Scuola Internazionale Superiore di Studi Avanzati – Trieste

Accuracy of Rats in Discriminating Visual Objects is Explained by the Complexity of Their Perceptual Strategy

Candidate:

Vladimir Djurdjevic

Supervisor:

Davide Zoccolan

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

Trieste, 2017

SISSA – Via Bonomea 265 – 34136 Trieste, Italy

Table of contents

Abstract

Despite their growing popularity as models of visual functions, it is widely assumed that rodents deploy perceptual strategies not nearly as advanced as those of primates, when processing visual objects. Such belief is fostered by the conflicting findings about the complexity of rodent pattern vision, which appears to range from mere detection of overall object luminance to view-invariant processing of discriminant shape features.

Here, we sought to clarify how refined object vision is in rodents, by measuring how well a group of rats discriminated a reference object from eleven distractors, spanning a spectrum of image-level similarity with the reference. We also presented the animals with random variations of the reference, and we processed their responses to these stimuli to obtain subject-specific models of rat perceptual choices.

These models captured very well the highly variable discrimination performance observed across subjects and object conditions. In particular, they revealed how the animals that succeeded with the more challenging distractors were those that integrated the wider variety of discriminant features into their perceptual strategy. Critically, these features remained highly subject-specific and largely invariant under changes in object appearance (e.g., size variation), although they were properly reformatted (e.g., rescaled) to deal with the specific transformations the objects underwent.

Overall, these findings show that rat object vision, far from being poorly developed, relies on the same kind of feature-based filtering (iterated across multiple scales, positions, etc.) that is at work in primates and is implemented in state-of-the-art machine vision systems, such as deep convolutional neural networks.

Chapter 1: Introduction

1.1 Complexity of the invariance

If, like humans, you belong to a species that has evolved to use vision as its main sensory input, your ability to effectively perform object recognition represents your mean of survival. Everything you do from differencing a predator from a pray to finding a potential mate depends on it. And this is the reason why "visually driven" species have developed incredibly efficient mechanisms in order to solve this task.

When you pick up a random paper or a review about object recognition you will usually, right at the beginning, find a sentence like this: "Achieving invariant (object) recognition represents such a formidable computational challenge that is often assumed to be a unique hallmark of primate vision" (Zoccolan, 2015). In other words, the task at hand is so complex that you need to be at the apex of the brain evolution (in terms of the brain/body ratio and therefore the raw computational power that you have at your disposal) to be able to solve it. And the complexity of the task can hardly be overstated. Even though humans are able to effortlessly perform this task in their everyday lives the number of calculations that our brain has to perform in order to fulfill the task is enormous (state of the art computer systems currently developed by the top tech companies for this purpose are still far behind). Having all this in mind it is not hard to understand why (invariant) object recognition presents one of the key open questions in visual neuroscience today.

And where, exactly, does this complexity come from? It comes from the organization of our visual system and the fact that our brain needs to reconstruct the physical 3D world from the 2D information it gets from our eyes.

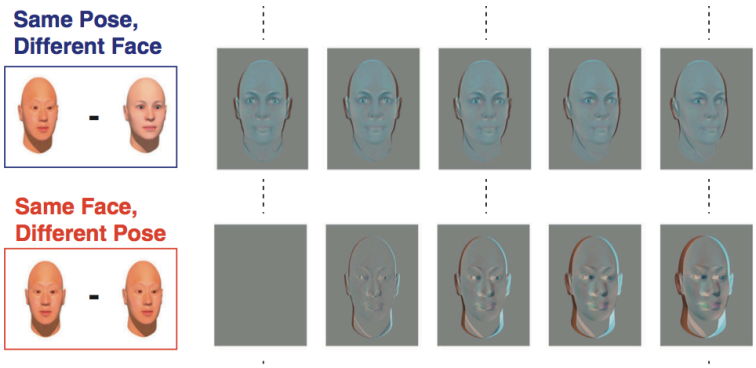

Figure 1. Pixel-level image variation caused by variation in viewing parameters for single object is often larger than the pixel differences between different objects. Here we show 3D-rendered images of the faces of two individuals undergoing a rotation through 20^0 in azimuth.

For example, when we compare the retinal images (pixel similarity) of different individuals viewed from the same angle they are more similar than the same individual viewed from different angles, even for relatively small angles (Figure 1, adopted from Cox, 2014).

And the rotation (in depth) is only one of the possible transformations. Same objects (depending on your view point) can vary in size, position, lighting and the presence of noise/background. Calculating all the possible combinations of these variables leads us to the conclusion that we are dealing with an effectively infinite number of different retinal patterns for every single object we encounter in our everyday lives. So how can a brain possibly be able to deal with so many variations and still be able to give us the answer to the question "what is that"? The solution is called "*invariance"* and it assumes that our visual representation of the outside world must be very *tolerant* to changes in the appearance of individual objects, while (at the same time) it must be *selective* enough to the features that define object identity to uniquely identify individual objects. In the primate visual cortex this problem has been solved in the *inferotemporal cortex* (IT), which represents the final stage of the so-called "ventral stream", the neural pathway primarily involved in the processing of the visual information necessary for the complex object recognition (for review see Logothetis and Sheinberg, 1996; Tanaka, 1996).

1.2 The anatomy of the ventral stream

1.2.1 Area V1

The ventral stream begins in the largest area dedicated to visual processing – primary visual cortex (area V1) in the occipital cortex. Our knowledge about the properties and functions of V1 neurons comes from the pioneering work of Hubel and Wiesel (for which they've received the Nobel Prize in Physiology or Medicine in 1981). Their work (Hubel and Wiesel, 1959) showed that: there is a *topographical map* of the visual field "contained" inside the visual cortex where neighboring neurons process information from nearby visual fields and that V1 region of the brain is mostly organized in the "*orientation columns*" - a column of neurons (spanning multiple layers) that fires only when exposed to the stimuli of a specific orientation or direction of motion. Orientation columns are composed of *simple* cells - small *receptive fields* (*RF* - fields of visual space where specific stimuli can elicit a neural response) and *complex* cells - usually summing the input of a couple of simple cells therefore having a wider receptive fields. The complex cells need a moving stimulus in order to elicit the best response. The original work also defined *hypercomplex* cells (also known as *end-stopped* cells) as a separate type of cells but later work (Dreher, 1972) showed that they could be categorized as subsets of either simple or complex cells. The primary role of the V1 region is to extract the so-called *low-level* information about the object properties (oriented edges, contrast, color) and to forward this pre-processed visual information to the adjacent visual areas that constitute the ventral stream.

1.2.2 Area V2

The stream continues with the V2 area. V2 is also known as *prestriate cortex,* since it is a direct continuation of the V1 region (Brodmann area 17 in humans), which is also known as the *striate cortex* (from Latin *stria* – line, channel) because of the tightly packed axons of the neurons from the *lateral geniculate nucleus* (LGN) which form stripes (*Line of Gennari*) easily noticeable on the microscope slices. Although this is a usual terminology in the visual neuroscience, it is worth mentioning

that most (if not all) non-primates lack this specific morphological characteristic. V2 receives two strong feedforward inputs – from V1 and from *pulvinar* (Marion, 2013). It projects further to areas V3, V4 and V5(*MT*) but it also sends feedback connections to V1. Most of the anatomical and functional characteristics of the V2 are similar to V1: retinotopy is preserved, orientation columns are present and cells are tuned for orientation, special frequency and color. Due to the convergent input from the V1, cells in V2 are able to code for some more sophisticated properties such as contours and textures. Approximately one-third of the neurons in V2 show inhibitory interactions that make them selective for combination of orientations, which is essential for the analysis of contours and textures (Anzai, 2007). It has been shown that V2 neurons also play a part in the discrimination between foreground and the background, combining stereoscopic information (the depth of field) with the global configuration of the contours (Gestalt factors) to interpolate 3D information from 2D images (Qiu and von der Heydt, 2005). Even though the simplified modular view is very popular in the presentation of how the visual system (and our entire brain works), it should be noted that all of the "visual" areas are actually multimodal and that they can also preform functions that are not directly related to sensory processing. Experiments in rats (López-Aranda et al, 2009) have shown the crucial role of the layer 6 of the V2 area in the processing of object recognition memory (ORM) – the expression of the protein RGS-14 led to the transformation of the short-term ORM into a long-term ORM while the selective destruction of this layer led to the complete loss of ORM.

Once we move from the (pre)striate cortex, things starts to complicate exponentially. The number of the *extrastirate areas* is very much species dependent and the structural mapping in humans is very limited, due to our inability to perform the usual histological and electrophysiological procedures we use in other spices. The introduction of the fMRI helped a lot but there is still substantial progress to be made – especially using anatomical MRI and diffusion tensor imaging (DTI). Since most of the research comes from the macaque monkeys as experimental models, both nomenclature and functional connectivity relies heavily on these finding.

It their paper, Felleman and Van Essen (1991) make a summary of the areas that are (almost) exclusively dedicated or connected to vision and they report "25 neocortical areas that are predominantly or exclusively visual in function, plus an additional 7 areas that we regard as visual-association areas on the basis of their

extensive visual inputs". This map is much less accurate/populated when it comes to humans (due to aforementioned limitations). Just last year a Nature paper also published by Van Essen's lab (Glasser, M. F. et al, 2016) added almost a 100 new areas to the map of neocortex with all likelihood that this number will increase with the new data that we obtain.

1.2.3 Area V3

Area V3 is also referred as a *third visual complex* because the questions of structural and functional subdivision (in macaques and humans, and much less in other species) are still not settled. This area was first characterized by Zeki (1969) as a single strip adjacent to V2 that has a mirror-symmetric visuotopic organization. Later studies showed that V3 actually consists of two areas that are physically separated by area V4 (Felleman and Van Essen, 1991; Burkhalter et al, 1986) and functionally divided between the dorsal (V3/V3d) and ventral (VP/V3v) pathway (Figure 2). However the most recent study published by Lyon and Kaas (2002) seems to confirm that V3d and V3v are subdivisions of a single V3 area. You can also notice how much the position and the size of these areas vary (Fig. 2) depending on the author, so additional data is needed to settle this argument. VP/V3v area is part of the ventral pathway and most of the cells are tuned to complex features like shape. Most of the projections from this area go to area V4. V3/V3d is part of the dorsal pathway and most of the cells are encoding a global motion perception.

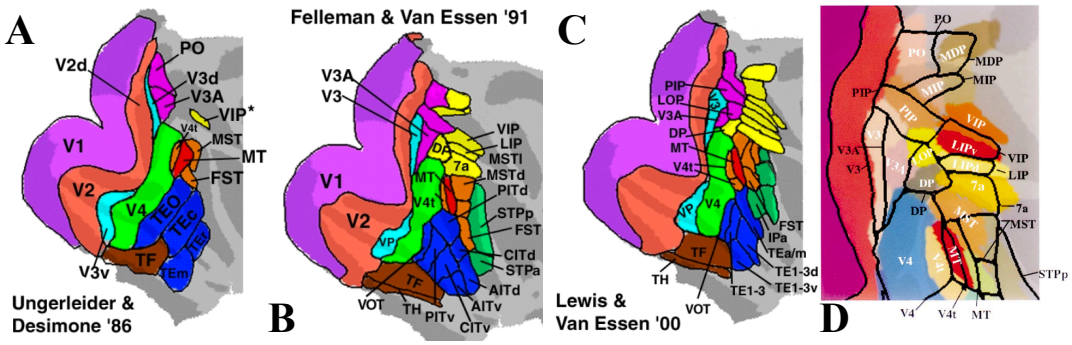

Figure 2. – Maps of the visual cortical areas in the macaques. Depending on the year of publication the maps get more and more refined. Still, because of the big intersubject variability it is very hard to produce a consistent map. A) Map published by Ungerleider and Desimone (1986). B) Map published by Felleman and Van Essen (1991). C) Map published by Lewis and Van Essen (2000). D) Overlay map showing the Lewis and Van Essen summary map (same as C) in colors with areas identified in white and the Felleman and Van Essen scheme (same as B) as outlines, with areas identified in black.

1.2.4 Area V4

Area V4 (in macaques) consists of a continuous strip adjacent to area V3 but is also functionally divided into several areas. The first division comes from the visuotopic organization where the ventral part of the area (V4v) carries a lower field representation while the dorsal part (V4d) carries an upper field representation. Both areas are also "separated" by their (left and right) hemifield representations (Fize et al. (2003). The area was first investigated by Zeki (1969) reporting projections from the central visual field representations of V2 and V3. New data has expanded this picture adding to the connection map of the area V4 the following regions – MT (medial temporal area)/V5, FSF (fundus of the superior temporal sulcus [STS]) and FEF (frontal eye field). Peripheral field representations of V4 are connected with occipitoparietal areas DP (dorsal prelunate area), VIP (ventral intraparietal area), LIP (lateral intraparietal area), PIP (posterior intraparietal area), parieto-occipital area, and MST (medial STS area), and parahippocampal area TF (Ungerleider et al, 2008). Major outputs of the area V4 are TEO (in posterior inferior temporal cortex) and TE (in anterior temporal cortex) both part of the IT cortex. Neurons in the V4 area are selective for the complex object features including color, texture and shape. While the homology between the earlier visual areas (V1, V2, V3) has been confirmed by fMRI (Serano, 1995) there's a lot of controversy about the human V4 homologue (Tootell and Hadjikhani, 2001; Goddard et al, 2011). Area V4 in humans was first considered to be the brain color center since lesions in this area of the occipitotemporal cortex produced color vision loss – achromatopsia (Damasio, 1980; Zeki 1990). However, current research shows that neurons that encode color are present right from the beginning of visual processing in the cortex – area V1 (Shapley and Hawken, 2011). It is shown that inside area V1 humans have two types of color sensitive neurons - single-opponent and double-opponent cells. Single-opponent neurons respond to large areas of color, which is especially useful in the perception of large color scenes and atmosphere. Double-opponent cells got their name from the ability to process the opposite inputs from different cone cells in the retina (very useful for identification of the contrasting colors such as green and red). The main role of the double-opponent cells is in the perception of the smaller color features like patterns, textures and color boundaries. It is also important to mention that area V4 is the first area of the ventral visual stream that shows a significant attention modulation (Moran and Desimone,

1985). Studies have shown that selective attention can change the firing rate of the neurons in V4 by up to 20%. Together with selectivity for color, neurons in V4 show selectivity for orientation and spatial frequency as well as object features of intermediate complexity (like shape).

