognition task (discriminate moving dots) with an orientation

recognition task (discriminate oriented gratings) and a shape recognition task. Lesioning V1 irreversibly impaired the performance at all three tasks. Lesions in extrastriate areas (V2ML/AM and V2MM/PM), instead, did not affect the performance. This study then concluded that V1 is fundamental for both stimulus orientation, shape encoding and direction of motion processing.

All in all, it is quite obvious that rats can perform visual recognition tasks. One could question, at this point, what is the nature of the information the rat must process in order to detect motion, and how shape and motion processing interact (or interfere) to produce successful behavior according to task demands. The issue will be addressed in the following sections (see Chapter 2).

# Chapter II | Advanced Shape Processing in Rats

## 1. Abstract

As discussed in the previous sections, the ability to recognize objects despite substantial variation in their appearance is such a formidable computational feat that it is commonly assumed to be unique of primates. Such an assumption has restricted the investigation of its neuronal substrates to primate studies, which allow only a limited range of experimental approaches. In recent years, the powerful array of optical and molecular tools that has become available in rodents has spurred a renewed interest for rodent models of visual functions. However, evidence of high-level visual object processing in rodents is still limited and controversial.

In this chapter, I report the results of a study comparing what strategies underlie rat discrimination of two different pairs of visual objects. The first pair consisted of objects made of highly distinctive features, therefore affording high discriminability. The objects in the second pair were made of less distinctive features and were designed to challenge the rat recognition with less discriminable stimuli. Two different groups of rats were tested, each with one object pair. For both groups, rat visual strategy was uncovered by applying an image masking method that revealed the features used by two groups of animals to discriminate two sets of objects across a range of sizes, positions, in-depth, and in-plane rotations.

Noticeably, when the task involved the first pair of objects (Stimulus Set 1 hereinafter), rat recognition relied on a combination of 2 to 3 multiple features that were mostly preserved across the transformations the objects underwent, and largely overlapped with the features that a simulated ideal observer deemed optimal to accomplish the discrimination task. Group analysis on the features used by each rat in each condition revealed that the same features tended to be used across rats.

In the case of the second pair of objects (Stimulus Set 2 hereinafter), the features the rats used in the recognition process were smaller, more numerous and only partially preserved across objects' transformations. Group analysis on the features showed a higher inter-individual variability in the rat choice of distinctive features, suggesting the use of different

strategies across all dimensions (subject-wise, stimulus-wise and transformation-wise) to solve the recognition task.

These results indicate that, in the presence of objects with very distinctive overall structure and local features, rats are able to process and efficiently use shape information, in a way that is largely tolerant to variation in object appearance. Such strategy is overall consistent with what is known in the vision science community as 'viewpoint-invariant' object recognition. In case the objects' shapes are less easily discriminable between one another, rats still maintain a 'feature-tracking' strategy, but do so on a less systematic basis, more consistent with a so-called 'viewpoint-dependent' strategy.

# 2. Introduction

One way to study object recognition from a behavioral point of view implies understanding how the diagnostic information afforded by different visual objects, within the context of a given discrimination task, is exploited by the subjects. Various kinds of *image classification* approaches (reviewed by Murray, 2011) can be applied to uncover the set of diagnostic, or critical, visual features (i.e., parts of the objects), whose correct detection and extraction is crucial for the recognition process to succeed.

When it comes to invariant recognition, some information that is irrelevant for the task must be discarded, while, at the same time, other relevant information must be preserved and processed. This raises the question of the existence of general priors a system would adopt to decide which part of the input signal must be honored, and which one disregarded. Such priors will determine the structure of an object representation in memory. Therefore, a theory of object recognition must specify not only the nature of the visual input, but also how perceptual representations of objects are encoded, processed and stored starting from that input. Within this domain, two main theories have emerged over the past couple of decades.

One class of theories, the so-called 'viewpoint-invariant' theories, assumes that the identity of an object can be recovered under almost all viewing conditions, thanks to specific view-invariant features, or 'object-centered' cues. Such theories assume either a complete three-dimensional description of an object (Ullman, 1989), or a structural description of the image that specifies the relationships among viewpoint-invariant volumetric primitives (Biederman, 1987; Marr, 1982).

In contrast, 'viewpoint-dependent' theories argue that no such general invariants exist

and that object features are represented much as they appeared when originally viewed, thereby relying on 'viewer-centered' representations. Such theories assume that the features visible in the input image are compared to the features of the object representations stored in memory, either by normalizing the input image to approximately the same viewing position as represented in visual memory (Edelman and Bülthoff, 1992; Tarr, 1995) or by computing a statistical estimate of the quality of match between the input image and the candidate representations (Perrett and Oram, 1998; Riesenhuber and Poggio, 1999).

The comparison between viewpoint-invariant and viewpoint-dependent approaches is important to our purposes, because each approach makes different predictions about how invariance is achieved (reviewed by Tarr and Vuong, in Pashler, 2002).

Viewpoint-invariant theories propose that recognition is itself invariant across transformations: as long as the critical features remain accessible, the response of the system remains constant. On the other hand, viewpoint-dependent theories state that recognition depends on the viewing parameters, which may produce changes in recognition performance, since objects are represented according to how they appeared when learned.

Humans display viewpoint-invariant recognition of familiar objects, but a viewpoint-dependent performance in recognition tasks with novel objects (Edelman and Bülthoff, 1992). However, viewpoint-dependent representations are often considered computationally implausible, since they would require a vast amount of memory to store all possible object views needed to achieve viewpoint invariance.

Nonetheless, it has been proposed that recognition performance ranges from viewpoint-invariant to viewpoint-dependent according to how easy it is to discriminate between the objects of reference (Hayward and Williams, 2000; Newell, 1998; Vuong and Tarr, 2006).

Also, as observed in Nielsen et al. (2008), under certain conditions, the effects of viewpoint on performance disappear, and even novel objects are recognized in a view-invariant manner: this seems to happen when objects can be identified by a number of very distinctive features, which remain 'diagnostic' of the stimulus identity despite changes in object rotation (Lawson, 1999; Tarr et al., 1997; Wilson and Farah, 2003).

The question of what object features are selected to recognize an object, and whether the same feature are 'tracked' as preferential markers of the object's identity has been addressed with the use of an algorithm for the masking of the visual stimulus, namely the Bubbles Method (Gosselin and Schyns, 2001).

The Bubbles Method involves the repeated presentation of stimuli occluded by an opaque mask punctured by Gaussian transparent windows, in such a way that only discrete

stimulus parts remain visible.

This method has been used with humans (Gosselin and Schyns, 2001; Nielsen et al., 2006, 2008), monkeys (Nielsen et al., 2006, 2008), pigeons (Gibson et al., 2005) and rats (Vermaercke and Op de Beeck, 2012).

In a couple of studies, the Bubbles Method was used to directly determine whether observers showing view-invariant performance use the same feature to identify an object, irrespective of its orientation. Monkeys usually show viewpoint-dependent behavior for rotated objects (Logothetis et al., 1994); however, it has been shown that training with multiple views can turn a viewpoint-dependent behavior into a viewpoint-invariant one (Wang et al., 2005). Nonetheless, although both humans and monkeys adopted view-invariant performance with objects rotated in the picture plane, they displayed different discrimination strategies: in the case of having to discriminate natural image sets, while humans tend to cover bigger portion of the image, monkeys select a few, confined image patches for correct identification (Nielsen et al., 2006). Moreover, in the case of having to discriminate artificially generated silhouettes of objects, humans tend to use the same features independent of shape orientation, monkeys use unique features for each orientation (Nielsen et al., 2008). Humans are able to generalize to a greater degree across orientation changes than rhesus monkey observers, who tend to re-learn separate problems at each object orientation, rather than flexibly apply previously learned knowledge to novel problems.

Therefore, the behavior of the human subjects is consistent with the viewpoint-invariant models, since subjects always used the same feature to identify an object, while monkeys' strategies can be better described by the viewpoint-dependent model, as it seems that the monkeys use different templates to identify the same shape at different rotation angles.

In order to assess whether an observer adopts a viewpoint-invariant *vs* viewpoint-dependent behavior, a critical aspect involves the intelligent choice of the visual stimuli. An object, in order to be correctly recognized, must contain sufficient structural, three-dimensional information to specify its structure. The use of objects such as two-dimensional geometrical shapes (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012) or real objects templates or silhouettes (Nielsen et al., 2008) can be misleading, since comparisons between similarity or distinctiveness of each object's features can be performed only on low-level criteria, based on luminance, contrast or contour information, hence speaking little, if at all, to the mechanisms of object recognition as it is accomplished in natural conditions.

In this chapter, I will describe a study aimed at addressing the issue of whether an object recognition strategy is more consistent with a view-invariant or a view-dependent processing

of the visual input, using rats as experimental subjects. This raises the question of the suitability of rats as models of high-level visual processing. Recently, a number of recent works have explicitly addressed the issue, reaching opposite conclusions.

Minini and Jeffery (2006) concluded that rats lack advanced shape processing and invariant recognition abilities and rely, instead, on low-level image cues to discriminate shapes (e.g., luminance in the lower half of the stimulus display). By contrast, two recent studies (Tafazoli et al., 2012; Zoccolan et al., 2009) have shown that rats can recognize objects despite remarkable variation in their appearance (e.g., changes in size, position, pose and lighting), thus arguing in favor of a sophisticated visual recognition strategy in this species. However, studies based on pure assessment of recognition performance cannot reveal the complexity of rat recognition strategy, i.e., they cannot tell: 1) whether shape features are truly extracted from the input image; 2) what these features are and how many; 3) what role they play in determining rat behavioral responses; and 4) whether they remain stable across the object transformations the animals face. In spite of a recent attempt at addressing these issues by Vermaercke and Op de Beeck (2012), who used a version of the same image classification technique (the Bubbles) we have applied in our study (described below), these questions remain largely unanswered. In fact, the authors' conclusion that rats are capable of using a flexible mid-level recognition strategy is affected by various limitations of their experimental design: the choice of the target shapes (a square and a triangle, which, again, lacks the structural complexity of three-dimensional objects made of multiple parts/features); the limited shape transformations tested (only position changes); and, more critically, the choice of the visual task (a two-alternative forced-choice procedure, with both target shapes equally shifted and simultaneously presented to the animals, which, de facto, prevented a true assessment of shape-based, position-invariant recognition).

In our work we applied the Bubbles Method on two sets of 3-dimensional rendered objects tested across a range of sizes, positions, in-depth rotations and in-plane rotations. The Bubbles allowed the identification of the 'salient' and 'anti-salient' image features underlying rat recognition strategy of the target objects (i.e., the features leading, respectively, to correct identification or misidentification of the target objects). Our results show that, when objects are easily identifiable, because of unique and distinctive (object-specific) structural properties, rats rely on largely viewpoint-invariant, multi-featural recognition strategy. On the other hand, when presented with structurally similar objects, which are harder to discriminate, rats rely on a viewpoint-dependent recognition strategy, with specific visual features used to recognize each view of an object.

# 3. Materials and Methods

## 3.1. Subjects

Twelve adult male Long Evans rats (Charles River Laboratories), split into two groups of six rats each, were used for behavioral testing. Animals were 8 weeks old at their arrival, weighted approximately 250 g at the onset of training and grew to over 600 g. Rats had free access to food but were water-deprived throughout the experiments, i.e. they were given with 1 hour of water *pro die* after each experimental session, and received an amount of 4-8 ml of pear juice as reward during the training. All animal procedures were conducted in accordance with the National Institutes of Health, International, and Institutional Standards for the Care and Use of Animals in Research and after consulting with a veterinarian.

## 3.2. Experimental Rig

The training apparatus consisted of six operant boxes. Each box hosted one rat, so that each group of six rats could be trained simultaneously, every day, for up to two hours. Each box was equipped with: 1) a 21.5" LCD monitor (Samsung 2243SN) for presentation of visual stimuli, with a mean luminance of 43 cd/mm$^2$ and an approximately linear luminance response curve; 2) an array of three stainless steel feeding needles (Cadence Science) ~10 mm apart from each other, connected to three capacitive touch sensors (Phidgets 1110) for initiation of behavioral trials and collection of responses; and 3) two computer-controlled syringe pumps (New Era Pump Systems NE-500), connected to the left-side and right-side feeding needles, for automatic liquid reward delivery.

A 4 cm diameter viewing hole in the front wall of each box allowed each tested animal to extend its head out of the box, so to frontally face the monitor (placed at ~30 cm in front of the rat's eyes) and interact with the sensors array (located at 3 cm from the opening). The location and size of the hole was such that the animal had to reproducibly place its head in the same position with respect to the monitor to trigger stimulus presentation. As a result, head position was remarkably reproducible across behavioral trials and very stable during stimulus presentation. Video recordings obtained for one example rat showed that the standard deviation of head position, measured at the onset of stimulus presentation across 50 consecutive trials, was ±3.6 mm and ±2.3 mm along the dimensions that were, respectively,

parallel (*x* axis) and orthogonal (*y* axis) to the stimulus display (with the former corresponding to a jitter of stimulus position on the display of ±0.69º of visual angle). For the same example rat, the average variation over 500 ms of stimulus exposure was $\Delta x$ = 2.5±0.5 mm and $\Delta y$= 1.0±0.2 mm (n = 50 trials), with the former corresponding to a jitter of stimulus position on the display of ~0.48º of visual angle. Therefore, the stability of rat head during stimulus presentation was close to what achieved in head-fixed animals. This guaranteed a very precise control over stimulus retinal size and prevented head movements from substantially altering stimulus retinal position (see Results and Discussion for further comments about stability of stimulus retinal position).

## 3.3. Visual Stimuli

Two groups of rats, group 1 and group 2 in the following, were trained on different stimulus sets, Stimulus Set 1 (containing Object 1 and 2) and Stimulus Set 2 (containing Object 3 and 4). For both sets, the objects were renderings of three-dimensional models that were built using the ray tracer POV-Ray (http://www.povray.org/). All objects were illuminated from the same light source location and, when rendered at the same in-depth rotation, their views were approximately equal in height, width and area (see Fig. 11A). Objects were rendered in a white, bright (see below for a quantification), opaque hue against a black background. Each object's default size was 35° of visual angle, and their default position was the center of the monitor (their default view was the one shown in Fig. 11A). As explained below, during the course of the experiment, transformed views of the objects were also shown to the animals (i.e., scaled, shifted, in-plane and in-depth rotated object views; see Figs. 11C-D).

Visual similarity or discriminability between the two objects in each given Stimulus Set (matching views) was measured in terms of pixel-wise Euclidean Distance (see Results).

## 3.4. Experimental Design

### 3.4.1. Phase I: Critical Features Underlying Recognition of the Default Object Views

Each group of rats was initially trained to discriminate the two default views of the

target objects (Fig. 11A). Animals initiated each behavioral trial by inserting their heads through the viewing hole in the front wall of the training box and licking the central sensor. This prompted presentation of one of the target objects on the monitor placed in front of the box. Rats learned to associate each object identity with a specific reward port (see Fig. 11B).

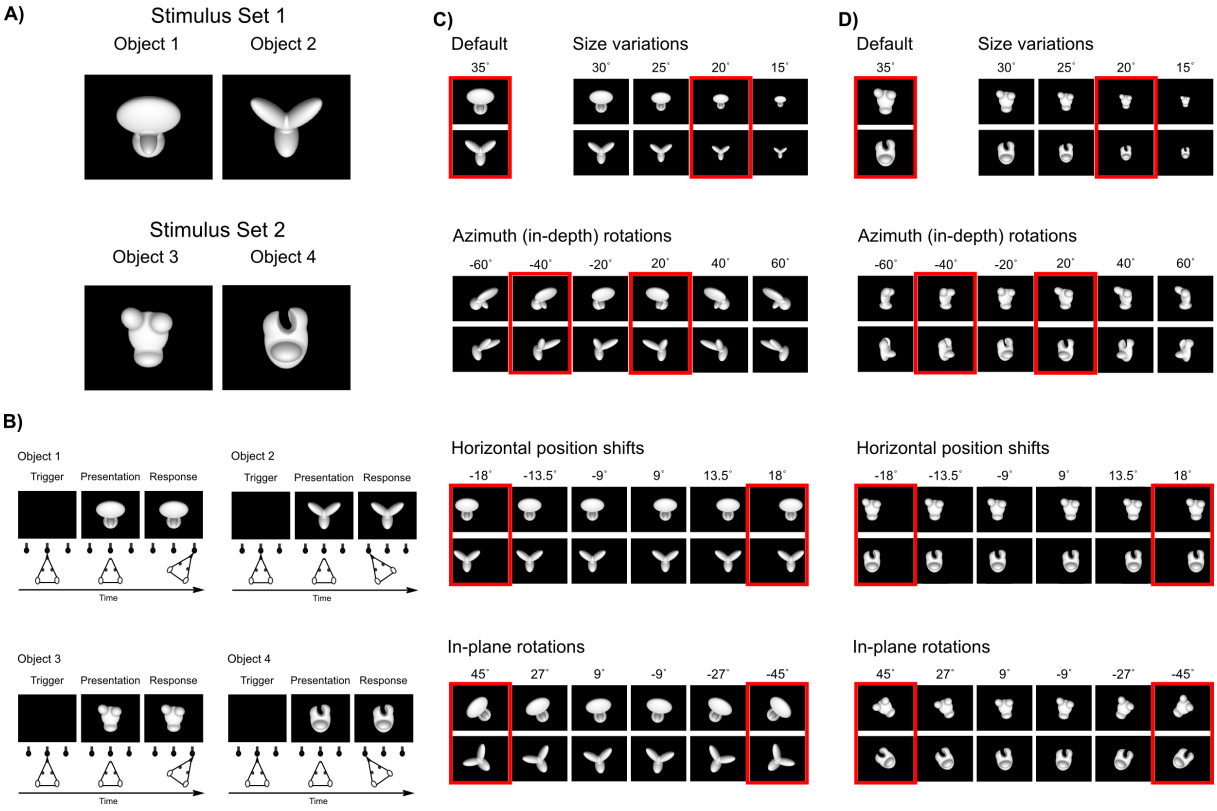


**Figure 11. Visual stimuli and behavioral task.**

**A)** Default views (0º in-depth and in-plane rotation) of the four objects that rats were trained to discriminate during Phase I of the study (each object default size was 35º of visual angle). Upper row: Stimulus Set 1; lower row, Stimulus Set 2.

**B)** Schematic of the object discrimination task. Rats were trained in an operant box that was equipped with an LCD monitor for stimulus presentation and an array of three sensors. The animals learned to trigger stimulus presentation by licking the central sensor, and to associate each object identity to a specific reward port/sensor (right port for Object 1/3 and left port for Object 2/4).

**C)** and **D)** Some of the transformed views of the two objects that rats were required to recognize during Phase II of the study. Left columns, Stimulus Set 1; right column, Stimulus Set 2. Transformations included: 1) size changes; 2) azimuth in-depth rotations; 3) horizontal position shifts; and 4) in-plane rotations. Azimuth rotated and horizontally shifted objects were also scaled down to a size of 30º of visual angle; in-plane rotated objects were scaled down to a size of 32.5º of visual angle and shifted downward of 3.5º. Note that each variation axis was sampled more densely than shown in the figure – sizes were sampled in 2.5º steps; azimuth rotations in 5º steps; position shifts in 4.5º steps; and in-plane rotations in 9º steps. The red frames highlight the subsets

of object views that were tested in bubbles trials.

In case of correct response, reward was delivered through the port and a reinforcement tone was played. An incorrect choice yielded no reward and a 1-3 s time out (during which a failure tone sounded and the monitor flickered from black to middle gray at a rate of 15 Hz). The default stimulus presentation time (in the event that the animal made no response after initiating a trial) was 2.5 s. However, if the animal responded correctly before the 2.5 s period expired, the stimulus was kept on the monitor for an additional 4 s from the time of the response (i.e., during the time the animal collected his reward). In the event of an incorrect response, the stimulus was removed immediately and the time-out sequence started. If the animal did not make any response during the default presentation time of 2.5 s, it still had 1 s, after the offset of the stimulus presentation and before the end of the trial, to make a response.

The rats were trained daily, and allowed us collect a number of 300 to 400 trials per session. Once a rat achieved ≥70% correct discrimination of the default object views (set as criterion performance, which typically required 3-12 weeks of training), an image masking method, known as the Bubbles (Gosselin and Schyns, 2001), was applied to identify the visual features the animal relied upon to successfully accomplish the task. This method consists in superimposing on a visual stimulus an opaque mask punctured by a number of circular, semi-transparent windows (or *bubbles*; Fig. 12A). When one of such masks is applied to a visual stimulus, only those parts of the stimulus that are revealed through the bubbles are visible. Hence, this method allows isolating the image patches that determine the behavioral outcome, for whether a subject (e.g., a rat) can identify the stimulus depends on whether the uncovered portions of the image are critical for the accomplishment of the recognition task.

In our implementation of the Bubbles method, any given bubble was defined by shaping the transparency (or alpha) channel profile of the image according to a circularly symmetrical, two-dimensional Gaussian (with the peak of the Gaussian corresponding to full transparency). Multiple such Gaussian bubbles were randomly located over the image plane. Overlapping of two or more Gaussians produced summation of the corresponding transparency profiles, up to the maximal level corresponding to full transparency. The size of the bubbles (i.e., the standard deviation of the Gaussian-shaped transparency profiles) was fixed to 2º of visual angle, while the number of bubbles was chosen so to bring each rat's performance to be ~10% lower than in unmasked trials. In the case of Stimulus Set 1, this typically brought the performance down from ~70-80% correct obtained in unmasked trials to 60-70% correct obtained in bubbles masked trials; for Stimulus Set 2, the performance was typically lower in

Bubbles trials, around 55-60% (see performance figures). Given the lower discriminability of the objects in Stimulus Set 2, this comes as no surprise; also, it must be pointed out that only 3 out of 6 rats in group 2 reached a suitable performance criterion, yielding data that was considered robust enough to be included in all analyses performed (again, because of the lower discriminability of the objects they had to distinguish).

The number of bubbles was randomly chosen, in each trial, among a fixed set of values that was set specific for each rat. These values ranged between 10 and 50 (in steps of 20) for top (i.e., displaying a stable 75-80% performance) performing rats, and between 50 and 90 (again, in steps of 20) for average (i.e., displaying a stable 70% performance) performing rats. Examples of objects occluded by masks with a different number of bubbles are shown in Fig. 12B. The latter range represented the typical range used for rats in group 2 (tested with Stimulus Set 2).

Trials in which the default object views were shown unmasked (named 'Regular Trials' in the following) were randomly interleaved to trials in which they were masked (named 'Bubbles Trials' in the following). The fraction of bubbles trials presented to a rat in any given daily session varied between 0.4 and 0.75. In order to obtain enough statistical power to extract the critical features underlying rat recognition, at least 3,000 bubbles trials for each object at each condition were collected over the course of 16.3 ± 4.4 sessions for group 1 (rat group average ± SD, $n = 6$), 20.0 ± 4.6 for group 2 ($n = 3$).

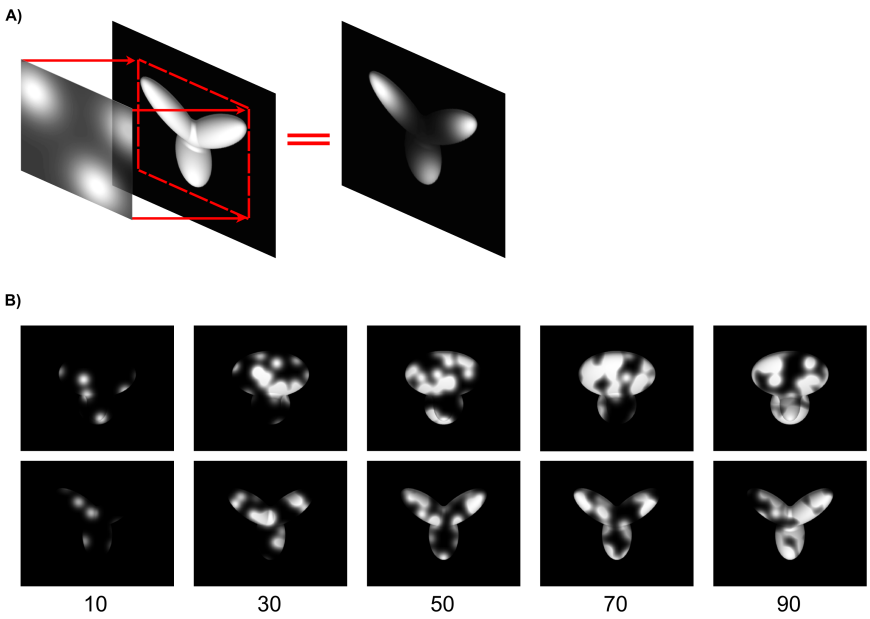


**Figure 12. The Bubbles Method.**

**A)** Illustration of the Bubbles method, which consists in generating an opaque mask (fully black

area) punctured by a number of randomly located transparent windows (i.e., the *bubbles*; shown as white, circular clouds) and then overlapping the mask to the image of a visual object, so that only parts of the object remain visible.

**B)** Examples, from Stimulus Set 1, of the different degrees of occlusion of the default object views that were produced by varying the number of bubbles in the masks.

### 3.4.2. Phase II: Critical Features Underlying Recognition of the Transformed Object Views

After having collected sufficient data to infer the critical features used by rats to discriminate the default object views, the animals were trained to tolerate variation in the appearance of the target objects along a variety of transformation axes. The goal of this training was to engage those high-level visual processing mechanisms that, at least in primates, allow preserving the selectivity for visual objects in the face of identity-preserving object transformations (DiCarlo et al., 2012; Li et al., 2009; Rust and Dicarlo, 2010; Zoccolan et al., 2005, 2007). Four different transformations were tested (see Fig. 11C), for each stimulus set, in the following order: 1) *size* variations, ranging from 35° to 15° visual angle; 2) *azimuth* rotations (i.e., in-depth rotations about the objects' vertical axis), ranging from -60° to 60°; 3) horizontal *position* changes, ranging from -18º to +18º visual angle; and 4) *in-plane* rotations, ranging from -45° to +45°.

Size transformations were the first to be applied, using an adaptive staircase procedure that, based on the animal performance, updated the lower bound of the range from which the object size was sampled (the upper bound was fixed to the default value of 35° visual angle). Once the sizes' lower bound reached a stable (asymptotic) value across consecutive training sessions (i.e., 15º of visual angle), a specific size (i.e., 20º of visual angle; see red frame in the top row of Fig. 11C-D) was chosen so that: 1) its value was different (lower) enough from the default one; and 2) most rats achieved about 65-70% correct recognition for that value. Rats were then presented with randomly interleaved regular trials (in which unmasked objects could be shown across the full 15º-35º size range) and bubbles trials (in which bubble masks were superimposed to the 20º scaled objects).

This same procedure was repeated for each of the other tested object transformations, which were included sequentially in the experimental design. For instance, after having trained size variations and having applied the Bubbles method to the 20º size-scaled objects, a staircase procedure was used to train the rats to tolerate the azimuth rotations. After reaching asymptotic azimuth values (i.e., ±60º), two azimuth rotations were chosen (using the same

criteria outlined above) for application of the Bubbles method: -40º and +20º (see red frames in the second row of Fig. 11C-D). Again, regular trials (in which unmasked objects could be shown across the full 15º-35º size range and the full -60º/+60º azimuth range) were then presented interleaved with bubbles trials (in which bubble masks were superimposed to either the -40º or the +20º azimuth rotated objects).

After the azimuth rotations, position changes were trained (with bubble masks applied to objects that were horizontally translated of ±18º of visual angle; see red frames in the third row of Fig. 11C-D) and then in-plane rotations (with bubble masks applied to objects that were rotated of ±45º; see red frames in the forth row of Fig. 11C-D). Note that, as explained above, while a new transformation was introduced, the full range of variation of the previously trained transformations was still shown to the animal, with the result that the task became increasingly demanding in terms of tolerance to variation in object appearance.

The staircase training along each transformation axis typically progressed very rapidly. On average, rats reached, for group 1: 1) the asymptotic size value in 1.2 ± 0.4 sessions (mean ± SD; $n$ = 6); 2) the asymptotic azimuth rotation values in 5.5 ± 1.0 sessions ($n$ = 6); 3) the asymptotic position values in 2.8 ± 1.9 sessions ($n$ = 5); and 4) the asymptotic in-plane rotation values in 2.0 ± 0.0 sessions ($n$ = 2).

Rats in group 2 required, on average, a slightly higher number of sessions: 1) asymptotic size value in 6.3 ± 2.0 sessions ($n$ = 3); 2) asymptotic azimuth rotation values in 4.3 ± 0.6 sessions ($n$ = 3); 3) the asymptotic position values in 5.0 ± 1.0 sessions ($n$ = 3); 4) asymptotic in-plane rotation values in 6.3 ± 4.0 sessions ($n$ = 3).

As for the default object views, transformed views also required a minimum of 3,000 bubbles trials to be collected for each of the transformed views that was tested with the Bubbles method.

In general, for each rat, bubbles trials could be collected only for a fraction of the seven transformed views we planned to test (see red frames in Fig. 11C-D).

This was because the overall duration of Experimental Phase II varied substantially among rats, depending on: 1) how many trials each animal performed per session (this number roughly varied between 250 and 500); 2) what fraction of trials were bubble trials (this number, which ranged between 0.4 to 0.75, had to be adjusted in a rat-dependent way, so to avoid the performance in bubbles trials to drop below ~10% less of the performance in regular trials); and 3) the longevity of each animal (some rats fell ill during the course of the experiment and had to be euthanized before being able to complete the whole experimental

phase).

All experimental protocols (from presentation of visual stimuli to collection of behavioral responses) were implemented using the freeware, open-source software package MWorks (http://mworks-project.org/). An ad-hoc plugin was developed in C++ to allow MWorks building bubbles masks and presenting them superimposed on the images of the visual objects.

## 3.5. Data Analysis

### 3.5.1. Computation of the saliency maps

The critical visual features underlying rat recognition of a given object view were extracted by properly sorting all the correct and incorrect bubbles trials obtained for that view. More specifically, saliency maps were obtained that measured the correlation between the transparency values of each pixel in the bubbles masks and the behavioral responses. That is, saliency map values $c^i$ for each pixel *i* were defined as:

$$c^i = \frac{\mathbf{x}^i \cdot \mathbf{B}}{\|\mathbf{x}^i\|_{L1}} \quad (1)$$

where $\mathbf{x}^i$ is a vector containing the transparency values of pixel *i* across all collected bubble trials for a given object view; **B** is a binary vector coding the behavioral outcomes on such trials (i.e., a vector with elements equal to either 1 or 0, depending on whether the object view was correctly identified or not); and $\|\mathbf{x}^i\|_{L1}$ is the $L_1$ norm of $x^i$, i.e.:

$$\|\mathbf{x}^i\|_{L1} = \sum_{n=1}^{N} |\mathbf{x}_n^i| \quad (2)$$

where *N* is the total number of collected bubbles trials. In the following, saliency maps are shown as grayscale masks superimposed to the images of the corresponding object views, with bright/dark pixels indicating regions that are salient/anti-salient, i.e., likely/unlikely to lead to correct identification of an object view when visible through the bubbles masks. For the sake of providing a clearer visualization, the saliency values in each map are normalized by subtracting their minimum value and dividing by their maximum value, so that all saliency values are bounded between zero and one.

To show which pixels, in the image of a given object view, had a statistically significant

correlation with the behavior, the following permutation test was performed. All the bubble trials that had been collected for that object view were divided in subsets, according to the number of bubbles that were used in a trial (e.g., 10, 30 or 50 for a top performing rat; see previous section). Within each subset of trials with the same number of bubbles, the behavioral outcomes (i.e., the elements of vector **B**) were randomly shuffled. Chance saliency map values $c^i$ were then computed according to eq. 1, but using the shuffled vector **B**. Among all the chance saliency values, only those corresponding to pixels within the image of the object view were considered (i.e., those corresponding to background pixels were discarded). This yielded an average of 28,605 chance saliency values per object view. This procedure was repeated 10 times, so to obtain a null distribution of saliency values for each object view.

Based on this null distribution, a one-tailed statistical test was carried out to find what values, in each saliency map, were significantly higher than what obtained by chance ($p < 0.05$), and, therefore, what pixels, in the image, could be considered as significantly salient. Similarly, significant anti-salient pixels were found by looking for corresponding saliency values that were significantly lower than what expected by chance ($p < 0.05$). Significantly *salient regions* of an object view will be shown in red, whereas *anti-salient regions* will be shown in cyan (e.g., see Figs. 13 and 14).

Group average saliency maps and significant salient and anti-salient regions were obtained, for Stimulus Set 1, using the same approach described above, but after pooling the bubble trials obtained for a given object view across all available rats (see Fig. 26B).

### 3.5.2. Ideal observer analysis

Rats' average saliency maps for Stimulus Set 1 were compared to the saliency maps obtained by simulating a linear ideal observer (Gibson et al., 2005; Gosselin and Schyns, 2001; Vermaercke and Op de Beeck, 2012). Given a bubble-masked input image, the simulated observer classified it as being either Object 1 or 2, based on which of the eight views of each object, to which the mask could have been applied (shown by red frames in Fig. 11C), matched more closely (i.e., was more similar to) the input image. In other words, the simulated ideal observer performed a template matching operation between each bubble-masked input image and the 16 templates (i.e., eight views for each object) it had stored in memory. The ideal observer was linear in that the template matching operation consisted in computing a normalized dot product between each input image and each template. For better consistency with the experiment, we chose as input images the bubbles trials presented to the

rat that could be tested with all the eight object views (i.e., rat 3; see Fig. 16B). Also, to better match the actual retinal input to the rats, each input image was low pass-filtered so that its spatial frequency content did not exceed 1 cycles per degree, i.e., the maximal retinal resolution of Long-Evans rats (Keller et al., 2000; Prusky et al., 2002).

Finally, to lower the performance of the ideal observer and bring it close to the performance of the rats, Gaussian noise (std = 0.5 of the image grayscale) was independently added to each pixel of the input images. This assured that potential differences between rats' and ideal observer's saliency maps were not merely due to performance differences. Crucially, this constraint did not force the recognition strategy of the ideal observer to be similar to the one used by rats (the ideal observer had no knowledge of how rats responded to the bubble-masked input images). This was empirically assessed by running the ideal observer analysis with different levels of noise added to the input images, and verifying that the resulting saliency maps did not substantially change as a function of noise level (i.e., as a function of the ideal observer's performance). Saliency maps and significant salient and anti-salient regions for the ideal observer were obtained as described above for the rats.

Each rat group average saliency map was compared to the corresponding map obtained for the ideal observer by computing their Pearson correlation coefficient. The significance of the correlation was assessed by running a permutation test, in which the behavioral outcomes of the bubble trials were randomly shuffled 100 times for both the average rat and the ideal observer, yielding 100 pairs of random rat-ideal saliency maps. Computation of the Pearson correlation coefficient between each pair of random maps yielded a null distribution of 100 correlation values, against which the statistical test was carried out with $p = 0.05$.

All data analyses were performed in MATLAB (http://www.mathworks.com).

## 4. Results

The goal of this study was to understand the visual processing strategy underlying rat ability to recognize visual objects in spite of substantial variation in their appearance (e.g., see Tafazoli et al., 2012; Zoccolan et al., 2009), and to assess to what extent rat recognition strategy was viewpoint-dependent or viewpoint-invariant, depending on the visual/structural similarity of the objects the animals had to discriminate. To this aim, 2 groups of rats were trained in an object recognition task that required them to discriminate two visual objects under a variety of viewing conditions. One group of rats was trained with structurally/visually

dissimilar objects (i.e., Object 1 and 2 shown in Fig. 11A, top row and Fig. 11C), while another group of rats was trained with a pair of objects that were structurally/visually more dissimilar (i.e., Object 3 and 4 shown in Fig. 11A, bottom row, and Fig. 11D). An image masking method, known as the Bubbles (Gosselin and Schyns, 2001) was then applied to a subset of the trained object views to infer what object features rats relied upon to successfully recognize these views. This approach not only revealed the complexity of rat recognition strategy, but also allowed tracking if and how such a strategy varied across the different viewing conditions the animals were exposed to.

## 4.1. Critical Features Underlying Recognition of the Default Object Views

During the initial experimental phase, two groups of Long-Evans rats (6 rats per group) were trained to discriminate the default views (or appearances) of two pairs of visual objects (shown in Fig. 11A): Object 1 vs. Object 2 (group 1) and Object 3 vs. Object 4 (group2). The animals were trained for 3-12 weeks until they achieved ≥70% correct discrimination performance. In group 1, all six animals reached this criterion performance. In group 2, only half of the animals reached the criterion and were able to maintain it in the subsequent tests of invariant recognition. Once the criterion was reached, regular trials (i.e., trials in which the default object views were shown unoccluded) started to be randomly interleaved with bubble trials (i.e., trials in which the default object views were partially occluded by opaque masks punctured by a number of circular, randomly located, semi-transparent windows; see Material and Methods and Fig. 12). The rationale behind the application of the Bubbles masks was to make it harder for the rats to correctly identify an object, by revealing only parts of it. Obviously, the effectiveness of a Bubbles mask at impairing recognition of an object depended on the position of the semi-transparent bubbles (thus revealing what object features a rat relied upon to successfully recognize the object), but also on their size and number. Following previous applications of the Bubbles method (Gibson et al., 2005; Gosselin and Schyns, 2001), in our experiments the bubbles' size was kept fixed (i.e., set to 2º of visual angle), while their number was adjusted so to bring each rat's performance in bubble trials to be, in the best cases, ~10% lower than in regular trials. For Stimulus Set 1, when the default object views were tested, rat average recognition performance dropped from ~75% correct in regular trials to ~65% correct in bubble trials (Fig. 13A). For Stimulus Set 2, rat average recognition on default views dropped from ~70% performance in regular trials to ~55%

correct in bubbles trials (Fig. 14A). This indicates that, as expected because of our stimulus design, objects in Stimulus Set 1 were easier to discriminate than object in Stimulus set 2 (especially when occluded by the bubble masks).

The critical visual features underlying rat recognition of the default object views were extracted by computing saliency maps that measured the correlation between bubbles masks' transparency values and rat behavioral responses. For each rat, the resulting saliency maps are shown as grayscale masks superimposed on the images of the corresponding object views (with the brightness of each pixel indicating the likelihood, for an object view, to be correctly identified when that pixel was visible through the bubbles masks). Whether a saliency map value was significantly higher or lower than expected by chance was assessed through a permutation test at $p < 0.05$ (see Materials and Methods). This led to the identification of significantly *salient* and *anti-salient regions* in the images of the default object views (shown, respectively, as red and cyan patches). These regions correspond to those objects' parts that, when visible through the masks, likely led, respectively, to correct identification and misidentification of the object views.

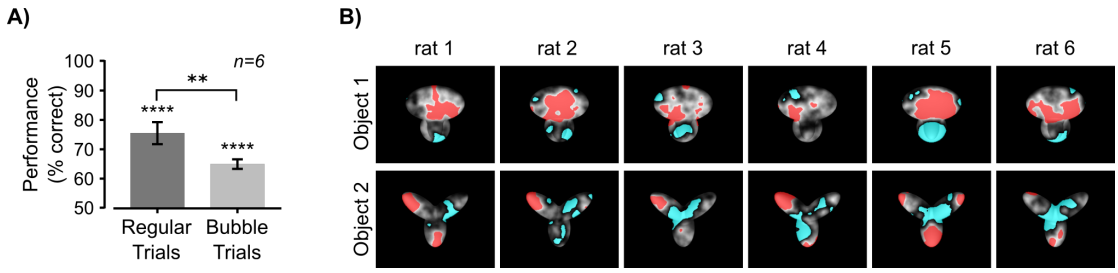


**Figure 13. Critical features underlying recognition of the default views of Stimulus Set 1.**

**A)** Group 1 average performance at discriminating the default object views was significantly lower in bubbles trials (light gray bar) than in regular trials (dark gray bar; $p < 0.001$; one-tailed, paired t-test). Both performances were significantly higher than chance (****$p < 0.0001$; one-tailed, unpaired t-test). Error bars are SEM.