1.2.5 IT cortex

The end of the ventral pathway is in the IT cortex and its topography also depends on the author and the year of publication (Figure 2). The temporal lobe is exclusive to primates and the fact that we present the mechanism of visual object recognition in the way that we do tells a lot about the history of research of the topic but also about our "anthropocentrism" when it comes to the scientific questions that we want answered. The subdivisions mostly used in the literature are TEO/(roughly equivalent to) PIT (posterior inferior temporal cortex) and TE/(roughly equivalent to) CIT + AIT (central and anterior inferior temporal cortex). The retinotopy of the IT cortex is very crude and the neurons have a very large receptive fields compared to the "earlier" regions which accounts for tolerance that is necessary to achieve the invariance (exceptions will be mentioned later). "High" visual areas are usually regarded as "nonretinotopic" but some authors argue that this is just a problem of non-adequate test stimuli. Recent papers by Levy et al (2001) and Hasson et al. (2003) have shown that orderly central and peripheral representation can be found all over the visual cortex. IT cortex has a lot of areas where neurons seem to be tuned for a special function like recognizing faces (*fusiform face area,* FFA – Sergent et al, 1992), recognizing places (*parahippocampal place area*, PPA – Epstein and Kanwisher, 1998), recognizing body parts (*extrastriate body area,* EBA – Downing et al, 2001) and discriminating between shapes and scrambled stimuli (*lateral occipital complex*, LOC - Grill-Spector et al, 2001). Even though these areas are specific for humans, we can assume that the brain has developed specific regions dedicated to critical evolutionary demands and we can also expect to find similar kind of areas related to different critical visual tasks in other spices (both in non human primates and other visually dominant spieces). IT cortex receives inputs from areas V2, V3 (both V3d and V3v) and V4 and it sends feedback projections back to these areas. It is also reciprocally connected to a number of STS areas, FTF in prefrontal cortex and

(most importantly) amygdala and hippocampus – connections responsible for establishing a "memory link" necessary for recognition and categorization of objects.

Figure 3 shows some of the complexity of the connections between the "visual" areas and the situation gets even more complicated as the new data comes in. We only outlined the major projections of the areas that are most involved in the object recognition and going into details would take the space that is bigger than this entire thesis. The primary goal was just to illustrate the immense complexity of the hardware responsible for the solution of a fascinatingly difficult problem.

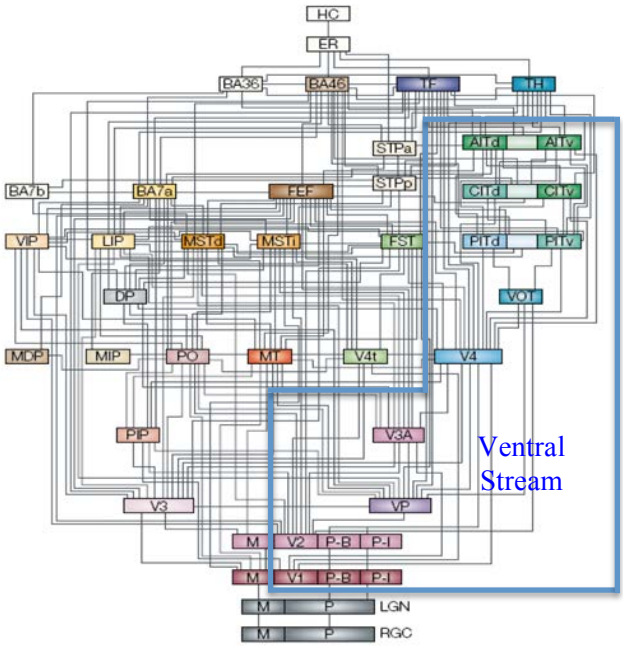

Figure 3. Hierarchy and connections between visual areas. This hierarchy shows 32 visual cortical areas, two subcortical visual stages (the retinal ganglion cell layer and the LGN) plus several non visual areas (area 7b of somatosensory cortex, perirhinal area 36, the ER, and the hippocampal complex). These areas are connected by 187 linkages, most of which have been demonstrated to be reciprocal pathways. M – magnocellular pathway, P – parvo-cellular pathway. P-B – wavelength selective stream, P-I – wavelength and (much less) orientation selective stream. VOT (Ventral occipitotemporal area) and V3A are not described in the text. Abbreviations referring to the parietal cortex are left out for clarity. Adapted from Felleman & Van Essen, 1991.

1.3 The functional properties of the ventral stream

The idea that there might be two separate streams of information inside the visual cortex comes from the experiments of Mishkin (Mishkin et al, 1983) which demonstrated a very different behavior in macaque monkeys depending on the region of the brain where the lesion was performed. In the first task, aimed at object discrimination, monkeys were first familiarized with a physical object and then presented with two other objects. The task was based on the principle of "non matching to sample", i.e., monkeys had to choose the object (from the other two) that was more different then the one they have learned to recognize (in order to get the reward). The bilateral removal of the area TE produced a severe impairment in their ability to solve the task (without interfering with their ability to precisely locate the objects). The other task tested monkey's spatial perception – they were rewarded if they chose a covered foodwell that was closer to the tall cylinder, which presented the "landmark". Bilateral removal of the posterior parietal cortex led to a sever impairment in their ability to solve this task but it did not interfere with their performance in the object recognition task. The behavioral results were confirmed using 2-deoxyglucose (a method developed by Sokoloff, 1982), which functions as a "metabolic encephalography" labeling the parts of the brain with the high glucose consumption as the one that are mostly active (very similar to oxygen consumption and fMRI).

The hypothesis of Mishkin is very much in accordance with the electrophysiological recordings of single neurons in different areas of the visual cortex. As we said in the previous part, the further away we move from area V1 the bigger the receptive fields of the neurons become (a necessary condition to achieve invariance). The same happens with the reaction times of the neurons. The latency (from the stimulus onset) rises from 40 ms in V1 to as much as 100 ms in IT cortex. And the complexity of preferred stimuli is increasing as well – the neurons in V1 are selective for very low level features (orientation, brightness, contrast) while the neurons in the IT cortex will be selective for very complex features like shape or the conjunction of features (Figure 4). All of this points to a hierarchically ordered system where every consecutive area summarizes information from the previous one until it reaches its end (in the IT cortex).

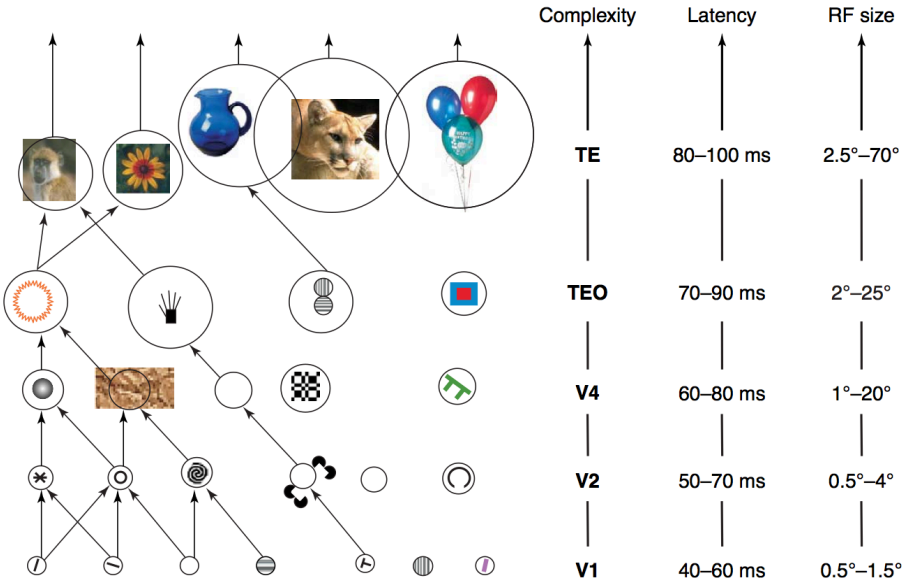

Figure 4. Through a hierarchy of cortical areas, from V1 through temporal-occipital cortex (TEO) to temporal cortex (TE), complex and invariant object representations are progressively built by integrating convergent inputs from lower levels. Examples of elements for which neurons respond selectively are represented inside receptive fields (RFs; represented by circles) of different sizes. Schematics on the right side present the progressive increase in the 'complexity' of the neuronal representations from V1 through temporal-occipital cortex (TEO) to temporal cortex (TE). 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)

As a result, DiCarlo and Cox (2007) put forward a theoretical framework suggesting a way in which the brain might solve the invariance problem. The framework is mostly computational and it doesn't address all the possible "complications" that can arise because of the complexity of the identification task but it provides a satisfying explanation of the behavioral experiments that demonstrate monkeys'/humans' ability to perform a recognition task in a "blink of an eye" – less than 300 ms (Thorpe et al.,1996; Delorme et al., 2000; Fabre-Thorpe, 2011), which they defined as "core" object recognition. We are going to explain this framework in more details.

1.3.1 Untangling the manifolds

When an image of an object is "projected" onto retina the result is the activation of ~100 million retinal photoreceptors which will then activate ~1 million of retinal ganglion cells. This type of activation can be represented in a high-dimensional extension of the Cartesian space where every axis represents a response

of a single ganglion cell (giving us a 1 million dimensional space). Every time we perceive this object it will produce a different pattern of activity depending on its position, size, orientation, lighting and background, which will lead to the formation of a low-dimensional curved surface (inside this high dimensional space) called an object "manifold" (Figure 5a). Different objects will produce different manifolds (Figure 5b-d). The framework proposes that disentangling of different manifolds happens by applying successive computational steps (moving from V1 to IT) of tresholding a sum of weighted synapses and then applying a decision function (linear classifier) to establish a separating hyperplane between two manifolds (Figure 5b).

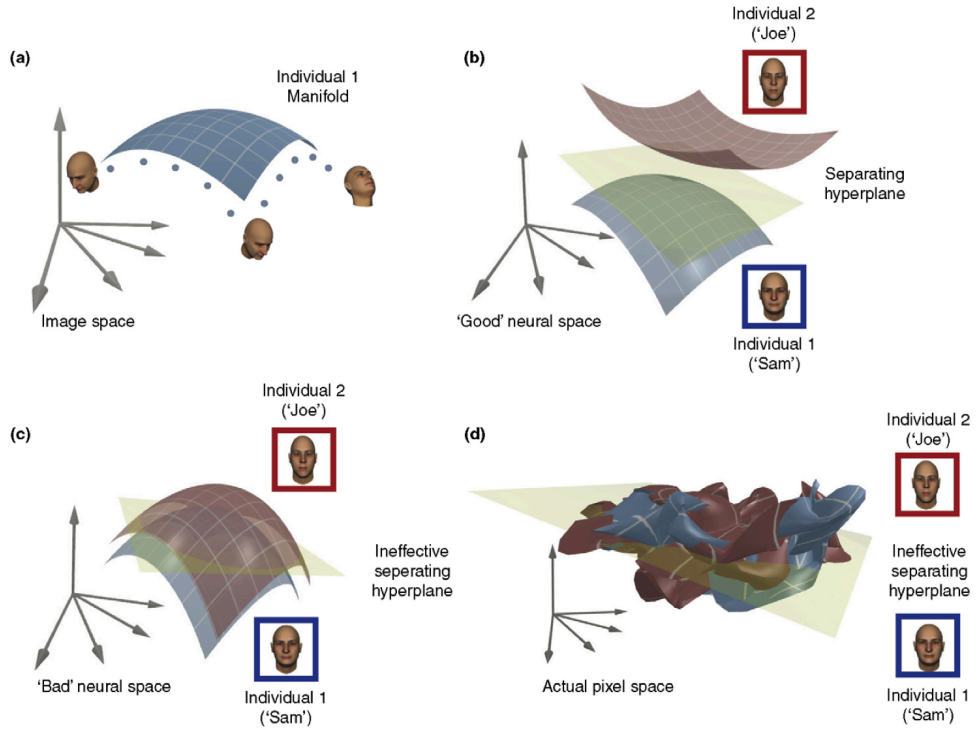

Figure 5. We can't accurately represent a high-dimensional space in a graphic form but a 3D approximation serves as a good illustration. a) The representation of an object (in this case a face) is "traveling" trough the high dimensional space acquiring different coordinates every time this object changes its parameters (in this case a pose). The manifold represents all of the possible variations in the object appearance. b) If two manifolds are disentangled enough a clear decision (hyper-) plane can be drawn between them. c) In case the manifolds are still entangled the decision plane can no longer separate the manifolds, no matter how it's tilted or translated. d) Manifolds generated using the real data (14,400-dimensions, 120x120 images) of two face objects (generated from mild variation in their pose, position, scale and lighting). Even with just a fraction of typical real-world variation, the object manifolds are hopelessly tangled. Adapted from DiCarlo and Cox, 2007

The goal of all the computations is to flatten the manifold as much as possible making the separation easier. What authors suggest is that flattening happens at every stage of signal propagation meaning that V1 neurons do a flattening of a manifold coming from the LGN, then V2 does the same to V1 input, until the signal reaches IT cortex where separation of the manifolds is possible. The claim is additionally supported by

the manifolds produced by simulated V1 cells (Gabor filters) and simulates IT cells (unimodal Gaussian functions). While V1 manifolds are super-entangled, IT manifolds tend to be quite flat and easy to separate. Even though it doesn't represent a full picture of what's happening in the brain and it is more suitable for computer vision than biological systems this framework is the best description of the way in which (we think) the brain solves the invariance problem.