**B)** For each rat, the saliency maps resulting from processing the bubbles trials collected for the default object views are shown as grayscale masks superimposed on the images of the objects. The brightness of each pixel indicates the likelihood, for an object view, to be correctly identified when that pixel was visible through the bubbles masks. Significantly salient and anti-salient object regions (i.e., regions that were significantly positively or negatively correlated with correct identification of an object; $p < 0.05$; permutation test) are shown, respectively, in red and cyan.

Visual inspection of the patterns of salient and anti-salient regions obtained for the two object pairs in the default condition revealed several key aspects of rat object recognition

strategy.

In Stimulus Set 1 (Fig. 13), in the case of Object 1, salient and anti-salient regions were systematically located within the same structural parts of the object (Fig. 13B, top row). Namely, for all rats, salient regions were contained within the larger, top lobe, while anti-salient-regions lay within the smaller, bottom lobes. Therefore, in spite of some variability in the size and compactness of the salient and anti-salient regions (e.g., compare the large, single salient region found for rats 2 and 5 with the smaller, scattered salient patches observed for rats 3 and 4.), the perceptual strategy underlying recognition of Object 1 was highly preserved across subjects.

In contrast, a substantial inter-subject variability was observed in the saliency patterns obtained for Object 2 (Fig. 13B, bottom row). Although the central part of the object (at the intersection of the three lobes) tended to be consistently anti-salient across rats, and the salient regions were always located within the peripheral part (the tip) of one or more lobes, the combination and the number of salient lobes varied considerably from rat to rat. For instance, rats 2 and 3 relied on a single lobe (the upper-left one), while rats 1, 4 and 6 relied on the combination of the upper-left and bottom lobes, and rat 5 relied on all three lobes. Moreover, some lobes (e.g., the bottom one) could be salient for some animal, but fully (rat 2) or partially (rat 4) anti-salient for some other.

The larger inter-subject diversity in the pattern of salient and anti-salient features that was found for Object 2, as compared to Object 1, is not surprising, given the different structural complexity of the two objects. In fact, Object 2 is made of three fully visible, clearly distinct and roughly equally sized lobes, while the three lobes of Object 1 are highly varied in size, with the smaller, bottom lobes that are partially overlapping and, therefore, harder to distinguish. As a consequence, Object 2 affords a larger number of distinct structural parts, as compared to Object 1, hence a larger number of “perceptual alternatives” to be used for its correct identification. As such, the saliency patterns obtained for Object 2 are more revealing of the complexity and diversity of rat recognition strategies.

As for Stimulus Set 2, the pattern of salient and anti-salient features displayed, in general, a higher variability between subjects (Fig. 14).

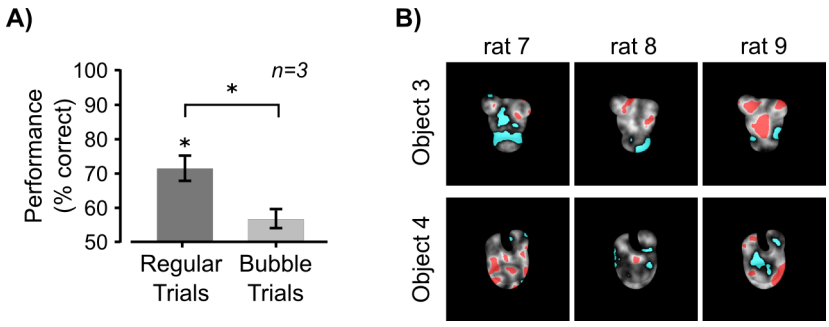


**Figure 14. Critical features underlying recognition of the default views of Stimulus Set 2.**

**A)** Group 2 average performance at default object views is shown for Regular trials and Bubbles trials (see Fig.13 for details).

**B)** Saliency maps for the three rats tested with Stimulus Set 2 (see Fig. 13 for details).

The upper lobes of Object 3 tended to be systematically selected as salient features by all three rats, whether it was both (rat 7 and 9) or only one (rat 8). For rat 8, one of the features, the rightmost one, did not cover the upper right lobe, rather was located slightly below it, reaching the central part of the object, which is broadly significant for rat 9. Rat 7 displayed, instead, an anti-salient region covering the central part of Object 3, while, for the other two rats, anti-salient regions were located along the lower/right margin. The central/top part of Object 4 was salient for two rats, in the guise of seven small, scattered patches for rat 7, and one single spot for rat 8. On the other hand, for rat 9, this same central/top part was anti-salient, along with the upper-right lobe. Similarly, the upper-right lobe was anti-salient for rat 7, and somewhat for rat 8 as well. The latter showed additional spots of anti-saliency towards the right and the left margins of the objects.

Hence, when it comes to Stimulus Set 2, where the two objects are structurally much more similar than in the case of Stimulus Set 1, a substantial inter-subject diversity was observed not only in terms of the number of salient features found in one object (as in the case of Object 2, in Stimulus Set 1), but also in terms of location, number and size of the salient and anti-salient regions. This is indicative of a larger variety of perceptual strategies used by the rats tested with Stimulus Set 2, as compared to Stimulus Set 1 (see next sections for a quantitative analysis).

## 4.2. Critical Features Underlying Recognition of the Transformed Object Views

After the critical features underlying recognition of the default object views were uncovered, each rat was further trained to recognize the target objects in spite of substantial variation in their appearance. Namely, objects were transformed along four different variation axes: size, in-depth azimuth rotation, horizontal position and in-plane rotation (the trained ranges of variation are shown in Fig. 11C-D). These transformations were introduced sequentially (i.e., size variation was trained first, followed by azimuth, then by position and finally by in-plane variation) and each of them was trained gradually, using a staircase procedure (see Materials and Methods). Once the animals reached a stable, asymptotic value along a given transformation axis, one or two pairs of transformed object views along that axis were chosen for further testing with the bubbles masks (these pairs are marked by red frames in Fig. 11C-D). Such views were chosen so to be different enough from the objects' default views, yet still recognized with a performance larger than chance by most animals. Rats were then presented with randomly interleaved bubbles trials (in which these transformed views were shown with superimposed bubble masks) and regular trials (in which unmasked objects were randomly sampled across all the variation axes tested up to that point). A total of seven different pairs of transformed object views were chosen for testing with bubbles masks, although, due to across-rat variation in longevity and fluency in the invariant recognition task (see Materials and Methods for details), not all animals undergo testing at all conditions.

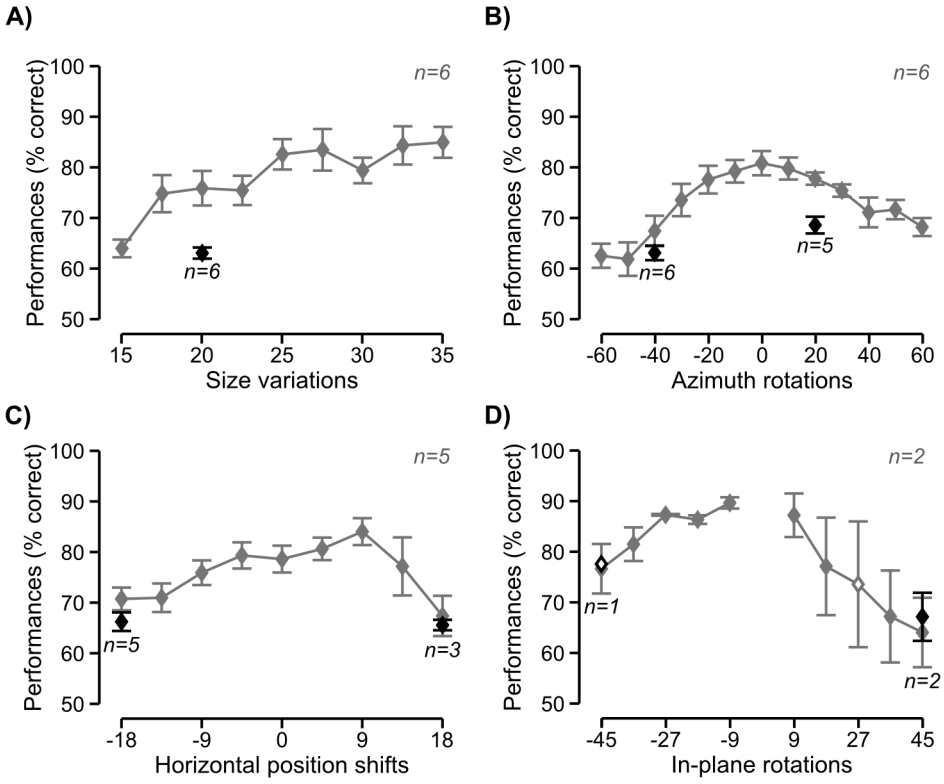


**Figure 15. Group 1 average recognition performance over the four variation axes along which the objects were transformed.**

Gray and black symbols show performances in, respectively, regular and bubbles trials that were collected over the course of the same testing sessions of Experimental Phase II (see Materials and Methods). Solid and open symbols indicate performances that were, respectively, significantly and non-significantly higher than chance ($p < 0.0001$ in A, B and C; $p < 0.05$ in D; one-tailed, unpaired t-test). Error bars are SEM.

For group 1, rat average recognition performance was significantly higher than chance for almost all tested object transformations, typically ranging from ~70% to ≥80% correct, and dropping below 70% correct only for some extreme transformations (Fig. 15, gray lines). This confirmed that rat recognition is remarkably robust against variation in object appearance, as recently reported by two studies (Tafazoli et al., 2012; Zoccolan et al., 2009).

As previously observed in the case of the default views (see Fig. 13A), rat performance at recognizing the transformed object views was generally 5-10% lower in bubbles trials than in regular trials (see black diamonds in Fig. 15).

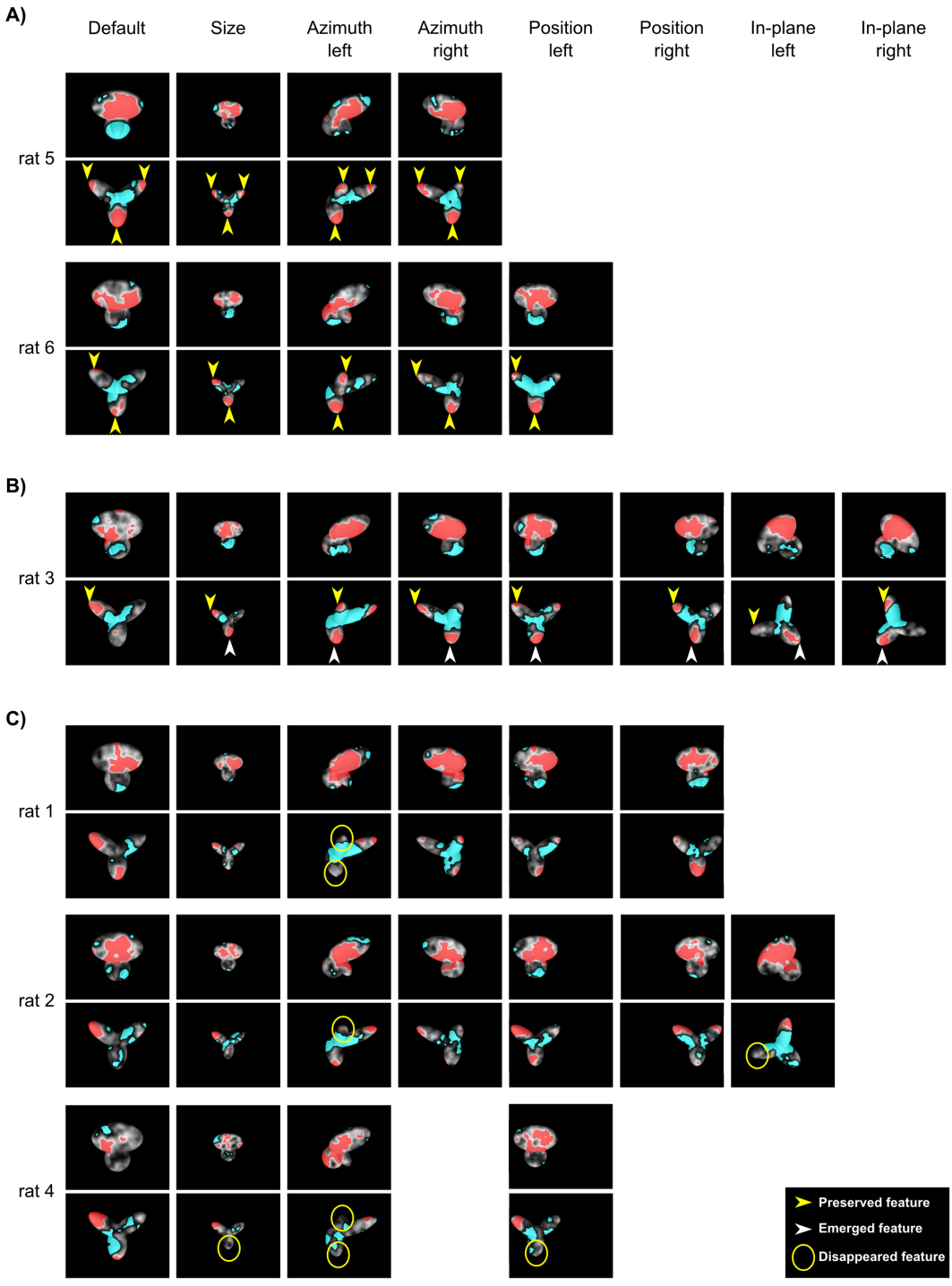


**Figure 16. Critical features underlying recognition of the transformed object views for Stimulus Set 1.**

For each rat, the saliency maps (with highlighted significantly salient and anti-salient regions; same color code as in Fig. 13B-14B) that were obtained for each transformed object view are shown. Maps obtained for different rats are grouped in different panels according to their stability across the tested views. A) For rats 5 and 6, the same pattern of salient features (i.e., lobes' tips) underlay recognition of all the views of Object 2 (see yellow arrows). B) For rat 3, one salient feature (i.e., the tip of the upper-left lobe) was preserved across all tested views of Object 2 (see

yellow arrows), while a second feature (i.e., the tip of the bottom lobe) became salient after the animal started facing variation in object appearance (see white arrows). C) For rats 1, 2 and 4, features that underlay recognition of Object 2's default view became no longer salient for some of the transformed views (see yellow circles) and were replaced by other salient features.

The critical features underlying rat recognition of a transformed view were extracted by properly processing all the correct and incorrect bubbles trials obtained for that view (see previous section and Materials and Methods). This yielded saliency maps with highlighted significantly salient and anti-salient regions that revealed if and how each animal recognition strategy varied across the different viewing conditions he was exposed to. The saliency maps for group 1 of rats are shown in Fig. 16.

As previously reported for the default object views (see Fig. 13), a larger inter-subject variability was observed in the patterns of critical features obtained for Object 2, as compared to Object 1. Namely, while for most rats a single and compact salient region was consistently found in the larger, top lobe of Object 1, regardless of the transformation the object underwent (see Figs. 16A-C, odd rows), in the case of Object 2 not only different rats relied on different combinations of salient lobes, but, for some rats, such combinations varied across the transformed object views (see Figs. 16A-C, even rows). Therefore, the saliency patterns obtained for Object 2 were more revealing of the diversity and stability of rat recognition strategies in the face of variation in object appearance.

For some rats, all the lobes used to discriminate the default view of Object 2 remained salient across the whole set of transformations the object underwent (see yellow arrows in Fig. 16A). This was particularly striking in the case of rat 5, which consistently relied on all three lobes of Object 2 as salient features across all tested transformations. Rat 6 showed a similarly consistent recognition strategy, although he relied only on two salient lobes (the upper-left and bottom ones). Also in the case of rat 3, the single salient lobe that was used for recognition of the default object view (the upper-left one) remained salient for all the subsequently tested transformed views (see yellow arrows in Fig. 16B). In this case, however, the bottom lobe, which only contained a pointlike hint of a salient patch in the default view, emerged as a prominent salient feature when the animal had to face size variations, and remained consistently salient for all the ensuing transformations (see white arrows in Fig. 16B). In still other cases, lobes that were originally used by a rat to discriminate Object 2's default view became no longer salient for some of the transformed views (see yellow circles in Fig. 16C) and were replaced by other salient lobes.

In the case of Stimulus Set 2, generally lower performances were observed, compared to Stimulus Set 2, in both regular and bubble trials (compare Figs. 15 and 17).

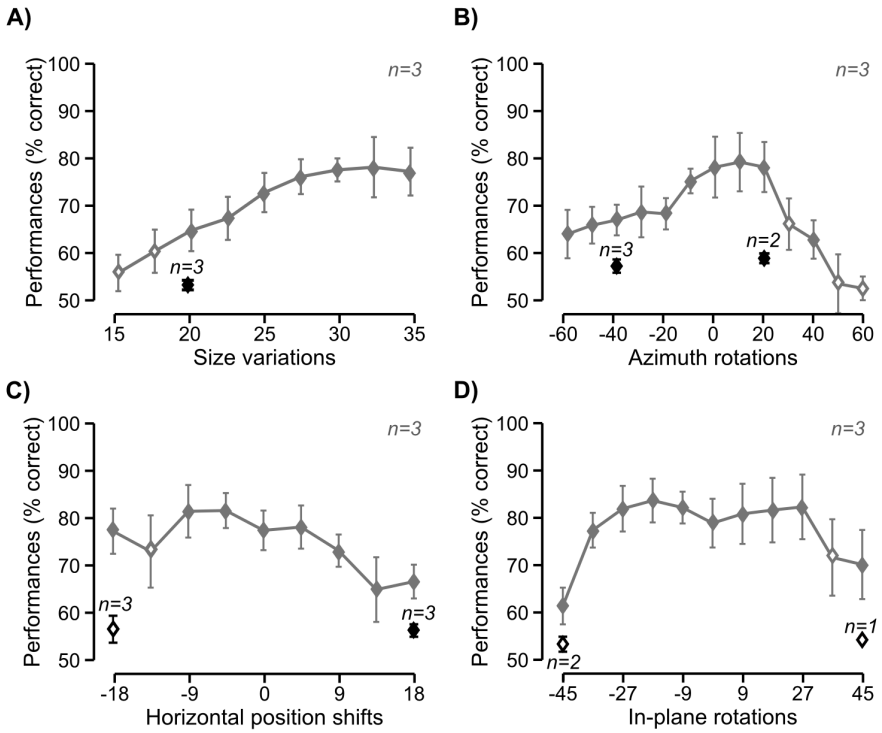


**Figure 17. Group 2 average recognition performance over the four variation axes along which the objects were transformed.**

See Figure 15 for details.

Still, even for Stimulus Set 2, rat average recognition performance was significantly higher than chance at almost all tested object transformations, mostly ranging from ~70% to ≥80% correct, although it decayed more sharply for small sizes (see Fig. 17A), compared to what observed for Stimulus Set 1 (see Fig. 15A). Moreover, rat performance at recognizing the bubbles-masked object views dropped much more dramatically than in the case of Stimulus Set 1 (compare black diamonds in Figs. 15 and 17), ranging between 55 and 60% correct, significantly below the performance on regular trials, although significantly above chance at 4 out of 7 conditions tested. Overall, this pattern of performance confirms that the object pair in Stimulus Set 2 was harder to discriminate than the object pair in Stimulus Set 1, especially when shape information was degraded by reducing the size of the objects or adding the bubbles masks.

In Fig. 18 the salient and anti-salient features for each object transformation of Stimulus

Set 2 are reported.

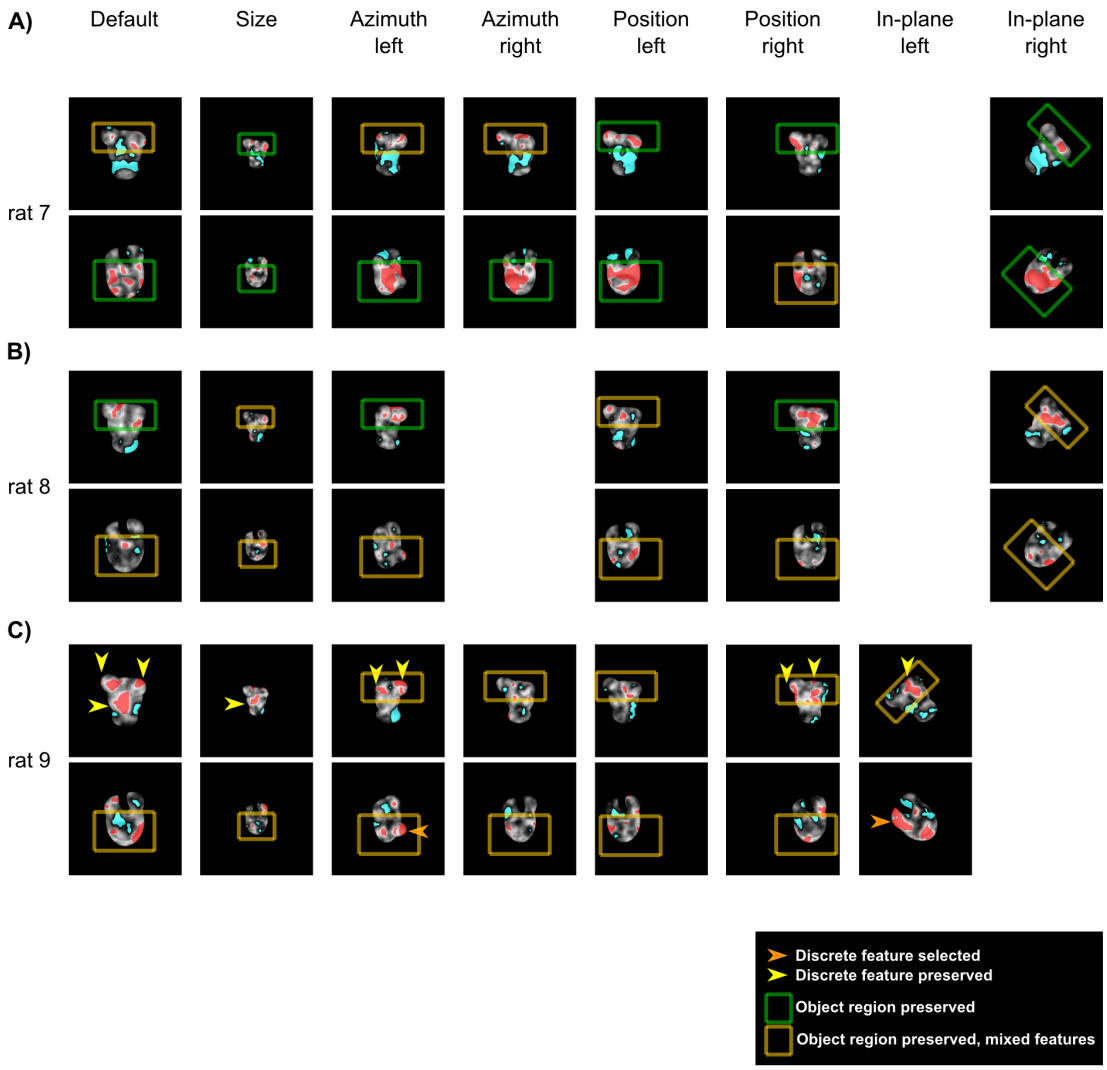


**Figure 18. Critical features underlying recognition of the transformed object views for Stimulus Set 2.**

For each rat, the saliency maps (with highlighted significantly salient and anti-salient regions; same color code as in Fig. 13B-14B) that were obtained for each transformed object view are shown. The green and ocher squares in A), B) and C) refer to object regions which were systematically selected across objects' view. The orange and yellow arrows in C) mark the selection, or tracking, of discrete objects' features, respectively.

Both Object 3 and 4 in Stimulus Set 2, just like Object 1 and 2 in Stimulus Set 1, are made of 'lobes'. However, such lobes are less protruded, hence less distinctive of the stimulus identity (quantitative cross-comparison between the two object pairs, in terms of analysis of size and number of salient features, is reported in the following sections). Consistent with this observation, we found a general tendency, for the diagnostic features of Object 3 and 4,

towards being distributed (often in a quite scattered way) over a region of the objects (i.e., top or bottom half) encompassing multiple lobes, rather than being precisely (and reproducibly) located in specific lobes (as found, for instance, in the case of Object 2; see Fig. 16).

For rat 7 (Fig. 18, upper rows), the salient features were located in the upper region of Object 3 throughout all conditions (green squares), although, in the case of the default view, were mixed with anti-salient spots (ocher squares), which only remained, as smaller points, in the azimuth-rotated views. The anti-salient regions covered preferentially the central and lower parts. A somewhat reversed pattern was observed for Object 4: the central/bottom part was largely salient across all tested views, starting with a combination of small patches in the default view, which reduced to a couple of small spots in the size-transformed condition, and finally merged into a big salient region for most of the remaining transformations.

Similarly, rat 8 (Fig. 18, central rows) displayed a preference for the upper region of Object 3 in most of the tested cases, with tiny patches of saliency appearing in two conditions (size transformed and horizontally shifted views). The anti-salient features generally covered the lower lobe, but extended to the central part of the object in three conditions (size transformed and horizontally shifted views) and to the upper-right lobe in one condition (horizontally shifted to the left). The salient features merged for the last condition tested (horizontally shifted to the right). It was again the central part of Object 4 its most salient region, but the salient patches remained small, few and scattered, and always mixed with anti-salient spots.

Rat 9 (Fig. 18, lower rows) displayed, in the beginning, i.e., for the default views, a more feature-selective preference for the upper lobes of Object 3, rather than the whole upper region (yellow arrows). The central region of Object 3 was selected as salient too, and remained salient for the following condition tested (size-transformed view). The upper lobes remained salient as well, but only as very small spots, to become prominent again in the azimuth left view. Then the lobe-located features scattered for the next tested views, to appear again in the last two conditions. In the case of Object 4, a highly variegate combination of salient features (and mixed with anti-salient spots) was found across the tested views, although discrete lobes were selected in two conditions (azimuth left, the central lobe or ‘nose’; in-plane left, the left lobe; orange arrows).

## 4.3. A Qualitative Comparison of Rat Recognition Strategies

As shown in Fig. 18, when facing objects that are hard to discriminate (as in the case of Stimulus Set 2), rats rely on a recognition strategy that is only partially preserved across the objects' transformations. Such a "preservation" appears certainly to be the case for rat 7, whose selection of the salient features of the Objects generally involved the upper lobes for Object 3, and the central part of Object 4, although the size and number of features largely varied across conditions (a reverse, similarly preserved strategy was observed for the pattern of anti-salient features). A similar trend was observed also for rat 8, although with smaller and more scattered salient and anti-salient regions. In the case of rat 9, some lobe-located salient features were 'tracked' across some views of Object 3. In general, however, all the rats tested with Stimulus Set 2, did not show a transformation-invariant, clear preference for well-defined structural parts of the objects (e.g., lobes).

On the contrary, half of the rats tested with Stimulus Set 1 showed a remarkably stable recognition strategy (Figs 16A, B) in the face of variation in object appearance, with the same combination of salient object parts (i.e., lobes) being relied upon across all (Fig. 16A) or most (Fig. 16B) object views. The other half of the rats showed a more variable recognition strategy, based on view-specific salient features' patterns (Fig. 16C), although less variable than found in the case of Stimulus Set 2.

Crucially, regardless of its stability across transformations, rat recognition strategy relied on a combination of at least two different salient features for most tested views of Object 2 (i.e., in 26 out of 34 cases). Since these features are located in structurally distinct parts of the object (i.e., distinct lobes), and, in all cases, in both its lower and upper half, this strongly suggests that rats are able to process global shape information and extract multiple structural features that are diagnostic of object identity.

The selection of both upper and lower object parts holds, to some extend, also for Stimulus Set 2, mainly for Object 4 (11 out of 20 cases), and less for Object 3 (5 out of 20 cases).

Such a shape-based, multi-featural processing strategy not only rules out a previously proposed low-level account of rat visual recognition in terms of luminance detection in the lower half of the stimulus display (Minini and Jeffery, 2006), but also suggests that rats are able to integrate shape information over much larger portions of visual objects (virtually, over a whole object) than reported by a recent study (Vermaercke and Op de Beeck, 2012).

However, having assessed that rats are able to process global shape information does not

imply, *per se*, that they are also capable of an advanced transformation-tolerant (or invariant) recognition strategy. In fact, having excluded one very low-level account of rat object vision (Minini and Jeffery, 2006) does not automatically rule out that some other low-level recognition strategies may be at work when rats have to cope with variation in object appearance. For instance, rats could rely on detection of some object feature(s) that is (are) largely preserved (in terms of position, size and orientation) across the tested object transformations. This would result in higher-than-chance recognition of the transformed object views, without the need, for rats, to form and rely upon higher-level, transformation-tolerant object features' representations. This is not a remote possibility, since recent computational work has shown that even large databases of pictures of natural objects (commonly used by vision and computer vision scientists to probe invariant recognition) often do not contain enough variation in each object appearance to require engagement of higher-level, truly invariant recognition mechanisms (Pinto et al., 2008).

Our study was designed to specifically tackle this issue, thus going beyond previous, purely performance-based accounts of rat invariant recognition abilities (Tafazoli et al., 2012; Zoccolan et al., 2009).

In fact, the existence of transformation-preserved features that are diagnostic of object identity would result in a large, systematic overlap between the significantly salient regions obtained for different object views. The fact that, in the case of Stimulus Set 1, the salient features underlying recognition of a given object "tracked" the object's transformations (i.e., changed in position, size, and orientation, as the object translated, shrunk and rotated in the animals' visual field) makes the existence of such a systematic overlap unlikely (see Fig. 16). This was quantitatively confirmed by the overlap analysis described in the following section.

## 4.4. Object-Centered or Image-Centered Strategy?

To further assess whether rat recognition strategy was more consistent with a high-level, transformation-tolerant (object-centered) representation of diagnostic features or, rather, with low-level detection of some transformation-preserved image patches (i.e., an image-centered strategy), and as a way to inspect if any diagnostic object feature existed that was consistently preserved across the tested object transformations, the saliency maps obtained for different pairs of object views were superimposed and the overlap between pairs of significantly salient regions assessed.

The overlap between the salient features obtained for all possible pairs of object views produced by affine transformations (i.e., all tested object views with the exclusion of in-depth azimuth rotations) was measured. The overlap was computed for both: 1) *raw* salient features' patterns, in which the image planes containing the salient features of the views to compare were simply superimposed (see second row of Figs. 19A, left plot); and 2) *aligned* salient features' patterns, in which the transformations that produced the two object views were "undone" (or reversed), so to perfectly align one view on top of the other (e.g., in the case of the comparison between the default and the horizontally translated views shown in Fig. 19A, the latter was shifted back to the center of the screen and scaled back to 35º, so to perfectly overlap with the default view; see second row of Fig. 19A, right plot). The overlap was quantified as the ratio between overlapping area and overall area of the significantly salient regions of the two object views (Nielsen et al., 2006) (e.g., as the ratio between the orange area and the sum of the red, yellow and orange areas in Fig. 19A, second row).

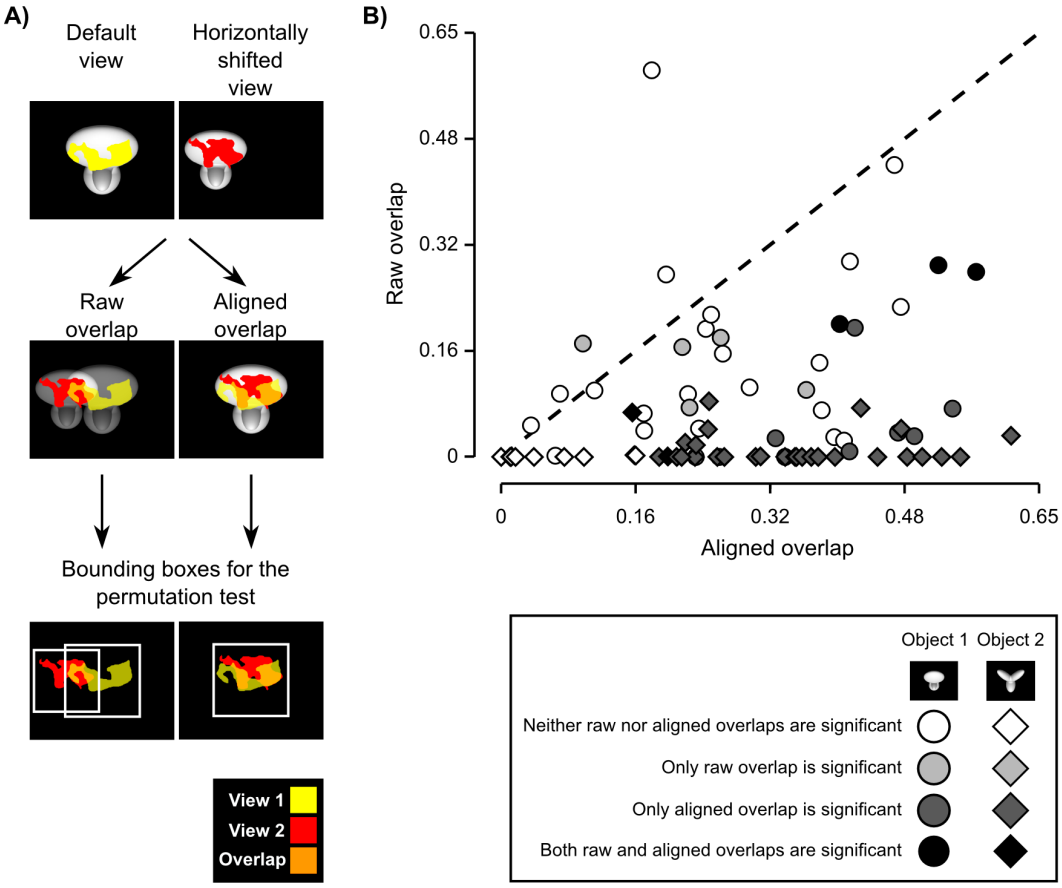


**Figure 19. *Raw* vs. *aligned* features' overlap for all pairs of object views of Stimulus Set 1.**
A) Illustration of the procedure to compute the *raw* and *aligned* overlap between the salient features' patterns obtained for two different views of an object. The default and the leftward

horizontally shifted views of Object 1 are used as examples (first row). To compute the raw features' overlap, these two object views (and the corresponding features' patterns) were simply superimposed (second row, left plot), as previously done in Figure 7. To compute the aligned features' overlap, the transformation that produced the leftward horizontally shifted view was reversed. That is, the object was shifted to the right of 18º and scaled back to 35º, so to perfectly overlap with the default view of the object itself (second row, right plot). In both cases, the overlap was computed as the ratio between the orange area and the sum of the red, yellow and orange areas. The significance of the overlap was assessed by randomly shifting the salient regions of each object view within the minimum bounding box enclosing each view (see Results for details). Such bounding boxes are shown as white frames in the third row of the figure, for both the raw and aligned views. B) The raw features' overlap is plotted against the aligned features' overlap for each pair of views of Object 1 (circles) and Object 2 (diamonds) resulting from affine transformations (i.e., position/size changes and in-plane rotations). The shades of gray indicate whether the raw or/and the aligned overlap values were significantly larger than expected by chance ($p < 0.05$; see caption).

The resulting pairs of raw and aligned overlap values obtained for all tested combinations of object views of Stimulus Set 1 are shown in Figure 19B (circles and diamonds refer, respectively, to pairs of views of Object 1 and 2). Similarly to what has been done by (Nielsen et al., 2006), the significance of each individual raw and aligned overlap was assessed through a permutation test, in which the salient regions of each object view in a pair were randomly shifted within the minimum bounding box enclosing each view. As illustrated by the example shown in Figure 19A, in the case of the raw overlap, the bounding boxes enclosing the two views partially overlapped (compare the white frames in the third row of Fig. 19A, left plot), while, in the case of the aligned overlap, by construction, the bounding boxes enclosing the two views were coincident (see the single white frame in the third row of Fig. 19A, right plot). Null distributions of raw and aligned overlap values were obtained by running 1,000 permutation loops, and the significance of the measured raw and aligned overlaps was assessed at $p = 0.05$ (significance is coded by the shades of gray filling the symbols in Fig. 19B; see caption).

In the case of Stimulus Set 1, for most pairs of object views (71 out of 76), the overlap between salient features was higher in the aligned than in the raw case (Fig. 19B). Namely, the average overlap between aligned views was 0.30 ± 0.01 (mean ± SEM), while the average overlap between raw views was 0.07 ± 0.01, with the former being significantly higher than the latter ($p < 0.0001$; significance was assessed through a paired permutation test, in which

the sign of the difference between aligned and raw overlap for each pair of views was randomly assigned in 10,000 permutation loops).

The larger overlap found for aligned vs. raw views was particularly striking in the case of Object 2, with most raw overlap values being zero, and the average overlap being one order of magnitude larger for aligned than raw views (i.e., 0.30 ± 0.02 vs. 0.01 ± 0.00; such a difference was statistically significant at $p < 0.0001$, according to the paired permutation test). Moreover, in the large majority of cases (30/38), the aligned overlap was significantly higher than expected by chance (see black and dark gray diamonds in Fig. 19B), while the raw overlap was significantly higher than chance only for a few pairs of object views (2/38; see black and light grey diamonds in Fig. 19B). This confirmed that the transformations Object 2 underwent were large enough to displace its diagnostic features in non-overlapping regions of the stimulus display (hence, the zero or close-to-zero salient features' overlap observed for the raw views), thus preventing rats from relying on any transformation-preserved feature to succeed in the invariant recognition task. At the same time, the large and significant salient features' overlap found for the aligned views of Object 2 indicates that the same structural parts were deemed salient for most of the object's views the rats had to face, thus suggesting that rats truly had to rely on some transformation-tolerant representation of these diagnostic structural features.

In the case of Object 1, in agreement with the examples shown in Figure 7, the overlap between raw views was considerably larger, as compared to what obtained for Object 2 (compare circles and diamonds in Fig. 19B). However, in most cases, the overlap between aligned views was higher than the corresponding overlap between raw views (i.e., 33 out of the 38 circles are in the lower quadrant in Fig. 19B) and, in several cases, the raw overlap was zero or close-to-zero. As a result, the average overlap was significantly larger for aligned than raw views (i.e., 0.30 ± 0.02 vs. 0.13 ± 0.02; $p < 0.0001$, paired permutation test). This suggests that, although a salient feature existed that was partially preserved across many tested views, rat strategy was nevertheless more consistent with "tracking" that feature (i.e., its position, size, orientation) across the transformations Object 1 underwent, rather than merely relying on the portion of that feature that remained unchanged across such transformations. This observation, together with the fact that the same feature was relied upon also when shifted in non-overlapping locations of the stimulus display (as in the case of the horizontally translated views), indicates that also recognition of Object 1 was more consistent with a high-level, transformation-tolerant representation of diagnostic features, rather than

with low-level detection of some transformation-preserved luminance patch. Finally, the fraction of overlap values that were significantly higher than expected by chance was similarly small for both the raw (8/38) and the aligned (9/38) views of Object 1 (see light gray, dark gray and black circles in Fig. 19B). This reflects the fact that relatively large overlaps were produced by chance in the permutation test (given the large area occupied by Object 1's salient regions), thus making the threshold to reach significance higher than in the case of Object 2. This confirms that the saliency regions/maps obtained for Object 2 were, in general, more powerful to understand the complexity of rat recognition strategy, as compared to the ones obtained for Object 1.

In the case of Stimulus Set 2, lower overlap values were found, in both aligned and raw case (Fig. 20). However, 37 pairs out of 60 displayed a higher overlap in the aligned as compared to the raw case, with the average aligned overlap being 0.08 ± 0.09 (mean ± SEM), and the average overlap between raw views being 0.03 ± 0.07 (such a difference was significant; $p < 0.01$; same test as for Stimulus Set 1). In 10 out of 60 cases zero overlap was found for both the raw and aligned cases.

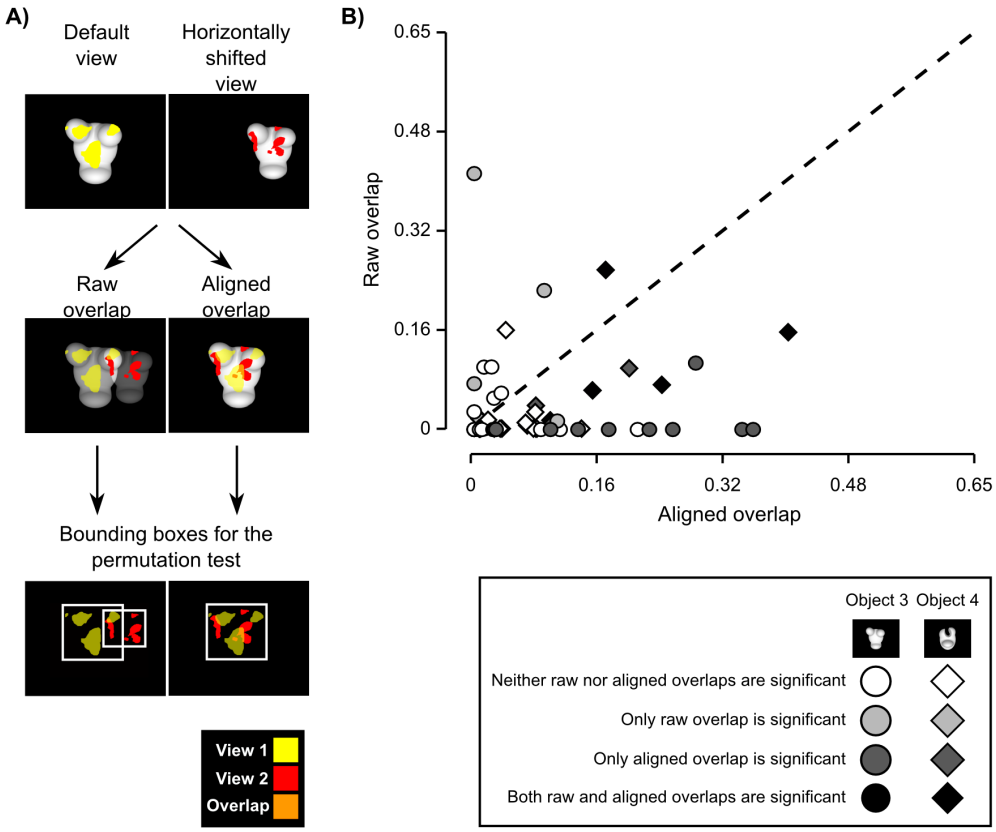


**Figure 20. *Raw* vs. *aligned* features' overlap for all pairs of object views of Stimulus Set 2.** See Figure 19 for details.

Taken individually, Object 3 and 4 displayed similar (not significantly different) values for both raw and aligned overlap in all cases. For Object 3, raw and aligned overlap values were significantly higher than chance in 5/30 cases and 9/30 cases, respectively. As for Object 4, the number of cases was 7/30 for the raw overlap and 11/30 for the aligned overlap. Nonetheless, for both Objects, the aligned overlap values were significantly higher than the raw values (Object 3: aligned, 0.08 ± 0.1 vs. raw, 0.04 ± 0.08. $p < 0.05$; Object 4: aligned, 0.07 ± 0.09 vs. raw, 0.03 ± 0.06. $p < 0.01$). However, only in less than the half of the cases (20/60), the aligned overlap was significantly higher than expected by chance (see black and dark gray diamonds in Fig. 20B), while the raw overlap was significantly higher than chance for 12/60 cases (see black and light grey diamonds in Fig. 20B).

Overall, the overlap analyses of Stimulus Set 1 shown in Figures 19 indicate that rat invariant recognition of visual objects does not trivially rely on detection of some transformation-preserved object features that are diagnostic of object identity across multiple object views. This can be stated, to some extent, also for Stimulus Set 2 (see Fig. 19). Although both raw and aligned overlap values were dramatically lower for Stimulus Set 2, still rat recognition strategy was more consistent with an object-based tracking of broadly-defined object saliency regions than an image-based detection of transformation-preserved, diagnostic image spots.

## 4.5. The Impact of Stimulus Discriminability on Rat Recognition Strategy

As mentioned in the introduction to this chapter, invariant recognition can be achieved through either a view-independent or a view-dependent strategy (i.e., by either relying on the same features, or by using newly learned features for different views of an object). In general, whether a subject will adopt one recognition strategy or the other will strictly depend on whether the objects to discriminate display the features that are critical for their identification at each tested appearance (although transformed). If this requirement is met, it has been shown that humans will tend to rely on a view-invariant strategy, but it is not obvious that other species will also do so. For instance, as it has been shown by Nielsen et al. (2008) using the 2D silhouette of a hand as a target shape, while humans tended to rely on the same features independent of shape orientation in the image plane (view-invariant behavior), monkeys used unique features for each orientation (view-dependent behavior). This, in spite of the fact that the object's critical features were visible at any moment, i.e. at any rotation

tested.

When it comes to our groups of rats, in case of Stimulus Set 1, the results are more consistent with the hypothesis that the animals relied on a view-invariant recognition strategy, i.e., they consistently used the same structural parts of the objects, across different views, as diagnostic features of object identity. On the other hand, the results obtained for Stimulus Set 2 are more consistent with a view-dependent strategy. Both a qualitative inspection of the saliency maps (see Fig. 18) and the lower values of the aligned overlap (see Fig. 20), as compared to Stimulus Set 1 (see Fig. 19) indicate an ad-hoc processing of the objects' identity at each tested view.

To check whether this difference between the strategies used by the two groups of rats depended on the different similarity of the pairs of objects the rats had to discriminate, we computed the pixel-wise Euclidean distance between matching views of each object pair. For a better comparison between Stimulus 1 and 2, the relative Euclidean distance (normalized to the square root of the number of pixels; i.e., the maximal possible distance between two points in our image space) was calculated on low pass-filtered versions of the object images (the same images used for Ideal Observer analysis, see next section), so that the spatial frequency content values did not exceed the maximal retinal resolution of Long-Evans rats. The relative Euclidean distance was quantified for all matching views of the two objects in one Set, but considering only the conditions on which the Bubbles Mask was applied, i.e. Default view (35˚), size transformed view (20˚), azimuth rotated to the left (-40˚) and to the right (20˚), in-plane rotated views (±45˚) and position shifted views (±18˚). The values for Stimulus Set 1 and 2 are reported in Table 1 and 2, respectively.

As expected, a significantly higher discriminability was quantified in the case of the objects in Stimulus Set 1 as compared to the objects in Stimulus Set 2 (paired t-test, $p<0.001$).

Having quantified the different discriminability of the object pairs, we further assessed how such a difference affected the recognition strategy of the two groups of rats by computing the average number (Fig. 21-22) and the average absolute and relative size (Fig. 23-24) of the salient features (across views) found for each object.

| | Default | Size | Azimuth left | Azimuth right | Positions | In-plane rotations |
|---|---|---|---|---|---|---|
| **Object 1 and 2** | **0.13** | **0.07** | **0.09** | **0.10** | **0.11** | **0.12** |

**Table 1. Relative Euclidean distance between objects of Stimulus Set 1.**

Pixel-wise Euclidean distance between Object 1 and 2 was calculated for matching views of the

conditions tested with Bubbles.

| | Default | Size | Azimuth left | Azimuth right | Positions | In-plane rotations |
|---|---|---|---|---|---|---|
| **Object 3 and 4** | **0.08** | **0.04** | **0.06** | **0.07** | **0.07** | **0.07** |

**Table 2. Relative Euclidean distance between objects of Stimulus Set 2.**
Pixel-wise Euclidean distance between Object 3 and 4 was calculated for matching views of the conditions tested with Bubbles.

The number of features was calculated, for all four Objects in the two stimulus sets and considering all views tested, as a function of feature size: the absolute size of the features ranged from a few pixels to possibly hundreds, when the features spanned over large fractions of the object (Fig. 21). The smallest features were considered a source of noise in the determination of the significant number of features and the determination of their average absolute and relative size. In order to have a control on this source of noise we considered a succession of thresholds on the minimum absolute size ranging from 1 to 100 pixels in steps of 1 pixel and computed the average number of salient features as a function of the these thresholds.

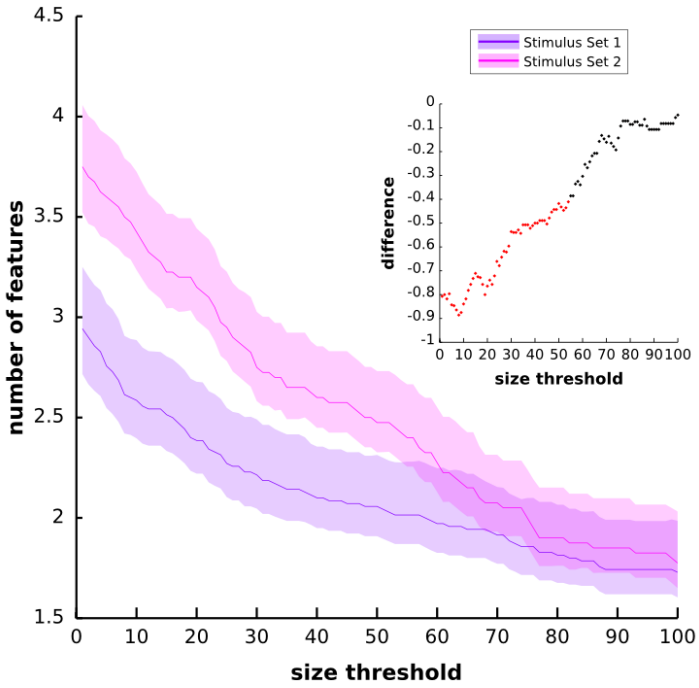


**Figure 21. Average number of features for Stimulus Sets.**
The average number of features is depicted, as group average for Stimulus Set 1 and 2 (see legends for color code), as a function of a threshold set on the features' size. Significance is plotted in inset

as dashed lines: black, no significance; red, significance at $p<0.05$. The higher the threshold on size considered, the smaller the average number of features, and the difference between them. Shaded regions: SEM.

We then considered, for each Stimulus Set first (Fig. 21), and then for each of the six possible pairs of objects (Fig. 22), the statistical significance (two-sample t-test at $p<0.05$) of the differences in the average number of features at each threshold (in Fig. 21-22, the red traces show the comparisons yielding a significant difference).

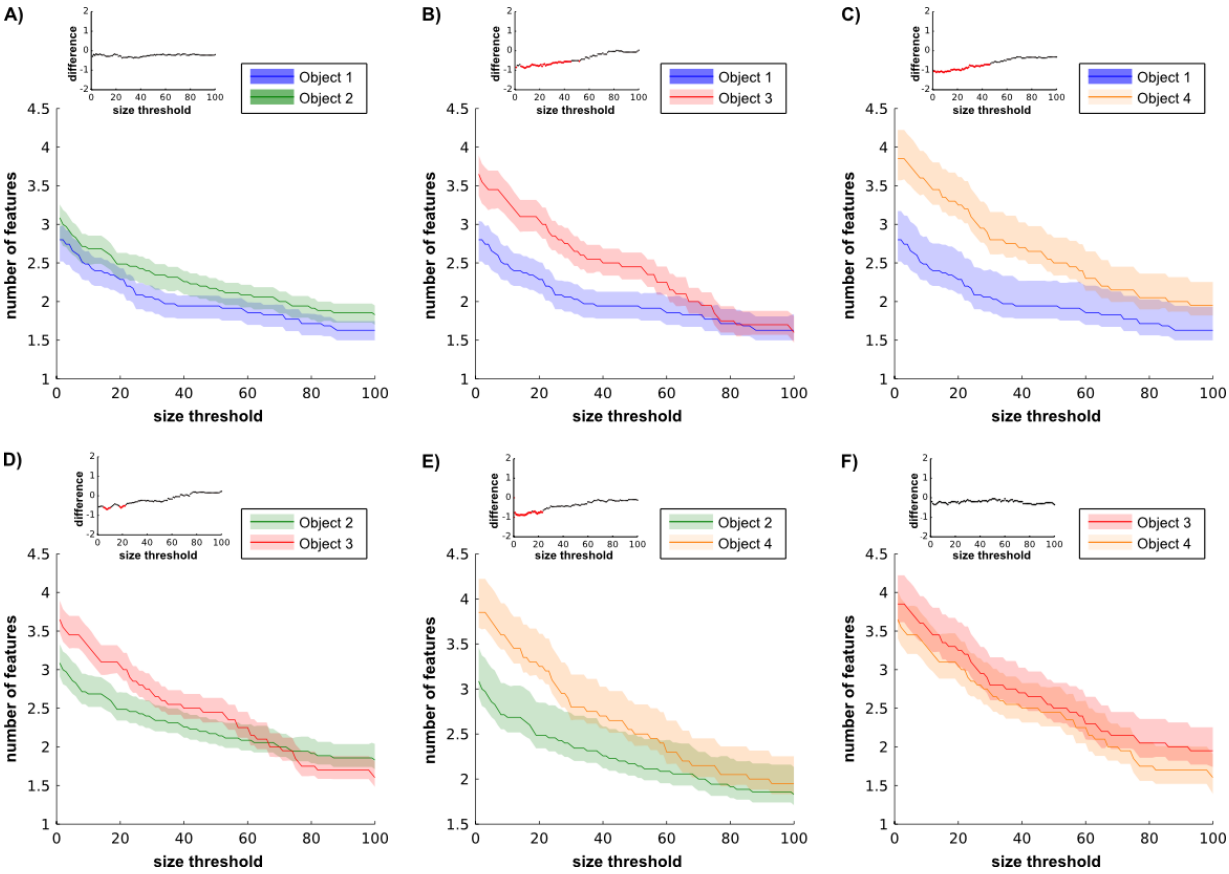


**Figure 22. Average number of features for individual objects.**

The average number of features is depicted, for all possible objects' couples (see legends for color code), as a function of a threshold set on the features' size. Significance is plotted in inset as dashed lines: black, no significance; red, significance at $p<0.05$. The higher the threshold on size considered, the smaller the average number of features, and the difference between them. Shaded regions: SEM.

**A)** Object 1 *vs* Object 2: no significant difference.

**B)** Object 1 *vs* Object 3: significant difference up to ~45-50 pixel size values.

**C)** Object 1 *vs* Object 4: significant difference up to ~45-50 pixel size values.

**D)** Object 2 *vs* Object 3: significant difference in the ranges between ~8-10 pixel size values and ~20-25 pixel size values.

**E)** Object 2 *vs* Object 4: significant difference up to ~25 pixel size values.

**F)** Object 3 *vs* Object 4: no significant difference.

We found that the average number of salient features for Stimulus Set 2 was higher (Fig. 21, pink line), as compared to Stimulus Set 1 (Fig. 21, purple line) and this difference was significant over a large range of threshold values on the features' size (from 1 to about 55 pixels; see red dots in the inset of Fig. 21). Only asymptotically (for very large features' sizes), the difference between the numbers of features found for the two stimulus sets became not significant (see inset in Fig. 21: the red dots become black). This is expected, given that, by construction, only a few large features covering big portions of the objects are considered, regardless of the stimulus set, when the threshold is very large.

Focusing on individual objects (Fig. 22), i.e. considering all possible pairs of the four objects, we found that the average number of salient features for Object 1 was significantly smaller than the one for Object 3 and 4 (Fig. 22, B and C), as long as the features' size did not cross the 45-50 pixel value (see dotted lines in insets), while it was never significantly different from the number of features of Object 2 (Fig. 22, A). Object 2, on the other hand, displayed a smaller difference, in terms of number of features, when compared to object 4 (significant up to ~20 pixels features' size, Fig. 22, E), and even smaller as compared to Object 3 (significant in the ranges between ~8 and ~10 and ~20-25 pixels, Fig. 22, D). No significant difference was found comparing Object 3 and 4 (Fig. 22, F).

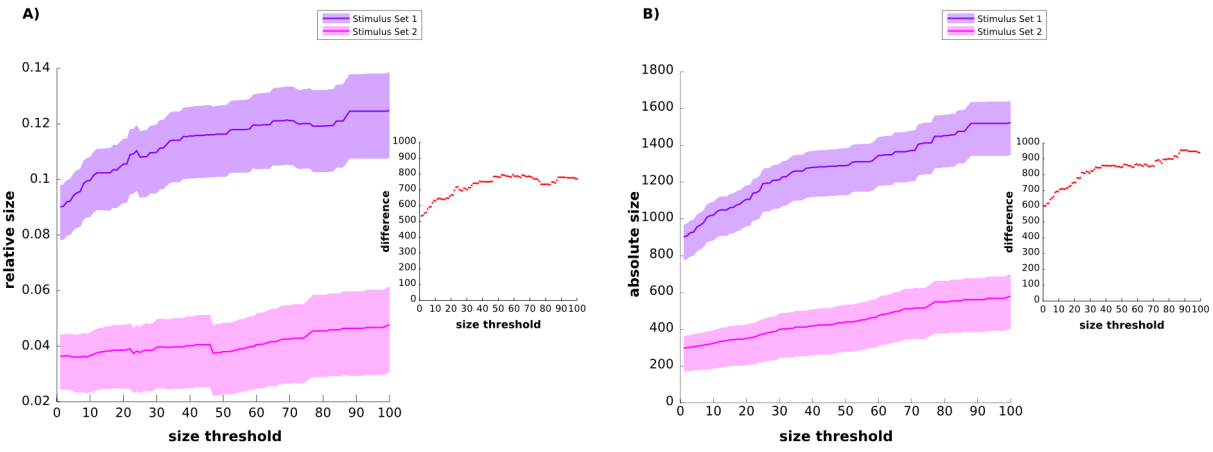


**Figure 23. Average features' relative and absolute size for Stimulus Sets.**

The average features' size is depicted, as group average for Stimulus Set 1 and 2 (see legends for color code), as a function of a threshold set on the features' size. A) Relative size; B) Absolute size. Significance in the difference is plotted in inset as dashed lines: black, no significance; red, significance at $p<0.05$. The difference is significant at all threshold values. Shaded regions: SEM.

Next, we compared the size of the salient features obtained for each pair of the four tested objects (both absolute and relative size were computed, with the latter computed relatively to the object's size in pixels). As shown in Fig. 23, a comparison between the two Stimulus Sets reveals that the two groups of rats tended, on average, to select bigger features for Stimulus Set 1, in terms of both absolute and relative size (the difference in the average feature's size between the two sets was significant at all threshold values; two-sample t-test at $p<0.05$; see red dots).

However, if we consider the differences between individual objects, in terms of their features' relative size (Fig. 24), we note that the only significant difference is between Object 1 of Stimulus Set 1 and all the other objects (Fig. 24, A, B, C). A certain degree of significance appears also, if we consider absolute size values (Fig. 25), for Object 2 with respect to Object 3 (Fig. 25, D).

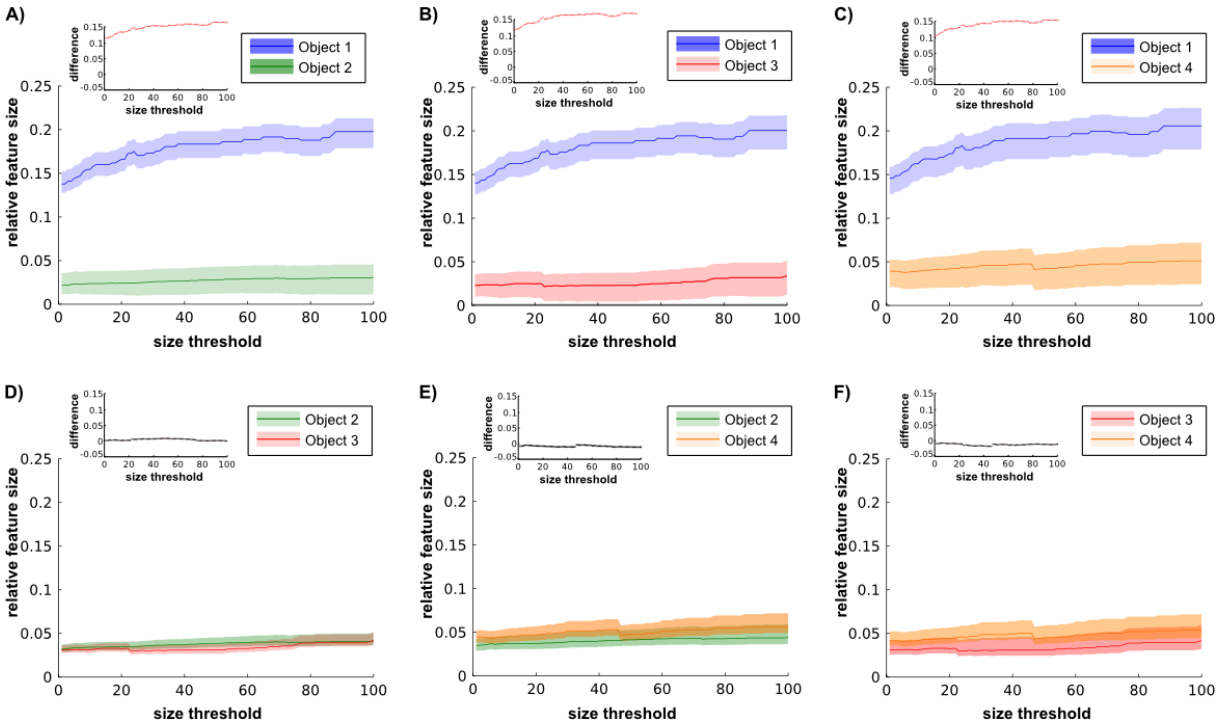


**Figure 24. Average features' relative size for individual objects.**

The mean features' relative sizes and the difference between them is depicted, for all possible objects' couples (A-F, same convention as in Fig. 22), as a function of a threshold set on the features' size. As in the insets of figure 22, significance in the difference is plotted as dashed line. The difference between Object 1 and all other objects is significant at all thresholds considered.

This finding, together with the analysis on the number of features, indicates that i) Stimulus Set 2 displays a higher number of salient features as compared to Stimulus Set 1; ii)

the relative size of the features tends to be smaller for Stimulus Set 1, but is not significantly different as compared to the one displayed by Object 2 in Stimulus Set 1.

Overall, the analyses presented in Figs. 21-25, along with the feature overlap analysis presented in Figs. 19-20, indicate that Objects 1 and 2 (in Stimulus Set 1) were discriminated using a smaller number of larger salient features that were more consistently relied upon across multiple views, compared to Objects 3 and 4 (in Stimulus Set 2). In other words, discrimination of Objects 1 and 2 is more consistent with a view-invariant strategy, while discrimination of Objects 3 and 4 is more consistent with a view-dependent strategy. These strategies, in turn, appear to depend on the visual and structural similarity of the objects to discriminate (much higher in the case of Stimulus Set 2; see Table 2).

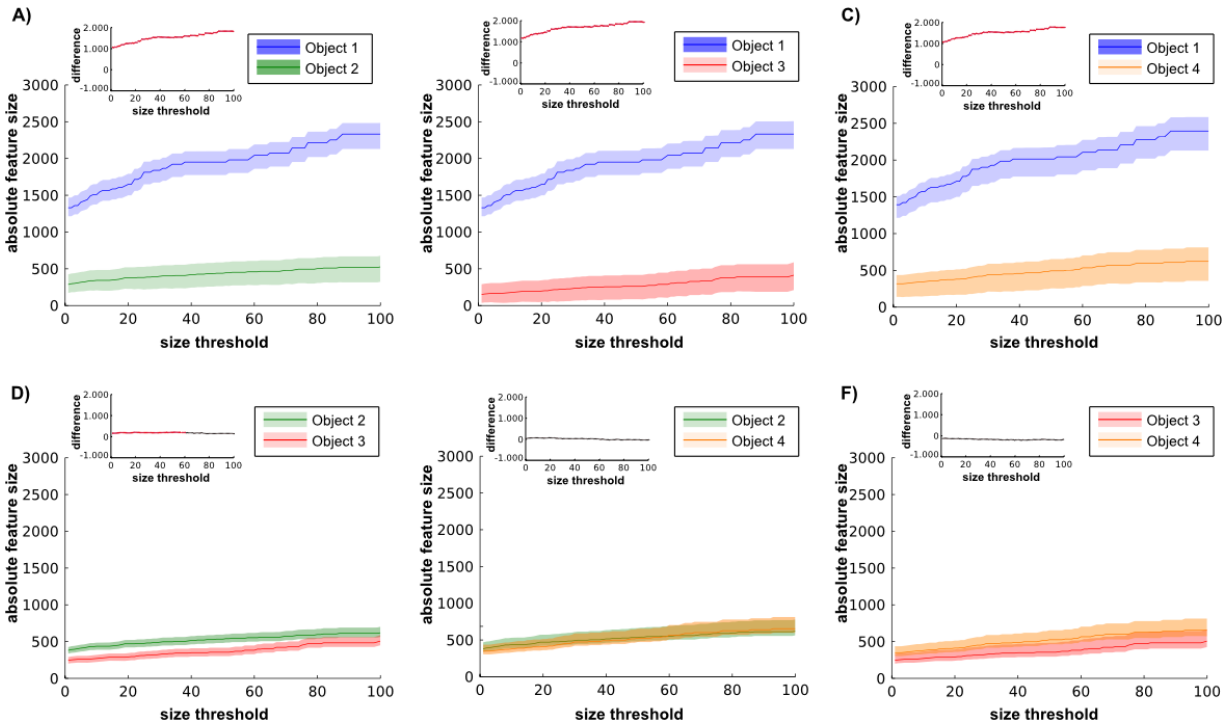


**Figure 25. Average features' absolute size for individual objects.**

The mean features' absolute sizes and the difference between them is depicted, for all possible objects' couples (A-F, same convention as in Fig. 22-23), as a function of a threshold set on the features' size. As figure 23, significance of the difference is plotted as dashed line. The difference between Object 1 and all other objects is significant at all thresholds considered. Also, the difference between Object 2 and Object 3 is significant up to a size threshold of ~60 pixels.

Objects that look different are recognized by a lower number of features, which follow the objects' distinctive structural parts across a variety of transformation axes and magnitudes

(view-invariant strategy). Objects that look similar are recognized through a more variable, more scattered and more numerous set of features (implicating that learning at each tested view is needed; viewpoint-dependent strategy).

## 4.6. Comparison Between the Average Rat and a Simulated Ideal Observer

Having found the possibility that rats are capable of an advanced, shape-based and transformation-tolerant recognition strategy raises the question of just how optimal such strategy is, given the amount of discriminatory information a pair of visual objects (each presented under many different viewing conditions) affords. To address this issue, we built a linear ideal observer and we extracted the critical features underlying its recognition of the same bubble-masked images that had been presented to one of the rats.

The simulated observer was *ideal*, in that it had stored in memory, as templates, the eight views each object could take (i.e., those marked by red frames in Fig. 11C), and was *linear*, since it classified each bubble-masked input image as being either one object or the other, based on which of these templates had the highest correlation with the image itself (see Materials and Methods for details). Given its full access to all possible appearances the objects could take, the ideal observer, by construction, was able to perform optimally in the invariant recognition task and, as such, its recognition strategy represents an upper, optimal bound.

We decided to compare the saliency maps obtained for the ideal observer with rat group average saliency maps, i.e., the maps obtained by pooling the bubble trials collected for a given object view across all available rats. The reason is that such group average maps summarizes rat invariant recognition strategy in a way that is more robust to noise (given the larger number of trials they were based on) and more suitable for comparison with the ideal observer, since idiosyncratic aspects of individual rat strategies get averaged out, while the features that are more consistently relied upon across subjects emerged more clearly.

Before preforming this analysis, we checked the across-rats consistency of performance for both groups of rats, by quantifying the overlap values for salient and anti-salient features in matching conditions (see section 4.3. for details). In other words, the pattern of salient and anti-salient features from one rat at one condition, e.g. default view, was compared to the pattern of features of another rat at the same condition. We did so for all the possible pairs of rats in one group. The distributions of overlap values for salient and anti-salient features of

both Stimulus Sets are depicted in Fig. 26.

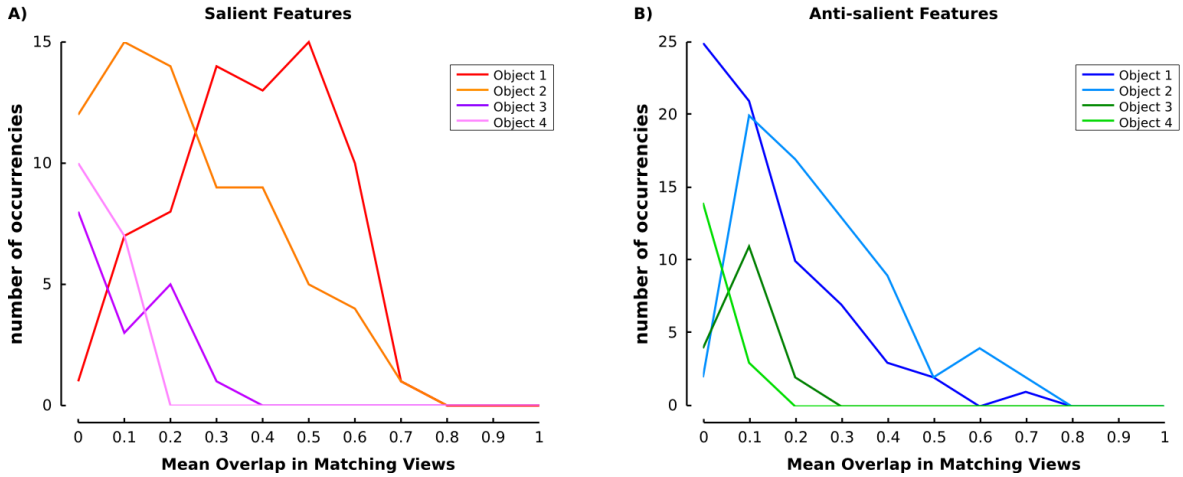


**Figure 26. Distribution of overlap value for salient and anti-salient features in matching views.**

The distribution of overlap values of salient and anti-salient features across rats and in matching views was computed separately for each of the four objects (see legend for color code).

**A)** Overlaps between salient features of the four tested Objects.

**B)** Overlaps between anti-salient features of the four tested Objects.

n=6 for Object 1 and 2 (Stimulus Set 1); n=3 for Object 3 and 4 (Stimulus Set 2).

Not surprisingly, the overlap values for both salient and anti-salient features were higher for group 1 (Stimulus Set 1) as compared to group 2 (Stimulus Set 2) ($p<0.001$, two-sample K-S Test). This indicates a higher between-subject reproducibility of the recognition strategy in the case of group 1. Given this, we deemed as meaningful to compute rat group average saliency maps only for Stimulus Set 1 and, as a consequence, the comparison between rat "average" recognition strategy and the ideal observer strategy was carried out only for Stimulus Set 1 (depicted in Fig. 27).

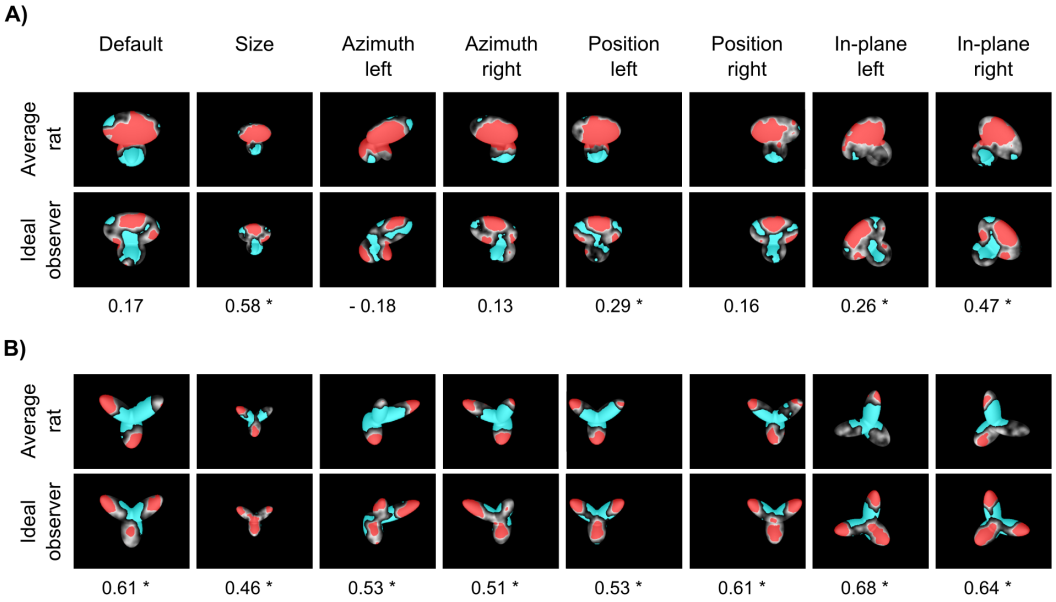


**Figure 27. Critical features' patterns obtained for the average rat and a simulated ideal observer.**

Rat group average saliency maps, with highlighted significantly salient (red) and anti-salient (cyan) features (top rows in **A** and **B**), are compared to the saliency maps obtained for a linear ideal observer (bottom rows in **A** and **B**). For each object view, the Pearson correlation coefficient between the saliency maps obtained for the average rat and the ideal observer is reported below the corresponding maps (* indicates a significant correlation at $p < 0.05$; permutation test).

The patterns of critical features extracted from the average saliency maps (see red and cyan patches in Figs. 27A, B, top rows) were a cleaner version of what observed at the level of individual rats (see Fig. 16). For Object 1, a large salient region (covering most of the upper lobe) and a smaller anti-salient region (covering the bottom part of the lower lobes) were found (Fig. 27A, top row). For Object 2, different combinations of salient features (located at the tips of the lobes) and a large anti-salient area (located at the lobes' intersection) were found (Fig. 27B, top row).

These patterns of critical features bore many similarities, but also some key differences, with those obtained for the ideal observer (Figs. 27A and B, bottom rows). The structural parts in which the salient and anti-salient features were located were largely the same for the rats and the ideal observer. However, in the case of Object 1, the salient region in the upper lobe was fragmented and smaller for the ideal observer, as compared to the average rat, while the anti-salient region was larger, extending from the bottom lobes to the upper one, and branching in two arms that resembled an outline of Object 2 (compare top and bottom rows in Fig. 27A). In the case of Object 2, the location and size of the salient features found for the

ideal observer and the average rat closely matched, although the salient region in the bottom lobe was larger for the ideal observer and typically extended over the lobes' intersection, at the expenses of the central anti-salient region, which was smaller and restricted to the base of the upper lobes (compare top and bottom rows in Fig. 27B). For any given object view, the similarity between the saliency maps obtained for the average rat and the ideal observer was quantified by computing the Pearson correlation coefficient (*r* values are reported under each pair of saliency maps in Fig. 27). Such a correlation was significantly higher than expected by chance for four out of eight views of Object 1 and for all the views of Object 2 ($p < 0.05$; permutation test; see Materials and Methods).

Overall, this comparison shows that rat recognition strategy was highly consistent with that of the ideal observer and, as such, relied on close-to-ideal use of the discriminatory information afforded by the two objects across their various appearances. At the same time, it is interesting to note that where rat strategy departed from the ideal one was mainly because it better parsed the structure of the objects. That is, objects' structural parts, such as the upper lobe of Object 1, were considered salient as a whole by rats, while the ideal observer carved the negative image of Object 2 out of Object 1, even if this operation resulted in a critical features' pattern that did not match the natural boundary of Object 1's upper lobe (see Discussion for possible implications).

| | **Default** | **Size** | **Azimuth left** | **Azimuth right** | **Position left** | **Position right** | **In-plane left** | **In-plane right** |
|---|---|---|---|---|---|---|---|---|
| **Average rat** | **- 0.80 *** | **- 0.66*** | **- 0.82 *** | **- 0.90 *** | **- 0.76 *** | **- 0.67 *** | **- 0.76 *** | **- 0.75 *** |
| **Ideal observer** | **- 0.43 *** | **- 0.61*** | **- 0.63 *** | **- 0.60 *** | **- 0.68 *** | **- 0.76 *** | **- 0.77 *** | **- 0.78 *** |

**Table 3. Phase opponency of the saliency maps obtained for matching views of Object 1 and Object 2.**

Pearson correlation coefficients between the saliency maps obtained for matching views of Object 1 and 2 (i.e., the same maps shown in Figure 10). For both the average rat (top row) and the ideal observer (bottom row), the correlation coefficients were all negative and significantly lower than expected by chance (* $p < 0.05$, permutation test).

In general, the pattern of critical features found for the view of a given object closely resembled the negative image of the pattern of visual features found for the matching view of the other object. This was more apparent for the ideal observer (because of the above-mentioned carving of the silhouette of Object 2 out of Object 1) but it was true also in the

case of rat recognition strategy. In fact, all the saliency maps obtained for matching views of the two objects in Figures 27 show a clear phase opponency. This was quantified, in the case of the ideal observer and the average rat, by computing the Pearson correlation coefficient between saliency maps of matching object views. Most correlation coefficients ranged between -0.6 and -0.8 (see Table 3) and were all significantly lower than expected by chance ($p < 0.05$; permutation test; see Materials and Methods), thus showing that the saliency maps of matching object views were strongly anticorrelated, for both the average rat and the ideal observer. Although the average correlation coefficient was larger for the average rat than for the ideal observer (-0.76 ± 0.03 vs. -0.66 ± 0.04), such a difference was not significantly larger than expected by chance ($p = 0.5$, paired permutation test). Overall, this suggests that the phase opponency of the saliency maps obtained for matching views of the two objects is a property of rat recognition strategy that is fully consistent with optimal extraction of discriminatory information afforded by the tested objects' views.

# 5. Discussion

The goal of this study was to uncover the perceptual strategy underlying rat invariant recognition of visual objects. By exploiting an image masking technique that has been previously applied to human (Gosselin and Schyns, 2001; Nielsen et al., 2008), monkey (Nielsen et al., 2008), pigeon (Gibson et al., 2005) and, very recently, rat vision studies (Vermaercke and Op de Beeck, 2012) we have extracted the patterns of critical features rats rely upon when faced with the challenge of discriminating two visual objects in spite of both affine (i.e., size/position changes and in-plane rotations) and non-affine transformations (i.e., azimuth in-depth rotations). Two different object pairs were used, one containing objects that were structurally/visually different (Stimulus Set 1) and another one containing objects that were structurally/visually similar (Stimulus Set 2).

This approach uncovered several key aspects of rat recognition strategy.

## 5.1. Summary and Implications of our Findings

Our first major finding is that, when it comes to discriminate objects with prominent, easily distinguishable structural parts (as in the case of Stimulus Set 1), rats are able to rely on most of such distinct structural parts (see Fig. 16). This implies that, contrary to what

previously proposed (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012), rats do process global shape information and make full use of the array of diagnostic elements an object is made of. On the other hand, for rats that faced a harder discrimination task (Stimulus Set 2), the diagnostic features were more scattered and more distributed over the whole surface of the objects, often without a clearing match with the objects' structural parts (i.e., lobes; see Fig. 18).

Our second major finding is that, for many rats tested with Stimulus Set 1, the recognition strategy was remarkably stable in the face of variation in object appearance. That is, in many cases, the combination of diagnostic structural parts a rat relied upon was the same across all or most of the object views the animal faced (Fig. 16 A, B). Moreover, the stability of rat recognition could not be accounted by trivial low-level strategies (e.g., relying on transformation-preserved diagnostic features; see Fig. 19). On the other hand, rats tested with Stimulus Set 2 showed a more variable pattern of diagnostic features across objects views, thus showing a more view-dependent recognition strategy (see Figs. 18 and 20).

Performance-wise, rat recognition was typically larger than chance over large extents of the tested transformation axes (see Fig. 15-17, grey diamonds), although not fully invariant, since a drop of performance was observed for extreme transformation values, especially in the case of Stimulus Set 2 (Fig. 17). As previously mentioned, viewpoint-invariant theories propose that, across changes in object view, there is no change in recognition performance: as long as the critical features remain accessible, the response of the system remains constant. In comparison, viewpoint-dependent theories hypothesize that changes in the object view reduces recognition performance, since objects are represented according to how they appeared when originally learned. Since both groups of rats displayed a modulation of recognition performance, one could argue that rats, in general, rely on a view-dependent recognition strategy. However, as shown in many human studies, perfectly invariant recognition performance is virtually never achieved. Moreover, our bubbles experiments allow going beyond what could simply be inferred based on performances. When the patterns of diagnostic obtained for the two groups of rats are analyzed, in many cases, for group 1, the combination of diagnostic structural parts a rat relied upon was the same across all or most the object views the animal faced (see Figs. 16A,B). Rats are therefore able of actively detect and extract discrete object features, which are relied upon irrespective of the transformations the objects may undergo. However, the crucial requirement for this ability to emerge appears to be related to the distinctiveness of the objects, in terms of their structural similarity and the presence of 'well affordable' object-specific features. This stability in the rat recognition

strategy is consistent with the existence of higher-level neuronal representations of visual objects that can tolerate substantial variation in the position, size and orientation of the objects' diagnostic features. This suggests that rats can make use of a view-invariant strategy to recognize objects, similarly to what has been reported for primates and humans, when engaged in the discrimination of familiar and easily discriminable objects (see Introduction fro references). Conversely, the greater across-view variability of the critical features' patterns found for some rats of group 1 (see Fig. 16C) and for virtually all the rats of group 2 (see Fig.18), suggests that such patterns may have been learned and stored in a view-dependent way, which is consistent with viewpoint-dependent accounts of transformation-tolerant recognition in primates (see Introduction fro references). In particular, the scattered nature of the salient features found for both objects of Stimulus Set 2 suggests the tendency to adopt of a novel strategy for each object view; such a strategy appears to typically make use of the objects' structure as a whole, rather than 'looking for' specific features, extracted in the previous phases of the learning process.