1.4 Limitations of the ventral stream framework

We are aware of the existence of feedback projections amongst all of the areas of the ventral stream, but the timescale acquired in the behavioral (<300 ms) and electrophysiological (~ 125 ms, Hung et al, 2005) experiments does not allow for any kind of long feedback loops to occur (like IT to V1 for example). Still, all of this is very much task dependent and it happens on a longer timeframe if it also includes a (prefrontal) top-down regulation like attention shift. There are also many other additional constraints to this generalized model. The "text-book" version of the ventral stream "postulates" the constant increase in the size of the receptive field when we move from area V1 to the IT cortex. But this is not always true. And one of the reasons is the so-called "binding problem" – if an IT neuron has a large receptive field and it is activated by the conjunction of elements A and B, it will be equally activated if AB or (for example) AC and DB are present in it's receptive field. In order to preserve the spatial information IT neurons must either have large but overlapping RF or some of them must have a much smaller RF. Both Ito et al (1995) and Op De Beeck and Vogels (2000) showed big variations in the size of RF inside the IT cortex. Ito reports RF sizes of 24.58° ± 15.78° while Op De Beck reports a minimal RF size of 2.8° and a maximal RF size of 26°. This variation might be explained by the finding that certain IT neurons change the size of RF depending on the size of the stimulus (DiCarlo and Maunsell, 2003) which also doesn't fit with the "canonical" description of IT neurons as invariant to changes in stimulus sizes. The changes are possible because of the complex interactions between the inhibitory and excitatory neurons inside the IT cortex (Wang, Y. et al., 2002). IT neurons are also usually presented as view-invariant but it might actually be a mixed population of view-selective and view-invariant neurons (Rolls, 2000). Things become additionally complicated in the real world situation where IT neurons get exposed to multiple stimuli and a complex background at the same time (unlike the typical lab settings – single stimuli, uniform background). In this case there's a competition between the stimuli present in the neurons' RF and the resulting activation can be either a weighted average of both stimulus presented alone (Chelazzi et al, 1998) or a MAX response function – the response to the two stimuli is equal to the response elicited by the most effective stimulus of the pair (Sato, 1989). Because complex objects tend to

be represented by distant columns inside the IT cortex (Tsunoda et al, 2001) there's an important role of lateral GABAergic inhibition. This inhibition might explain the competitive interaction between the representations of spatially nearby objects as well. Taken together all of these evidence paint to a much more complex picture than just a straight feedforward mechanism the ventral stream framework suggests. If we can see the ventral stream (form V1 to IT) as a trunk, there are also a lot of side branches that are needed to complete the entire tree (which would represent the object recognition) leaving a lot of areas to explore and a lot of questions to be answered.

Chapter 2: Rat as an experimental model for vision research

2.1 Introduction

The monkeys (aka. non-human primates, NHP) have been the predominant choice when it comes to vision research. The reason is our desire to know how our own vision (brain) works and there's no better substitute. But the use of NHP as experimental models has a lot of difficulties connected to it. Depending on the species they usually require large storage facilities, an optimal space specially designed as their playground, they are difficult and cumbersome to handle, they can be aggressive and even dangerous for the experimenter. Application of the most recently discovered experimental methods, like optogenetics, is severely limited, and there are almost no protocols for the use of genetic manipulation (knock in/out varieties) in NHP. All of this also comes with an additional pressure from the society to completely exclude monkeys from the experimental research. As of 2006 Austria, New Zealand (restrictions on great apes only and not a complete ban), the Netherlands, Sweden, and the UK had introduced either de jure or de facto bans (Langley, 2006). Spain became the first country to announce that it will extend (some) human rights to the great apes. An all-party parliamentary group advised the government to write legislation giving chimpanzees, bonobos, gorillas, and orangutans the right to life, to liberty, and the right not to be used in experiments. An incident involving a hidden camera video recently forced Nikos Logothetis, one of the most prominent scientists in the field of vision research, to close down his lab. It is obvious that the "glory days" of vision research on NHP are behind us and that there's an urgent need for an alternative experimental model.

Since rodents (rats and mice) represent the most widespread laboratory animals (80% in EU – Burn, 2008) the logical question would be: "Why not use them?". The problem lies in our uncertainty of rodents' ability to perform "higher order" visual tasks. Rodents are mostly nocturnal creatures with a very low visual acuity (~1 cyc/deg in pigmented rats – Lashley, 1930; Dean, 1981) compared to humans and monkeys which have the visual acuity of 30–60 cyc/deg (Campbell and Gubisch, 1966; Hirsch and Curcio, 1989; Merigan and Katz, 1990) and the other assumption is that they rely mostly on smell (Uchida and Mainen, 2003) and touch

(Diamond et al., 2008). Another big difference is the organization of the neurons in the rodents' primary visual cortex – rodents don't have structures like orientation columns that are typical for primates (Bonin et al, 2011). The reason behind the most of the research performed on rodents' visual system was easy anatomical accessibility compared to the brains of monkeys or cats (either in the developmental studies or in the studies of learning and memory). The rediscovery of rodents' vision can be seen as a coincidence of circumstances mentioned before – development of sophisticated technics for very precise manipulation of neural circuitry in rodents' and an every growing number of obstacles in working with non-human primates. Before reviewing the behavioral evidence on rodent's vision let us first explore in more details the visual system of a rodent which is very prominent in the current literature on vision – the rat.

2.2 Rats' eyes

Unlike primates, rats have their eyes set on the side of their head which enables them to have a broader field of vision (from -40 to 60 degrees in horizontal plane and more than 100° in sagittal plane for each eye - Adams and Forrester, 1968) but much narrower binocular vision, necessary for depth perception (estimated to be around 40-60°, Figure 6).

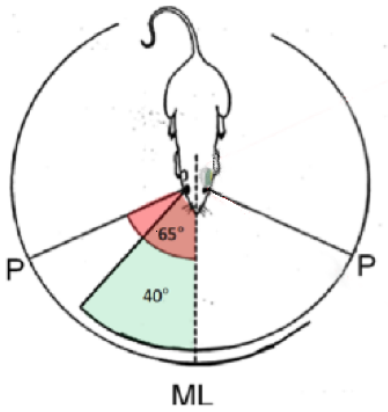

Figure 6. The presumed fields of view of the rat's eyes in the horizontal plane. ML is the meridian which lies in the sagittal plane of the rat. The red angle represents the angle between the sagittal plane of the animal and the fixation point of the pupil of the right eye P. The green angle represent the binocular region of the visual field of the right eye. The left primary visual cortex of the animal is shown (green patch). Adapted from Adams and Forrester, 1968.

We already mentioned the very low visual acuity in pigmented rats but the situations gets even worse in the albino varieties. There has been no systematic analysis of the rat strains until Prusky (2002) compared six varieties (three pigmented, three albino) and found out that some strains (Fisher-Norway) have a substantially higher visual acuity (1,5 c/d), while others (all albino strains) have substantially smaller visual acuity (0.5 c/d), raising an interesting research question – "How did the mutations leading to albinism reduced the visual acuity?". Rats' lens lacks the ability to focus the light (unlike primates) but experiments done by Hughes (1977) demonstrated that constriction of the pupil can considerably increase their depth of perception. Rats' eye (unlike primates) doesn't have a fovea and the receptors are approximately uniformly distributed across the retina. Since rats are nocturnal animals they are much better adapted than primates to low light conditions. They are able to distinguish "total darkness" from the light source of only 0.107 lux (Campbell and Messing, 1969). Almost all of their photoreceptors (99%) are rods (discriminating light from darkness) but they do have cones (discriminating colors) as well (Szel & Rohlich, 1992). Jacobs et al. (2001) performed the analysis of the cones using electroretinograms and behavioral tests and concluded the following: Around 93% of the cones respond maximally to blue–green light (around 510 nm), while the remaining 7% respond to ultraviolet (UV) (around 360 nm) light. Cone responses are normally distributed, so rats actually perceive hues ranging from ultraviolet (400 nm)

to orange-red (around 635 nm) but they are most responsive to colors near their peak sensitivities. And the differences don't end there. The way in which the rats' eyes move is also completely different from the way the primates' eyes function. In the first experiment of its kind, a team of scientists from Max Plank Institute managed to attach a very light (~1 gram) but very precise camera on the rats head and observe the motion of its eyes while they were freely exploring (Wallace et al, 2013). Primates always move their eyes in the typical fashion, whether they compensate for the head movements or when they search around with their gaze – both eyes move together and always follow the same object. The rats' eyes are almost routinely misaligned by as much as 60 degrees. This means double vision is normal for rats. Rats are able to see both above their heads and behind their backs allowing them to have the best possible surveillance of the environment.

2.3 Rats' primary visual cortex

Montero et al. (1973) distinguished seven visuotopically organized areas in the occipital cortex of the rat. The largest of these areas – their primary visual area corresponds well to area 17 (V1) distinguished on the basis of cytoarchitectonic criteria (Krieg, 1946). The neurons in the area V1 are clustered according to their inputs into two populations – mononuclear neurons (integrate information from one eye) and binuclear neurons (integrate information from both eyes). These two populations form two distinct cortical areas – monocular V1 (V1M, positioned medially) and binocular V1 (V1B, positioned laterally). The percentage of binocular cells in the rat primary visual cortex is about 80% (Sefton et al, 2004, Figure 7). Area V1 in rats has the same characteristics as in other types of mammals used for the visual research – their neurons are tuned for direction, orientation, contrast, spatial frequency etc. (Girman et al., 1999). Both cortical areas in V1 (V1M and V1B) have rethinotopic organization with the difference in the size of the RF depending on their position. RF size decreases at the border of the vertical meridian (lower than 3°) but in the periphery it can reach 20°. The average size of the RF is around 13° (Espinoza end Thomas, 1983). Girman experiments showed that almost 95% of the neurons in V1 show some sort of orientation selectivity (either to a flashing on-off stationary stimuli or to the moving gratings). Most of the neurons in V1 are very sharply tuned (with orientation tuning – bandwidth at half height – of ~ 60° or sharper). The orientation selectivity spans a wide range of spatial frequencies ranging from 0 to 1.2 c/d with the peak at 0.08 c/d.

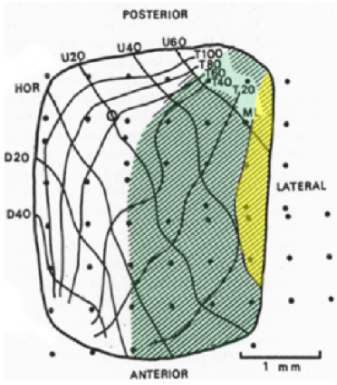

Figure 7. The projection of the visual field on the left primary visual cortex. Abbreviations: HOR: the horizon; U20: a parallel 20° above the horizon; D20: a parallel 20° below the horizon; white circle: position of the optic disc (which is not itself represented). The binocular area, to which the left eye projects as well as the right, is green. Adapted from Adams and Forrester, 1968.

These experiments also demonstrated the existence of both simple and complex cells. All these evidence support the claim that rats' primary visual cortex can perform a very precise visual perception with low spatial resolution.

2.4 Rats' extrastriate cortex

Even though we have the first accounts of the rodents' (mice) exstrastirate areas from the end of 1920's (Rose, 1929), the loss of interest in the visual scientific community for rats and mice as an experimental model meant that we had to wait up until the 70's for a systematic study of rats' exstrastriate areas (Montero, 1973). In their experiments, using electrophysiological recordings, Montero and colleagues partitioned the rats' visual cortex into seven visuotopically-organized areas (as we mentioned before). This was followed by the work of Espinoza and Thomas (1983) that established the most prevalent nomenclature used today for the labeling of the rats' visual cortex so we will further elaborate this one.

The regions are labeled according to their position. Since the cytoarchitecture of the exstrastirate areas in rats can be separated in two regions (labeled 18a and 18b), the regions that belongs to the area 18a (also called *lateral extrastriate cortex*) are: *lateromedial* (LM), *anterolateral* (AL), *laterointermediate* (LI) and laterolateral (LL). The regions belonging to 18b (also called *medial exstrastriate cortex*) are: *anteromedial* (AM) and *posteromedial* (PM) – Figure 8.