In summary, taking into account the larger stability of both the recognition performance and the recognition strategy observed for Stimulus Set 1, as compared to Stimulus Set 2, we can conclude that rat recognition relies on a combination of view-invariant and view-dependent strategies, with the former being prominent in the case of discrimination of structurally dissimilar objects. This is in agreement with a recent report (Tafazoli et al., 2012), demonstrating how rats can spontaneously generalize their recognition to novel object views, without the need of any training (viewpoint-invariant strategy), although the accuracy of the discrimination improves when training is provided (viewpoint-dependent strategy).

Our third major finding is that the critical features' patterns underlying rat recognition strategy at its best (group 1, Stimulus Set 1) closely (although not fully) matched those obtained for a simulated ideal observer engaged in the same invariant recognition task (see Fig. 27). This implies that rats are capable of extracting object discriminatory information, across the various object views they encounter, in a way that is close to ideal. This finding is in disagreement with the study of Vermaercke and Op de Beeck (2012), in which only a small, non-significant overlap between the diagnostic stimulus regions used by rats and by a simulated ideal observer was found (but see further discussion below). Finally, our finding that the salient diagnostic features were confined within the boundaries of the object structural parts (i.e., the objects' lobes) for the rats, but not for the ideal observer, suggests that rat recognition naturally tends to parse visual objects into their more prominent structural

elements, and to process such elements as a whole (although this may lead to a slightly sub-optimal use of the pixel-level object discriminatory information). This further argues for the existence of transformation-tolerant representations of diagnostic object features, which, for most three-dimensional objects, will naturally correspond to the objects' more prominent structural parts.

## 5.2. Validity and Limitations of our Findings

As mentioned before, one limitation of our study is that we did not probe pure generalization of rat recognition to novel object views. This would require presenting the transformed object views (in both regular and bubbles trials) without giving feedback (e.g., reward) to the rats about the correctness of their responses, so to avoid any potential learning of the new views. This was not possible in our study, since the recognition task the rats performed was already exceptionally demanding (given the presence of the occluding masks), and we could not avoid giving reward on bubbles trials, since these trials typically represented more than half of the total in each session (see Materials and Methods). However, we have solid reasons to believe that rat recognition of the transformed object views mainly resulted from generalizing rather than learning/memorizing each individual view. This speculation is based on three arguments. First, two previous studies (Tafazoli et al., 2012; Zoccolan et al., 2009) have rigorously established that rats do spontaneously generalize their recognition to novel appearances of visual objects across many different transformations axes and ranges (including those tested in this study). Second, rat progression along each transformation axis during the staircase training was very quick (see Materials and Methods). For instance, for some rats, it took a single session to reach the asymptotic value along the size axis (i.e., to achieve ≥70% correct discrimination at size 15º of visual angle). Third, although only 8 image pairs were tested with bubbles, bubbles trials were randomly interleaved with regular trials, in which objects could be presented in any of the transformations that a rat had faced up to that point. For instance, when bubbles were applied to in-plane rotations, rats could be presented, in each regular trial, with unmasked objects sampled from the size, position, in-depth rotation and in-plane rotation axes. This amounts to 78 different object views for each Stimulus Set, which makes highly unlikely that rats would memorize each of them. In addition, the bubbles masks themselves produced large changes in the objects' appearance. Crucially, masks varied randomly from trial to trial, thus making even more unlikely (if not impossible) for a rat to

rely on a recognition strategy based on memorizing each object appearance.

Another potential limitation of our study is that rats were not head-fixed and their gaze was not monitored through an eye-tracker. Importantly, however, we designed our study so that the lack of precise eye control did not affect our conclusions. First, rat recognition strategy was recovered by relying on a method that, by its very nature, is alternative to eye-tracking in obtaining saliency maps. As a matter of fact, with the exception of a few monkey studies (Nielsen et al., 2006, 2008), none of the studies that have previously applied the bubbles method (in humans, avian, and rodents) has made use of eye-tracking or other ways of controlling eye position, not even fixation dots (e.g., see Gibson et al., 2005; Gosselin and Schyns, 2001; Vermaercke and Op de Beeck, 2012). Second, most of the transformations the objects underwent in our study were such that rats could not *undo* them through compensatory eye movements. In fact, by construction, saccades cannot possibly nullify size variations, in-plane rotations and azimuth rotations. In fact, by construction, saccades cannot possibly nullify size variations, in-plane rotations and azimuth rotations, although the former two could, in principle, be compensated by head movements (i.e., retractions/protractions and rotations of the head). Importantly, the head of the rats (albeit not rigidly fixed) was restrained during stimulus presentation. As a consequence, head position was highly reproducible across trial onsets and very stable during exposure to the visual objects (see Materials and Methods for a quantification). This guaranteed full control over stimulus size and in-plane rotation. In-depth azimuth rotation was equally well controlled, given that view-point changes were virtual, i.e., resulted from rendering different views of 3D object models on a 2D stimulus display. Therefore, the only object transformations that rats could, in principle, compensate through eye movements were the horizontal position shifts. This would be possible only if the animals made target-oriented saccades to bring the horizontally shifted stimuli always in the same retinal position. However, this is a very remote possibility, since rats do not have a fovea (George Paxinos, 2004), saccade much less frequently than primates do (Chelazzi et al., 1989; Zoccolan et al., 2010) when they are not moving to explore an environment (Wallace et al., 2013), and no evidence of target-oriented saccades has even been reported in rodents. In summary, our experimental design allowed obtaining reliable saliency maps for the various object transformations we tested, without the need of monitoring rat gaze direction.

# 6. Conclusions

Over the past five years, a new tide of studies, encompassing behavior (Busse et al., 2011; Histed et al., 2012; Meier et al., 2011; Tafazoli et al., 2012; Vermaercke and Op de Beeck, 2012; Zoccolan et al., 2009), imaging (Andermann et al., 2011; Bonin et al., 2011; Greenberg et al., 2008; Marshel et al., 2011; Sawinski et al., 2009) and electrophysiology/anatomy (Gao et al., 2010; Van Hooser and Nelson, 2006; Kerr and Nimmerjahn, 2012; Niell and Stryker, 2008; Wang et al., 2011) has reignited the interest for the use of rodent models in vision research. Taken together, these studies show that both mice and rats possess more advanced visual processing abilities than previously appreciated.

In spite of these major advances in our understanding of visual processing in rodents, it remains unclear to what extent rodent higher-level, shape-coding areas are comparable to those found in the primate ventral stream, in terms of their power to represent complex combinations of visual shape features in a way that is tolerant to changes in position, size, orientation, etc. To our knowledge, our study is the first to provide an answer to this question, by uncovering the patterns of diagnostic features underlying rat recognition of structurally complex visual objects that were presented under a variety of viewing conditions. Our results strongly suggest that rats are capable, under conditions of high object discriminability, to process visual objects through rather sophisticated, shape-based, transformation-tolerant mechanisms. As such, given the powerful array of experimental approaches that are available in rats, this model system will likely become a valuable tool in the invasive study of the neuronal mechanisms underlying object vision and, possibly, other higher-level visual functions.

# Chapter III | Motion and Shape Processing in Rat Visual Cortex

## 1. Abstract

In the recent years, researchers have shed light on the rodents' suitability as a useful animal model for the study of middle/high-level cognitive functions. For instance, it has been shown that rats can adopt quite sophisticated strategies in performing highly demanding tasks, as in the case of having to recognize different complex objects under challenging viewing conditions.

While several recent studies have explored cortical visual processing in mice, with the goal of discovering the possible rodent homologues of the primate ventral and dorsal visual streams, we addressed the issue in rats.

First, we run a behavioral study to investigate rat ability to extract *motion vs shape* information from identical stimuli (i.e., moving shapes). These experiments were meant to provide behavioral support to the physiology study described next. Our results show that rats can extract different types of visual information from the same input, and generate behaviors accordingly. This suggests the presence of distinct cortical representations for the same visual input, one conveying information about shape, the other one conveying information about motion.

Then, we carried out extracellular electrophysiological recordings from 5 distinct visual areas of the rat cortex, including V1. Receptive field and tuning profiles were characterized using a variety of visual stimuli (objects, gratings, bars and random dots), presented either as static patterns or in motion at different velocities and directions.

Both approaches are part of a still ongoing study, whose preliminary results are the following:

1) Consistent with the presence of a functional hierarchy, RF size and latency of neuronal response increases along an occipital-temporal pathway starting in V1 and ending in the deeper visual stages of temporal cortex.

2) Tolerance to stimulus transformations and object selectivity increase in the same fashion.

3) Consistent with previous reports, RF size negatively correlates with latency to stimulus onset.

4) Orientation tuning is stronger in the most temporal areas, as compared to the early stages.

6) As shape selectivity increases, responsiveness to motion decreases in the same bottom-up fashion.

7) Direction tuning, computed using moving patterns of random dots, decreases from V1 to the most temporal areas.

8) Consistent with previous reports, RF size positively correlates with motion responsiveness.

Taken together, these findings argue in favor of a bottom-up tuning for visual stimulus complexity in the rat's visual cortex.

## 2. Introduction

The question of the presence, or lack thereof, a discrete streams of processing in the rat's brain, where functionally specialized modules contribute to the processing of shape and motion information, is scientifically interesting, and challenging, for different reasons.

As previously mentioned, both the ventral and dorsal streams exhibit, in humans and primates, a hierarchical organization, where the structural and functional properties of neurons become more and more complex, so as to allow the visual system to eventually solve higher-order, computationally hard problems, such as the Invariance and the Aperture problems (see Introduction, 1. The Ventral and Dorsal streams of processing). Most of the attempts at modeling the accomplishment of high-level shape and motion processing have taken into account such a hierarchy (see sections 1.1.2 and 1.2.3). Taking into account a hierarchy in the accomplishment of a function implies questioning the position that a visual area holds, i.e. its role, the type of information it conveys to and gathers from the other stages.

As previously discussed, many reports (Pack and Born, 2001; Pack et al., 2003; Tinsley et al., 2003) have proposed an independence of V1 from the other stages in the dorsal stream for the solution of the Aperture Problem (thanks to its 'end-stopped' cells, see section 1.2.4). A lesion study has proposed, on the other hand, a fundamental role of rat V1 in both motion processing and orientation discrimination (Petruno et al., 2013, see section 2.4). However, a number of reports have reported specialized (though inconclusive) roles of discrete

extrastriate areas in the mouse brain, which closely mimics the ones found in the rat, for both anatomical and retinotopic organization, in both shape and motion information processing (reviewed by Niell, 2011; see Introduction, 2.2. Is there a ventral/dorsal stream in rats?). However, as stated before, investigating motion processing, or shape processing, by characterizing neuronal responses according to their tuning for spatial and temporal frequencies can be misleading, since real-world objects contain multiple spatial and temporal frequencies. As a consequence, processing of motion and shape information for natural objects might be quite different from what tested in laboratory conditions. Indeed, the response of cortical neurons can be dramatically dependent on the stimulus structure, i.e., the neuron's response to a complex stimulus can be different from what would be predicted by looking at the response to the spatial (or temporal) frequencies of which it is composed (see, for V1: Carandini and Heeger, 1994; Dean and Tolhurst, 1986; Dean et al., 1982; Movshon et al., 1978; Reid et al., 1987; for V1 and V2: Orban et al., 1986; for MT: Jazayeri et al., 2012; Priebe et al., 2003; for IT: Brincat and Connor, 2004; DiCarlo and Cox, 2007; Logothetis and Sheinberg, 1996; Tanaka, 1996).

Optimal choice of visual stimulation aside, although it is well established that rats are able to solve both invariant shape recognition (Tafazoli et al., 2012; Zoccolan et al., 2009; see Introduction, 2.3 and Chapter 2.Advanced Shape Processing in rats) and motion tasks (Douglas et al., 2006; Hupfeld and Hoffmann, 2006; Petruno et al., 2013, see Introduction, 2.4.), and that their visual machinery displays all the fundamental functions of other visual mammals (see Introduction, 2.1. The rat's visual system), the ability of these animals to extract shape and motion information from the same stimulus set, and the possible differential role of rat striate and extrastriate areas in the processing of these two types of information, remains, to date, unexplored. In fact, although we know that rats can learn to recognize complex objects in spite of substantial variation in their appearance, (see Chapter 2 and Tafazoli et al., 2012; Zoccolan et al., 2009), no attempt at investigating object recognition with moving objects has ever been made. Moreover, if rats can segregate between motion and shape information from the same visual input, the implications are quite interesting: such ability possibly points at the presence of discrete neural processing mechanisms for shape and motion in the rat's brain, whose action can be either enhanced or suppressed according to task demands.

In this chapter, I will describe the results of a neurophysiology investigation of visual motion and shape processing in rat visual cortex, with a preliminary behavioral experiment, designed as a sort of introductory 'proof of concept' to the electrophysiological study.

# 3. Materials and Methods

## 3.1. Behavioral Task: Animal Training

Our behavioral paradigm was designed to assess rat ability to: 1) discriminate the identity of two visual objects, when presented either as static images (transformed along a variety of axes, as done in Chapter 1) or moving shapes; and 2) discriminate the direction of motion of two moving objects (same as in the shape discrimination task) and of various patterns of moving dots. Note that the moving object conditions were the same in both tasks, but, in one case, rats had to respond to the shape of the object, while in the other case they had to respond to the direction of motion. Two different groups of rats were used, each tested with one of the two tasks.

### 3.1.1. Visual Stimulation and Experimental Design

Six Long-Evans Rats were placed into six operant boxes in front of a monitor. The box carried a hole to allow the rat to stick out his head, face the monitor and interact with three feeding needles (also acting as touch sensors); a central one for the triggering of the stimulus presentation and two lateral ones for the behavioral outcome and reward collection (same as described in Chapter 1. Materials and Methods for details).

Visual stimulation consisted in:

1) Two complex shapes, or objects (same used as Stimulus Set 1 for the Bubbles study, see Fig. 11A, top row), which could be presented as *static* or *moving* conditions. Static conditions consisted of interleaved transformed views of the two objects (see Fig. 11C) , i.e., size changes (3 different sizes: 15°, 25° and 35°, the default), position shifts (6 different positions: -18, -12, -6, 6, 12, 18), in-plane rotations (6 different values: -45, -30, -15, 15, 30, 45) and in-depth azimuth rotations (6 different values: -60, -40, -20, 20, 40, 60). An overall of 21 views of each object were used.

2) Three patterns of 400 moving random dots. Each pattern consisted of white bright and opaque dots on a black background. Each dot was 2° wide. The dot patterns were moving either to the left or to the right with a 100% coherence and a velocity of 40°/s. To create a moving pattern, a static image of a dots pattern was shifted along the screen in steps of 0.65° per frame. The initial and final positions of the dot pattern were -15°/15° for a right moving pattern, and 15°/-15° for a left moving pattern. Overall the stimulus presentation time was

~800 ms. All patterns were presented full-screen. Velocity, position and presentation time values were the same applied to objects in moving conditions.

Two experimental designs were used for 2 groups of 3 rats each:

*1) Group 1 (Shape Discrimination).*

Rats had to learn to discriminate two objects presented in both default (35˚ size, 0.0˚ position, 0˚ azimuth angle; see Fig. 11A, top row) and transformed conditions in interleaved trials.

After the rats had reached a stable (i.e., a performance they could keep over 2-3 days of training) 65-70% performance, the objects were presented in motion, i.e. moving either to the left or to the right. The moving objects were presented in the default view (full contrast, 0˚ in-depth rotation, 35° visual angle) and shifted from -15˚ to 15˚ or 15˚ to -15˚ along the horizontal axis.

*2) Group 2 (Motion Detection),*

Rats had to learn to discriminate the direction of motion of 3 random dot patterns shifted from -15˚ to 15˚ (rightward) or 15˚ to -15˚ (leftward) along the horizontal axis.

After the rats had reached a stable 65-70% performance, the moving shapes were added in interleaved trials, and rats had to perform the same kind of judgment (on the direction of motion).

In other words, the same two default objects were used. Only, in Group 1, a correct response is associated with the object's shape: *right*, for Object 1, *left* for Object 2. In Group 2, a correct response is associated with the object's motion direction: *right*, if the object is moving from left to right, or *left*, if the object is moving from right to left.

This design allowed us to assess whether two groups of rats are able to extract different kinds of information out of structurally identical visual stimulation.

## 3.2. Physiology: Animal Preparation and Surgical Procedures

All animal procedures were conducted in accordance with the National Institutes of Health, international and institutional standards for the care and use of animals in research, and after consulting with a veterinarian. Naïve Long-Evans male rats (from Charles River Laboratories or in-house born rats) were used for extracellular recordings in different visual cortical areas. All the recordings were performed after the complete development of the rat visual system (Berardi et al., 2000). Animals weight ranged from 300 to 700 grams. Rats were

anesthetized with an intraperitoneal injection of a 0.3 g/kg of Fentanyl (Fentanest®, Pfizer, 0.1 mg/2 ml) and 0.3 g/kg Medetomidin (Domitor®, Orion Pharma, 1 mg/ml) solution. The level of anesthesia was monitored by checking the absence of tail, ear and paw reflex. As soon as the paw reflex started to show up again, the maintenance anesthesia was administered (0.1 g/kg/h Fentanyl and 0.1 g/kg/h Medetomidin solution) and kept throughout the whole recording session. Body temperature was maintained at 37°C with a thermostatically controlled heating pad to avoid anesthesia-induced hypothermia. Heart rate and oxygen level were monitored through a pulse oximeter, and a constant flow of oxygen was delivered to the rat in order to avoid hypoxia. After the first shot of anesthesia produced its effect, rats were positioned on the stereotaxic apparatus to measure the exact location of the craniotomy and the penetration site. During surgical procedures, the rat's eyes were protected from direct light and kept in wet conditions using ophthalmic solution Epigel (Ceva Vetem).

Once the necessary measurements were taken with the stereotaxic arm, the craniotomy was performed and the dura mater was removed from the exposed brain.

## 3.3. Recording Procedure

Recordings were performed targeting 5 visual cortical areas (V1, AL, LM, LI and LL), through a single shank, 32 channels Michigan probe (NeuroNexus Technologies, Ann Arbor, MI, USA) with two typical configurations: 1) a 5 mm shank with densely packed recording area (775 μ) and 25 μm distance between recording sites for tight recording of one area (Fig. 29A); 2) a 6 mm shank with a recording area spanning 1.5 mm and 50 μm distance between recording sites, for recording of more than one area (Fig. 29B).

The craniotomy was made over left hemisphere at different coordinates according to the targeted areas. When targeting V1/AL, the coordinates of penetration were ML (i.e. in the medio-lateral axis) ~4.4-4.6, AP ~6 (i.e., in the anterio-posterior axis, or 6 mm posterior from bregma; Fig. 29C, black squares over red squares in D); when targeting LM/LI/LL, the coordinates were ML ~4.6-5, AP ~7.5 (Fig. 29E, black squares over red squares in F). The craniotomy consisted in a small square of 2x2 mm around the coordinates of penetration (Fig. 29, red squares in D and F).

Before penetration, the rat was placed on a 21 cm-high base, in order to place the horizontal plane of the rat's eye at the center of the medial horizontal axis of the monitor, at a distance of 30 cm from it (see Fig. 28). The base was placed on a rotating platform, which

allowed us to optimize the position of the rat eye, relative to the monitor, once the probe was already inserted in the brain, avoiding mechanical damage. The left eye was covered with a black tape and the right eye (contralateral to the hemisphere from which we recorded) was immobilized using an eye-ring anchored to the stereotaxic apparatus. The elevation of the pupil was then adjusted around 0° on the vertical meridian and rotated to 60° azimuth. The stereotaxic apparatus was at this point rotated 45° toward the left, causing the pupil to point 0° on the vertical meridian and 15° azimuth, leading the center of the right eye binocular visual field to be aligned with the center of the monitor.

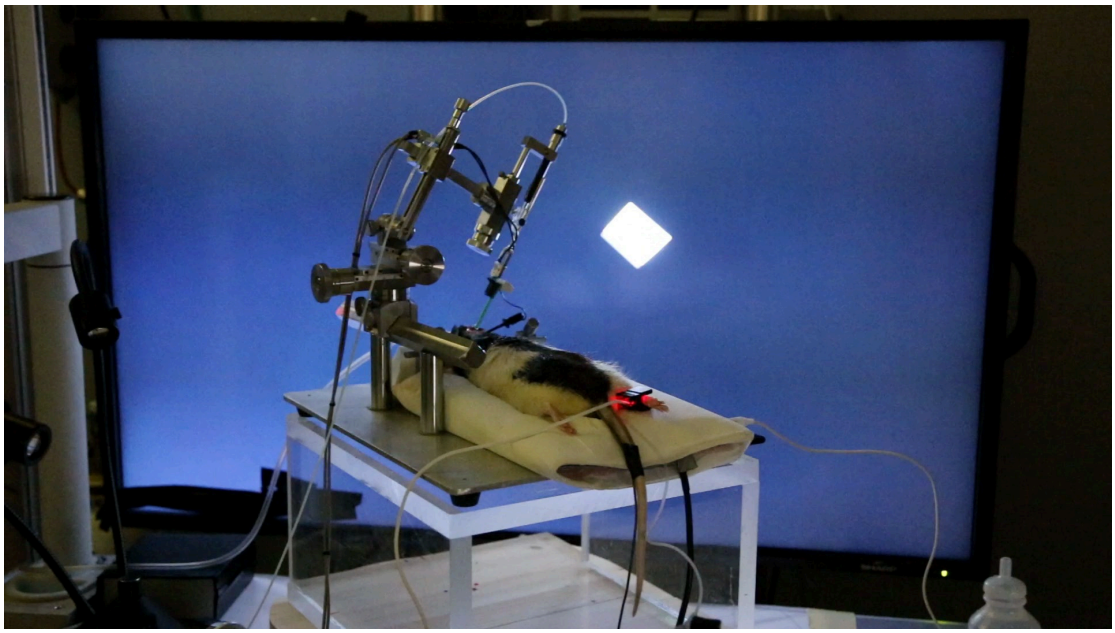

**Figure 28. The recording setup.**

The rat was placed on a 21 cm-high base, in order to place the horizontal plane of the rat's eye at the center of the medial horizontal axis of the monitor, with a distance of 30 cm from it. The left eye was covered with black tape while the right eye (contralateral to the hemisphere from which we recorded) was immobilized using an eye-ring anchored to the stereotaxic apparatus. The elevation of the pupil was then adjusted around 0° on the vertical meridian and rotated to 60° azimuth. The stereotax apparatus was at this point rotated 45° toward the left, causing the pupil to point 0° on the vertical meridian and 15° azimuth, leading the center of the right eye binocular visual field to be aligned with the center of the monitor. On the screen, a frame of the RF mapping protocol is shown (10˚ moving bar, see 3.2).

The probe was coated with Vybrant® DiI cell-labeling solution (Invitrogen, Oregon, USA), a lipophilic membrane stain that diffuses laterally to stain the cells, weakly fluorescent until incorporated into membranes. This orange-red fluorescent dye is often used as a long-

term tracer for neuronal and other cells, to eventually recover the shank position through histological procedures (Blanche et al., 2005; DiCarlo et al., 1996).

After grounding (by wiring the probe to the animal's head skin), the electrode was manually moved using a microdrive (allowing a speed of insertion of ~10 μm/s) into the cortical tissue according to the chosen coordinates, and moved until the chosen depth was reached. For each recording session two or more depths were typically chosen, i.e. two or more areas to gather data from in one session.

The eye and cortex were periodically irrigated using saline solution throughout the whole procedure.

At the end of the experiment, animals were transcardially perfused with phosphate-buffered saline (PBS) 0,1 M followed by 4% paraformaldehyde. The brain was then removed from the skull and kept in 4% paraformaldehyde for 24 h at 4°C, followed by immersion in cryoprotective solution (15% w/v sucrose in PBS 0,1 M, then 30% w/v sucrose in PBS 0,1 M) for at least 48 h each at 4 °C.

Finally, the brain was sectioned in 20 or 25 μm thick coronal slices by a microtome (Leica SM2000R, Nussloch, Germany). The slices were photographed visualizing both the DiI fluorescence left from the electrode and the bright field image, with a digital camera adapted to a Leica microscope (Leica DM6000B-CTR6000 DiI filter, Nussloch, Germany).

The sections that presented the electrode fluorescent track were Nissl stained using Cresyl Violet Acetate solution; this allowed us to reconstruct the electrode position during the recording session.

## 3.4. Visual Stimulation

Each recording session lasted between 6 and 8 hours, and was divided in multiple blocks of 2 hours and 15 minutes. Stimuli were displayed on a 47 inch LCD monitor (SHARP PN-E471R, 1920X1080 pixel resolution; 60 Hz refresh rate; 9 ms response time; 700 cd/m$^2$ maximal brightness; 1.200:1 contrast ratio), positioned at a distance of 30 cm from the right eye, spanning a visual field of 120° azimuth and 90° elevation. Recordings were performed with a Tucker-Davies Technologies (TDT) 32 channels amplification/acquisition system. The signal was recorded at a 25 kHz and 1.5 kHz sampling rate for the spiking activity and local field potential respectively, and filtered at a 300 to 3000 Hz bandwidth (10 to 300 for LFPs).

Rats were exposed to passive viewing of a battery of objects against a black background.

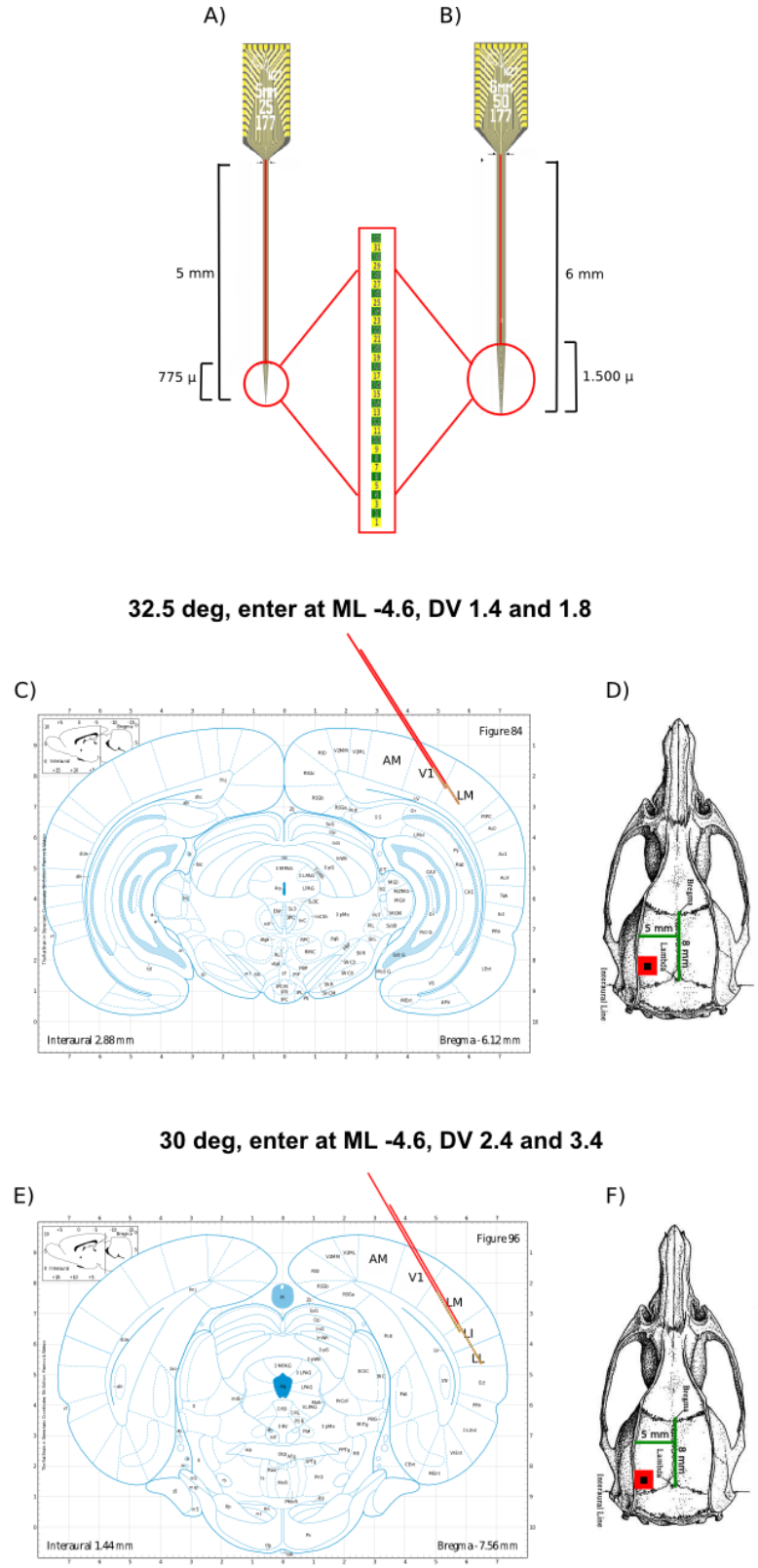


**Figure 29. The coordinates of penetration.**

**A)** The typical probe model used for tight recordings from a single area was a 5-mm long

Michigan probe with densely packed recording area spanning the 775 µm end of the shank tip and 25 µm spacing between recording sites.

**B)** The typical probe model used for smooth recordings from different cortical areas was a 6-mm long Michigan probe with the recording area spanning the 1.500 µm end of the shank tip and 50 µm spacing between recording sites.

**C)** Typical coordinates for targeting V1 and AL in one sessions (two depth of penetration are depicted (DV coordinates)). The corresponding craniotomy coordinates are shown in D) (red squares). The black squares mark the point of penetration.

**E)** Typical coordinates for targeting LM, LI and LL in one sessions (two depth of penetration are depicted (DV coordinates), the last one crossing the two temporal-most areas). The corresponding craniotomy coordinates are shown in F) (red squares). The black squares mark the point of penetration.

A typical experimental session consisted into two main protocols; before running the actual experimental protocol, receptive field (RF) position and size was mapped (see Fig. 34 for details). To this aim, 10˚ wide moving bars at various orientations and 66 different positions on the monitor were shown (6 rows spanning vertically from -20° to +30°, 11 columns spanning horizontally from -50° to +50°) for 15 minutes, so that a total of ~10 trials per condition could be collected.

For this protocol, a tangent screen approximation was implemented to avoid distortions in the stimulus size at each retinal position (i.e., the bars were shown as if they were painted on flat, planar surfaces tangent to a sphere with radius equal to the distance of the eye from the position on the monitor in front of the eye itself – 30 cm). Each bar was displayed for 300 ms and followed by a 250 ms blank ISI.

The actual motion and shape protocol was run right after the RF mapping, and lasted for a period of around 2 hours, so that a total of ~20 trials per condition could be collected.

The visual stimulation that was subsequently used for all data analysis other than RF mapping consisted of 3 main categories: static, fast moving and slow moving conditions. Additionally, two types of blank stimulation were used: i) a black frame (‘black blank’) and ii) a white frame (‘white blank’). They were both presented full screen, and their time could be 250 ms, 800 ms or 1 second to match the other visual conditions (see next sections).

Each condition was presented in random order, with an ISI of 250 ms (see Fig. 33).

### 3.4.1. Static conditions

The static conditions were presented as static frames of 250 ms, and consisted in 3

groups of stimuli.

1) Four 3D artificial objects, chosen on the basis of the rat's ability to discriminate between them, as previously reported (see Chapter 1 and Tafazoli et al., 2012; Zoccolan et al., 2009). The 4 objects were presented at different viewing conditions, i.e. changing size (Fig. 30B), in-plane orientation (Fig. 30C), azimuth rotation (Fig. 30D) and position (Fig. 30E). Changes in size, orientation and rotation were always presented at the center of the screen (position 0.0), while for the position changes the objects were presented at a default size (35˚ visual angle) and default orientation and azimuth rotation (0˚).

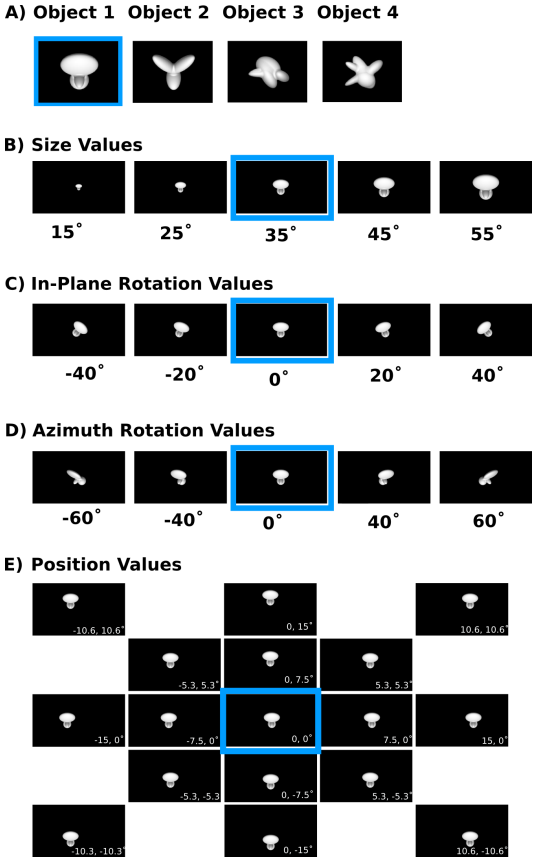


**Figure 30. Static conditions: Objects.**

**A)** The four objects used in static (and moving, see Fig. 32) conditions. The cyan square frames an object presented in equal (default) conditions (35 ˚ visual angle, orientation 0˚, azimuth rotation 0˚, position 0.0˚).

**B)** Size values for the possible views of one exemplar object.

**C)** Orientation values for the possible views of one exemplar object.

**D)** Azimuth values for the possible views of one exemplar object.

**E)** Positions an exemplar object could assume on the screen.

2) 35˚ wide oriented bars (Fig. 31), also subjected to changes in size (Fig. 31A), in-plane orientation (Fig. 31B), and position (Fig. 31C and see Fig. 32 for the axes convention). Changes in size and orientation were always presented at the center of the screen (position 0.0), while for the position changes the objects were presented at a default size (35˚ visual angle) and a range of different orientations (0:45:90) according to the position (see Fig. 31 D) to match the corresponding moving condition (see Fig. 32D).

3) Gratings at 4 different spatial frequencies (SF, 0.03, 0.05, 0.1, 0.4), presented at 4 different orientation values (0, ±45, 90). SF values were chosen according to both electrophysiological (Girman et al., 1999) and behavioral (Meier et al., 2011) reports.

### 3.4.2. Fast and slow moving conditions.

The same stimuli presented in static conditions were presented as movies (Fig. 32). These movies could be fast (45˚/s) or slow (15˚/s) and followed 8 different directions of motion (conventionally referred to as 0˚, 45˚, 90˚, 135˚, 180˚, 225˚, 270˚, 315˚, see Fig. 32B). Moreover, we added to the moving stimuli a battery of two different patterns of random dot kinematograms at 100% coherence. Each dot subtended 2˚ of visual angle, and each frame contained a total of 800 dots at random positions (Fig. 32B). Both gratings (in both static and moving conditions) and random dot patterns were presented as full screen stimulation.

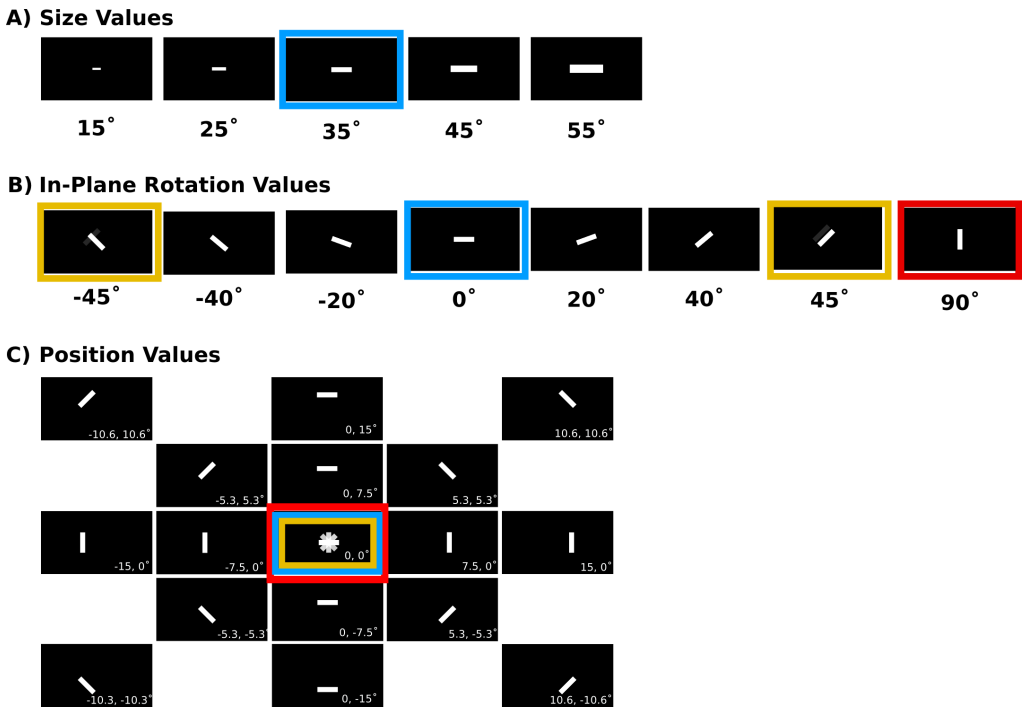


**Figure 31. Static conditions: Bars.**

See Fig. 30 for details. Colored squares frame bars presented in equal conditions: cyan, size 35˚,

orientation 0˚; red, size 35˚, orientation 90˚; ocher, size 35˚, orientation ±45˚.

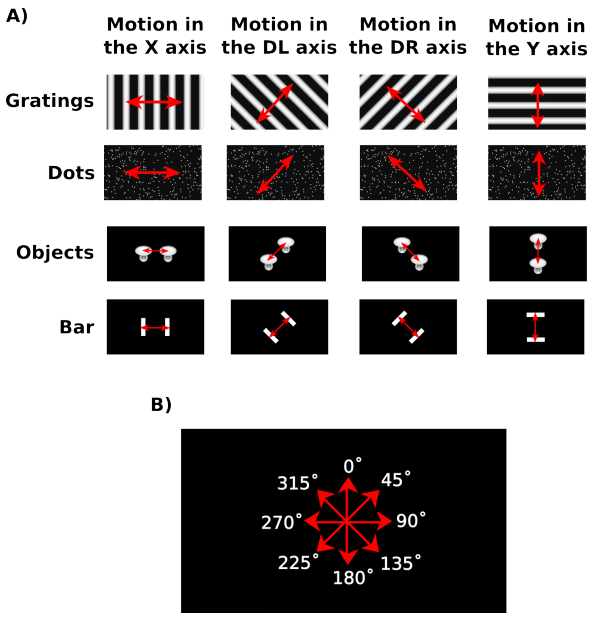


**Figure 32. Moving Conditions.**

**A)** The column title report the conventional labels used for different direction of motion. Rows depict the four groups of stimuli used in our protocol.