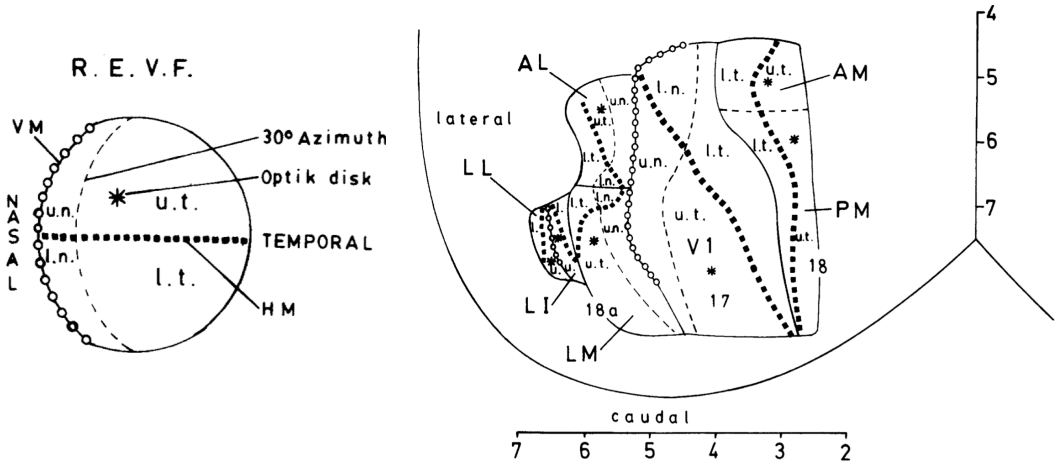

Figure 8 – On the left: R.E.V.F – right eye visual field. (u.n. – upper nasal, l.n. – lower nasal, u.t. – upper temporal, l.t. – lower temporal). HM – horizontal meridian (bold dashed line), VM – vertical meridian (circle dashed line). On the right: dorsal view of the left posterior cortex (and projections of the horizontal/vertical meridian and azimuth) - locations of the areas containing representations of the right eye visual field; primary visual area (V1), lateromedial (second) visual area (LM), anterolateral (third) visual area (AL), laterointermediate (LI) and laterolaterat (LL) visual areas, anteromedial (AM) and posteromedial (PM) visual areas. Asterisk represents the projection of the optic disk (on both sides). Size is in millimeters. Adopted from Espinoza and Thomas, 1983.

In LM the upper VF is represented caudally and the nasal VF medially, being thus a mirror image of V1. In AL 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. In the medial part in AM, the upper temporal VF is medial and the lower temporal VF, lateral (the extreme temporal field is rostral). AM is therefore 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 reversal of VF is one of the hallmark features of the organization of the rats' visual cortex and it's extensively used in electrophysiology to determine the position of the electrode.

As we discussed earlier, topography and partialization are never without issues. For example, a paper by Miller and Vogt (1984) about direct connections of rat visual cortex with sensory, motor, and association cortices (in which they applied anterograde, autoradiographic, and retrograde, horseradish peroxidase, labeling techniques) has no mentioning of the separate extrastriate areas (and only refers to cytoarchitectonic areas 18a and 18b). Malach (1989) used fluorescent tracer bis-bisbenzimide and concluded that "extrastriate band adjoining striate cortex has a single, global map organization" and suggested that "the global map may constitute the rat homolog of area V2 in cat and monkey". He does mention that within this global map "clear modular organization was evident". And, as in the case of monkeys and humans, new data and new experimental technologies lead us to the better understanding of the problem.

One of the ways in which the problem of connectivity and partialization can be improved is to use more than one tracer at the same time. Which is exactly what Montero (1993) did. Using fluorogold (FG) – retrograde tracer (yellow), fast blue (FB) - retrograde tracer (blue), rhodamine dextran – (mostly) anterograde tracer (red) and rhodamine-labeled "beads" – also a retrograde tracer, he was able to track the mutual connections between the striatum and exstrastriate areas by performing simultaneous injections of all of them at once (Figure 9A). By carefully injecting the tracers within specific areas of V1, Montero was able to follow the projections of the different parts of the V1 retinotopic field confirming the previously established "quadratic" rethinotopy (nasal-temporal and upper-lower axes) in the exstrastriate zones. This experiment made an addition of a few more zones to the previous map of Espinoza and Thomas, namely – *anterior* (A) in 18b (titled by Montero as Oc2M) and *rostrolateral* (RL) and *posterior* (P) in 18a (titled by Montero as Oc2L) expanding the number of rats extrastriate areas to 10 (Figure 9B).

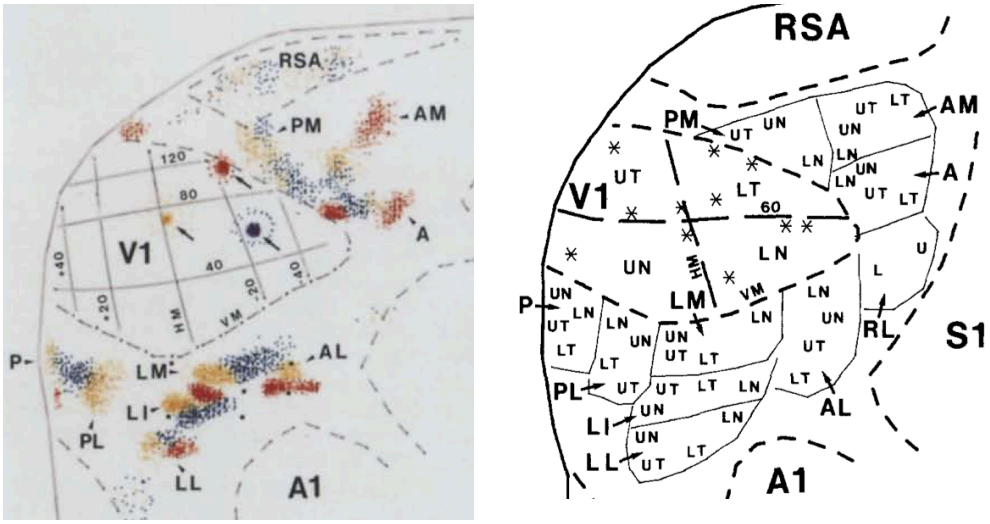

Figure 9. A) Distribution of labeled cells and fibers in the rat extrastriate cortex, after multiple injections of fluorescent tracers in the striate cortex (V1). The position of the injection sites (arrows) of fluorogold (FG, in yellow), fast blue (FB, in blue), and rhodamine dextran (RD, in red) are indicated by arrows (dots indicate maximum extent of injection sites). The distribution of cells and fibers labelled with FG, FB and RD, are indicated with dots of corresponding colors to the tracers injection sites. B) Extrastriate visual areas in the rat - location and quadrantic retinotopy of ten extrastriate visual areas defined by their distinct pattern of striate retinotopic connections. Horizontal meridian (HM) and 60° vertical meridian in V1 define the upper-nasal (UN), upper-temporal (UT), lower-nasal (LN), and lower-temporal (LT) quadrants of the visual field in the primary visual cortex. RSA – retrosplenial agranular cortex, A1 – auditory cortex. S1 – somatosensory cortex. Adopted from Montero, 1993.

As more and more data was collected about the rats' visual cortex, another rodent – the mouse, was being used in order to "fill the gaps" in our knowledge. Studies on mice had a big advantage on their side – the availability of the molecular biology protocols already developed for other type of studies as well as a huge number of different strains. The main downside - poor visual acuity which is around 0.5 c/d (in line with albino rats – Prusky et al, 2000). Considering how phylogenetically close rats and mice are this allowed us to have complementary studies that are even more related than human-monkey studies. Before the experiments of Wang and Burkhalter (2007) the most widely used map of mouse visual cortex was the one published by Wagor et al. (1980) which contained only two lateral areas (labeled V2 and V3) and two smaller medial areas labeled Vm-r and Vm-c. Combining tracing with electrophysiology Wang and Burkhalter substantially changed the parcelization of the extrastriate visual cortex. Their map included 9 exstrastriate areas (Figure 10) and showed a very high level of similarity between the mice and rats (compare figures 9 and 10). On the medial side beside AM and PM (abbreviations are the same as in Montero, 1993) we also have MM (mediomedial) area, while on the lateral side beside RL, AL, LM, LI and P we also have POR

(postrhinal area) – a homologue of PL. Area LL has been re-assigned to area 36p (posterior part of temporal association cortex) while in the front part we also find area A thus completing the picture.

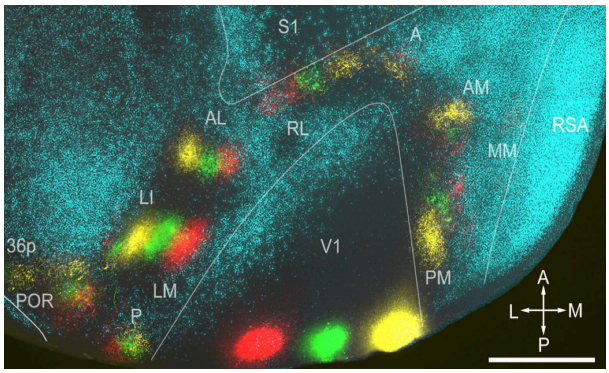

Figure 10. Distribution of labeled cells and fibers in the mouse extrastriate cortex, after multiple injections of fluorescent tracers in the striate cortex (V1). The map was generated by making three simultaneous injections of fluororuby (FR; red), fluoro-emerald (FE; green), and biotinylated dextran amine (BDA; yellow) into V1 and triple anterograde tracing of intracortical connections. Shown in blue are bisbenzimide-labeled callosal connections (marking the boarders between the reagions). Adopted from Wang and Burkhalter, 2007.

2.5 Functional division in rats' extrastriate cortex

But what about the functional roles of these areas? Is there something inside the extrastriate cortices of rat and mouse that would resemble the "streams" we find in the primate vision. There are a few questions to answer in order to make that conclusion. First, let's look at the connectivity among these areas. Since early tracer studies, it became clear that a hierarchy could be established among rat visual cortical areas, based on their pattern of connectivity. At the most qualitative level, this was suggested by the fact that, although V1 was reciprocally connected with all the extrastriate areas (Olavarria and Montero, 1981), the strength of these connections was not uniform: for instance, V1 projections were much denser to the adjacent lateral areas AL and LM than to the far lateral area LL (Olavarria and Montero, 1984; McDonald and Mascagni, 1996). A more rigorous analysis determined that V1, LM, AL and the medial areas (AM and PM) occupy increasingly higher ranks in the visual cortical hierarchy, while the status of the more lateral areas (LI and LL) remained more ambiguous (Coogan and Burkhalter, 1993, 1990). On the other hand, McDonald and Mascagni found that the most lateral extrastriate region (presumably LL) was the only visual area with direct projections to the amygdala, which, in primates, receives strong inputs from the temporal areas of the ventral stream. The same study also reported projections to temporal, perirhinal and entorhinal cortex from both lateral and medial extrastriate areas, as well as projections to parietal areas, especially from the anterior medial region (presumably AM).

When it comes to studies in mice, two anatomical studies that quantified the strength of the reciprocal connections among mouse visual areas confirmed (and additionally strengthened) these findings. The first one (Wang et al., 2011) has shown that LM projects more strongly than AL to lateral extrastriate areas (LI, POR and P), to temporal association cortex (e.g., area 36p) and to lateral entorhinal cortex (LEC), but is also heavily connected to AL itself and to some medial visual areas (AM and PM). Conversely, AL projections are stronger to medial and parietal extrastriate areas (AM, RL and A), as well as to medial entorhinal cortex (MEC), motor cortex and somatosensory areas. These findings suggests that LM and AL play the role of gateways to, respectively, the ventral and dorsal streams, although the strong connections of LM to AL, AM and PM extended these conclusions, by showing that

mouse visual cortical areas can be segregated into two groups of strongly reciprocally connected areas (Wang et al., 2012). The first group includes V1 and the lateral areas (LM, LI, P and POR), and is strongly linked to ventral targets, such as temporal cortex, hippocampus and parahippocampal cortex. The second group includes AL, as well as the medial and parietal areas (AM, PM, RL and A), and is strongly connected to somatosensory, motor and prefrontal regions.

These findings go very well with the assumption of the existence of ventral and dorsal stream in rodents but what about the other crucial requirements – tolerance and selectivity? As we stated earlier one of the "classical" features of the ventral stream as defined in the primate studies is the (general) increase of the neurons' RF sizes. This has been reported both for the rats (Espinoza and Thomas, 1983; Vermaercke et al., 2014; Tafazoli et al 2017) and for the mice (Wang and Burkhalter, 2007; Van den Bergh et al., 2010). Another indicator of hierarchical processing is the increase of response latency, from V1 toward progressively more lateral extrastriate areas, which has been found (again) in both rats (Vermaercke et al., 2014; Tafazoli et al 2017) and mice (Polack and Contreras, 2012). The existence of the putative dorsal stream has been demonstrated in mice by two imaging studies (Andermann et al., 2011; Marshel et al., 2011) showing that neurons in AL, AM and RL have preference for higher temporal frequencies (TFs) and lower spatial frequencies (SFs), as well as sharper direction tuning compared to V1. And when in comes to the ventral stream the most recent results in rats (Vermaercke et al., 2014, Tafazoli et al 2017) lead to a conclusion that there's a very promising candidate for this role - V1→LM→LI→LL.

Taking all of this into consideration, it is safe to conclude that rodents do represent a good model for the analysis of the neural circuitry that underlies complex visual processing. One thing that should be stated is that the investigation of high-order shape and spatial processing by rodent cortical areas is still in its infancy (primate studies have more than half a century "advantage" and even there we still don't have "clear cut" answers) and that we need a lot more data in order to fortify the rodents status as a model of choice for visual studies. We can, however, make a claim that we are on the right track.

2.6 Behavioral experiments on rats' vision

Although he was not the first to try to use the rats in visual experiments, Karl Lashley was the first to systematically explore the visual acuity of different strains of rats (albino vs pigmented, 1930a) but also the first to clearly demonstrate rats' ability to recognize shape (1930b, 1938). Lashley's experiments established that: rats can differentiate different shapes like triangle, circle, and cross (to name a few) either in direct comparison with each other or intermingled within the cluttered background. Experiments also showed that this ability doesn't suffer from the increase of luminosity in figure or background as long as there is no inversion in luminosity of these two. The conclusion was that: "there must be some primitive generalization of form which goes beyond the recognition of identical elements" (Lashley, 1938). Lashley also tested rats' tolerance to in-plane rotation (which he reported as "limited") but also how the occlusion of certain portion of the images impacts rats' decision making – he concluded that rats can use local features like the distance of the object from the frame and ratio between the frame and the stimulus. These experiments set the narrative for the further research of "higher" visual capabilities of the rats that persist up to this day. The key questions we are asking ourselves today are the same Lashley tried to answer: "What are the key features underlying the discrimination of visual patterns in rats?", "Do rats only use low-level features of the images (brightness, contrast) when they perform a recognition task?" and "Do rats form some sort of a Gestalt-like perception in their brain". And all of these questions are very complicated to answer.

Before being able to answer some (any) of these questions, scientists had to acquire a much brother knowledge about the ways in which the rats' visual system works – not available to researchers like Lashley and Krechevsky (1938a, 1938b). They also had to construct much better experimental rigs that would allow the rats to perform a much bigger number of trials (unlike the jumping apparatus that Lashley developed). But all of this would have to wait for quite some time since (with some rare exceptions – Sutherland et al. in the 60's – 1961, 1962a, 1962b and Gaffan et al. at the end of the 90's – 1996, 2000) the scientist involved in the exploration of vision completely lost interest in the rats as potential experimental models switching to monkeys and small carnivores (cats) instead. The "big comeback" of the rodents

happened with the introduction of molecular manipulations like optogenetics (Deisseroth, 2011) and two-photon imaging (Okhi et al, 2005; Greenberg et al, 2008) which coincided with previously mentioned problems with monkeys. This "new wave" of behavioral research on rats' vision was in a much better position to provide the answers to the questions we stated above. Our understanding of the functioning of rodents' visual system came a long way since the first half of the 20th century and the all-pervasive digitalization enabled us to construct experimental rigs where rats can perform a big number of trials giving us a much needed statistical rigor for our conclusions. But the answers remain inconclusive and there's still a lot of debate within the scientific community. The most convincing evidence that rats are able to perform high-level visual tasks comes from the work of Zoccolan and his group. In the first paper titled "A rodent model for the study of invariant visual object recognition" (2009), Zoccolan and colleagues systematically addresses some of the major shortcomings of the previous experiments:

1) Presenting 2 stimuli at the same time in the two-alternative forced choice task. This has been a long known problem, which severely limited the conclusions of the previous studies. When rats are presented with two stimuli at the same time, this gives them the opportunity to directly compare these stimuli and use low-level features (brightness, luminosity, contrast) in order to solve the task. Presenting only one stimulus at the time (on the screen) mitigates this problem – whatever kind of strategy the rats are using in order to solve the task it's happening on the level of memory, not on the level of retina.

2) The design of the experimental rig allowed for high throughput experiments, giving the rats a chance to perform a substantial number of trials in every training session (hundreds of trials in the space of an hour, comparable with monkeys' performance) while (at the same time) insuring the high consistency in every performed trial - previous visual experiments with rats didn't control for rats' head position allowing for a possibility that transformation tolerance might occur because of the visual compensations that rats would achieve by changing their viewpoint. In this experiment rats had to assume the (almost) exact same position on the beginning of every trial insuring that the stimuli they were presented with were always seen from the same viewpoint.

3) Because of the possibility to test the rats on a very big number of trials, this provided an opportunity to test a variety of different transformations necessary to

substantiate the claim that rats can perform ”higher-order” visual tasks. Experiment included size transformations, rotations in plane and in depth as well as elevation and lightning variations. Most importantly, the calculated differences (on the pixel level) were much bigger for the transformations of the same stimulus then between the default stimuli.

All of this allowed the author to conclude - “Our study provides systematic evidence that rats are capable of invariant visual object recognition, an advanced visual ability that has only been ascribed to a select few species.”

Still, one of the “shortcomings” of the paper was that it didn’t offer any suggestions about the way in which the rats achieve this transformation tolerance. And this was the next thing to be done. In the follow-up of this paper Alemi-Neissi and Rosselli (2013) used the same experimental procedure as Zoccolan in order to try and establish the strategy the rats apply in order to solve this complicated visual task. Here’s the key difference – once the rats were sufficiently trained to do the recognition task (together with all the variations included in the previous experiment), they were exposed to the masked versions of the stimuli they’ve already encountered. The masks were “punctuated” by a number of transparent “patches” (with a Gaussian distribution of transparence – i.e. the further away from the center of the patch the less transparent the patch). The number of patches was regulated by the rats’ performance (making the task easier/harder by increasing/decreasing the number of patches). In the end, masked trials were separated according to the rats’ answers (object 1 or object 2) and then summed up in order to obtain a (so called) *Classification Image*. This way the authors were able to “track down” the parts of the objects (aka features) that were mostly present in the case of correct/incorrect recognition. The features that were the most “diagnostic” (present in the case of the correct answer) for a certain stimuli were labeled as “salient” while those features that led the rats to incorrectly identify the stimulus were labeled as “antisalient”. The analysis of the salient and antisalient features provided an insight into rats’ recognition strategies and was able to show the following: rats are able to develop a very sophisticated recognition strategy that can’t possibly be explained by the use of low-level features. Salient regions were (to a large extent) preserved across different transformations leading to a conclusion that rats do form some sort of Gestalt-like concept of the object in their memory. Also, these salient regions weren’t located on a specific position on the screen, thus eliminating the possibility that rats’ strategy could be based only on paying attention to the

specific bright/dark patches instead of a full stimulus. Position of the saliency features demonstrated that rats (equally well) use both upper and lower part of their visual field, which is in a direct contrast with the previous findings of Minini and Jeffery (2006) and Vermaercke and Op de Beeck (2012). Another important finding was the diversity of rats' strategy – some rats had more and some had fewer features they were relaying on. This experiment effectively replicated the results of the Zoccolan's 2009 paper but also gave us many additional informations about the underlying strategy of the rats' object recognition.

Although very robust, these findings did receive some criticism, especially when it comes to the claim that rats do indeed use shape for object recognition (Bossens and Op de Beeck, 2016). There are two reasons for this: first – the term shape is "ill-defined" at best and there's no agreed upon definition that everyone in the visual community adheres to (for discussion see Baldassi et al, 2016), and second – the *Bubbles method* (Gosselin and Schyns, 2001), i.e. the classification image approach used in this experiment, has its own limitations. In his paper from 2004 (aptly named "Troubles with Bubbles") Murrey discusses the limitations of the Bubbles Method and concludes that: "bubbles (classification) image does not completely recover an observer's template, but only the parts that correspond to nonzero locations in the ideal template" and it also points that a "more serious problem with the bubbles method is that showing only small fragments of a stimulus will often change an observer's behavior compared to when the stimulus is shown intact" while "a bubbles (classification) image, which is calculated from responses to small fragments of a stimulus, may not only provide an incomplete characterization of a system's behavior, but a misleading one." And we tried (to the best of our abilities) to address these issues in our experiment.

2.7 The goals of our experiment

The first thing that we wanted to address in our study (that was lacking in the previous studies) is the rats' ability to recognize the *target* stimuli (in a two-alternative forced choice task) when exposed to a big number of *distractors* (instead of just one used so far). We first wanted do see if the rats will be able to do the task at all and then what will be their recognition performance depending on a given distractor. The stimuli in the distractor group were carefully selected in order to span a different amount of overlap with the target stimulus (some had a very small overlap while the most complex stimulus had an almost complete overlap with the target stimulus – see Materials and methods). We also wanted to compare these new findings with the results of the previous paper from Zoccolan's group (Alemi-Neissi et al, 2013; Rosselli et al, 2015) to see if the more complex task produces a more robust recognition strategy in rats. Since this paper already observed a significant inter-subject variability in rats' performance and recognition strategies, we hypothesized that a more complex recognition task will make this distinction even more explicit. Next, we wanted to address the previously mentioned problems with the "Bubbles method" and in order to do this we developed a specifically designed classification image protocol that required an identification of a deformed reference stimulus as either a target or a distractor (see Materials and methods) which accomplished two things at the same time:

1) It didn't add any external noise to the stimuli allowing the rats to see a complete 3D structure in order to make a discrimination, and
2) Instead of just analyzing the visual space limited by the borders of the original stimuli this protocol allowed us to test the entire visual space which proved to be essential for the recovery of the recognition strategy (see Results)

To assess the robustness of the rats' strategy we introduced three different variations of the transformed target stimuli (at 30° of visual angle, at 25° of visual angle and a contour version of the transformed target at 30° of visual angle – see Materials and methods). This allowed us to produce three different classification images in order to compare the consistency of the rats' strategy under various conditions (see Results).

Finally, we wanted to be able to quantify and predict rats' recognition performance. In order to do that we modeled rats' performance using logistic regression and compared these results with the actual performance (see Results).

Chapter 3: Materials and Methods

3.1 Subjects

Six adult male Long-Evans rats (Charles River Laboratories) were used for behavioral testing. At the arrival, animals were around 8 weeks old and weighted approximately 250g. During the course of the experiment, animals grew to over 500g. Dietary routine was the following: rats had free access to food but were water deprived. Within the time of the training session (approximately 1-1.5 hours) they would receive 4-8 ml of pear juice as a reward, as well as 1 hour of water ad libitum post training. All animal procedures were in agreement with international and institutional standards for the care and use of animals in research and were approved by the Italian Ministry of Health: project N. DGSAF 25271 (submitted on Dec. 1, 2014) was approved on Sep. 4, 2015 (approval N. 940/2015-PR); project N. 933-X/10 (submitted on Feb. 216, 2012) was approved according to the legislative decree 116/ 92, article 7.

3.2 Experimental rig

The experimental rig consisted of six operant boxes (i.e., two racks with three boxes each) allowing the parallel training of 6 subjects in daily sessions of about 1.5 h. Each box was equipped with: (1) a 21.5 inch LCD monitor (Samsung 2243SN) for presentation of visual stimuli, with a mean luminance of 43 cd/mm2 and an approximately linear luminance response curve; (2) an array of three stainless steal needles (Cadence Science) connected to three capacitive touch sensors (Phidgets 1110) that were used for collection of behavioral responses and delivery of reward; and (3) two computer-controlled syringe pumps (New Era Pump Systems NE-500), connected to the left and right feeding needles, for automatic liquid reward delivery (Figure 11). Access to the sensors was allowed trough a hole in the wall (3cm in diameter) facing the stimulus display that enabled the rat to extend his head out of the box and frontally face the monitor at approximately 30cm from its eyes. The position of the sensors' array enabled us to reliably constrain the distance between the head of the rat and the display, insuring that the retinal size of the stimuli was almost constant across all of the trials. This was confirmed by a video recording of the rat performing

the task, as reported in a previous paper from our group using the same experimental rig (Alemi-Neissi et al, 2013).