**B)** Conventional values for motion direction.

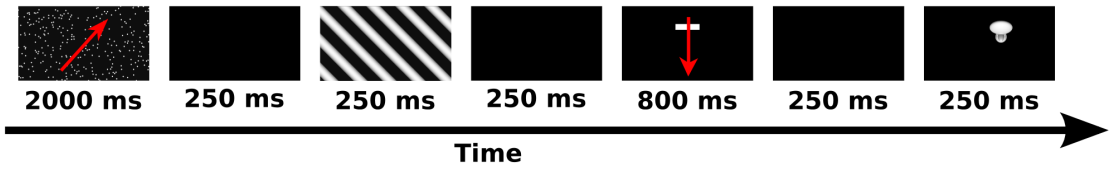


**Figure 33. Timeline of the recording session.**

All visual conditions were presented randomly interleaved. Different presentation times refer to different conditions (250 ms=static, 800 ms=fast moving, 1 s=slow moving). ISI=250 ms.

## 3.5. Data Analysis

All experimental protocols for visual stimuli presentation were implemented using the freeware, open-source software package MWorks (http://mworks-project.org/).

The four objects were renderings of three-dimensional models that were built using the ray tracer POV-Ray (http://www.povray.org/).

Azimuth rotation renderings, gratings and random dot patterns were generated in

MATLAB (http://www.mathworks.com), where also all data analyses were performed.

### 3.5.1. Spike Sorting

It is important to highlight that our data are collected by means of extracellular recording; i.e., we are capturing action potentials of neurons nearby each of the 32 channels of our probes. The raw data consists in the extracellular potential measured in the surroundings of the channel. Extracellular voltage is influenced by many neurons, therefore one single channel could be recording the activity of more than one cell at the same time.

In line with a previous report by Vaiceliunaite et al. (2013), since only sub-optimal spike-sorting strategies are currently available for high-density multielectrode arrays, we did not attempt to perform systematic spike sorting; hence, it is very plausible that in certain cases (e.g. when detecting a firing rate crossing the threshold of 100 spikes/s) what we name 'unit' is actually a 'multiunit', hence we are considering multi-unit activity as support for our results.

We chose Waveclus (Quiroga et al., 2004) as a spike sorter algorithm, a method which combines the wavelet transform, which localizes distinctive spike features, with superparamagnetic clustering, which allows automatic classification of the data without assumptions such as low variance or gaussian distributions. In other words, the algorithm takes a selection of wavelet coefficients to define the clusters of different spikes, and then applies superparamagnetic clustering to group the activity of the recorded neurons in the feature space. The end result of the algorithm is the sequence of spike times, the cluster membership and the spike shapes (Quiroga, 2012).

### 3.5.2. Receptive field size and position

As mentioned before, receptive field size for each neuron was calculated from its firing rate in response to 10˚ wide drifting bars, in four different orientations at 66 different positions on the screen, spanning horizontally from -50° to +50°, vertically from -20˚ to 30˚ (see Fig. 34A).

The response for each orientation was gathered at each position on the screen, and was then used to build firing rates as a function of bar position. The firing rate at each position was fitted with a two-dimensional Gaussian (with independent x and y axis) (see Fig. 34B). The RF size was calculated as the double of the mean between the width of the Gaussian in the x and y axes expressed in degrees. The position of the RF was defined as the center of the fitting. The goodness of fit was taken as a quality measure of the RF observables.

The process was repeated for all the directions and averaged across all conditions. The RF mapping was used as principal marker to identify the different visual areas, by looking at the RF track direction reversals (see next section).

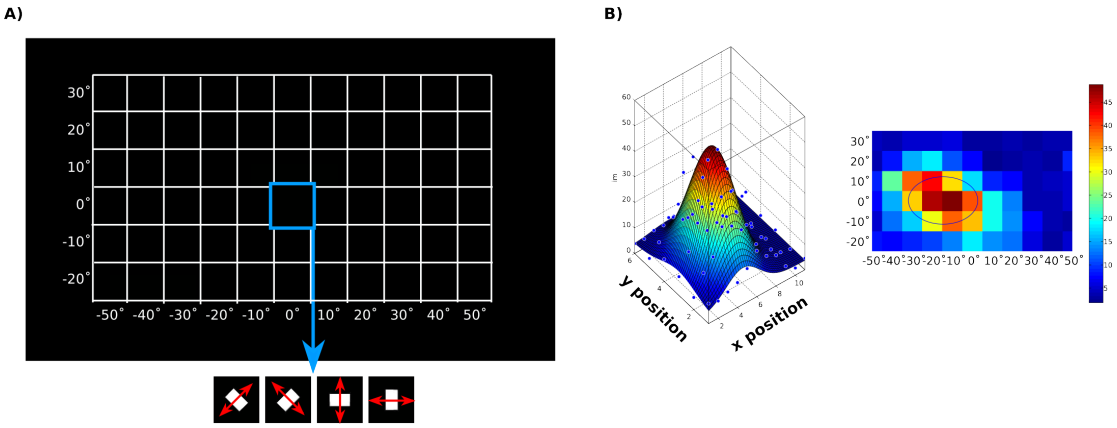


**Figure 34. Receptive Field mapping.**

Receptive field size for each neuron was calculated from its firing rate in response to drifting bars, in four different orientations at 66 different positions on the screen.

**A)** The grid marks the part of the screen that was used for RF mapping. Each square represents 10°. The cyan square marks position 0.0, and points at the possible orientations the bar could assume in that position.

**B)** The mean firing response to each moving bar at each position and orientation was then fitted with a two-dimensional Gaussian (left panel) and revealed RF size and position in the visual field (right panel). The firing rate in response to the drifting bar at each quadrant is color coded (see colorbar).

## 3.5.3. Identification of distinct visual areas by RF reversal

As previously described by Espinoza and Thomas (1983), the rat's visual cortex is composed of seven retinotopically organized visual areas with the cortical topography that is mirrored moving from one area to the next (see Fig. 35F and G). For instance, in LM and LL the retinotopy is the mirror of the one found in V1, with the medial border of each area representing the nasal portion of the visual field, and the lateral border representing its temporal portion. On the other hand, the topography in LI is the same as in V1: nasal to temporal RF positions correspond to lateral to medial recording sites. Crucially, this precise pattern of transitions from one retinotopic map to the next, while crossing multiple visual cortical areas, can be used to identify the area from which each neuron was recorded.

As we have seen in the previous sections (see Materials and Methods, Fig. 29), two main

points of penetrations were used in our recordings: one more anterior in the AP axis, targeting V1 and AL, another more posterior in the AP axis, targeting LM, LI and LL.

When targeting LM, LI and LL, the typical scenario in the reconstruction of the RF location was consistent with 3 (or 4) 'traces' with specific directions and positions in the visual field. An example of one of these cases is depicted in Fig. 35.

The recording session consisted in 2 consecutive blocks of recordings, preceded by RF mapping protocols of 15 minutes each (see section 3.4). Two depths were chosen (marked by green and ocher probe schematics, respectively, in Fig. 35A): one for Block 1, targeting areas LM and LI (DV 2.4 mm), and the other one for Block 2, targeting areas LI and LL (DV 3.6 mm). After RF mapping, the RF track was reconstructed by plotting the position on the screen of the neuronal RFs that had the best fitting (Fig. 35B and D).

The RF track outlined in Block 1 revealed a RF reversal around 110˚ in the azimuth axis (Fig. 35B), consistent with a scenario in which the most superficial recording sites (down to channel no. 9, see white numbers next to the green dots) were located in an area in which the direction of the RF positions was rightward, typical of V1 bordering areas, such as LM and AL (see red arrows Fig. 35F and G). On the other hand, the tip of the probe (from channel no. 9 on) was recording from an area in which the direction of the RF positions was leftward, typical of V1 (see cyan arrow in Fig. 35F) and LI (see cyan arrow Fig. 35G). When going deeper during Block 2, at least one more area was found, indicated by the fact that RF reversed again around -10/-20 in the azimuth axis (Fig. 35D). From this point on (channel 12), the RFs started moving rightwards: we were deep into area LL (see Fig. 35G). As sometimes happens when reaching such cortical depths, another area was found, starting at the reversal around 110˚ in the azimuth axis (see magenta spots in Fig. 35D). We named this area LLb, since the presence of another visual area beyond LL has no reports in literature.

Although very few units were recorded from this area, since it is very hard to target, and although their RF position was rarely at the center of the screen (see Fig. 38F), some of their RF sizes was big enough to encompass part of the objects presented in the central position; therefore, these neurons are included in the following analyses, but the properties they show must not be considered as conclusive.

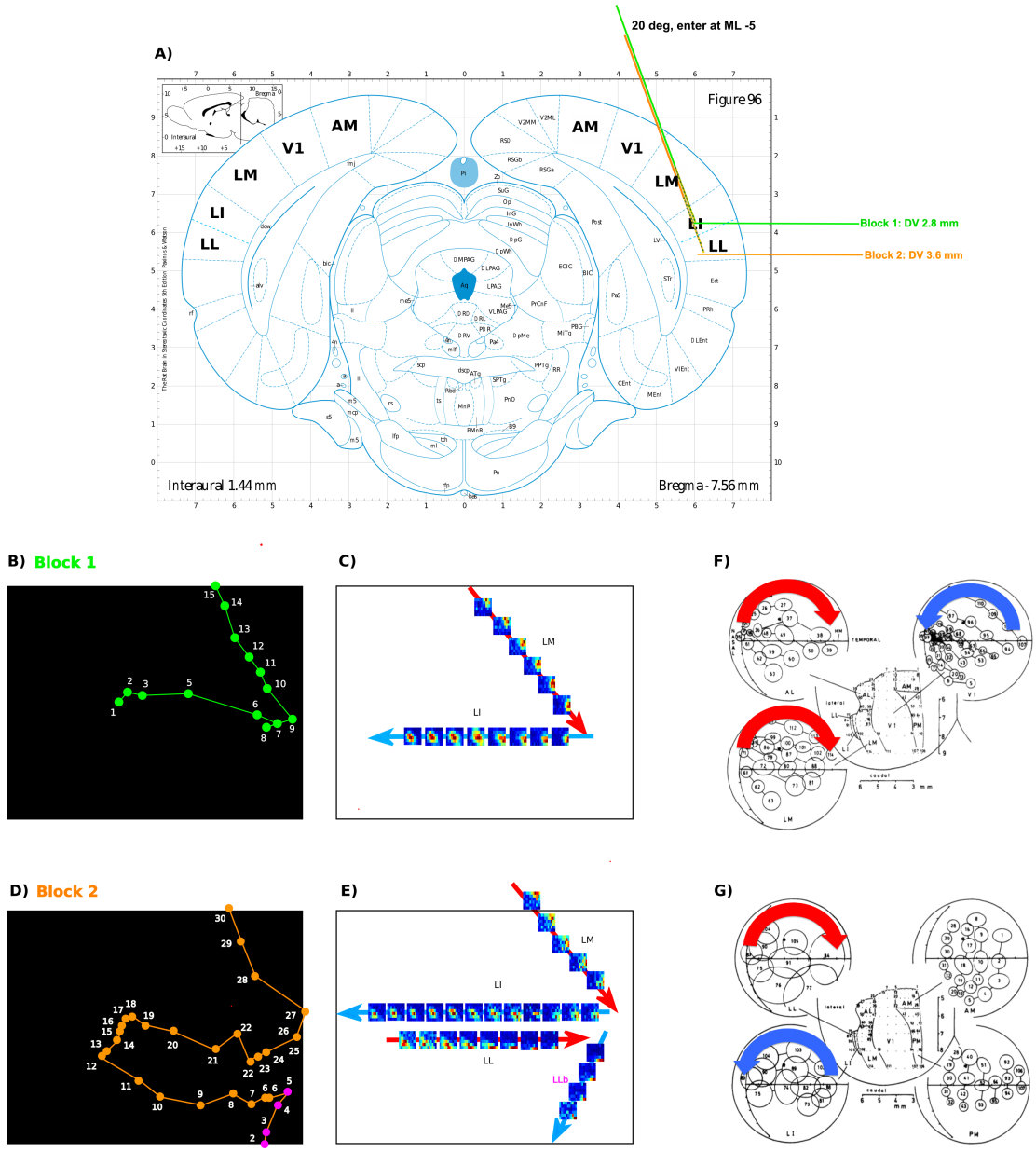


**Figure 35. Retinotopic map reverse marks the spanning of at least 3 different visual areas.** RF track was used as marker of the area we were recording from.

**A)** An example of the coordinates of penetration for one session. The session consisted of 2 main blocks at two different depths (marked by green and ocher lines, representing the probe).

**B)** RF Track for Block 1 for the best fitted RFs. The depth of penetration and the RF direction of reversals were consisted with the targeting of areas LM and LI. White numbers mark the number of recording channels.

**C)** Schematic reconstruction, with the real RF fittings, of the reversals for Block 1.

**D)** RF Track for Block 2 for the best fitted RFs. The depth of penetration and the RF direction of reversals were consisted with the targeting of one more area, namely LL, plus another one, LLb (marked by magenta dots).

**E)** Schematic reconstruction, with the real RF fittings, of the reversals for Block 2.

**F)** and **G)** Representation of RF position, size and direction of reversal of V1 and the extrastriate

temporal areas. The arrows have the same color code as the arrows in C) and E). Adapted from Espinoza and Thomas, 1983.

### 3.5.4. Spike counting window and latencies of response onsets.

In order to measure the neuronal response we computed the optimal spike counting window (SCW) for each neuron following the procedure described in Zoccolan et al. (2007). In order to apply this method, the true epoch of the entering of a stimulus into the RF is required. Since for moving conditions this epoch cannot be determined (response onset can vary widely, depending on the relative position between RF and starting point of the moving object), we had to modify substantially the approach in order to cover these cases. We describe the original approach and its adaptation to the motion case in the following.

*1) Static conditions.* For each neuron and each condition the responses were first computed over a common spike counting window of 0-250 ms, where the zero represent the time of stimulus onset. For each neuron and each condition we first computed the average firing rate FR_obj(t) across the trials. The number of trials was always in the range [18-20] depending on the condition. The FR were computed in overlapping time bins of 25 ms shifted in time steps of 1 ms. Then for each neuron we ranked all the conditions depending on the spikes count in the fixed common window (computed between 0 and 250 ms after stimulus onset). We selected the 10 best conditions (i.e. the 10 conditions that displayed the highest peak in the evoked response) for each neuron and averaged the FR_obj(t), obtaining <FR_obj(t)>_obj = FR(t). Over these conditions we averaged the response in the time window [-50,0] ms in order to obtain the background rate FR_bkg(t) of the neuron (we observed that other choices for this time window, such as [0,50] ms are not safe because for V1b and other areas the onset times can be smaller). We subsequently defined the stimulus-averaged driven firing rate (SADFR) profile as the difference between the firing rate averaged over the best conditions and the corresponding background: SADFR(t)=FR(t)-FR_bkg(t). From this profile we could extract the region corresponding to the spike counting window with a procedure that we will call in the following ‘invasion’. We selected as a ‘pivot’ point the peak of SADFR(t) in the window 0-250 ms, and set the onset and offset of the spike counting window as centred on the peak. Then we considered as a threshold value the 20% of the peak amplitude. We then moved right the offset until the SADFR(t) passed through the threshold. Then we moved the onset to the left in an analogous way. The points of threshold crossing defined the extremes of a connected component containing the peak where the

SADFR(t) was consistently larger than the threshold (i.e., > 20% of the peak). If other components where SADFR was larger than the threshold were found at a distance smaller than 25 ms, the components were merged together and the extremes of the merged component were taken as the onset and offset of the spike counting window (see Fig. 36A). Other information, such as the amplitude of the peaks, their time location, the length of the windows, and the areas of the neurons where stored for further statistical analysis.

*2) Motion conditions.* The fast and slow motion conditions were treated on the same footing, the only difference between the two being the preliminary spike counting window in which we measured the responses, in order to rank all the conditions. These windows were respectively [0-800] ms and [0-2000] ms for the fast and slow motion conditions. In order to determine the optimal spike counting window for each neuron we proceeded as follows. The FR_obj(t) profiles and the rankings were computed in exactly the same way as for the static conditions in their respective windows. Then, for each neuron we selected the 10 best conditions and computed a condition-dependent spike counting window with the invasion algorithm applied to the single FR_obj(t).

With this procedure we obtained a condition-dependent set of onsets and offsets for the spike-counting window, and in order to reduce the uncertainty of this estimate, we decided to determine the offsets as the sum of the onsets plus a duration which was set common for all conditions. The procedure by which we selected a common duration implied the selection of the 10 best conditions as in the static case and we aligned their profiles FR_obj(t) based on their onsets (see Fig. 36B). We then averaged these profiles and subtracted the background computed in the same way as in the static case, obtaining the analogue of the SADFR(t) in the motion case (see Fig. 36C). We than applied the ‘invasion’ algorithm on this SADFR(t) in order to compute the duration of the response. Then we went back to the FR_obj(t) and for each profile we could finally compute the offset adding to the onset the duration of the response obtained.

Alternative and interesting methods for a precise computation of latencies were developed in DiCarlo and Maunsell (2005), but we could not apply directly these methods here due to the different spatiotemporal nature and complexity of the stimuli.

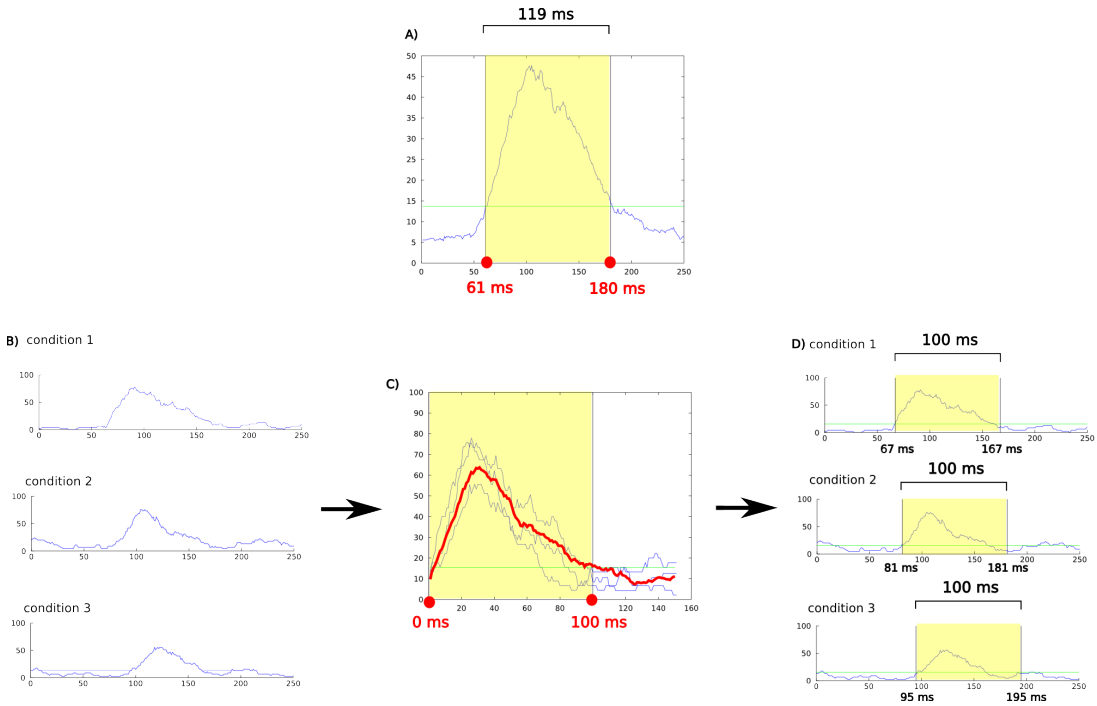


**Figure 36. Selection of the spike counting window for static and moving conditions.**

**A)** Spike counting windows for static conditions. The neuron's background rate was computed over the 50 ms before stimulus onset, and it is marked by the green line. The 10 best conditions were selected for each neuron, averaged and the background subtracted, thereby obtaining the neuron's stimulus-averaged driven firing rate (SADFR, blue trace). The spike counting window was centered around the peak, considering as a threshold value the 20 % of the peak amplitude. The window is highlighted by a yellow patch. Onset and offset values are reported in red; the time window duration is reported in black.

**B-D)** Spike counting windows for moving conditions. Three exemplar traces with different onset values were aligned in the time axis and averaged **(C)**, and a fixed duration of the window chosen. The window is highlighted by a yellow patch, with initial onset and offset values reported in red. We then went back to the original traces, and the offset was defined as the original onset plus the duration time **(D)**. Onset and offset values are reported in black.

## 3.5.5. Selectivity and tolerance indexes

Object selectivity and tolerance to object transformations were calculated on static conditions.

A tolerance indexe measures how invariant is the response of a neuron to transformations of a given stimulus. Indexes for Position Tolerance (PTI), Size Tolerance (STI), Azimuth Rotation Tolerance (ATI), and Orientation Tolerance (OTI) were measured with the same method for each neuron as we describe in the following for the case of Position

Tolerance (PTI). Given a tuning curve for a generic transformation (for example the size) we first normalized the corresponding tuning curve $TC$ dividing by its maximal value. Then we defined the tolerance for the transformation as the average over all the values that were less than 1: $Tol = \langle TC \rangle^*$ (we denote the averaging procedure with the angular brackets and the star). The tolerance index ranges from zero to 1 by construction, these two values corresponding to the extreme situations in which the neuron is extremely tuned or perfectly tolerant to the transformation. For the case of position tolerance we defined first a global tuning curve for position transformations concatenating the four tuning curves corresponding to the four axes of position transformations.

A selectivity index measures the preference of a neuron for a particular stimulus with respect to other stimuli. A typical measure of selectivity is sparseness (Olshausen and Field, 2004), which has proven to be useful when the number of stimuli is large (Zoccolan et al., 2007), but has the limitation of being based only on average firing rates, without taking into account the variability of the responses and, therefore, the actual discriminability of the stimuli. In our case, to overcome these limitations (also given the small number of stimuli in our object set), we decided to use a measure of selectivity based on the d' index. We recall that the d' index between two normally distributed random variables X and Y is defined as $\frac{\mu_x - \mu_y}{\sqrt{\frac{1}{2}(\sigma_x^2 + \sigma_y^2)}}$ where $\mu_x, \mu_y$ are the means and $\sigma_x, \sigma_y$ the standard deviations respectively. In other words the d' index between two random variables is the difference in their expectation values normalised by the square root of their average variance. This measure is preferable to sparseness in that it accounts explicitly for the statistical fluctuations in firing rates: the difference of the means being equal, the discriminability of two signals is higher when they have a low variance and this is captured by the d' index. We computed the selectivity for each neuron as follows. For each pair of objects we computed the maximal d' over all the tested 17 positions and we repeated the computation for all the six pairs of objects. We then defined the selectivity of a neuron as the maximal value obtained among the six couples.

### 3.5.6. Orientation and direction selectivity indexes

The Orientation selectivity index (OSI) was defined as $(R_{pref} - R_{ortho})/(R_{pref} + R_{ortho})$, where $R_{pref}$ is the mean response in the preferred orientation. $R_{ortho}$ was defined as the response evoked by the orthogonal orientation. With this index, perfect orientation selectivity would give OSI = 1; an equal response to all orientations would have OSI = 0, and 3:1

selectivity corresponds to OSI = 0.5. Similarly, the Direction selectivity index (DSI) was considered as a marker of motion selectivity for an area, and calculated separately for fast and slow moving conditions. The DSI was defined as $(R_{pref}-R_{opp})/(R_{pref}+R_{opp})$, where $R_{opp}$ is the response in the direction opposite to the preferred direction. With this index, perfect direction selectivity would give DSI=1; an equal response to all directions would have DSI=0, and 3:1 selectivity corresponds to DSI=0.5. Highly and poorly tuned neurons were defined as neurons with an OSI/DSI>0.5 and OSI/DSI<0.5, respectively.

In order to avoid the confound of artifacts due to RF position (see section 4.5), OSI sand DSIs were quantified for full-field conditions only.

# 4. Results

## 4.1. Rats Can Segregate Shape and Motion Information

The performance in the behavioral task described in Materials and Methods, showed that rats are able not only to discriminate static visual shapes (Fig. 37A; blue bar) and motion direction (Fig. 37B; green bars), as previously reported (see Douglas et al., 2006; Petruno et al., 2013; Tafazoli et al., 2012; Zoccolan et al., 2009), but also to correctly discriminate either the identity (Fig. 37A; red bars) or the motion direction (Fig. 37B; red bars) of moving shapes. Being able to extract different kinds of visual information, related to either shape or motion, from the same visual input, and generate behaviors accordingly, points at rat's ability to tell motion from shape, and shape from motion. This suggests the presence of distinct cortical representations for the same visual input, one that must convey its shape features, and another one that must convey its motion features.

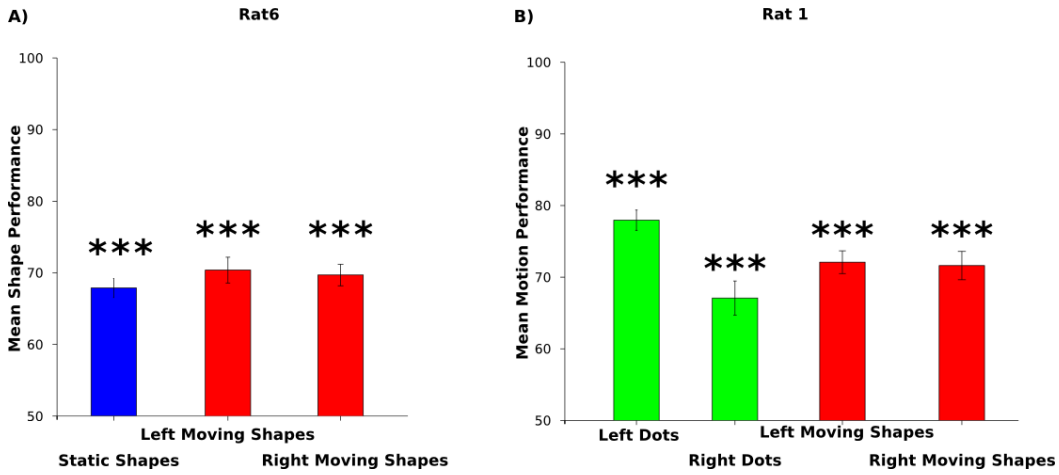


**Figure 37. Performances of the two best rats.**

**A)** Performance of the best rat of group 1 (shape task). The performance was around ~70% of correct trials at each condition. Number of sessions=106.

**B)** Performance of the best rat of group 2 (motion task). The performance was around ~70% of correct trials at the moving shapes trials, with a slight bias to the right at the moving dots conditions. Number of sessions=113.

***=p<0.001, right-tailed T test.

## 4.2. RF Size Increases towards the Most Temporal Areas

The mean RF size per area was quantified after filtering out the worst units, by setting a threshold on the quality and position of each RF in each area. In other words, RFs that had a bad Gaussian fitting (i.e. yielded a coefficient of determination lower than 0.5), or were located off the visual field, were removed from this analysis, and all the analyses which explicitly required the RF quantitative features. Note that, for neurons belonging to area LLb, only a threshold on goodness of fit on the RF was applied as a criterion of selection, since the center of their RFs was typically off the the screen (see Fig. 38F). These neurons were nevertheless considered in the following analyses, because, having very large RFs, they afforded the possibility to catch some of their tuning properties from the stimulus conditions falling into the lower/lower right quadrant of the screen (occasionally, since their RF size was very wide, sometimes even the conditions falling in the center of the screen could be encompassed). As mentioned before, this area yielded very few units; applying a threshold on the RF positions as well would prevent us from having enough data to, albeit weakly, characterize it.

The RF centers and sizes found in each area (before thresholding) are depicted in Fig. 38 (blue circles and red dots, respectively). The white circle represents the typical area that a

default object (35˚ size, position 0.0˚) covered in the visual field.

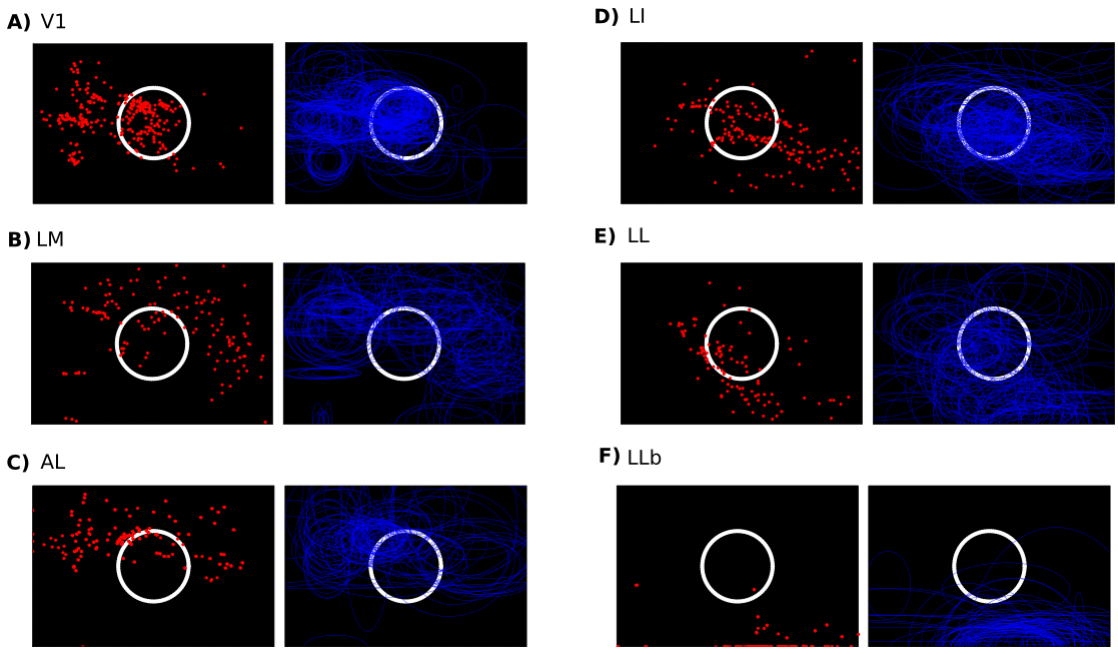


**Figure 38. RF sizes and positions for each area.**

RF size (blue circles) and positions (red dots) are depicted for each area, from V1 (**A**) to LLb (**F**). The red dots located on the edges in F) represent off-screen RFs.

Consistently with Espinoza and Thomas (1983), we found an increase in the RF size going from V1, to AL/LM, to LI/LL (and LLb) (Fig. 39).

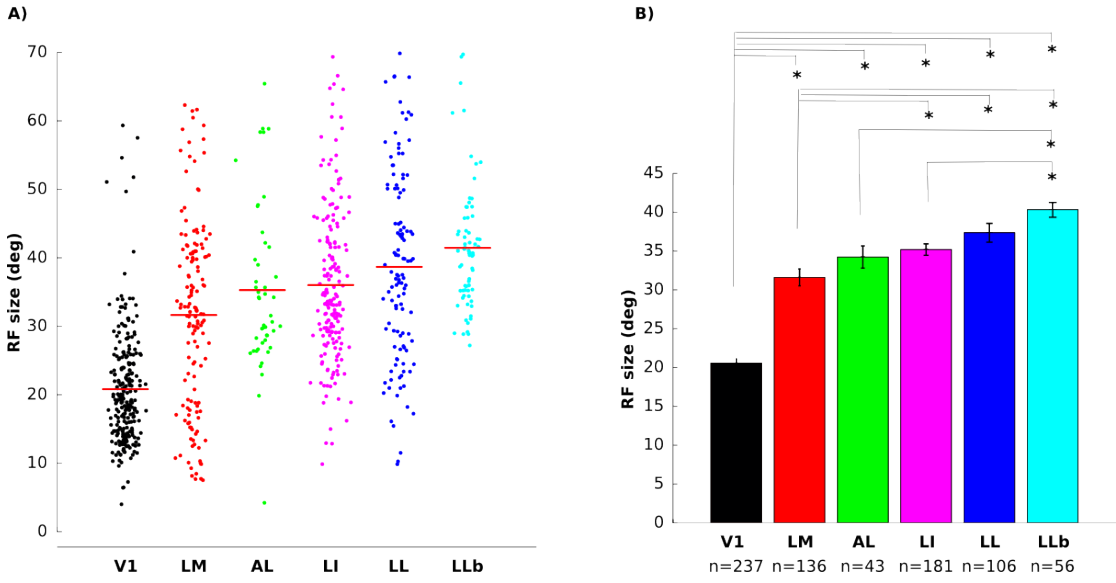


**Figure 39. Mean RF size per area.**

RF size increases as a function of area, from the smaller ones in V1 to the bigger ones.

**A)** RF size distribution of each area. Red lines: mean value.

**B)** Average RF size for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

The statistical difference between one area and the next was computed by running a one-way ANOVA (Tukey post-hoc test, $p<0.05$) on each population. V1 population turned out to be significantly different from all the other areas. The other significant difference where between LM and LI, LL and LLb; LLb and AL and LI.

## 4.3. Spontaneous and Luminance Driven Firing Rate Profiles

Firing rate profiles were calculated as previously described (see Materials and Methods, section 3.5.3.). Spontaneous activity was quantified considering the neurons' response to black blank conditions (blank screen) (Fig. 40). We also considered the white blank conditions as a measure of the average area response to luminance (Fig. 41). For both blanks, all the three stimulus presentation times used for these conditions (250ms, 800 ms and 1 second) were considered and averaged. When tested with full field white blank stimulation, a decreasing tendency to fire appeared across the occipital-temporal axis: higher firing rate profiles were observed for LM and V1 (LM firing was significantly higher than all other areas), followed by AL and LI. LL and LLb appeared to be the areas with the least sensitivity to full luminance stimulation.

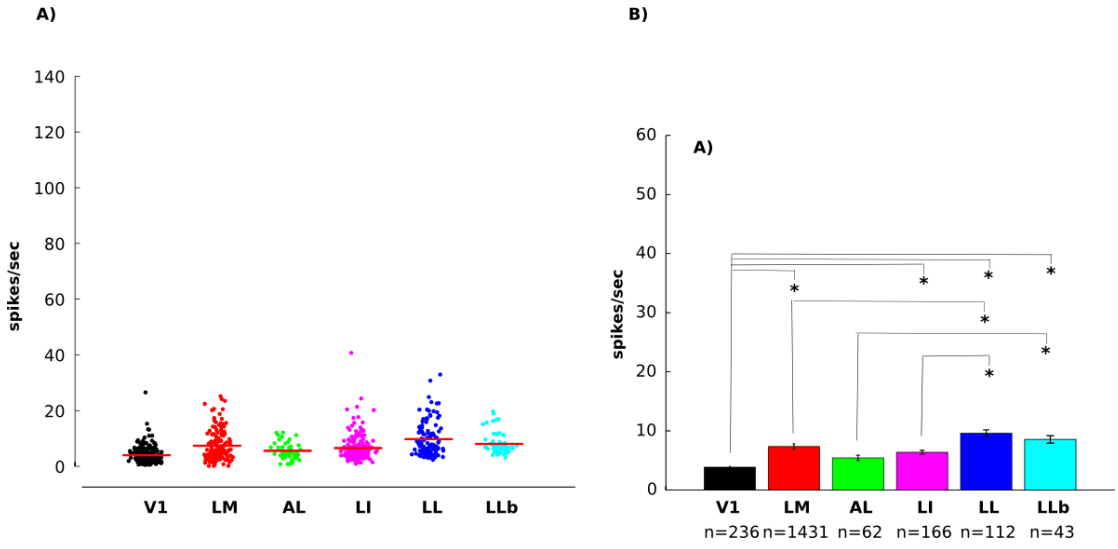


**Figure 40. Spontaneous firing rates per area.**

**A)** Spontaneous firing rate values distribution of each area. Red lines: mean value.

**B)** Average spontaneous firing rates for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

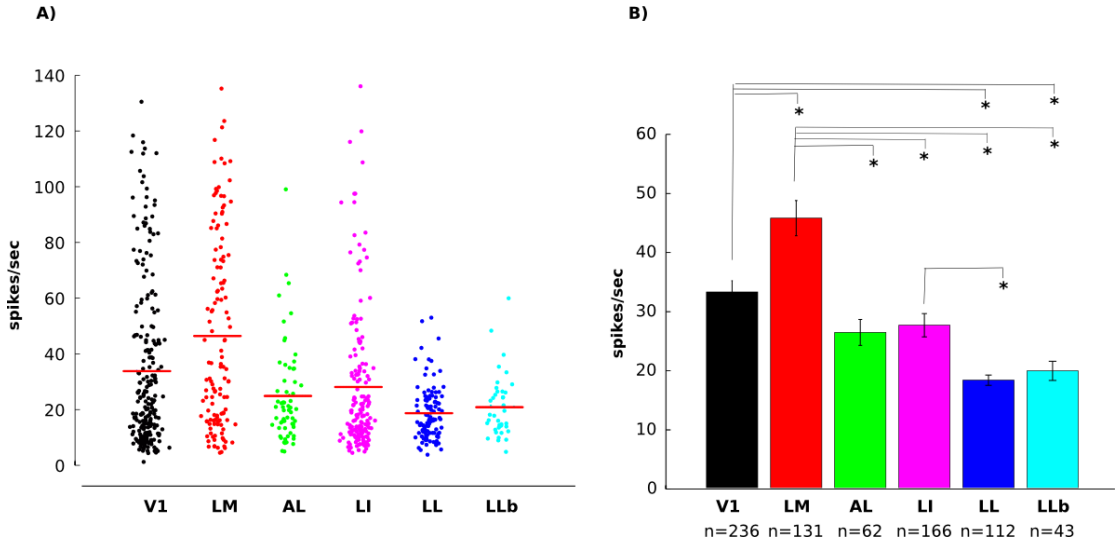


**Figure 41. Average firing rate profile in response to white blanks.**

**A)** Evoked firing rate values distribution of each area in response to full-field white blanks. Red lines: mean value.

**B)** Average firing rates for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

## 4.4. Shape Tuning

Shape tuning was quantified for each area as the response of its neurons to static conditions. Tuning curves were quantified as the average response of neurons to different transformation values of size, position, azimuth rotation and orientation.

A variety of tuning profiles was found for each area. Focusing only on areas V1 (Fig. 42 and 43) and LL (Fig. 44 and 45), and taking as examples units displaying small and big RF sizes, we can make some general considerations on how tuning might change from the first to the last stage of processing. We see that: i) independently of RF size, LL units are more tuned for smaller objects as compared to V1 (compare panels E in Fig. 42-43 and 44-45); ii) consistently with Fig. 41, V1 is more sensitive to luminance full-field stimuli (compare light grey lines in panels E-J in Fig.42-43 and Fig. 44-45); iii) when neurons with similar RF sizes in V1 and LL are compared, the responses of V1 and LL neurons to their most effective (or preferred) objects are similarly tolerant to position changes (compare Fig. 42C to Fig. 44C),

but LL neurons show a stronger power to discriminate the preferred from the non preferred object across the visual field (compare Figs. 44C and D), as compared to V1 neurons (compare Figs. 42C and D), thus suggesting a lower object selectivity for the latter area.