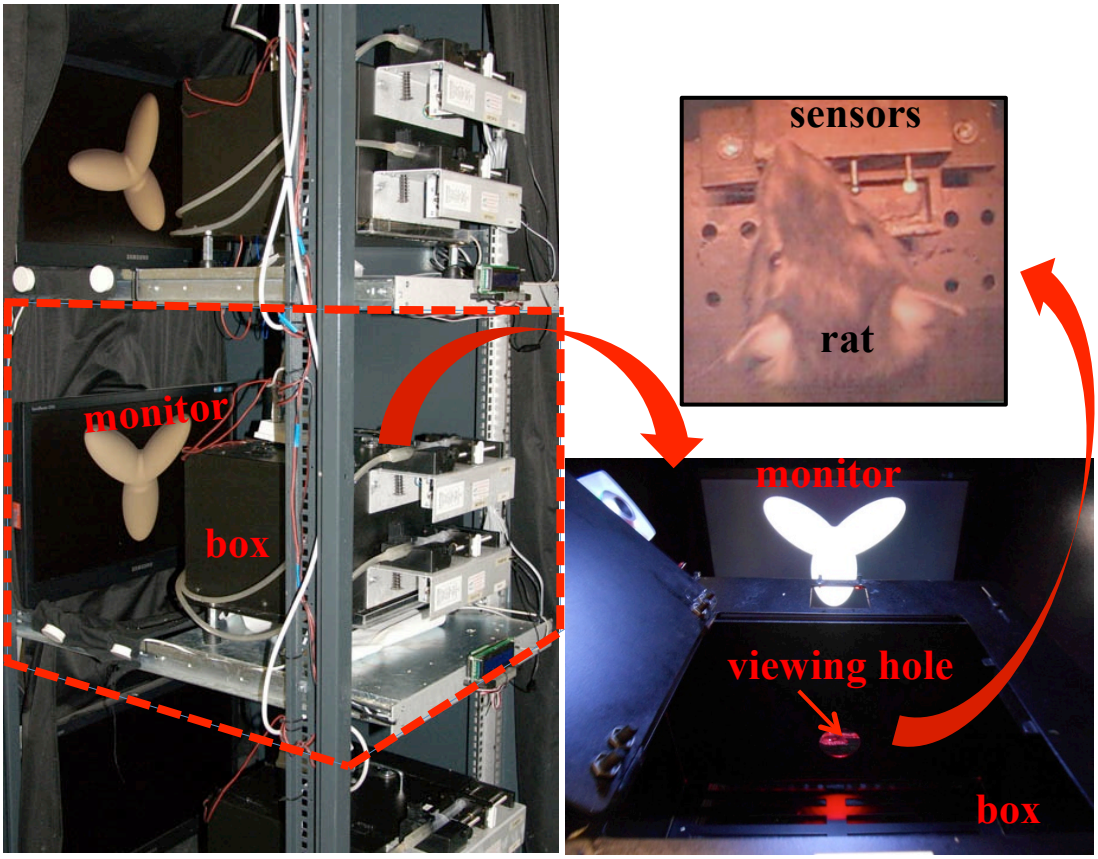

Figure 11. The experimental rig – two cabinets with three operand boxes each. One cabinet is presented here. The side view on the left demonstrates the spatial distribution of the components while on the right we see a top-down view of the operand box with a viewing hole in the middle. Rat extends his head trough the hole and interacts with the sensors.

3.3 Visual stimuli

In this experiment, the rats were trained to discriminate a reference object from 11 distractor objects. Both the reference and the distractors were stimuli previously used by our group either in behavioral (Zoccolan et al. 2009; Alemi-Neissi et al., 2013; Tafazoli et al. 2012) or neurophysiological (Tafazoli et al, 2017) investigations of rat object vision (Figure 12A). The *reference* was an artificial shape made of three lobes, approximately equally sized and arranged in a tripod-like, Y-shaped configuration. The *distractors* were a mix of artificial shapes and renderings of computer-graphics models of natural objects. Objects were rendered using a ray tracer program POV-Ray (http://www.povray.org). Light source position parameter in the rendering of all the objects was kept constant, and the objects were rendered

approximately equal in size (i.e., diameter of a bounding circle) along either the vertical (height) or horizontal (width) dimension. All the objects were rendered against the black background. The default size of each object (i.e., the size used for the initial training of the rats) was 35° of visual angle and they were all presented in the center of the screen. The distractors were chosen based on the (broad) range of image-level similarity with the reference, i.e. how much they overlapped with the tripod object (Figure 12B). This allowed distinguishing the rats based on their level of success with the various distractors. In addition to the default size, both the reference and distractors stimuli were also shown at different sizes (ranging from 15° to 35° in steps of 2.5°). As mentioned in the introduction, random variations of the reference (tripod) object were also shown to the animals in the later stages of the experiment. These stimuli (named *random tripods*) will be described in the section devoted to Phase IV of the experimental design.

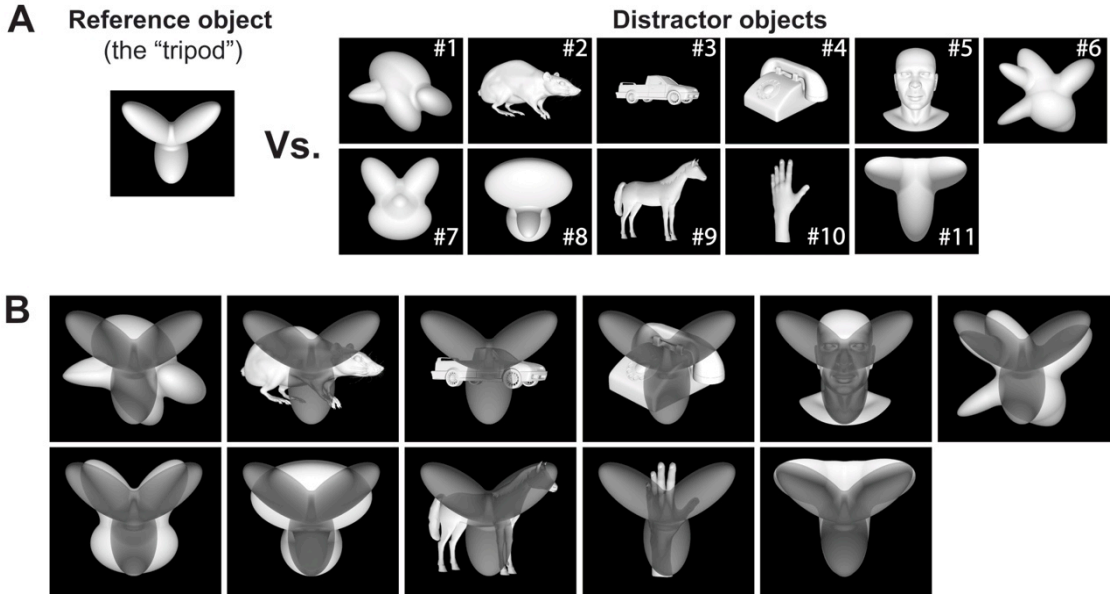

Figure 12. A) Reference object (the "tripod") and eleven distractor objects (a combination of artificial and natural stimuli) were used in the experiment. Since this was a two forced-choice task the rats had to learn to distinguish the tripod stimuli from "everything else". B) Distractor stimuli were chosen based on the amount of overlap with the tripod stimuli. Some had a very small amount of overlap (like distractors #2 and #3) while others completely overlapped with one of the tripods lobes (like distractor #9) or all of them (like distractor #11).

3.4 "Shaping" procedure

One of the most important aspects in the behavioral experiments with animals is their introduction to the experimental setup. Experimental setups represent highly artificial environment that can be stressful for the animals, especially at the beginning

of the training. Even though the "shaping" procedure is often omitted from the "Materials and Methods" section, it can have a crucial impact on the successful training of animals in a perceptual task (especially at the onset of the training), possibly leading to low recognition performances, as well as the small number of performed trials, if not done properly (see Results, Phase I). Not reporting details about the "shaping" procedure can also lead to the assumption that animals were completely naive at the beginning of data collection (which is almost never the case). Therefore, it is worth explaining such procedure in details in this section of the thesis.

Rats were first given a few days to introduce themselves to the environment, in order to know that they can feel safe inside the operant box. The next task was to introduce the animals to the sensor's array. To this aim, the central needle was covered with a sugary treat, so that rat could learn to associate the licking of the central needle with the stimulus presentation. To help them form an association between stimulus identity and the response needle (either left or right), a free reward was delivered (on that needle) as soon as the stimulus was presented (keeping the stimulus on screen to strengthen the association). During the initial shaping phase, to emphasize the existence of two categories of stimuli, the reference stimulus (the "tripod") was kept on full luminance at all times, while the first two distractors the rats were introduced to were dimmed out starting every session at 10% luminance contrast and then gradually increasing during the session (by 10% every 10 trials, reaching full luminance by the end of the session). As the unconditional (free) reward was gradually removed, we introduced a reinforcing stimulus on incorrect trials (consisting of a failure-signaling tone, paired to the monitor flickering from black to middle gray at a rate of 15 Hz). This stimulus was initially aversive for the rats, leading some of them to stop performing. Once the rats were sufficiently "shaped up" and accustomed to the reinforcing stimulus, the data recording begun. Since the "shaping" procedure lasted a certain amount of time (10-15 sessions) during which the stimuli were presented, the rats already acquired some level of stimulus discrimination before the data recording started.

3.5 Experimental design

3.5.1 Phase I

The primary task of the Phase I was to teach the rats that there is more than one stimulus assigned to one port. In order to make the task easer at the beginning, only three stimuli were chosen (the tripod – reference object and two distractors - #1 and #8, Figure 12A) since we knew from the previous experiments of our group (Zoccolan et al, 2009; Tafazoli et al, 2012; Alemi-Neissi et al, 2013), that rats could learn to distinguish these stimuli relatively fast. The rats were trained until they've reached a 70% of discrimination performance (stability of performance was assessed across several days). Each trial would begin by the licking of the central sensor (Figure 13), which would initiate the presentation of the stimuli on the screen. As a result of the training the rats learned to associate the right feeding needle with the Tripod stimulus and the left feeding needle with these two initial "Distractors" stimuli. Giving a correct response would yield a reward (pear juice) and was accompanied with a "success" sound, while an incorrect response would result in the failure-signaling stimulus mentioned in the previous section (a flickering screen from black to gray at 15Hz interval, for 1-3 seconds, paired to a "failure" sound). The standard presentation time was set at 3 seconds during which the rat had to make a response. In the case of a correct reply, the stimulus was kept on the screen for a total time of 4 seconds, while, in the case of an incorrect response, the stimulus would disappear and the flickering would begin.

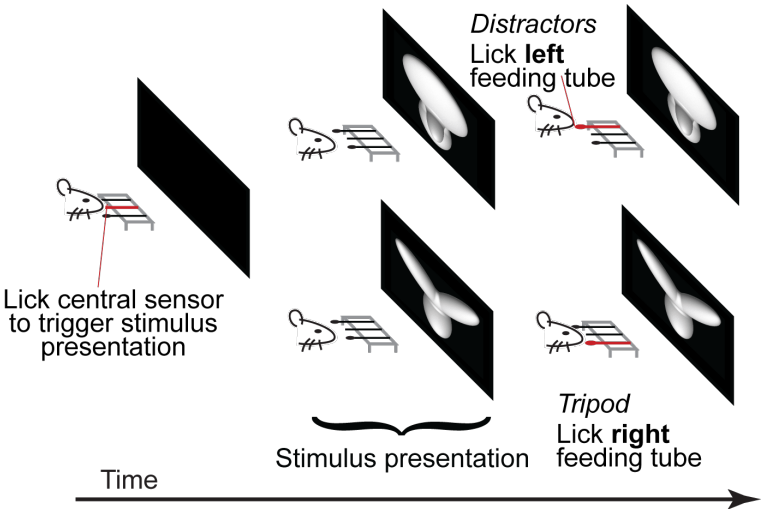

Figure 13. Schematic of the object discrimination task. Rats were trained to initiate the task by licking the central sensor. This would trigger the presentation of the stimuli on the screen. In order to correctly perform the task and earn the reward, rats had to learn to associate the right feeding tube with the reference (tripod) object and to associate the left feeding tube with the distractor objects.

If the rat didn't make a reply in the 3 seconds timeframe, the trial was classified as ignored. To prevent the rats from making very quick, impulsive responses, a trial was aborted if the animal's reaction time was lower than 300 ms. In such cases, neither reward or time-out was administered, the stimulus was immediately turned off, and a brief tone was played. Out of the three stimuli the tripod was presented 50% of the time while the two distractors were presented 25% of the time each (50% total) – Figure 14A. The rats were trained daily and performed (on average) 400-500 trials in a session lasting ~1.5 hours.

3.5.2 Phase II

The goal of Phase II was to introduce the rest of the distractor stimuli. To make this easier, at the beginning of this phase (Phase IIA) we kept the proportion of the previously presented distractors at 25% total (12,5% each) and equally shared the remaining 25% among the new distractors (~2.8% each), Figure 14B. As a result, the overall performance in the task remained above the 70% allowing the rats to keep improving their performance on the "old" stimuli while learning the "new" ones at their own pace. Once the rats achieved >70% of performance on at least half of the "new" distractors, we equaled the percentage of presentation for all of the distractors (~4.55%; Phase IIB, Figure 14C). We continued to train the rats until they've reached a stable performance (over the course of five consecutive sessions) of 70% overall.