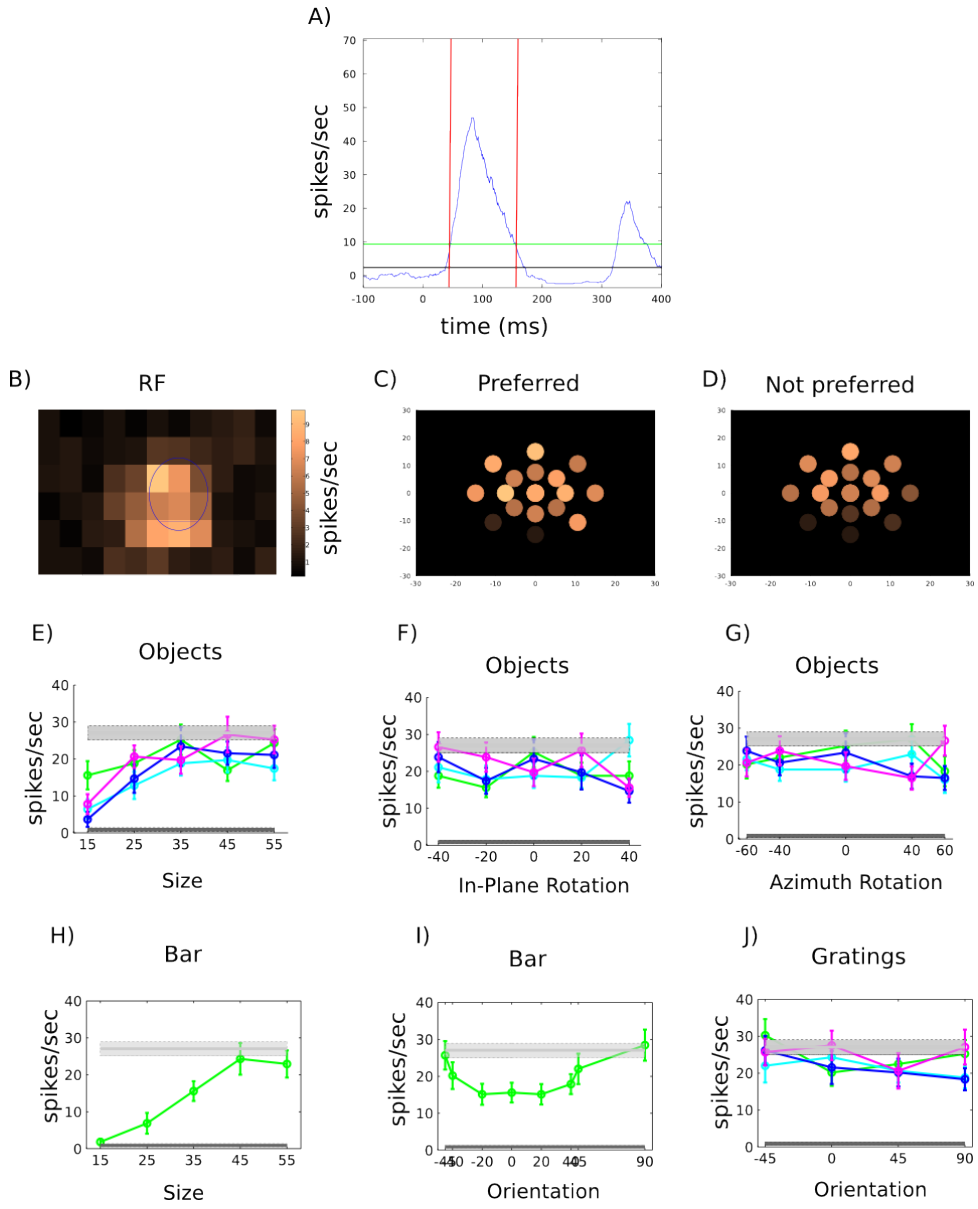


**Figure 42. A V1 neuron with wide RF.**

**A)** Mean PSTH computed considering the neuron's mean response to the 10 best conditions. The two red lines outline the spike counting window. The green line marks the onset threshold. The black line marks the average background activity.

**B)** The neuron's RF. Size=26.67°.

**C)** Position tuning of the best object. The dots mark the different positions a stimulus can assume on the screen. Brighter dots mark spots of higher sensitivity to that position. The bigger the RF, the broader the part of screen covered (the higher the number of bright dots).

**D)** Position tuning of the worst object. Same convention as in C.

**E-G)** Tuning curves to objects undergoing size, orientation and azimuth rotation transformations. Green line=object 1; cyan line=object 2; blue line=object 3; magenta line=object 4. The neuron shows tuning for bigger object areas. No important tuning for orientation or rotation in the azimuth plane. Error bars indicate SEM.

**H-I)** Tuning curves to bars undergoing size and orientation transformations. Consistently with object tuning, the neuron shows tuning for bigger bar areas. A tuning for -45˚ and 90˚ orientations appears.

**J)** Tuning curves to oriented gratings of different spatial frequencies. Green line=0.03; cyan line=0.05; blue line=0.1; magenta line=0.4. No significant tuning appears, except for SF 0.03 (green line), which matches the bars tuning.

In all tuning curves plots, the dark grey and light grey shaded lines refer to black and white blank mean response, respectively. The response to full field white blank condition is high, typical of V1 neurons (see Fig. 41).

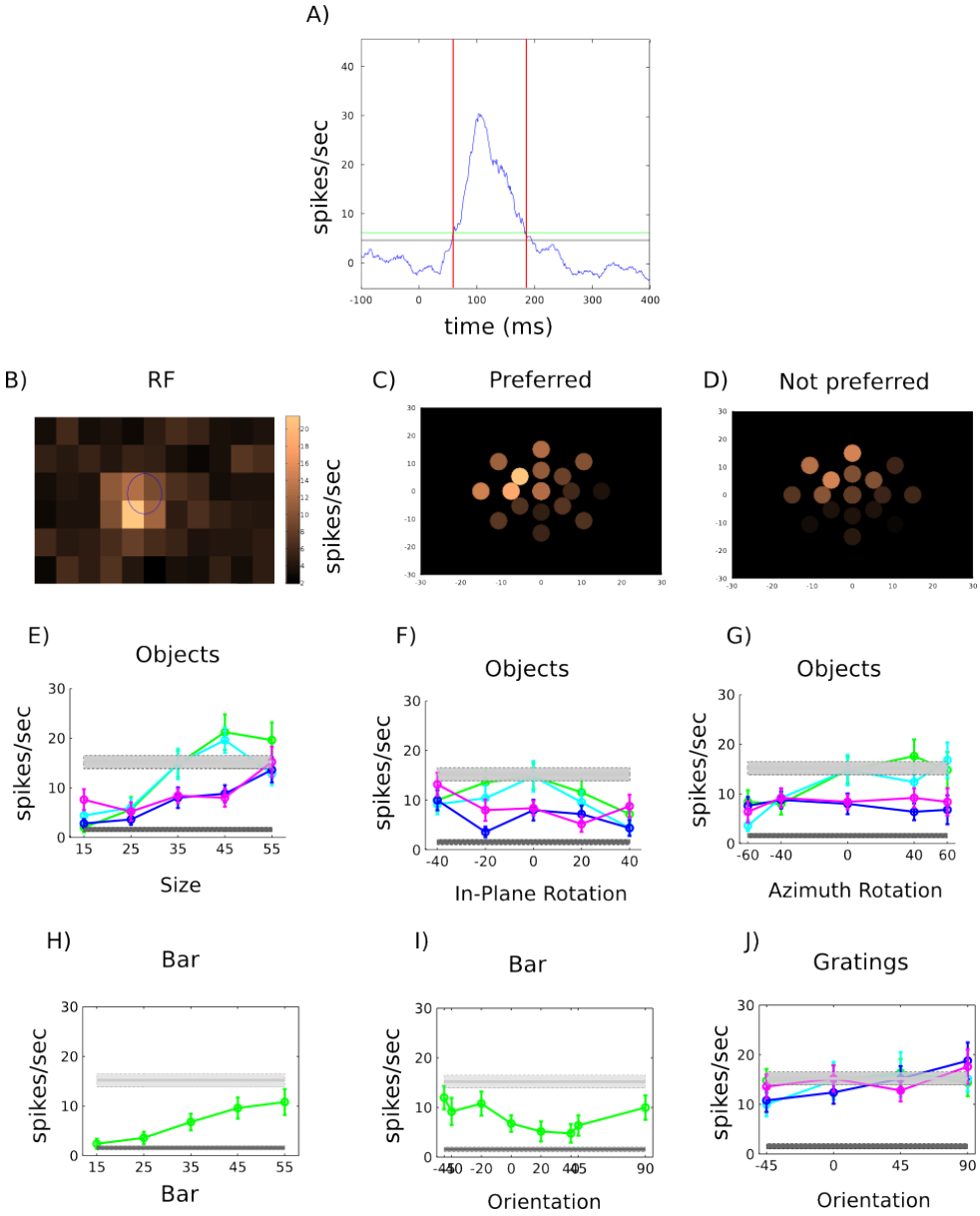


**Figure 43. A V1 neuron with small RF.**

**A)** Mean PSTH computed considering the neuron's mean response to the 10 best conditions. The two red lines outline the spike counting window. The green line marks the onset threshold. The black line marks the average background activity.

**B)** The neuron's RF. Size=15°.

**C)** Position tuning of the best object. Brighter dots mark spots of higher sensitivity towards the central/left side of the screen, consistently with RF position.

**D)** Position tuning of the worst object. Same convention as in C.

**E-G)** Tuning curves to objects undergoing size, orientation and azimuth rotation transformations. Green line=object 1; cyan line=object 2; blue line=object 3; magenta line=object 4. The neuron shows tuning for bigger object areas, and a preference for Object 1 and 2 is shown, which reflects in the tuning at 0° orientation and at the more positive values in the azimuth tuning. Error bars indicate SEM.

**H-I)** Tuning curves to bars undergoing size and orientation transformations. Consistently with object tuning, the neuron shows tuning for bigger bar areas. A slight tuning for -45˚, -20˚ and 90˚ orientations appears.

**J)** Tuning curves to oriented gratings of different spatial frequencies. Green line=0.03; cyan line=0.05; blue line=0.1; magenta line=0.4. A slight tuning for 90˚ orientation is shown.

In all tuning curves plots, the dark grey and light grey shaded lines refer to black and white blank mean response, respectively. The response to full field white blank condition is high, typical of V1 neurons (see Fig. 41).

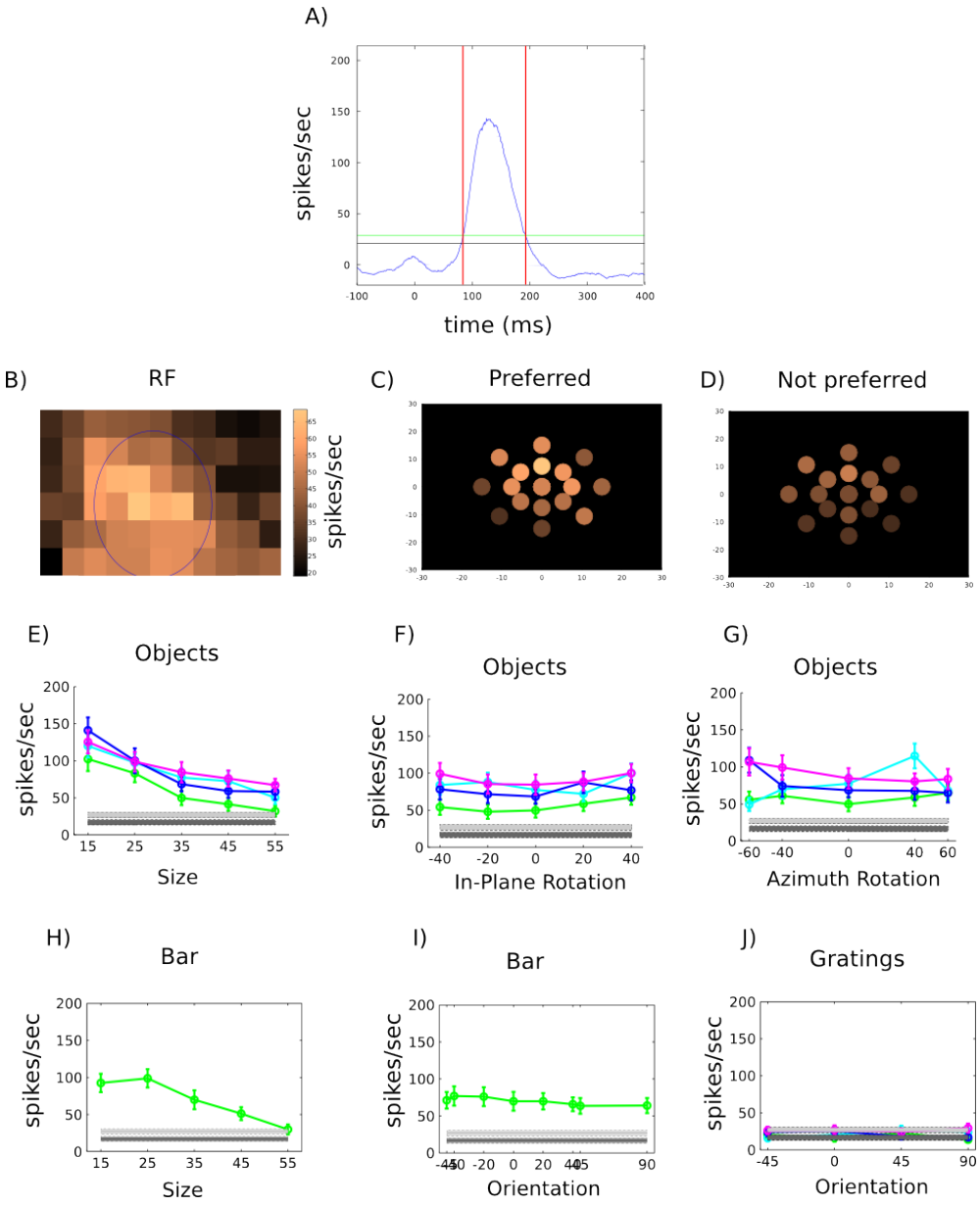


**Figure 44. A LL neuron with wide RF.**

**A)** Mean PSTH computed considering the neuron's mean response to the 10 best conditions. The two red lines outline the spike counting window. The green line marks the onset threshold. The

black line marks the average background activity.

**B)** The neuron's RF. Size=53.52˚.

**C)** Position tuning of the best object. Brighter dots mark spots of higher sensitivity at all positions.

**D)** Position tuning of the worst object. Same convention as in C.

**E-G)** Tuning curves to objects undergoing size, orientation and azimuth rotation transformations. Green line=object 1; cyan line=object 2; blue line=object 3; magenta line=object 4. The neuron shows tuning for smaller object areas, typical of temporalmost areas. No important orientation tuning. A slight tuning to 40˚ azimuth rotation for Object 2 is shown. Error bars indicate SEM.

**H-I)** Tuning curves to bars undergoing size and orientation transformations. Consistently with object tuning, the neuron shows tuning for smaller bar areas. No significant orientation tuning appears.

**J)** Tuning curves to oriented gratings of different spatial frequencies. Green line=0.03; cyan line=0.05; blue line=0.1; magenta line=0.4. No significant tuning appears.

In all tuning curves plots, the dark grey and light grey shaded lines refer to black and white blank mean response, respectively. The response to full field white blank condition is low, typical of the most temporal areas (see Fig. 41).

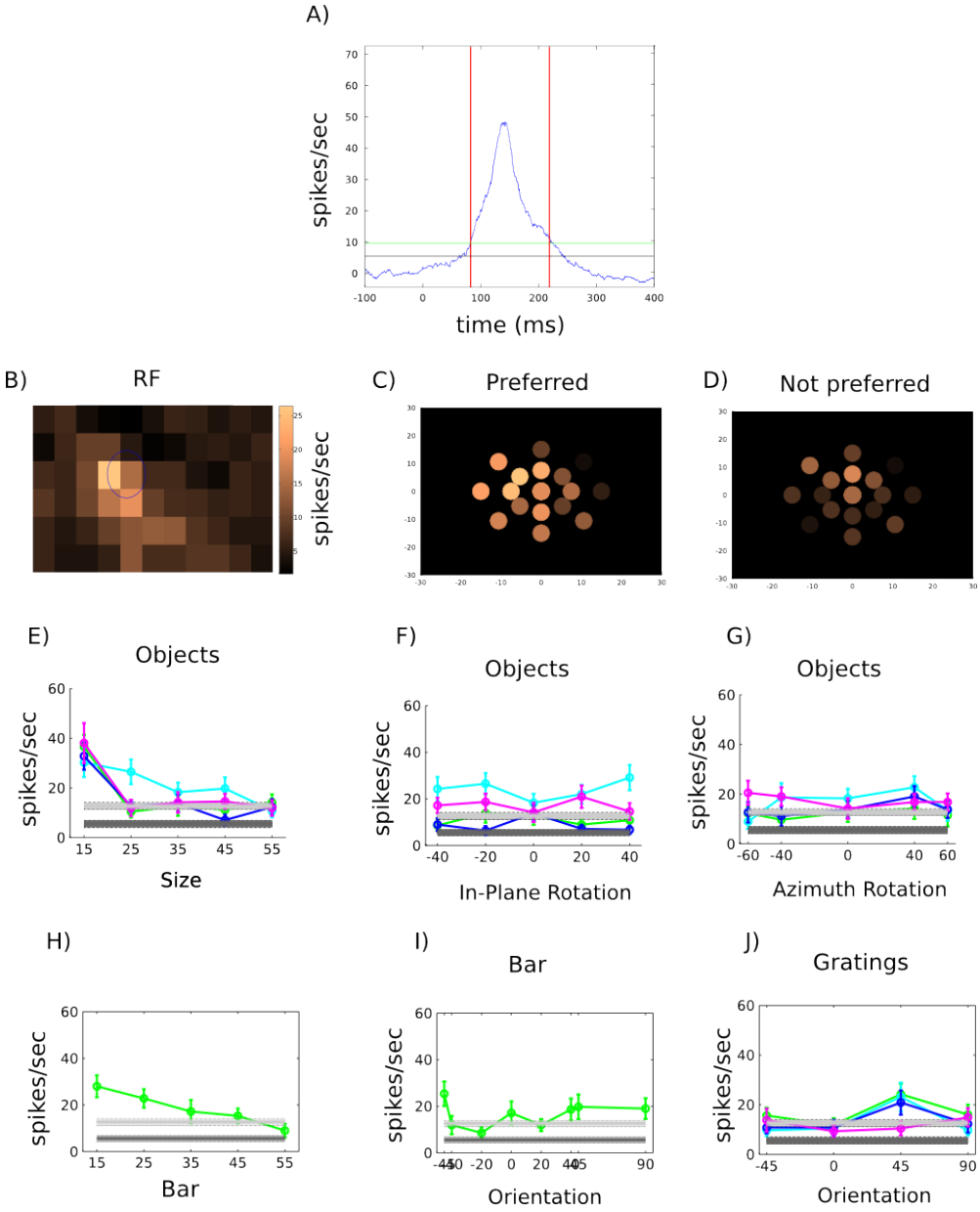


**Figure 45. A LL neuron with small RF.**

**A)** Mean PSTH computed considering the neuron's mean response to the 10 best conditions. The two red lines outline the spike counting window. The green line marks the onset threshold. The black line marks the average background activity.

**B)** The neuron's RF. Size=17.23˚.

**C)** Position tuning of the best object. Brighter dots mark spots of higher sensitivity towards the central/left part of the screen, consistent with RF position.

**D)** Position tuning of the worst object. Same convention as in C.

**E-G)** Tuning curves to objects undergoing size, orientation and azimuth rotation transformations. Green line=object 1; cyan line=object 2; blue line=object 3; magenta line=object 4. The neuron shows tuning for smaller object areas, typical of temporalmost areas. No significant orientation or azimuth rotation tuning. Error bars indicate SEM.

**H-I)** Tuning curves to bars undergoing size and orientation transformations. Consistently with

object tuning, the neuron shows tuning for smaller bar areas. An orientation tuning for -45˚ appears, probably due to the falling of the uppermost part of the bar into the RF.

**J)** Tuning curves to oriented gratings of different spatial frequencies. Green line=0.03; cyan line=0.05; blue line=0.1; magenta line=0.4. Slight tuning to 45˚ orientation appears for 3 different SF.

In all tuning curves plots, the dark grey and light grey shaded lines refer to black and white blank mean response, respectively. The response to full field white blank condition is low, typical of the most temporal areas (see Fig. 41).

### 4.4.1. Latency Profiles

Latencies per area were computed considering static conditions only (see Materials and Methods). Stimulus onset latencies increased along the visual hierarchy, from a mean of ~55 ms in V1 to a mean of ~80 ms in area LL and ~100 ms in area LLb (Fig. 46).

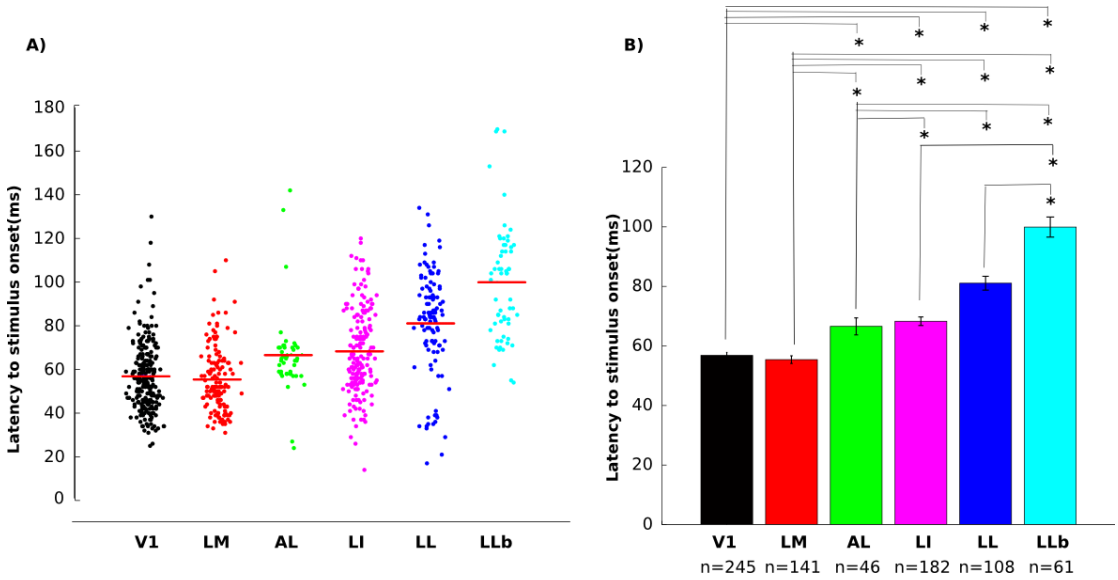


**Figure 46. Latencies to stimulus onset per area.**

**A)** Latencies distribution of each area. Red lines: mean value.

**B)** Average latency value for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

A correlation between RF size and latency was calculated for each area (Fig. 47). For each area we fitted a linear model and estimated the parameters of the fit by a leave-one-out bootstrap (red lines in Fig. 47). RF size and latency to stimulus onset negatively correlate at all instances but LM and AL, significantly in V1 and LI (Fig. 47B and C).

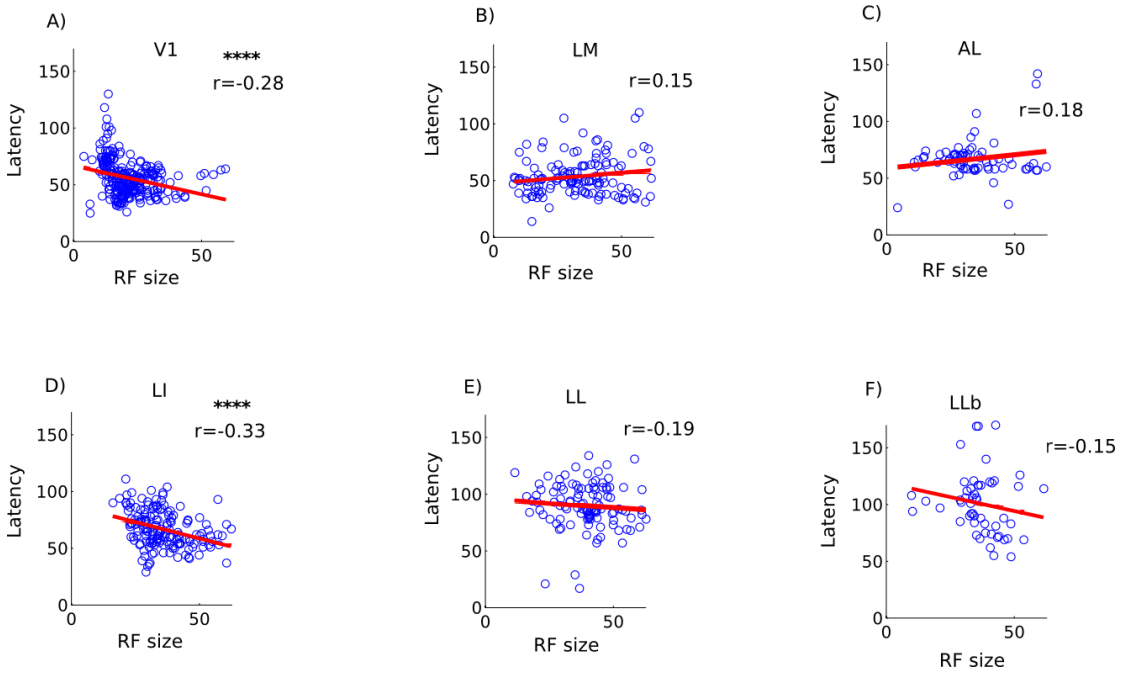


**Figure 47. Negative correlation between RF sizes and latencies.**

The correlation between RF size and latency to stimulus onset is depicted for each area, from V1 to LLb (A-F). Red lines represent the linear fit, whose parameters' variability was estimated by leave-one-out bootstrap. RF size and latency significantly trade off in V1 and LI. The correlation coefficients (r) for each area are reported.

**** = $p<0.0001$.

## 4.4.2. Firing Rate Profiles in Response Static Stimulation

Evoked firing rates were computed in response to all static stimuli presented as static conditions. Specifically, for each neuron, the peak of the mean responses to the 10 best static conditions was selected. We observed that area AL fired lower than all other areas, while all other areas appear to fire at equal magnitude to static stimulation (Fig. 48).

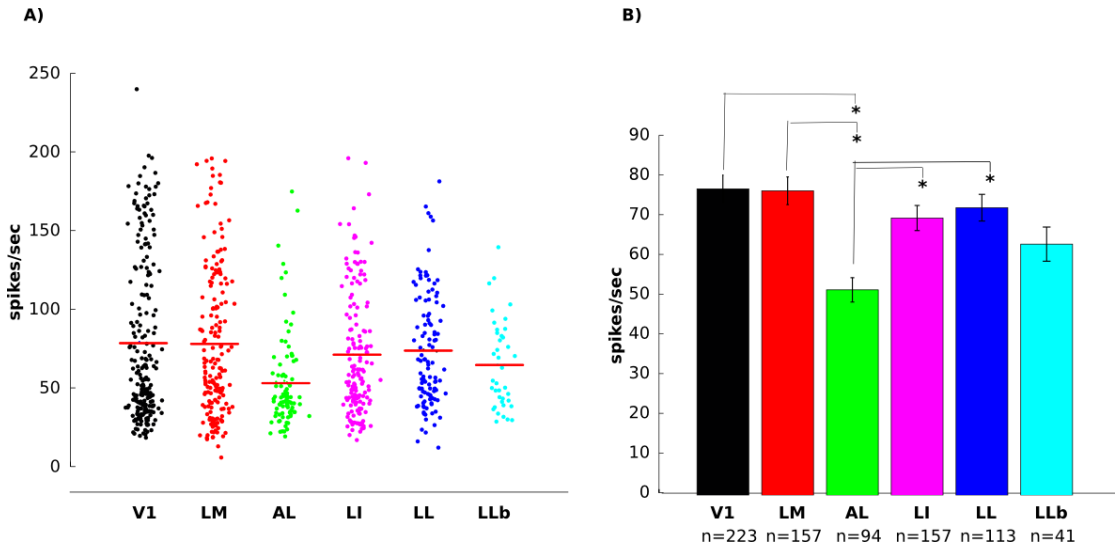


**Figure 48. Average firing rate profiles evoked by static stimulation.**

**A)** Evoked firing rate values distribution of each area in response to static stimulation. Red lines: mean value.

**B)** Average evoked firing rates for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

### 4.4.3. Shape Selectivity and Tolerance to Stimulus Transformations Increase along a Medial-Lateral Axis

Tolerance to various transformations in object appearance were calculated as described in Materials and Methods. Tolerance metrics were computed as a measure of how stable was the response of a neuron to its preferred object over size (size tolerance index, STI), in-plane rotation (OTI), azimuth rotation (ATI) and position (PTI) changes (Fig. 51-52). A mean index of tolerance, considering all object transformations together, was also computed (Fig. 49-50).

In general we found that tolerance increased significantly from the earlier stages of the putative rat ventral pathway (i.e., V1/LM/AL) to its later stages (i.e., LI, LL and LLb). This was observed for each of the tested transformation axes (see Fig. 51) and when all transformations were considered at once (see Fig. 49). Note that the analysis on position tuning yielded overall lower values as compared to the other indexes. This because position tuning indexes were computed by means of a concatenation of the four position axes, thereby averaging across a higher number of points as compared to the other tolerance metrics (see Material and Methods).

This analysis was also repeated after imposing a constraint on the maximal distance between the center of a RF and the center of the stimulus display. That is, only neurons with a

RF within 20º from the center of the stimulus display were considered. This was done to restrict the computation of the tolerance metrics only to neurons with a RF covering most of the tested object conditions, discarding those neurons with very peripheral RFs. This allowed comparing neuronal populations in different visual areas on a more equal basis, since only populations with similar coverage of the stimulus conditions were included. This minimized the possible impact of across-areas variations in the distributions of the RF centers (see Fig. 38) on the computation of the tolerance metrics. Note that, in the case of LLb, being the number of units much lower as compared to the other areas, and all the RF very large but very peripheral, a filter on the RF size was applied instead, to ensure that only RFs covering the stimulus conditions were included (only neurons with RF size larger than 30˚ were included). As shown by comparing Figure 48 to Figure 49 and Figure 51 to Figure 50, the increase of tolerance observed for the most temporal/lateral areas, compared to the more medial, was still observed (with the exception of position tolerance, which contemplated the use of RF spanning the whole visual field by construction; see Materials and Methods), when the neuronal populations were subsampled according to the criteria outlined above.

Not surprisingly, this increase in the tolerance of neuronal responses to variation in the size, location, etc. of their preferred objects correlated positively with the size of the neuronal RFs (Fig. 53). That is, on average, neurons with larger RFs tended to be more tolerant to variation in object appearance across a variety of transformation axes.

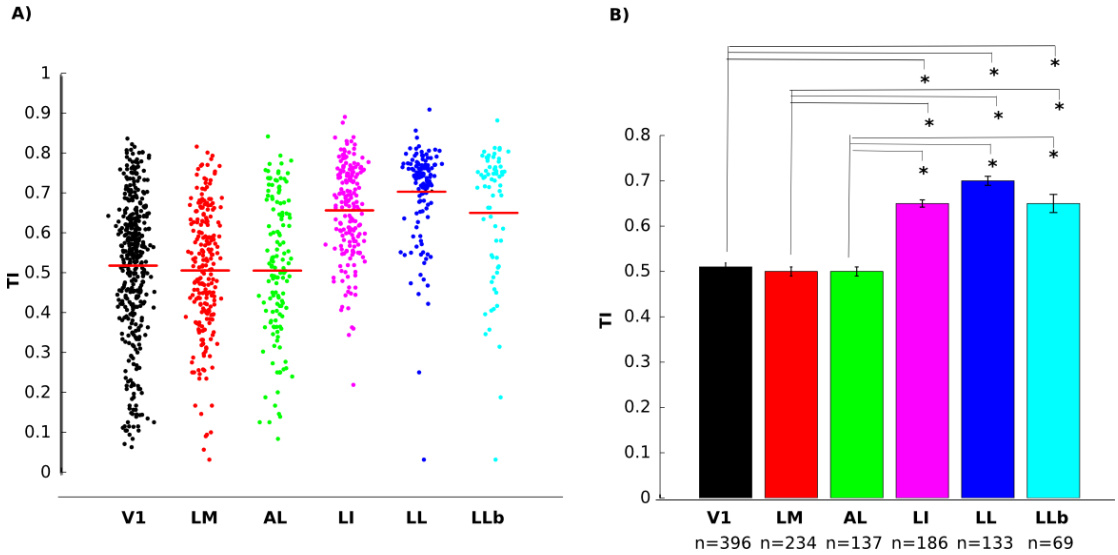


**Figure 49. Mean tolerance by area with no filter on RF positions.**

**A)** Distributions of mean tolerance indexes for each area. Red lines: mean value.

**B)** Mean tolerance for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

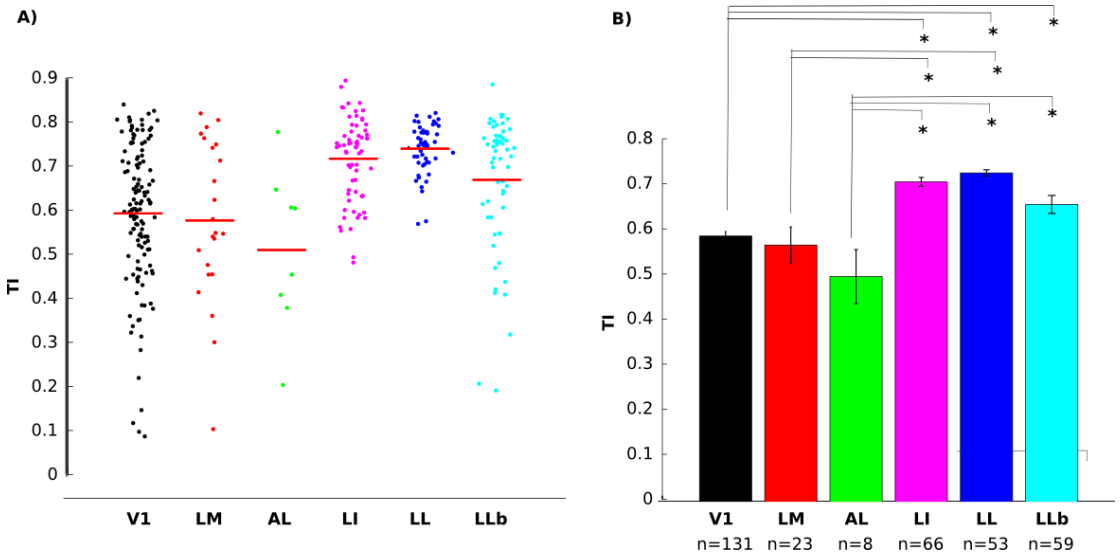


**Figure 50. Mean tolerance by area with filter on RF positions.**

See Fig. 49 for details.

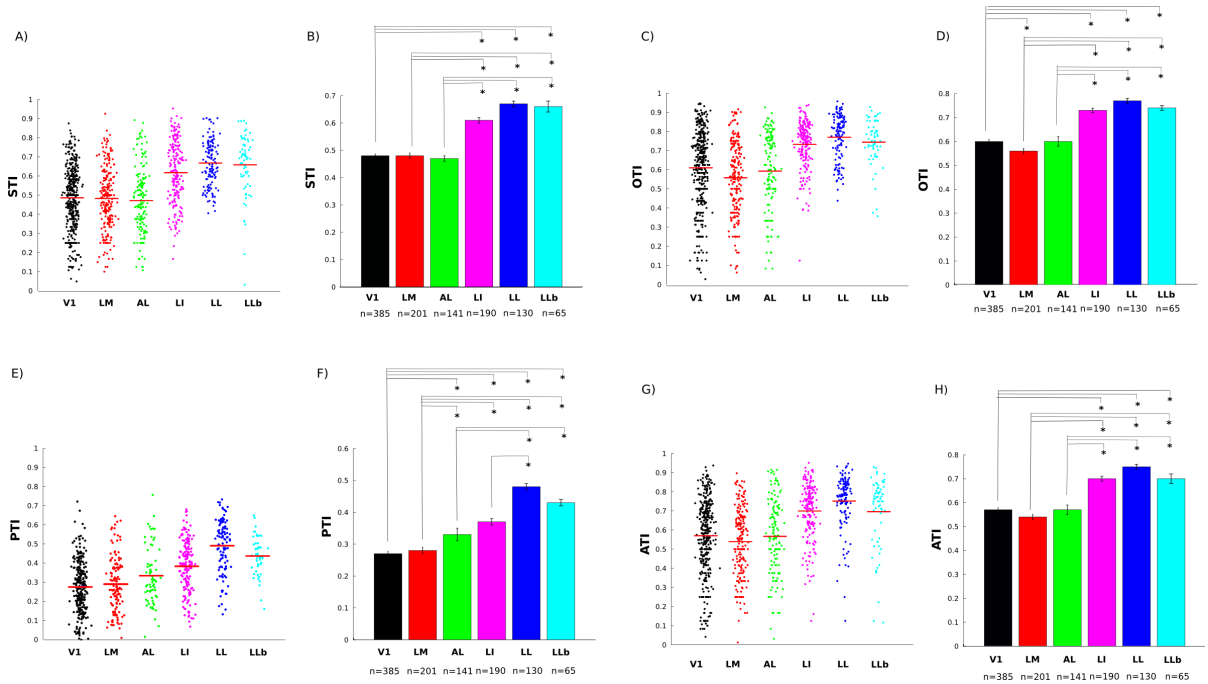


**Figure 51. Tolerances by area with no filter on RF positions.**

**A, C, E, G)** Distributions of indexes of tolerance to size, orientation, position and azimuth rotation for each area. Red lines: mean value.

**B, D, F, H)** Mean tolerance indexes for each transformation and each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

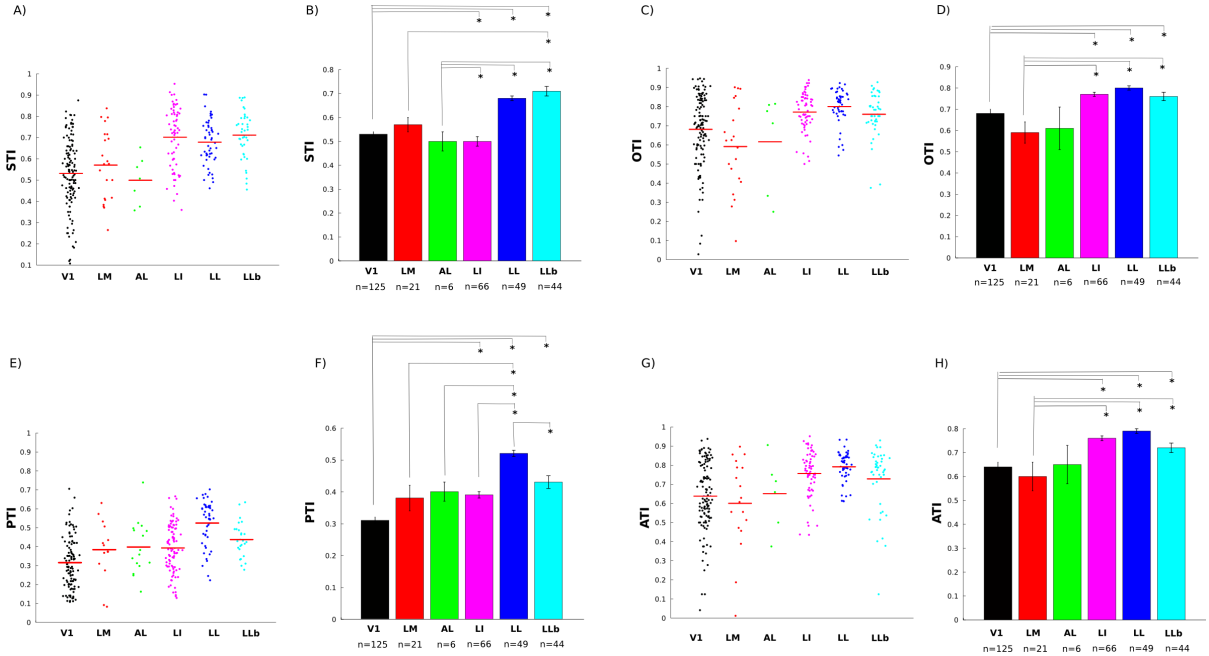


**Figure 52. Tolerances by area with filter on RF positions.**

See Fig. 51 for details.

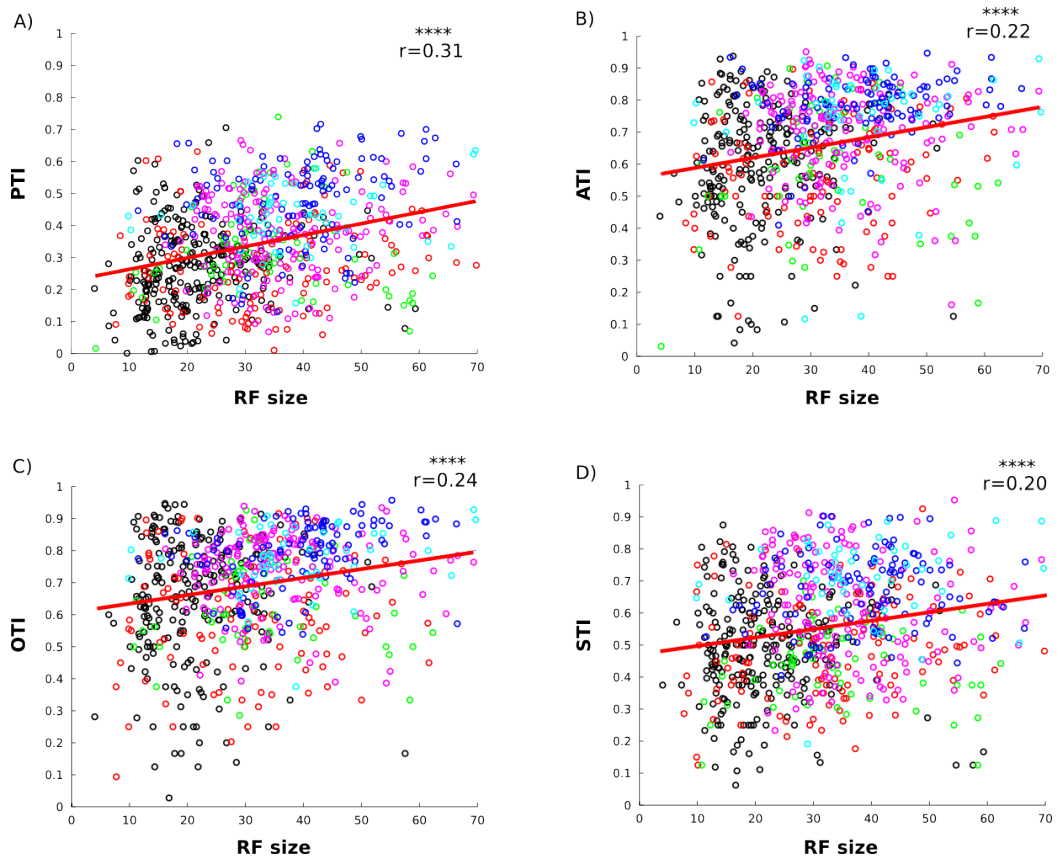


**Figure 53. Correlation between RF sizes and tolerance indexes.**

The correlation between RF size and position (A), azimuth rotation (B), orientation (C) and size (D) tolerance is depicted for all areas (see Fig. 52 for color code). Red lines represent the linear fit,

whose parameters' variability was estimated by leave-one-out bootstrap. A positive correlation between RF size and tolerance to all transformations tested is shown. The correlation coefficients (r) for each area are reported.

**** = p<0.001.

In addition to tolerance, we also compared how selectivity for visual objects changed across rat visual areas. Object selectivity was computed as described in the Materials and Methods, i.e., by measuring the d' for all object pairs at each tested position and taking the maximum. This provides an estimate of the maximal discriminability power of each neuron under the tested shape/position conditions. As in the case of tolerance, we also observed a significant V1-to-LL gain in selectivity (Fig. 54).

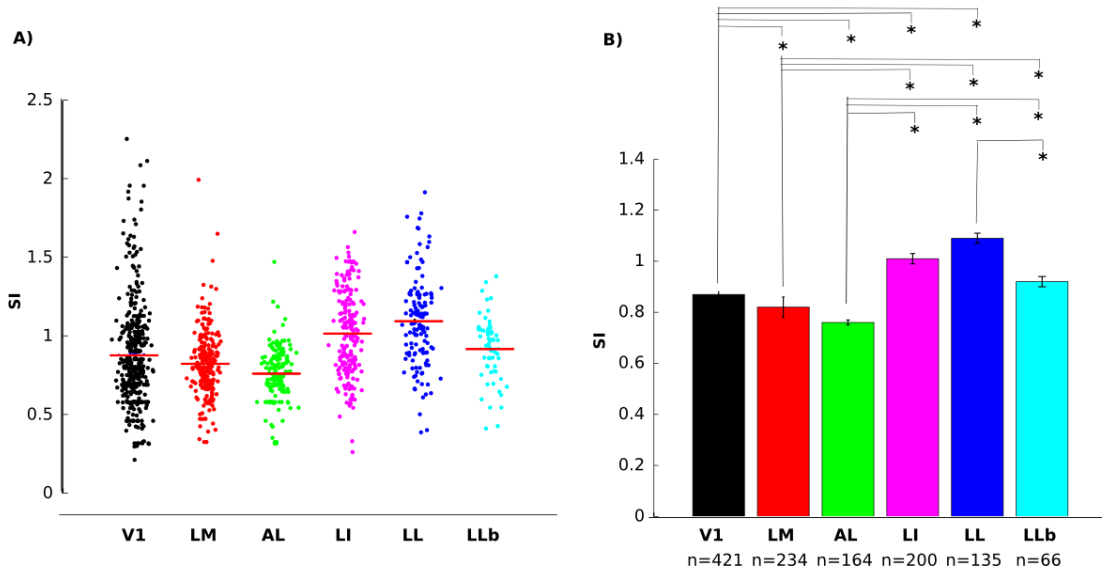


**Figure 54. Object selectivity by area.**

A) Distributions of selectivity indexes for each area. Red lines: mean value.

B) Mean selectivity for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

Finally, we checked if the observed increase in shape/object selectivity could be accounted by a parallel increase in orientation tuning. The response to static gratings of 4 different orientations were used to compute the orientation tuning indexes (OSI) for each area. Results show a V1-to-LL increase in orientation selectivity, with LL being significantly more selective as compared to V1 and LM (Fig. 55).

In summary, the analysis of the static object conditions revealed a concomitant increase of both selectivity for visual objects/shapes and tolerance to variation in their appearance,

which is consistent with the possible role of LI, LL and LLb as terminal stages of a rat shape processing pathway.

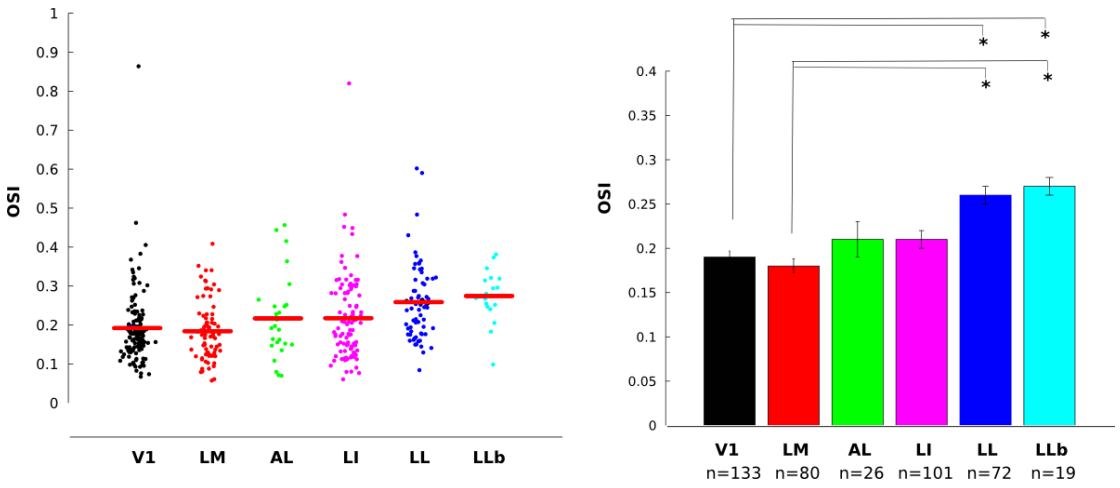


**Figure 55. Orientation tuning by area.**

**A)** Distributions of orientation indexes for each area. Red lines: mean value.

**B)** Mean OTI for each area. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

## 4.5. Motion Tuning

As mentioned before, a variety of moving stimuli were used to estimate the tuning properties of neurons recorded across rat visual areas for motion direction and velocity. These stimuli (see Fig. 32A) included both full-field patterns (such as gratings and random dot fields) and localized shapes (i.e., object sand bars). When it comes to the latter, it is important to point out a potentially dangerous confound in quantifying a neuron's direction tuning: the position of its RF (see Fig. 56L). This is illustrated in Fig. 56, for an example V1 neuron with a very peripheral RF. Since, in our experiments, many neurons were simultaneously recorded, it was not possible to "center" the motion, as well as the static, stimulus conditions on the RF center of each neuron. Rather, all stimulus conditions were displayed in the same (absolute) visual field locations for each neuron (i.e., relative to the center of the stimulus display). This means that a neuron with a very peripheral RF would be unable to respond to many directions of motion (when tested with localized stimuli such as object and bars), simply because a stimulus moving along these directions would not enter its RF (see Fig. 55L). This would result in an apparent sharp tuning for motion direction, when tested with object and bars (see

Figs. 55B-E), even if the neuron is not at all tuned for motion direction, as revealed by the full field grating conditions (see Figs. 55F-G).

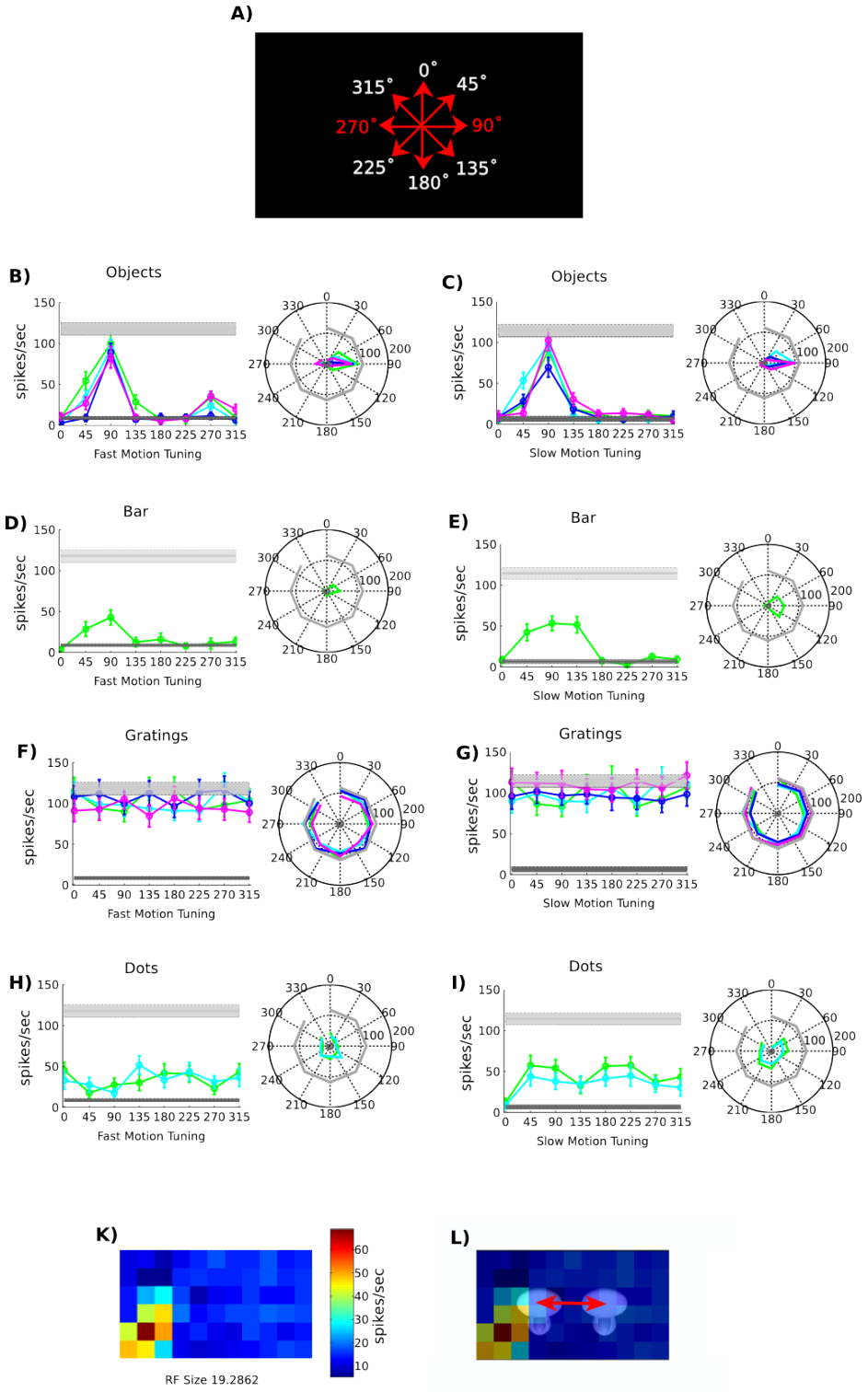


**Figure 56. A V1 neuron with 'fake' direction tuning.**

**A)** Conventional values for the 8 directions of motion used (marked in red).

**B, C)** Direction tuning for Objects moving at fast and slow motion, respectively. All objects appear tuned for direction towards 90˚, but not the orthogonal one (270˚), suggesting direction

tuning.

**D, E)** Moving bars tuning in fast and slow motion, respectively. The tuning reflects Object tuning.

**F, G)** Flat motion tuning curves for gratings moving at slow and fast motion, respectively.

**H, I)** Flat motion tuning curves for random dots moving at slow and fast motion, respectively. The two line colors refer to the two patterns used.

**K)** The neuron's RF is 19.28° wide and covers the lower-left part of the screen.

**L)** The objects' left border touches the RF at its starting and finishing position, but the neuron fires only at the objects' starting position.

As indicated by the light grey line in B-I, the response to white full-field blank is high, typical of V1 (see Fig. 41). Together with the observation on the starting position on the moving object, this suggest that the neuron might respond to abrupt changes in contrast falling into the whole RF or part of it.

All tuning curves are displayed as both carthesian axes and polar plots.

Such an apparent motion tuning would be exacerbated by the fact that most neurons respond much more strongly to abrupt changes in luminosity (i.e., when a stimulus suddenly appears within a neuron's RF) than to gradual changes (e.g., when a stimulus moves into a neuron's RF), likely because of some adaptation mechanisms. Therefore a neuron would typically fire more strongly for a stimulus departing from a position close to the center of its RF than for a stimulus approaching that position. This would result in an apparently large difference in the response of the neuron to orthogonal motion directions (compare the peaks with opposite directions in Fig.56B-E). In order to avoid such artifacts in the analysis of motion tuning, we restricted the computation of direction selectivity to full-field conditions only (i.e. gratings and dots; see Figs. 56F-I).

### 4.5.1. Firing Rate in Response to Moving Stimuli

Similarly to what we did for firing rates evoked by static stimulation, we selected the peak of the response of each neuron to the 10 best stimuli presented as fast and slow moving conditions.

Neuronal firing rates in response to the fast moving conditions decreased in the most temporal areas (i.e., LI, LL), as compared to the more medial areas (i.e., V1, and LM; see Fig. 57). This marks a very different pattern, compared to what observed for firing rates in response to static conditions (shown in Fig. 48 and also reported, for better comparison, in the inset of Fig. 57), where no significant difference between LI/LL and V1/LM was found. With slow moving stimuli, the evoked firing activity in area LI was similar to the one observed for

V1/LM (and similar to what found for static conditions), while the activity in area LL was still significantly lower than what observed in the more medial areas (i.e., V1, LM and LI) and a similar trend was found for LLb (see Fig. 58, and compare it to the inset).

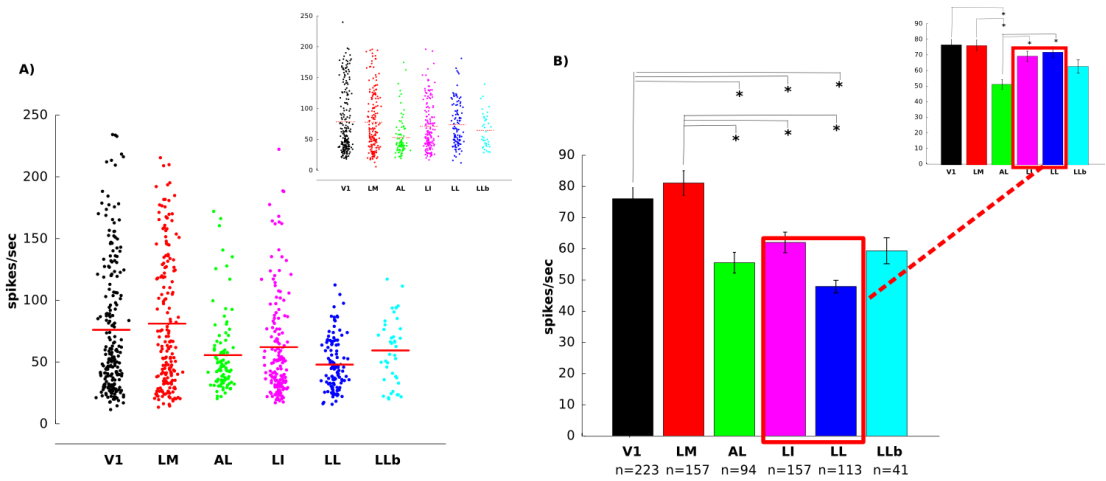


**Figure 57. Average firing rate profiles evoked by fast moving stimulation.**

**A)** Evoked firing rate values distribution of each area in response to fast moving stimulation. Red lines: mean value. The evoked firing rate distribution for static stimulation is shown in inset.

**B)** Mean firing rates for each area. The areas which display a decrease in their firing as compared to static stimulation are marked in red frames, and the mean firing rate distribution for static stimulation is shown in inset.

The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

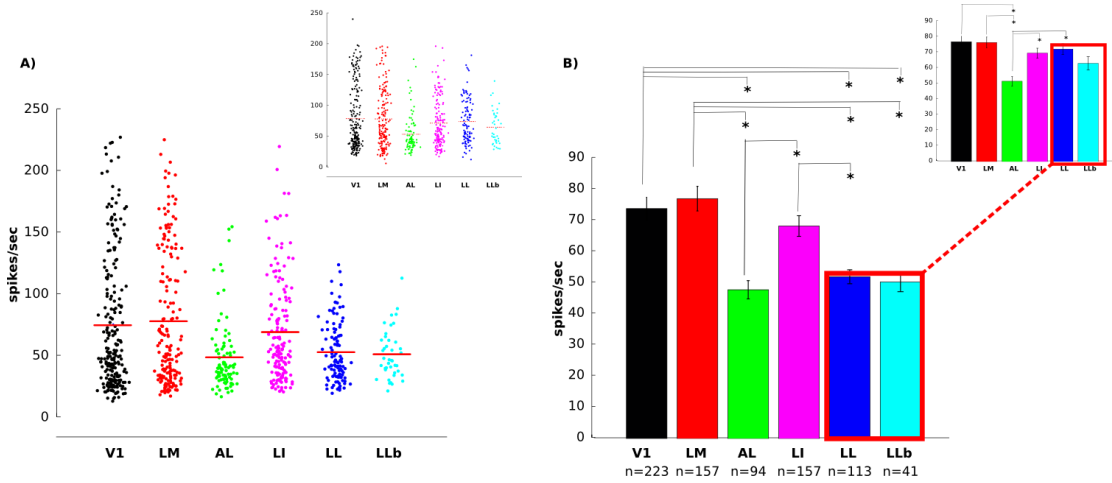


**Figure 58. Average firing rate profiles evoked by slow moving stimulation.**

**A)** Evoked firing rate values distribution of each area in response to slow moving stimulation. Red lines: mean value. The evoked firing rate distribution for static stimulation is shown in inset.

**B)** Mean firing rates for each area. The areas which sensibly decreases their firing as compared to static stimulation are marked in red frames, and the mean firing rate distribution for static

stimulation is shown in inset.

The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

When restricting the analysis to drifting gratings only, the same tendency was observed (see Fig. 59): the magnitude of the response evoked by drifting gratings was lower in areas LI and LL, again as compared to V1/LM and as compared to areas LI and LL themselves when tested with static stimuli (compare to the insets).

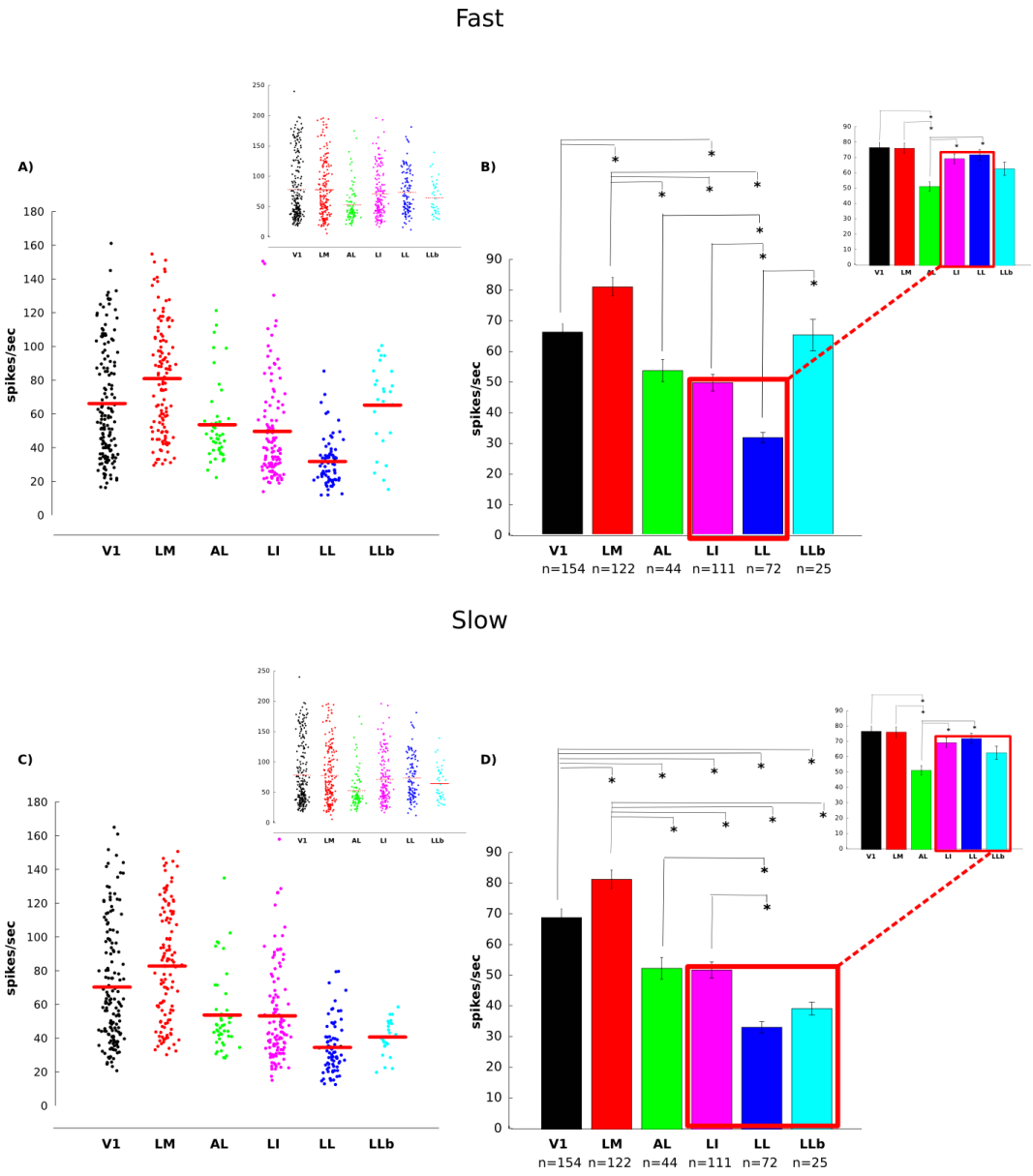


**Figure 59. Average firing rate profiles evoked by fast and slow drifting gratings.**

**A, C)** Evoked firing rate values distribution of each area in response to fast and slow drifting, respectively. Red lines: mean value. The evoked firing rate distribution for static stimulation is

shown in inset.

**B, D)** Mean firing rates for each area in response to fast and slow drifting gratings, respectively. The areas which sensibly decrease their firing as compared to static stimulation are marked in red frames. The mean firing rate distribution for static stimulation is shown in inset for comparison. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

When restricting the analysis to moving dots only (see Fig. 60), the evoked response magnitude of all areas was lower, as compared to what observed with gratings (see Fig. 59) and static stimulation (see insets in Fig. 60). Also, area LI was slightly more responsive to moving dots as compared to other areas in slow conditions (Fig. 60C and D).

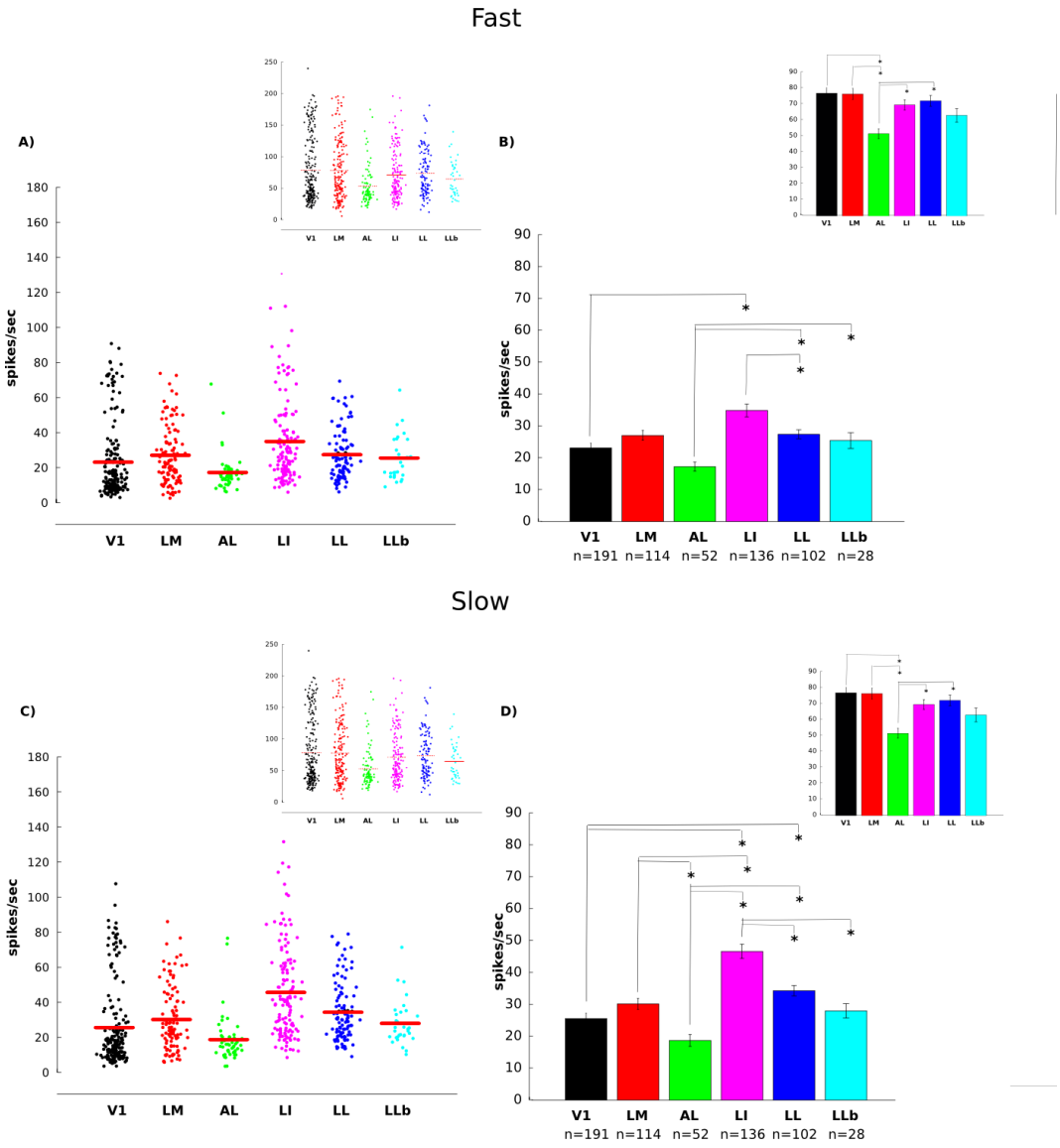


**Figure 60. Average firing rate profiles evoked by fast and slow moving random dots.**

**A, C)** Evoked firing rate values distribution of each area in response to fast and slow moving dots,

respectively. Red lines: mean value. The evoked firing rate distribution for static stimulation is shown in inset.

**B, D)** Mean firing rates for each area in response to fast and slow moving dots, respectively. The areas which display a decrease in their firing as compared to static stimulation are marked in red frames. The mean firing rate distribution for static stimulation is shown in inset for comparison. The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

Thus, we checked for a possible correlation between RF size and evoked response magnitude to moving random dot fields, as it has been reported for motion-driven areas like MT (see for example Mikami et al., 1986). Indeed, we found positive correlations, in both fast (Fig. 61) and slow (Fig. 62) conditions for all areas. The same correlation was found in the case of drifting gratings, but significant only in V1 (r=0.30, p<0.001) for fast conditions, and in V1 and LI (r=0.32, p<0.001; r=0.20, p<0.05, respectively) for slow conditions (data not shown).

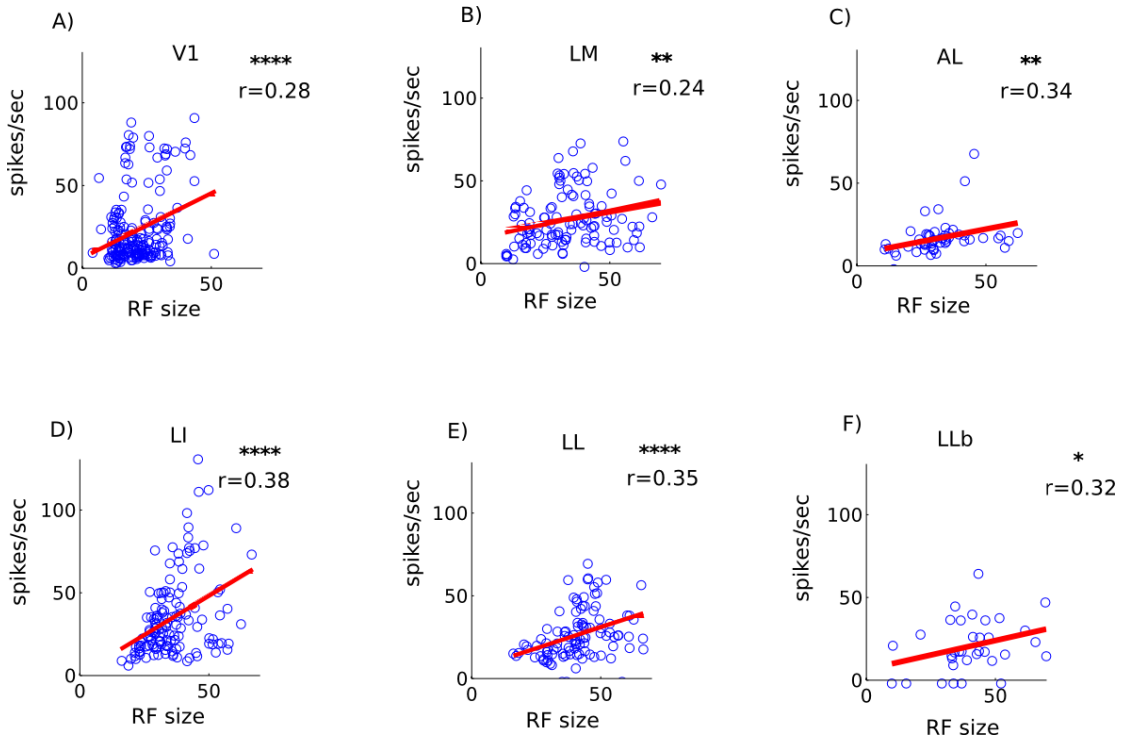


**Figure 61. Correlation between RF size and responsiveness to fast moving dots.**

The correlation between RF size and responsiveness to fast moving dots is depicted for each area, from V1 to LLb (A-F). Red lines represent the linear fit, whose parameters' variability was estimated by leave-one-out bootstrap. The correlation coefficients (r) for each area are reported. Selectivity and tolerance significantly correlate in all areas, but trade off in area LI.

**** = p<0.0001, *** = p<0.001, ** = p<0.01, * = p<0.05

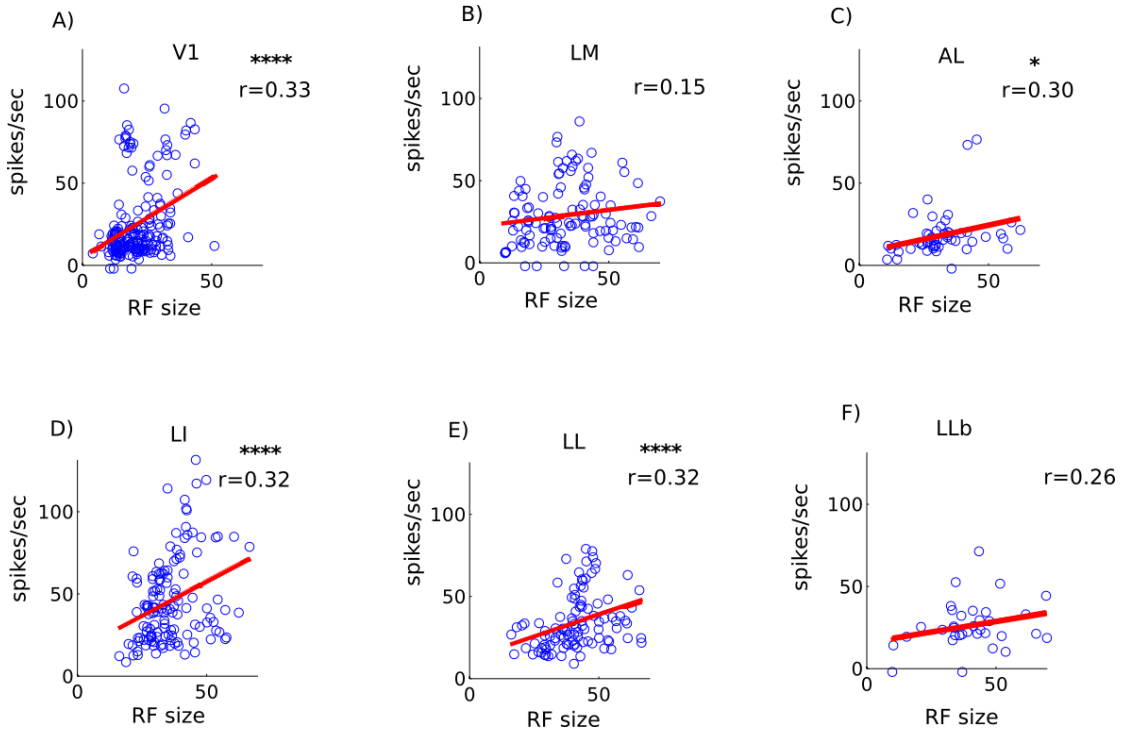


**Figure 62. Correlation between RF size and responsiveness to slow moving dots.**

See Fig. 61 for details.

## 4.5.2. Direction Tuning

Direction selectivity was computed as described in Materials and Methods for each tested visual area. First, in order to investigate the areas' responsiveness to motion, limiting the influence of shape information to the responsiveness, we focused on pure motion stimuli, i.e. moving dot fields. Since this kind of stimulation produced a generally lower firing rate in all areas (see Fig. 60), we considered only units displaying a minimum firing rate of 10 spikes/s. As shown in Fig. 63, the distribution of the Direction Selectivity Index (DSI) showed a decreasing trend towards the most temporal areas, in both fast and slow conditions, suggesting a V1-to-LL/LLb loss of direction selectivity.

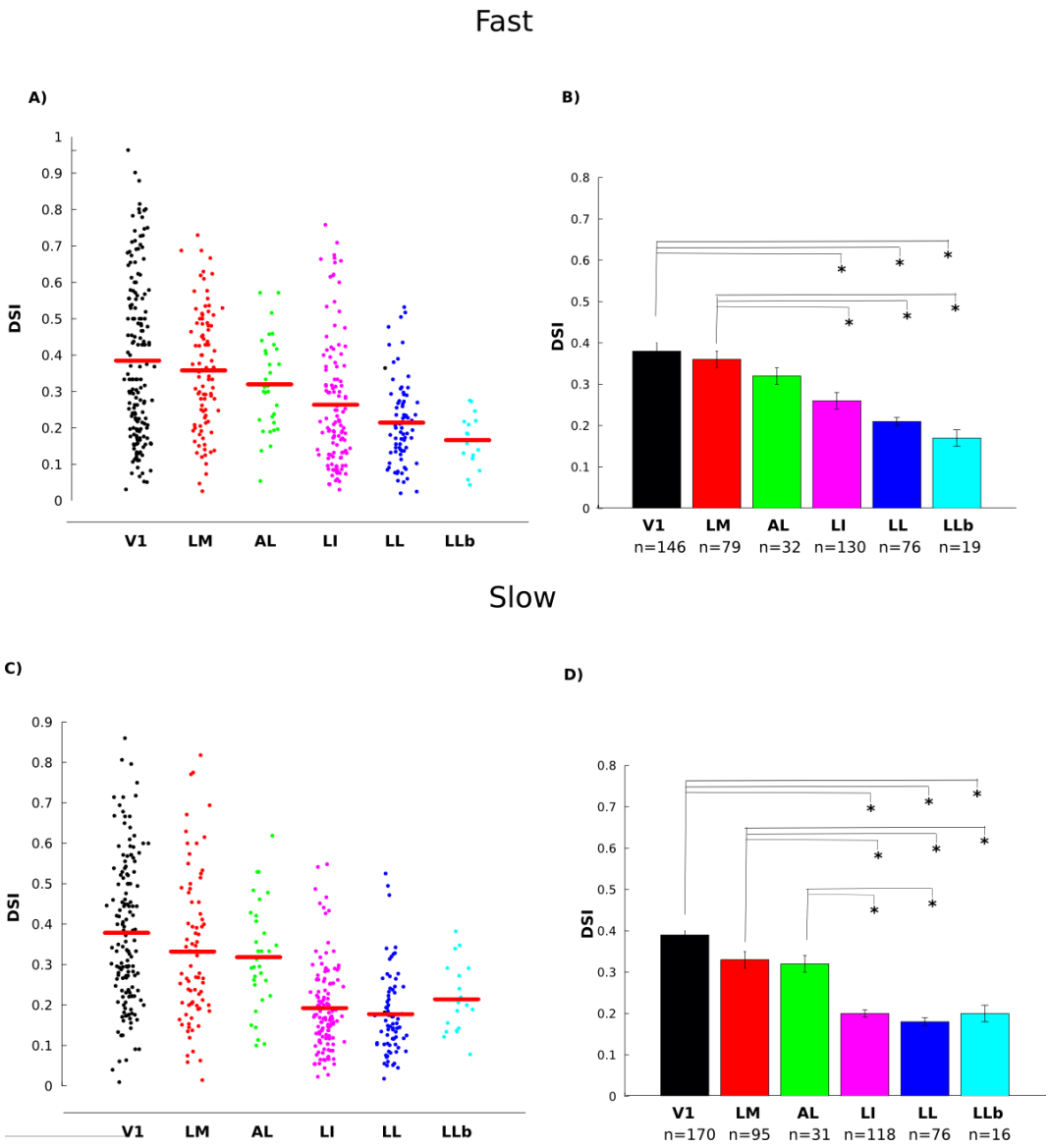


**Figure 63. Direction tuning indexes in different visual areas calculated with moving dots.**

**A, C)** DSI distribution of each area as calculated for fast and slow moving dots, respectively. Red lines: mean value.

**B, D)** Mean DSI for each area as calculated with fast and slow moving dots, respectively.

The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

When analyzing direction tuning profiles using drifting grating stimulation, a different trend appeared: temporal areas like LI and LL showed a higher direction selectivity than more medial, although this difference was significant only for the slow conditions (Fig. 64), with the exception of LLb, for which direction tuning was larger with fast than slow moving gratings.