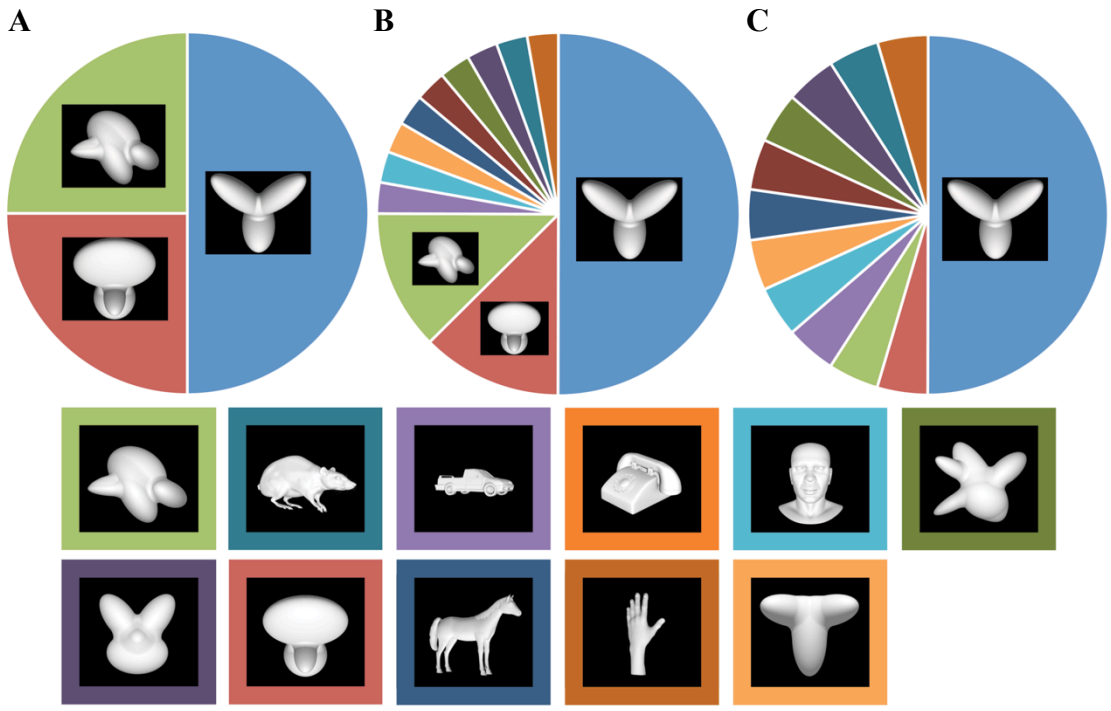

Figure 14. Distribution of the stimuli presentation in the Phase I and the Phase II of the experiment. A) In the Phase I of the experiment the tripod stimuli was presented 50% of the time, while the two distractors equally shared the remaining 50% (25% each). B) At the beginning of the Phase II (Phase IIA) the two "old" distractors were presented 25% of the time (12,5% each) while the other 25% reserved for the distractors was equally shared between the 9 "new" distractors (~2.8% each). C) Once the rats learned to recognize at least half of the "new" distractors with >70% accuracy we equalized the presentation percentage for all of the distractors (~4,55% each, Phase IIB). Stimuli are "color coded" for clarity.

3.5.3 Phase III

In phase III, the size transformations were introduced. As previously shown by Zoccolan et al. (2009), Alemi-Neissi et al. (2013) and Rosselli et al. (2015), rats have a very robust tolerance to size transformations. Since all of these previous studies used only two different objects for the discrimination task, the goal of this phase was to establish how well the rats were able perform when they encounter a high number of stimuli which are size transformed. All of the stimuli were presented in the range of sizes from 35° of visual angle to 15° degrees of visual angle in the steps of 2.5° (Figure 15). As already mentioned in the introduction, rats have very low spatial acuity. This led to some interesting results at the sizes lower than 22.5°, where rats vision reaches it's limits (see *Results*).

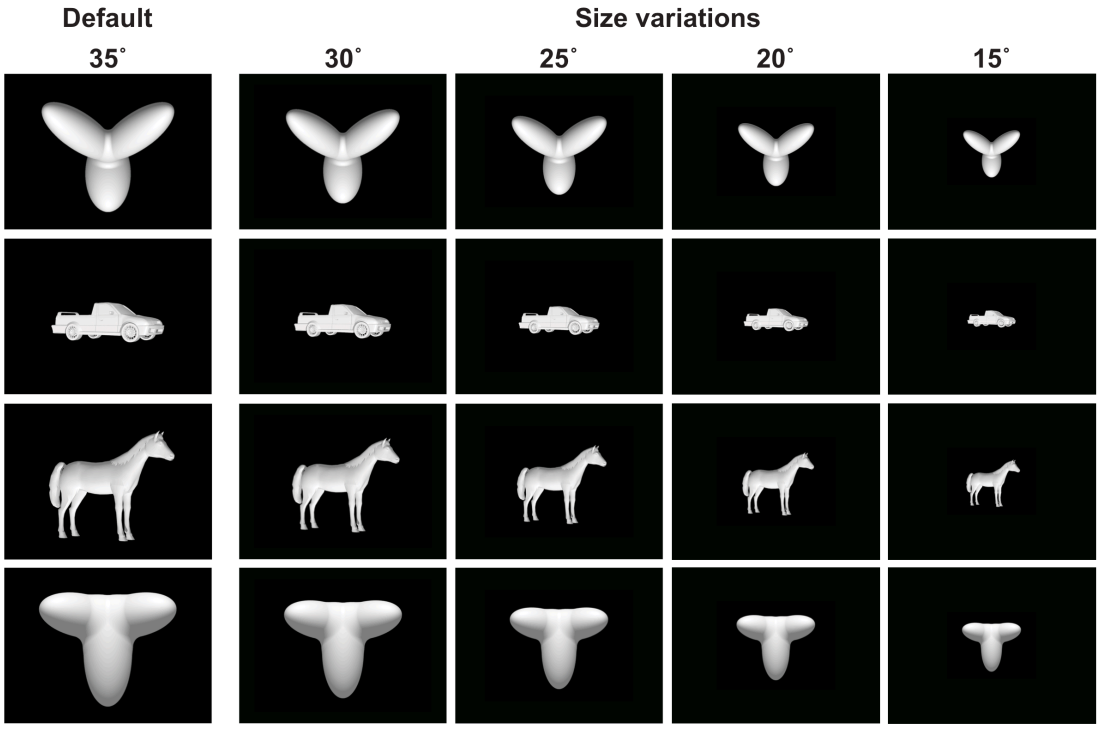

Figure 15. Preview of the default view and the size transformations. The "half sizes" (.5) and the rest of the distractors are left out for clarity.

Considering that we had eleven distractors and that 50% of the stimulus presentation was dedicated to the reference object, heaving additional nine size variations meant that we had 99 different conditions in the space of 200 – 250 distractor trials the rats would perform during an average session. This meant that some of the size transformations were presented only once or twice during a typical session, which, in turn, made necessary to perform a large number of sessions in order to get a

statistically robust assessment of the performance. To avoid overwhelming the rats by presenting all of these conditions at once, we implemented a staircase procedure, where we started from the default view size of 35° of visual angle and then allowed the rats to "determine" their own pace of progress down the size transformation ladder. The staircase procedure was implemented in the following way – the rats would start the task at a given size (35° at the beginning of the Phase III) and if they responded correctly to 7 out of 10 consecutive trials, the complete set of stimuli at a size that was 2.5° smaller was included in the presentation pool. To ensure that rats would gradually learn the task and reach the bottom of the size range, we took the size that rats have reached in their current session, increased it by two steps (5°), and set that as the beginning point for the next session. For example, if the rats reached 25° at the end of a session in a given day, we would set size 30° as the beginning point for the next session (next day). Once the rats finally reached size 15° for five consecutive sessions, we stopped the staircase procedure and equalized the frequency of all sizes. The training continued until the rats achieved >70% performance, at which point we continued to Phase IV.

3.5.4 Phase IV

After successfully answering one of our main questions – will the rats be able to perform a discrimination task when faced with a large number of different stimuli that, in addition, are also size transformed (Phase III), we continued our experiment trying to understand the perceptual strategy deployed by the rats, when performing this discrimination task. As discussed in the introduction, the Bubbles method used by Alemi-Neissi et al. (2013), Rosselli et al. (2015) and Vermaercke et al. (2012) is a very powerful classification image approach to infer perceptual strategies, but has a few limitations that are inherent to the method itself and cannot be circumvented by the experimental design. In order to find alternative ways to tackle this issue, we looked back at the work of Ahumada (1996), which introduced the concept of classification image into the literature. In a typical two-forced choice experiment that implements the classification image approach, the stimulus in each trial is one of two possible images (signals) in a Gaussian noise field that varies from trial to trial (Figure 16a). The participant (a human or an animal) tries to correctly discriminate which signal is presented. The participant processing strategy can be summed up by the following equation (Figure 16b):

$$c = (\overline{\boldsymbol{n}}^{12} + \overline{\boldsymbol{n}}^{22}) - (\overline{\boldsymbol{n}}^{11} + \overline{\boldsymbol{n}}^{21})$$

where $\overline{\boldsymbol{n}}^{ab}$ and $\overline{\boldsymbol{n}}^{ba}$ are the averages of the noise fields over all misidentified trials (where the stimulus contained signal **a** but was identified as signal **b** or it contained the signal **b** but it was identified as signal **a)** while $\overline{\boldsymbol{n}}^{aa}$ and $\overline{\boldsymbol{n}}^{bb}$ are the averages of the noise fields over all correctly identified trials (where the stimulus contained signal **a** and was identified as signal **a** and where the stimulus contained signal **b** and it was identified as signal **b**). The reason to sum up the stimuli with the different signal but the same response from the observer is the following: although the noise is random, the structure of the signal isn't. Therefore, on some trials, the additional noise might contribute to the correct perception and, on other trials, to incorrect perception. Since there was "something" in the noise to contribute to the perception of the stimuli as signal one (and dissimilar from signal two), it seems likely that $\overline{\boldsymbol{n}}^{11}$ and $\overline{\boldsymbol{n}}^{21}$ will show what are the features leading to this decision (because they represent the averages of noise fields over trials identified as signal one). If this is the case, we can sum up this two in order to reduce the sampling noise. The same will apply for $\overline{\boldsymbol{n}}^{12}$ and $\overline{\boldsymbol{n}}^{22}$, when it comes to signal two. And since these sums represent the averages of

noise fields over trials leading to the opposite responses (therefore the photographic negatives of each other) we can further add the first sum ($\overline{\boldsymbol{n}}^{12} + \overline{\boldsymbol{n}}^{22}$) to the negative value of the second sum ($\overline{\boldsymbol{n}}^{11} + \overline{\boldsymbol{n}}^{21}$) to reduce sampling noise even further, ending up with our original equation.

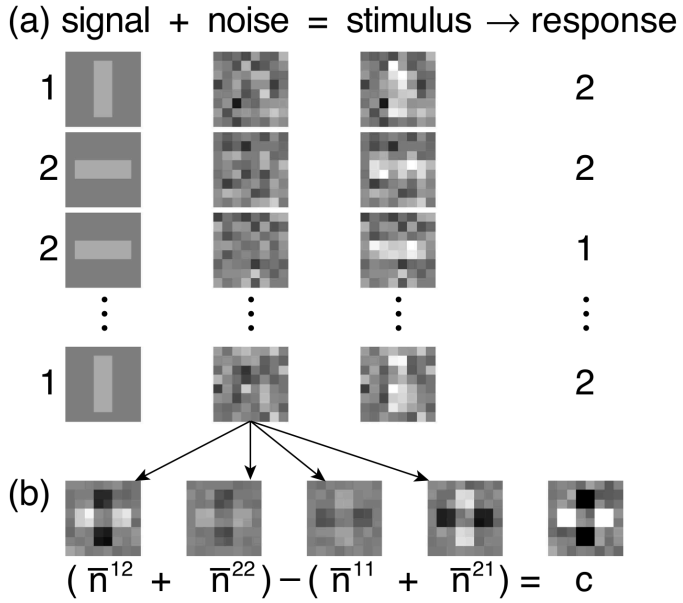

Figure 16. The standard method of calculating a classification image. (a) The experiment: on each trial, a signal and a noise image are summed to produce the stimulus, and the observer generates a response. Sometimes the response is the correct one, sometimes not.
(b) The analysis: the noise fields from each signal - response category of trials are averaged together, and the averages are combined according to the equation we described in order to produce the classification image. Adopted from Murray (2011).

As discussed in the introduction about the Bubbles method, addition of the external noise can change the perception of the stimuli (and even systematically lead to the wrong identification). To avoid this problem, we decided to introduce the "noise" in a structural way, by changing the parameters (size, rotation and translation) of the lobes of our reference (tripod) object (Figure 17). This allowed sampling a region of the shape space, centered over the tripod, in a rather homogeneous way, always presenting whole objects (i.e., structurally altered tripods), rather than partially degraded or masked versions of the reference (i.e., random tripods). Thanks to this design, the resulting classification images were not constrained within the boundaries of the reference or distractor objects (as it would be the case with the Bubbles method) but extended over the whole image plane. The random tripods were randomly interleaved in the protocol used for Phase III with the presentation frequency of 20% total (leaving the remaining 80% of trials equally divided among size transformed reference object and distractors). To maximize the generalization and avoid any learning effect, the trials involving the random tripods were not rewarded or punished in any way (no feedback trials).

Unlike the regular trials, the random tripods were not kept on the screen after the response was obtained and the rats could start the new trial immediately.