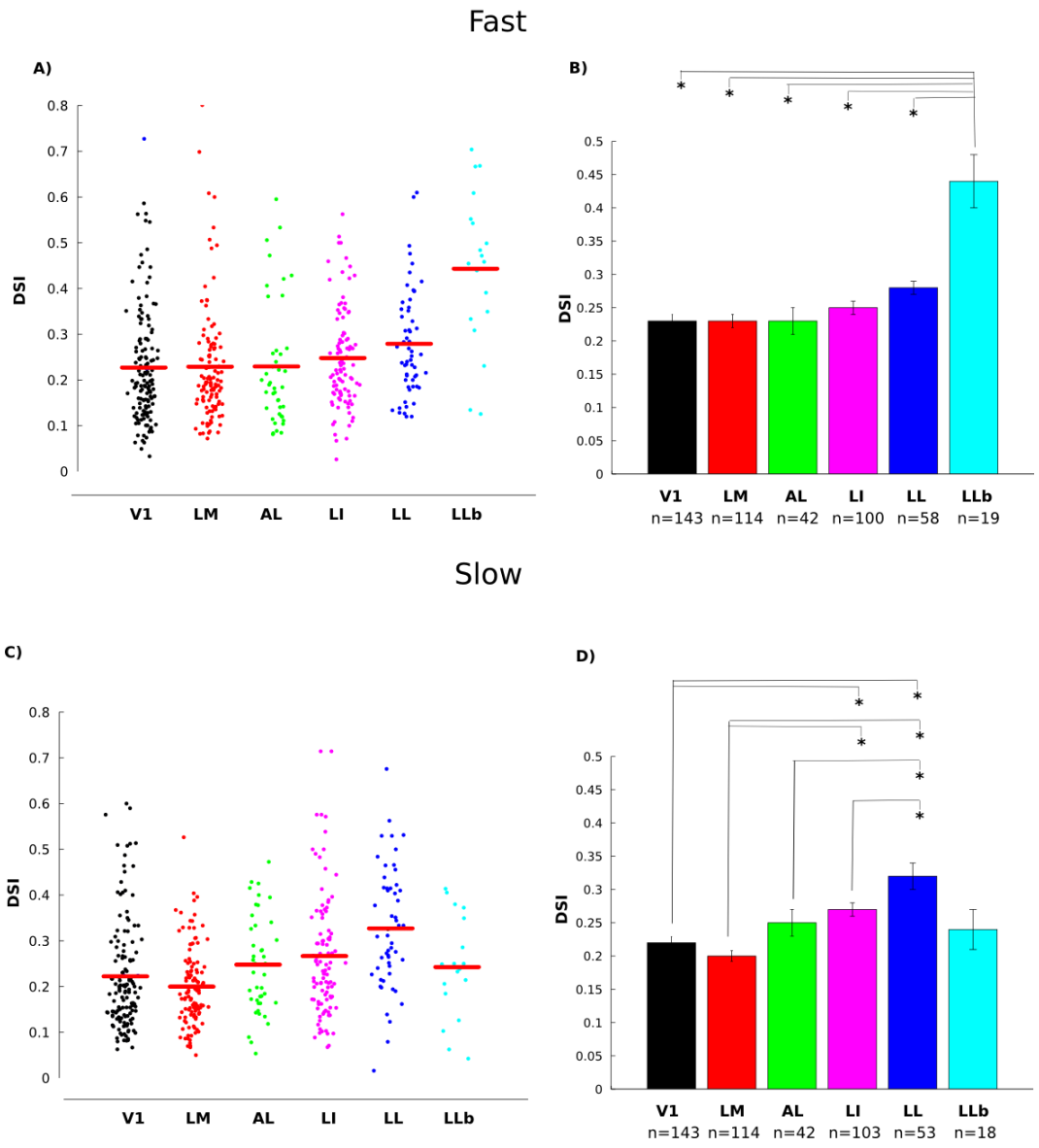


**Figure 64. Direction tuning indexes in different visual areas calculated with drifting gratings.**

**A, C)** DSI distribution of each area as calculated for fast and slow drifting gratings, respectively. Red lines: mean value.

**B, D)** Mean DSI for each area as calculated with fast and slow drifting gratings, respectively.

The number of units of each area is reported as x labels. Errorbars: SEM. *=p<0.05, one-way ANOVA.

# 5. Discussion

Behavioral investigation of rats' ability to process shape and motion information in independent ways yielded positive results, suggesting the presence of discrete visual networks for the segregate processing of motion and shape information. Further investigation of such ability might take into account previously documented task-switching capacity of these animals (see for example Floresco et al., 2008) in order to shed light on the possible mechanisms of task-dependent attentional modulation of the same visual input, as it has been shown in the monkey's area V4 (see for example Mirabella et al., 2007).

In line with previous reports (Girman et al., 1999) we found retinotopically organized areas in the rat visual cortex, consistent with the presence of a hierarchy, which is not only anatomical (Coogan and Burkhalter, 1993) but also functional.

It is important to note that, due to our method of probe insertion (i.e. oblique penetrations spanning different areas), different layers from different areas were likely being recorded from. This might lead to confounding effects in all functional properties tested under this design. Fine-scale histology methods must, and will be applied in order to reconstruct the exact location of each recording site, thereby allowing a more robust across-areas comparison between neuronal properties belonging to the same cortical layer.

The observation of an increase in response latency, together with a V1-to-LL/LLb gain in tolerance (invariance) to object transformations and object selectivity, suggest a putative homology of the areas investigated with the primate ventral stream of processing (see for example Rousselet et al., 2004). Tolerance gain appears to be accounted for by an increase in RF size (which positively correlates with all tolerance metrics tested), affording a decrease in the sensitivity to object transformations.

The picture emerging from the analysis on motion is harder to interpret.

Areas that appear to be more object selective appear to be the least responsive to motion, as demonstrated by both the decrease in the firing rate in response to moving stimulation and the narrowing of the direction tuning indexes as calculated with random dots. Moreover, consistently with previous reports (Albright and Desimone, 1987; Maunsell and Van Essen, 1987; Mikami et al., 1986), RF size correlates positively with motion responsiveness, suggesting that motion information is processed by neurons with big RF size.

However, the DSIs calculated with random dots stimulation do not match with the ones calculated with drifting gratings: in the latter case, for slow moving conditions, an increasing trend is shown from V1 to the most temporal areas. As a speculation, this could be due to an

interference of shape information into the processing of motion. More generally, differences between direction tuning values observed for gratings and dots could be due to the relatively small number of tested directions/orientations in our study. An enrichment/adjustment of the stimulus set will likely be necessary for future recordings, in order to fine-tune the full-field conditions for an optimal quantification of direction tuning.

Further investigation of the response properties to dynamic stimuli needs to be performed for a thorough characterization of a cortical hierarchy in motion processing, e.g., in terms of: i) integration of different stimulus velocities (for which mammalian V1 displays 'paradoxical' features of non-linearity, see for example Dean et al., 1982; Orban et al., 1986); ii) end-stopped properties in rat's V1 (see, for example, Pack and Born, 2001); iii) higher-level integration of motion patterns in areas AM and PM (proposed as the putative 'gateways' for a dorsal stream in the mouse brain by Marshel et al., 2011; Wang et al., 2011).

# 6. Conclusions

RF mapping of 6 cortical areas in the rat occipital and temporal cortex confirmed their previously demonstrated retinotopic organization. The increase in RF size and latency from one area to the next supports the hypothesis of a hierarchical organization of rat visual cortex, which functionally culminates in higher tolerance and selectivity properties in LI, LL and LLb.

The late stages of such hierarchy decrease their firing when facing moving stimulation, suggesting a more shape-based processing strategy as compared to the previous stages, and a possible role of these areas as late integration steps in a ventral-like visual stream.

Analysis with moving stimuli did not reveal a similar progress in the integration of motion information, suggesting a more parallel, widespread network as compared to the one found for shape processing.

Future perspectives will involve analysis of so far unexplored areas, i.e. the medial ones (AM and PM) to investigate a possible role of these sites in motion processing, as it has been suggested by some reports in mice.

Also, we plan to use different methods for robust assessing of the presence of a split visual network, from immunohistochemistry techniques to chronic recordings in awake and trained rats, implanted in different areas. This will also allow us to assess the weight of experience and awakeness on the quality of electrophysiological data.

Thus far, our results support the use of the rat as a good model to study invariant object

recognition, and strengthens the hypothesis of an homology between the area investigated in the rat and primates ventral visual stream.

# Conclusions

*'We have to remember that what we observe*
*is not nature herself, but nature*
*exposed to our method of questioning.'*
*W. K. Heisenberg*

We have demonstrated that rats hold a finely functioning visual system, allowing these animals to successfully perform highly engaging visual tasks, from invariantly recognizing objects to tell shape from motion information and vice versa.

We must remember that rats do not preferentially explore the environment using the visual modality (Cox, 2014). However, their potential as exploiters of visual information comes to light when tested under highly motivating (reward-related) conditions. Rats that are exposed to visual information, and are 'forced' to represent such information in order to generate successful behavior, will reveal such a potential. Also, the more the visual inputs can be segregated in the phenomenological space, the stronger such a potential will be expressed.

Our study over the neural underpinnings for such abilities reveals, for the areas investigated, the presence of a functional hierarchical organization in the rat visual cortex, consistent with the possibility of shape information processing at different integration stages, where the properties of object selectivity and tolerance to object transformations become more and more prominent up in the stream. These properties suggest the presence of a neural homologues of the primate ventral-like stream in the rat cortex (see for example Rousselet et al., 2004).

The analysis of moving patterns yielded results which are harder to interpret: the inherent limitations of our approach made it difficult to coherently extract the fine tuning of neurons in response to motion. However, the analysis on random dots shows that those areas that are more responsive to shape are the least responsive to pure motion, and the correlation found between RF size and motion responsiveness is in line with previous reports on the primate dorsal stream (Albright and Desimone, 1987; Maunsell and Van Essen, 1987; Mikami et al., 1986).

We must consider that when investigating the neural substrates of visual abilities in rats, through extracellular recordings techniques, the scenario we are facing is dramatically

different from a behavioral approach.

First, the method we are using, i.e. extracellular electrophysiological recordings, is inherently underconstrained: we collected the activity of many different neurons at one time, from different areas which have not been properly characterized in the rat (V1 aside, see Girman et al., 1999). We are basically moving over an unexplored territory, where most likely the only map we can follow is retinotopic organization (Bonin et al., 2011).

Second, it is not trained rats we are dealing with, but naïve subjects. Naïve rats, meaning rats that didn't undergo a visual training, likely have a lower experience with visual statistics. Before considering this fact as a limitation, recordings on trained (either awake or anaesthetized) rats should be performed and used as supporting argument.

However, it is not illogical to believe that fundamental experience/plasticity-driven differences might arise between the two cases. It has been shown, for example, that even visual acuity, which we could refer to as a 'hard-wired' function, is subjected to the mechanisms of experience-dependent visual plasticity (Prusky et al., 2000b).

Also, it wouldn't be implausible to find differences between the optimal stimulus parameters for behavioral and electrophysiological testing (e.g. rats best discrimination of spatial frequencies peaks at 0.22 behaviorally (Meier et al., 2011), and at 0.08 from electrophysiological investigation of V1 (Girman et al., 1999)).

All in all, given the increasing consensus over the suitability of rats to serve as robust experimental models in vision research, the number of methods to be exploited is high, and so are the cognitive and neural mechanisms still to be explored.

# References


Adams, A.D., and Forrester, J.M. (1968). The projection of the rat's visual field on the cerebral cortex. Q. J. Exp. Physiol. Cogn. Med. Sci. *53*, 327–336.

Adelson, E.H., and Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. Nature *300*, 523–525.

Aggleton, J.P., Albasser, M.M., Aggleton, D.J., Poirier, G.L., and Pearce, J.M. (2010). Lesions of the Rat Perirhinal Cortex Spare the Acquisition of a Complex Configural Visual Discrimination Yet Impair Object Recognition. Behav. Neurosci. *124*, 55–68.

Akula, J.D., Lyubarsky, A.L., and Naarendorp, F. (2003). The sensitivity and spectral identity of the cones driving the b-wave of the rat electroretinogram. Vis. Neurosci. *20*, 109–117.

Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. J. Neurophysiol. *52*, 1106–1130.

Albright, T.D., and Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Exp. Brain Res. *65*, 582–592.

Alemi-Neissi, A., Rosselli, F.B., and Zoccolan, D. (2013). Multifeatural shape processing in rats engaged in invariant visual object recognition. J. Neurosci. Off. J. Soc. Neurosci. *33*, 5939–5956.

Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L., and Reid, R.C. (2011). Functional specialization of mouse higher visual cortical areas. Neuron *72*, 1025–1039.

Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J. Neurosci. Off. J. Soc. Neurosci. *11*, 168–190.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. Psychol. Rev. *94*, 115–147.

Birch, D., and Jacobs, G.H. (1979). Spatial contrast sensitivity in albino and pigmented rats. Vision Res. *19*, 933–937.

Blanche, T.J., Spacek, M.A., Hetke, J.F., and Swindale, N.V. (2005). Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. J. Neurophysiol. *93*, 2987–3000.

Bonin, V., Histed, M.H., Yurgenson, S., and Reid, R.C. (2011). Local diversity and fine-

scale organization of receptive fields in mouse visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *31*, 18506–18521.

Born, R.T., and Bradley, D.C. (2005). Structure and function of visual area MT. Annu. Rev. Neurosci. *28*, 157–189.

Brincat, S.L., and Connor, C.E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat. Neurosci. *7*, 880–886.

Brooks, D.I., Ng, K.H., Buss, E.W., Marshall, A.T., Freeman, J.H., and Wasserman, E.A. (2013). Categorization of photographic images by rats using shape-based image dimensions. J. Exp. Psychol. Anim. Behav. Process. *39*, 85–92.

Busse, L., Ayaz, A., Dhruv, N.T., Katzner, S., Saleem, A.B., Schölvinck, M.L., Zaharia, A.D., and Carandini, M. (2011). The detection of visual contrast in the behaving mouse. J. Neurosci. Off. J. Soc. Neurosci. *31*, 11351–11361.

Bussey, T.J., Muir, J.L., and Aggleton, J.P. (1999). Functionally dissociating aspects of event memory: the effects of combined perirhinal and postrhinal cortex lesions on object and place memory in the rat. J. Neurosci. Off. J. Soc. Neurosci. *19*, 495–502.

Carandini, M., and Heeger, D.J. (1994). Summation and division by neurons in primate visual cortex. Science *264*, 1333–1336.

Chagas, C., Gattass, R., Gross, C., Aedibvs, E., In, A., Vaticana, C., Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1986). The Analysis Of Moving Visual Patterns.

Chelazzi, L., Rossi, F., Tempia, F., Ghirardi, M., and Strata, P. (1989). Saccadic Eye Movements and Gaze Holding in the Head-Restrained Pigmented Rat. Eur. J. Neurosci. *1*, 639–646.

Chelazzi, L., Ghirardi, M., Rossi, F., Strata, P., and Tempia, F. (1990). Spontaneous Saccades and Gaze-Holding Ability in the Pigmented Rat. II. Effects of Localized Cerebellar Lesions. Eur. J. Neurosci. *2*, 1085–1094.

Cicerone, C.M. (1976). Cones survive rods in the light-damaged eye of the albino rat. Science *194*, 1183–1185.

Coogan, T.A., and Burkhalter, A. (1993). Hierarchical organization of areas in rat visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *13*, 3749–3772.

Cox, D.D. (2014). Do we understand high-level vision? Curr. Opin. Neurobiol. *25*, 187–193.

Davies, M., Machin, P.E., Sanderson, D.J., Pearce, J.M., and Aggleton, J.P. (2007). Neurotoxic lesions of the rat perirhinal and postrhinal cortices and their impact on

biconditional visual discrimination tasks. Behav. Brain Res. *176*, 274–283.

Dean, A.F., and Tolhurst, D.J. (1986). Factors influencing the temporal phase of response to bar and grating stimuli for simple cells in the cat striate cortex. Exp. Brain Res. *62*, 143–151.

Dean, A.F., Tolhurst, D.J., and Walker, N.S. (1982). Non-linear temporal summation by simple cells in cat striate cortex demonstrated by failure of superposition. Exp. Brain Res. *45*, 456–458.

DeAngelis, G.C., and Newsome, W.T. (1999). Organization of disparity-selective neurons in macaque area MT. J. Neurosci. Off. J. Soc. Neurosci. *19*, 1398–1415.

Deegan, J.F., 2nd, and Jacobs, G.H. (1993). On the identity of the cone types of the rat retina. Exp. Eye Res. *56*, 375–377.

Derrington, A.M., and Badcock, D.R. (1992). Two-stage analysis of the motion of 2-dimensional patterns, what is the first stage? Vision Res. *32*, 691–698.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. J. Cogn. Neurosci. *3*, 1–8.

Desimone, R., and Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. Brain Res. *178*, 363–380.

Diamond, M.E., von Heimendahl, M., Knutsen, P.M., Kleinfeld, D., and Ahissar, E. (2008). “Where” and “what” in the whisker sensorimotor system. Nat. Rev. Neurosci. *9*, 601–612.

DiCarlo, J.J., and Cox, D.D. (2007). Untangling invariant object recognition. Trends Cogn. Sci. *11*, 333–341.

DiCarlo, J.J., and Maunsell, J.H.R. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. J. Neurophysiol. *93*, 2974–2986.

DiCarlo, J.J., Lane, J.W., Hsiao, S.S., and Johnson, K.O. (1996). Marking microelectrode penetrations with fluorescent dyes. J. Neurosci. Methods *64*, 75–81.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434.

Douglas, R.M., Neve, A., Quittenbaum, J.P., Alam, N.M., and Prusky, G.T. (2006). Perception of visual motion coherence by rats and mice. Vision Res. *46*, 2842–2847.

Dubner, R., and Zeki, S.M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res. *35*, 528–532.

Dürsteler, M.R., Wurtz, R.H., and Newsome, W.T. (1987). Directional pursuit deficits

following lesions of the foveal representation within the superior temporal sulcus of the macaque monkey. J. Neurophysiol. *57*, 1262–1287.

Edelman, S., and Bülthoff, H.H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Res. *32*, 2385–2400.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? Neuron *23*, 115–125.

Epstein, R., Graham, K.S., and Downing, P.E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. Neuron *37*, 865–876.

Espinoza, S.G., and Thomas, H.C. (1983). Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. Brain Res. *272*, 137–144.

Van Essen, D.C., Maunsell, J.H., and Bixby, J.L. (1981). The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. J. Comp. Neurol. *199*, 293–326.

Van Essen, D.C., Newsome, W.T., Maunsell, J.H., and Bixby, J.L. (1986). The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. J. Comp. Neurol. *244*, 451–480.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex N. Y. N 1991 *1*, 1–47.

Floresco, S.B., Block, A.E., and Tse, M.T.L. (2008). Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. Behav. Brain Res. *190*, 85–96.

Forwood, S.E., Bartko, S.J., Saksida, L.M., and Bussey, T.J. (2007). Rats spontaneously discriminate purely visual, two-dimensional stimuli in tests of recognition memory and perceptual oddity. Behav. Neurosci. *121*, 1032–1042.

Friedman, L.J., and Green, D.G. (1982). Ganglion cell acuity in hooded rats. Vision Res. *22*, 441–444.

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature *360*, 343–346.

Gao, E., DeAngelis, G.C., and Burkhalter, A. (2010). Parallel input channels to mouse primary visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *30*, 5912–5926.

Geesaman, B.J., Born, R.T., Andersen, R.A., and Tootell, R.B. (1997). Maps of complex motion selectivity in the superior temporal cortex of the alert macaque monkey: a double-label 2-deoxyglucose study. Cereb. Cortex N. Y. N 1991 *7*, 749–757.

George Paxinos, C.W. (2004). The Rat Brain in Stereotaxic Coordinates - The New

Coronal Set, Fifth Edition.

Gibson, B.M., Wasserman, E.A., Gosselin, F., and Schyns, P.G. (2005). Applying bubbles to localize features that control pigeons' visual discrimination behavior. J. Exp. Psychol. Anim. Behav. Process. *31*, 376–382.

Girman, S.V., Sauvé, Y., and Lund, R.D. (1999). Receptive field properties of single neurons in rat primary visual cortex. J. Neurophysiol. *82*, 301–311.

Gosselin, F., and Schyns, P.G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res. *41*, 2261–2271.

Graziano, M.S., Andersen, R.A., and Snowden, R.J. (1994). Tuning of MST neurons to spiral motions. J. Neurosci. Off. J. Soc. Neurosci. *14*, 54–67.

Green, D.G., and Powers, M.K. (1982). Mechanisms of light adaptation in rat retina. Vision Res. *22*, 209–216.

Greenberg, D.S., Houweling, A.R., and Kerr, J.N.D. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat. Neurosci. *11*, 749–751.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

Grill-Spector, K., Knouf, N., and Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. Nat. Neurosci. *7*, 555–562.

Gross, C.G. (1992). Representation of visual stimuli in inferior temporal cortex. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *335*, 3–10.

Gulyás, B., Orban, G.A., Duysens, J., and Maes, H. (1987). The suppressive influence of moving textured backgrounds on responses of cat striate neurons to moving bars. J. Neurophysiol. *57*, 1767–1791.

Gur, M., and Snodderly, D.M. (2007). Direction selectivity in V1 of alert monkeys: evidence for parallel pathways for motion processing. J. Physiol. *585*, 383–400.

Hawken, M.J., Parker, A.J., and Lund, J.S. (1988). Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. J. Neurosci. Off. J. Soc. Neurosci. *8*, 3541–3548.

Hayward, W.G., and Williams, P. (2000). Viewpoint dependence and object discriminability. Psychol. Sci. *11*, 7–12.

Von Heimendahl, M., Itskov, P.M., Arabzadeh, E., and Diamond, M.E. (2007). Neuronal activity in rat barrel cortex underlying texture discrimination. PLoS Biol. *5*, e305.

Hildreth, E.C., and Koch, C. (1987). The analysis of visual motion: from computational theory to neuronal mechanisms. Annu. Rev. Neurosci. *10*, 477–533.

Histed, M.H., Carvalho, L.A., and Maunsell, J.H.R. (2012). Psychophysical measurement of contrast sensitivity in the behaving mouse. J. Neurophysiol. *107*, 758–765.

Van Hooser, S.D., and Nelson, S.B. (2006). The squirrel as a rodent model of the human visual system. Vis. Neurosci. *23*, 765–778.

Hubel, D.H. (1995). Eye, Brain, and Vision (Henry Holt and Company).

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. J. Physiol. *195*, 215–243.

Hughes, A. (1977). The refractive state of the rat eye. Vision Res. *17*, 927–939.

Hughes, A. (1979). A schematic eye for the rat. Vision Res. *19*, 569–588.

Hughes, A., and Wässle, H. (1979). An estimate of image quality in the rat eye. Invest. Ophthalmol. Vis. Sci. *18*, 878–881.

Hupfeld, D., and Hoffmann, K.-P. (2006). Motion perception in rats (Rattus norvegicus sp.): deficits in albino Wistar rats compared to pigmented Long-Evans rats. Behav. Brain Res. *170*, 29–33.

Huxlin, K.R., and Goodchild, A.K. (1997). Retinal ganglion cells in the albino rat: revised morphological classification. J. Comp. Neurol. *385*, 309–323.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. J. Exp. Psychol. [Hum. Learn.] *6*, 1–12.

Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. J. Neurophysiol. *73*, 218–226.

Jacobs, G.H., Fenwick, J.A., and Williams, G.A. (2001). Cone-based vision of rats for ultraviolet and visible lights. J. Exp. Biol. *204*, 2439–2446.

James, T.W., Culham, J., Humphrey, G.K., Milner, A.D., and Goodale, M.A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain J. Neurol. *126*, 2463–2475.

Jazayeri, M., Wallisch, P., and Movshon, J.A. (2012). Dynamics of macaque MT cell responses to grating triplets. J. Neurosci. Off. J. Soc. Neurosci. *32*, 8242–8253.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. Off. J. Soc. Neurosci. *17*, 4302–4311.

Keller, J., Strasburger, H., Cerutti, D.T., and Sabel, B.A. (2000). Assessing spatial vision - automated measurement of the contrast-sensitivity function in the hooded rat. J. Neurosci.

Methods *97*, 103–110.

Kerr, J.N.D., and Nimmerjahn, A. (2012). Functional imaging in freely moving animals. Curr. Opin. Neurobiol. *22*, 45–53.

Kim, J.G., Biederman, I., Lescroart, M.D., and Hayworth, K.J. (2009). Adaptation to objects in the lateral occipital complex (LOC): shape or semantics? Vision Res. *49*, 2297–2305.

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

Kolb, B., and Walkey, J. (1987). Behavioural and anatomical studies of the posterior parietal cortex in the rat. Behav. Brain Res. *23*, 127–145.

Kolb, B., Pittman, K., Sutherland, R.J., and Whishaw, I.Q. (1982). Dissociation of the contributions of the prefrontal cortex and dorsomedial thalamic nucleus to spatially guided behavior in the rat. Behav. Brain Res. *6*, 365–378.

Kourtzi, Z., and Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. J. Cogn. Neurosci. *12*, 48–55.

Kourtzi, Z., Bülthoff, H.H., Erb, M., and Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. Nat. Neurosci. *5*, 17–18.

KRIEG, W.J.S. (1946a). Connections of the cerebral cortex; the albino rat; topography of the cortical areas. J. Comp. Neurol. *84*, 221–275.

KRIEG, W.J.S. (1946b). Connections of the cerebral cortex; the albino rat; structure of the cortical areas. J. Comp. Neurol. *84*, 277–323.

Kriegeskorte, N., and Kievit, R.A. (2013). Representational geometry: integrating cognition, computation, and the brain. Trends Cogn. Sci. *17*, 401–412.

Kriegeskorte, N., Sorger, B., Naumer, M., Schwarzbach, J., van den Boogert, E., Hussy, W., and Goebel, R. (2003). Human cortical object recognition from a visual motion flowfield. J. Neurosci. Off. J. Soc. Neurosci. *23*, 1451–1463.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. Acta Psychol. (Amst.) *102*, 221–245.

Ledgeway, T., and Hess, R.F. (2002). Rules for combining the outputs of local motion detectors to define simple contours. Vision Res. *42*, 653–659.

Li, N., Cox, D.D., Zoccolan, D., and DiCarlo, J.J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? J. Neurophysiol. *102*, 360–376.

Liu, J., and Newsome, W.T. (2003). Functional organization of speed tuned neurons in

visual area MT. J. Neurophysiol. *89*, 246–256.

Livingstone, M.S., and Hubel, D.H. (1984). Specificity of intrinsic connections in primate primary visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *4*, 2830–2835.

Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. Annu. Rev. Neurosci. *19*, 577–621.

Logothetis, N.K., Pauls, J., Bülthoff, H.H., and Poggio, T. (1994). View-dependent object recognition by monkeys. Curr. Biol. CB *4*, 401–414.

Lueschow, A., Miller, E.K., and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. Cereb. Cortex N. Y. N 1991 *4*, 523–531.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (New York, NY, USA: Henry Holt and Co., Inc.).

Marshel, J.H., Garrett, M.E., Nauhaus, I., and Callaway, E.M. (2011). Functional specialization of seven mouse visual cortical areas. Neuron *72*, 1040–1054.

Maunsell, J.H., and Van Essen, D.C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. J. Comp. Neurol. *266*, 535–555.

Maunsell, J.H., and Newsome, W.T. (1987). Visual processing in monkey extrastriate cortex. Annu. Rev. Neurosci. *10*, 363–401.

Meier, P., Flister, E., and Reinagel, P. (2011). Collinear features impair visual detection by rats. J. Vis. *11*.

Mikami, A., Newsome, W.T., and Wurtz, R.H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. J. Neurophysiol. *55*, 1328–1339.

Minini, L., and Jeffery, K.J. (2006). Do rats use shape to solve "shape discriminations"? Learn. Mem. Cold Spring Harb. N *13*, 287–297.

Mirabella, G., Bertini, G., Samengo, I., Kilavik, B.E., Frilli, D., Della Libera, C., and Chelazzi, L. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. Neuron *54*, 303–318.

Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. Trends Neurosci. *6*, 414–417.

Missal, M., Vogels, R., Li, C.Y., and Orban, G.A. (1999). Shape interactions in macaque inferior temporal neurons. J. Neurophysiol. *82*, 131–142.

Montero, V.M. (1981). Comparative Studies on the Visual Cortex. In Multiple Visual Areas, (Humana Press), pp. 33–81.

Montero, V.M. (1993). Retinotopy of cortical connections between the striate cortex and extrastriate visual areas in the rat. Exp. Brain Res. *94*, 1–15.

Morel, A., and Bullier, J. (1990). Anatomical segregation of two cortical visual pathways in the macaque monkey. Vis. Neurosci. *4*, 555–578.

Movshon, J.A., and Newsome, W.T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. Off. J. Soc. Neurosci. *16*, 7733–7741.

Movshon, J.A., Thompson, I.D., and Tolhurst, D.J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J. Physiol. *283*, 53–77.

Movshon, J.A., Lisberger, S.G., and Krauzlis, R.J. (1990). Visual cortical signals supporting smooth pursuit eye movements. Cold Spring Harb. Symp. Quant. Biol. *55*, 707–716.

Muntz, W.R. (1967). A behavioural study on photopic and scotopic vision in the hooded rat. Vision Res. *7*, 371–376.

Murray, R.F. (2011). Classification images: A review. J. Vis. *11*.

Nassi, J.J., and Callaway, E.M. (2009). Parallel processing strategies of the primate visual system. Nat. Rev. Neurosci. *10*, 360–372.

Newell, F.N. (1998). Stimulus context and view dependence in object recognition. Perception *27*, 47–68.

Newsome, W.T., Wurtz, R.H., Dürsteler, M.R., and Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. J. Neurosci. Off. J. Soc. Neurosci. *5*, 825–840.

Niell, C.M. (2011). Exploring the next frontier of mouse vision. Neuron *72*, 889–892.

Niell, C.M., and Stryker, M.P. (2008). Highly selective receptive fields in mouse visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *28*, 7520–7536.

Nielsen, K., Logothetis, N.K., and Rainer, G. (2006). Discrimination strategies of humans and rhesus monkeys for complex visual displays. Curr Biol *16*, 814–820.

Nielsen, K.J., Logothetis, N.K., and Rainer, G. (2008). Object features used by humans and monkeys to identify rotated shapes. J. Vis. *8*, 9.1–15.

Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P., and Reid, R.C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature *433*, 597–603.

Olshausen, B.A., and Field, D.J. (2004). Sparse coding of sensory inputs. Curr. Opin. Neurobiol. *14*, 481–487.

Op De Beeck, H., and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. J. Comp. Neurol. *426*, 505–518.

Orban, G.A., Kennedy, H., and Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. J. Neurophysiol. *56*, 462–480.

Pack, C.C., and Born, R.T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature *409*, 1040–1042.

Pack, C.C., Livingstone, M.S., Duffy, K.R., and Born, R.T. (2003). End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. Neuron *39*, 671–680.

Parnavelas, J.G., Burne, R.A., and Lin, C.S. (1981). Receptive field properties of neurons in the visual cortex of the rat. Neurosci. Lett. *27*, 291–296.

Pashler, H.E. (2002). Stevens' handbook of experimental psychology (New York: J. Wiley).

Paxinos, G., and Watson, C. (2005). The rat brain in stereotaxic coordinates (Elsevier Academic Press).

Perrett, D.I., and Oram, M.W. (1998). Visual recognition based on temporal cortex cells: viewer-centred processing of pattern configuration. Z. Für Naturforschung C J. Biosci. *53*, 518–541.

Perry, V.H., and Cowey, A. (1979). Changes in the retino-fugal pathways following cortical and tectal lesions in neonatal and adult rats. Exp. Brain Res. *35*, 97–108.

Petruno, S.K., Clark, R.E., and Reinagel, P. (2013). Evidence that primary visual cortex is required for image, orientation, and motion discrimination by rats. PloS One *8*, e56543.

Pettigrew, J.D., Dreher, B., Hopkins, C.S., McCall, M.J., and Brown, M. (1988). Peak density and distribution of ganglion cells in the retinae of microchiropteran bats: implications for visual acuity. Brain. Behav. Evol. *32*, 39–56.

Pinto, N., Cox, D.D., and DiCarlo, J.J. (2008). Why is real-world visual object recognition hard? PLoS Comput. Biol. *4*, e27.

Priebe, N.J., Cassanello, C.R., and Lisberger, S.G. (2003). The neural representation of speed in macaque area MT/V5. J. Neurosci. Off. J. Soc. Neurosci. *23*, 5650–5661.

Prusky, G.T., West, P.W., and Douglas, R.M. (2000a). Behavioral assessment of visual acuity in mice and rats. Vision Res. *40*, 2201–2209.

Prusky, G.T., West, P.W., and Douglas, R.M. (2000b). Experience-dependent plasticity of visual acuity in rats. Eur. J. Neurosci. *12*, 3781–3786.

Prusky, G.T., Harker, K.T., Douglas, R.M., and Whishaw, I.Q. (2002). Variation in

visual acuity within pigmented, and between pigmented and albino rat strains. Behav. Brain Res. *136*, 339–348.

Prusky, G.T., Douglas, R.M., Nelson, L., Shabanpoor, A., and Sutherland, R.J. (2004). Visual memory task for rats reveals an essential role for hippocampus and perirhinal cortex. Proc. Natl. Acad. Sci. U. S. A. *101*, 5064–5068.

Purves, D., Wojtach, W.T., and Lotto, R.B. (2011). Understanding vision in wholly empirical terms. Proc. Natl. Acad. Sci. U. S. A. *108 Suppl 3*, 15588–15595.

Quiroga, R.Q. (2012). Spike sorting. Curr. Biol. CB *22*, R45–46.

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687.

Reid, R.C., Soodak, R.E., and Shapley, R.M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. Proc. Natl. Acad. Sci. U. S. A. *84*, 8740–8744.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025.

Rolls, E.T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron *27*, 205–218.

Rolls, E.T. (2012). Invariant Visual Object and Face Recognition: Neural and Computational Bases, and a Model, VisNet. Front. Comput. Neurosci. *6*, 35.

Rousselet, G.A., Thorpe, S.J., and Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? Trends Cogn. Sci. *8*, 363–370.

Rubin, B.D., and Katz, L.C. (2001). Spatial coding of enantiomers in the rat olfactory bulb. Nat. Neurosci. *4*, 355–356.

Rust, N.C., and Dicarlo, J.J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J. Neurosci. Off. J. Soc. Neurosci. *30*, 12978–12995.

Rust, N.C., Mante, V., Simoncelli, E.P., and Movshon, J.A. (2006). How MT cells analyze the motion of visual patterns. Nat. Neurosci. *9*, 1421–1431.

Sanderson, K.J., Dreher, B., and Gayer, N. (1991). Prosencephalic connections of striate and extrastriate areas of rat visual cortex. Exp. Brain Res. *85*, 324–334.

Sawinski, J., Wallace, D.J., Greenberg, D.S., Grossmann, S., Denk, W., and Kerr, J.N.D. (2009). Visually evoked activity in cortical cells imaged in freely moving animals. Proc. Natl. Acad. Sci. U. S. A. *106*, 19557–19562.

Seymoure, P., and Juraska, J.M. (1997). Vernier and grating acuity in adult hooded rats: the influence of sex. Behav. Neurosci. *111*, 792–800.

Sia, Y., and Bourne, J.A. (2008). The rat temporal association cortical area 2 (Te2) comprises two subdivisions that are visually responsive and develop independently. Neuroscience *156*, 118–128.

Simoncelli, E.P., and Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Res. *38*, 743–761.

Snodderly, D.M., and Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (Macaca fascicularis): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. J. Neurophysiol. *74*, 2100–2125.

Snowden, R.J., Treue, S., and Andersen, R.A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. Exp. Brain Res. *88*, 389–400.

Sun, B., Lv, B., Dong, Q., Wang, S., and Chai, Z. (2009). Watching moving images specifically promotes development of medial area of secondary visual cortex in rat. Dev. Neurobiol. *69*, 558–567.

Szél, A., and Röhlich, P. (1992). Two cone types of rat retina detected by anti-visual pigment antibodies. Exp. Eye Res. *55*, 47–52.

Tafazoli, S., Di Filippo, A., and Zoccolan, D. (2012). Transformation-tolerant object recognition in rats revealed by visual priming. J. Neurosci. Off. J. Soc. Neurosci. *32*, 21–34.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. *19*, 109–139.

Tanaka, K., and Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. J. Neurophysiol. *62*, 626–641.

Tarr, M.J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychon. Bull. Rev. *2*, 55–82.

Tarr, M.J., Bülthoff, H.H., Zabinski, M., and Blanz, V. (1997). To What Extent Do Unique Parts Influence Recognition Across Changes in Viewpoint? Psychol. Sci. *8*, 282–289.

Tees, R.C. (1999). The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. Behav. Brain Res. *106*, 55–73.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.

Tinsley, C.J., Webb, B.S., Barraclough, N.E., Vincent, C.J., Parker, A., and Derrington,

A.M. (2003). The nature of V1 neural responses to 2D moving patterns depends on receptive-field structure in the marmoset monkey. J. Neurophysiol. *90*, 930–937.

Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., and Belliveau, J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci. Off. J. Soc. Neurosci. *15*, 3215–3230.

Tovee, M.J., Rolls, E.T., and Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J. Neurophysiol. *72*, 1049–1060.

Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat. Neurosci. *4*, 832–838.

Uchida, N., and Mainen, Z.F. (2003). Speed and accuracy of olfactory discrimination in the rat. Nat. Neurosci. *6*, 1224–1229.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. Cognition *32*, 193–254.

Uttal, W.R., Spillmann, L., Stürzel, F., and Sekuler, A.B. (2000). Motion and shape in common fate. Vision Res. *40*, 301–310.

Vaiceliunaite, A., Erisken, S., Franzen, F., Katzner, S., and Busse, L. (2013). Spatial integration in mouse primary visual cortex. J. Neurophysiol. *110*, 964–972.

Vermaercke, B., and Op de Beeck, H.P. (2012). A multivariate approach reveals the behavioral templates underlying visual discrimination in rats. Curr. Biol. CB *22*, 50–55.

Vogels, R., and Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. Cereb. Cortex N. Y. N 1991 *12*, 756–766.

Vuong, Q.C., and Tarr, M.J. (2004). Rotation direction affects object recognition. Vision Res. *44*, 1717–1730.

Vuong, Q.C., and Tarr, M.J. (2006). Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. Perception *35*, 497–510.

Wallace, D.J., Greenberg, D.S., Sawinski, J., Rulla, S., Notaro, G., and Kerr, J.N.D. (2013). Rats maintain an overhead binocular field at the expense of constant fusion. Nature *498*, 65–69.

Wallis, G., and Bülthoff, H.H. (2001). Effects of temporal association on recognition memory. Proc. Natl. Acad. Sci. U. S. A. *98*, 4800–4804.

Wallis, G., and Rolls, E.T. (1997). Invariant face and object recognition in the visual system. Prog. Neurobiol. *51*, 167–194.

Wang, G., Tanifuji, M., and Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. Neurosci. Res. *32*, 33–46.

Wang, G., Obama, S., Yamashita, W., Sugihara, T., and Tanaka, K. (2005). Prior experience of rotation is not required for recognizing objects seen from different angles. Nat. Neurosci. *8*, 1768–1775.

Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *31*, 1905–1918.

Van Wezel, R.J.A., and van der Smagt, M.J. (2003). Motion processing: how low can you go? Curr. Biol. CB *13*, R840–842.

Wilson, K.D., and Farah, M.J. (2003). When does the visual system use viewpoint-invariant representations during recognition? Brain Res. Cogn. Brain Res. *16*, 399–415.

Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A.N., and Tanifuji, M. (2006). Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. J. Neurophysiol. *96*, 3147–3156.

Zangenehpour, S., and Chaudhuri, A. (2005). Patchy organization and asymmetric distribution of the neural correlates of face processing in monkey inferotemporal cortex. Curr. Biol. CB *15*, 993–1005.

Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., and Frackowiak, R.S. (1991). A direct demonstration of functional specialization in human visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *11*, 641–649.

Zoccolan, D., Cox, D.D., and DiCarlo, J.J. (2005). Multiple object response normalization in monkey inferotemporal cortex. J. Neurosci. Off. J. Soc. Neurosci. *25*, 8150–8164.

Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J.J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J. Neurosci. Off. J. Soc. Neurosci. *27*, 12292–12307.

Zoccolan, D., Oertelt, N., DiCarlo, J.J., and Cox, D.D. (2009). A rodent model for the study of invariant visual object recognition. Proc. Natl. Acad. Sci. U. S. A. *106*, 8748–8753.

Zoccolan, D., Graham, B.J., and Cox, D.D. (2010). A self-calibrating, camera-based eye tracker for the recording of rodent eye movements. Front. Neurosci. *4*, 193.