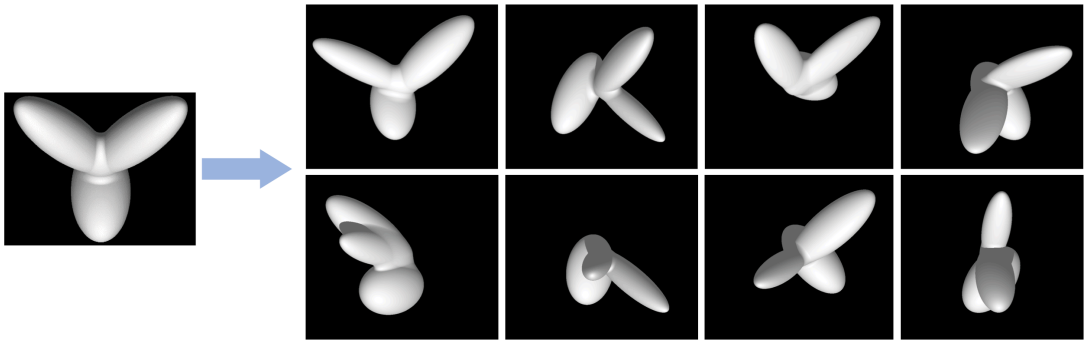

Figure 17. Examples of the random tripods. Depending on the parameters the resulting stimuli could be a slightly modified version of the reference object or the tripod stimulus can be completely "deformed"

In the first "version" of the Phase IV protocol we've set the size of the random tripods to 30° of visual angle. The rats performed 32 sessions (~16000 – 18000 trials) so we could have enough trials to preform a robust statistical analysis. The original equation $c = (\bar{n}^{12} + \bar{n}^{22}) - (\bar{n}^{11} + \bar{n}^{21})$ can be simplified in our case since the signal is only one – the tripod, which can then be classified either as the reference stimulus or a distractor stimulus. Therefore our classification image was computed using the following equation:

$$\mathbf{CI} = \langle \mathbf{I}_{\mathbf{tripod}} \rangle - \langle \mathbf{I}_{\mathbf{not\ tripod}} \rangle$$

(the pixel-wise average of all the stimuli that were classified as a tripod minus the pixel-wise average of all the stimuli that were classified as not tripod, Figure 18).

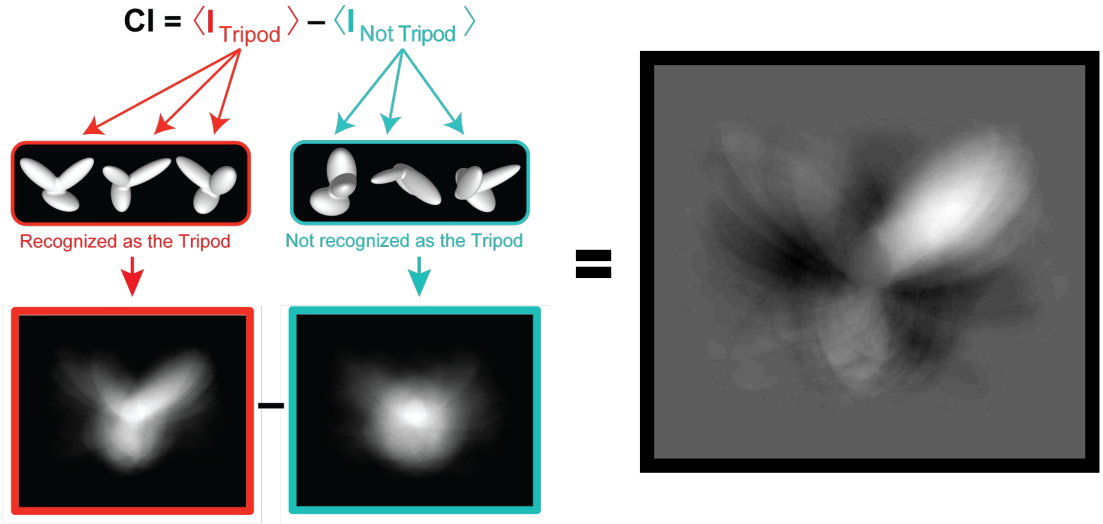

Figure 18. Calculation of the *Classification Image*. We summed up (on the pixel level) all of the stimuli presented in the trials classified as tripod and all the stimuli classified as not tripod and then we subtracted these sums. The resulting classification image is characterized by the black and white regions where the black regions are highly present in the stimuli identified as not tripod while the white regions are highly present in the stimuli identified as tripod.

After computing the classification images, we established their statistical significance at the pixelwise level, by performing a permutation test in the following

way. We obtained 100 randomly shuffled versions of the rat responses and, for each one of these, we computed a 'fake' classification image in the same way we already described. We observe that, in this way, the proportion of the rat responses in the two categories is preserved. We then stacked the fake classification images on top of each other and for each pixel we computed the empirical distribution of its intensity values. We then computed for each pixel the mean and standard deviation of its empirical distribution and fitted the latter with a Gaussian with the same mean and standard deviation. In this way we defined a pixel-wise null distribution against which we could test the corresponding pixel intensity values in the original classification image at a significance level of 0.01. Pixels belonging to the right tail of these distributions where assigned to the "*salient*" part of the classification image and pixels belonging to the left tail where assigned to the "*anti-salient*" part of the classification image, following the "convention" used by previously studies of our group (Tafazoli et al, 2012; Alemi-Neissi et al, 2013, Rosselli et al. 2015).

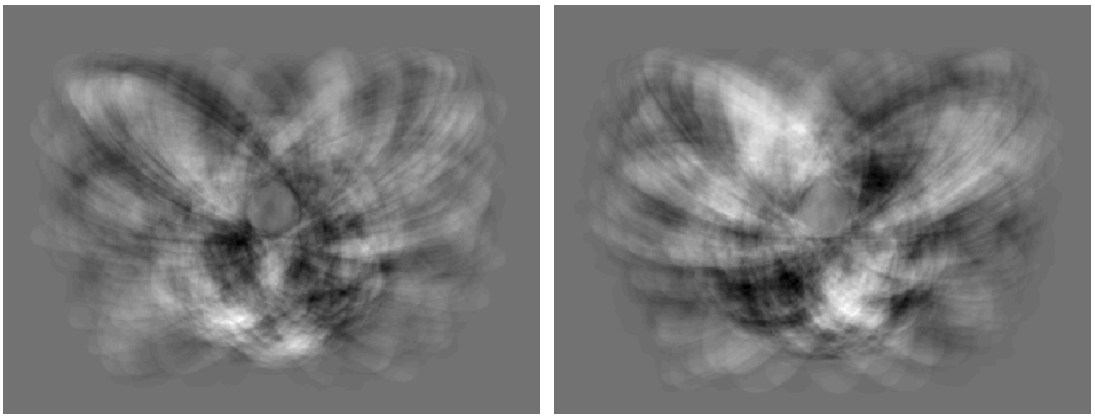

Figure 19. Examples of 'fake' classification images obtained randomly shuffling the rat responses. As expected, classification images produced in this way do not show any kind of emerging "structure" unlike the real classification images

After we finished the sessions with the random tripods at 30° of visual angle, we presented the rats with the same stimuli in their outline version (also at 30° of visual angle, Figure 20, left). The rats also performed additional 32 sessions with the random tripods at 25° of visual angle (Figure 20), completing this phase of the experiment. Based on the computed classification images, we were able to qualitatively infer for the rats' perceptual strategy (see *Results)*, but our ultimate goal was to quantify this strategy, building a predictive model of rat perceptual choices and then test this models' accuracy.

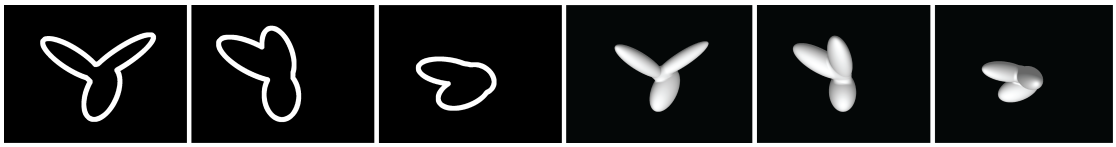

Figure 20. Outline version of the random tripods (at 30° of visual angle) and random tripods at 25° of visual angle.

In order to do that, we combined the classification image approach with logistic regression to construct a generalized linear model to predict the probability of rat choice, when faced with an incoming visual stimulus. We modeled the probability of rat *i* to classify an arbitrary input image *x* as being the tripod, as:

$$p_i(y = 1|x) = \sigma(\theta_0 + \theta_1 \mathrm{CI}_i \cdot x),$$

where y is a binary variable indicating the choice of the animal (1 = "tripod"; 0 = "everything else" category), CI_i is the classification image obtained for rat i, $\mathrm{CI}_i \cdot x$ is its dot product with the input image (Figure 21), θ_0 and θ_1 are the regression parameters, and σ is the logistic regression function. The argument of the logistic function models the evidence that the animal has acquired about the presence of the tripod, based on a bias term θ_0 and the matching between the input image and the perceptual template provided by the classification image (properly weighted by a gain factor θ_1)

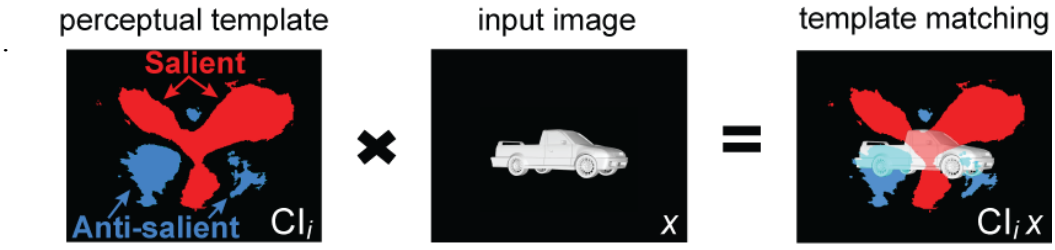

Figure 21. Calculating the dot product between the classification image (perceptual template CI_i) and the input image. The pixel value of the salient regions is intensity based and positive, while the anti-salient regions have intensity based negative values. The background has the value of zero.

The logistic function translates this evidence into the probability of a "tripod" choice (Figure 22). This model was fitted to the responses that the rat gave to the random tripods, using a two-step procedure. First, the classification image CI_i was obtained, as described in the previous section and then the CI_i was plugged into the above equation and the regression parameters θ_0 and θ_1 where found by minimizing the logloss (aka logarithmic loss or cross-entropy loss) using a gradient-based method

(implemented in Matlab (Mathworks) using the *glmfit* function of the *Statistics and Machine Learning Toolbox* with *binomial* distribution). The logloss is defined as the normalized negative log-likelihood of the true responses, given the probabilistic outcomes of the model's prediction predictions on the same input images (Bishop, 2006).

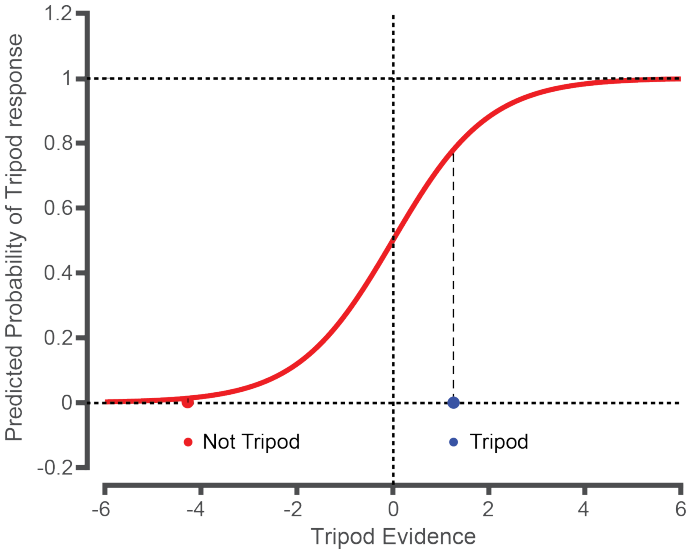

Figure 22. Predicting the probability of a Tripod response after all the parameters are known. Tripod evidence on the x-axis is the solution of the equation $\theta_0 + \theta_1 \mathrm{CI}_i \cdot x$ and presents the amount of "evidence" the rat has obtained in order to perform a classification. In case of a big negative value (red dot) the rat will almost certainly classify the presented stimuli as "Not Tripod". In case of a big positive value (blue dot) the rat will reply "Tripod" with a high probability. Zero represents the point of equivalency – the rat has a 50% chance of replying "Tripod" or "Not Tripod".

The equation for the *logloss* is the following:

$$l(\theta_0, \theta_1) = -\frac{1}{N}\sum_{k=1}^{N} y^k \log p(y^k = 1|x^k) + \ (1 - y^k) \log\left(1 - p(y^k = 1|x^k)\right)$$

where N is the number of samples used to train the model, k is a running index and the dependency of the probability on θ_0, θ_1 is not made explicit. We observe that this equation is composed of two terms. The first term is positive when the response of the rat is 1 ("tripod") and vanishes when the response is 0 "not tripod"; vice versa for the second term, so that the two terms take care of the two possible responses of the rat independently.

When the model predicts a very high probability of responding 1 and the real response was indeed 1, the first term adds a very small contribution to the loss (it vanishes only if the predicted probability is exactly 1, which does not happen in practice, since the sigmoid function reaches the extreme values only when its argument grows or decrease indefinitely). On the contrary, if the model predicts a small probability, the contribution to the loss is high, reflecting the fact that the model is wrong. The more the model is confident in a wrong prediction, the higher is the

contribution on the total loss. Similar considerations can be done for the second term, when the real response of the rat was 0 ("not tripod").

This model was used to fulfill three different goals:

First, we verified that each rat used a unique, subject-specific perceptual strategy. This analysis was carried out by applying a 10-fold cross-validation procedure, where only 9/10 of the responses to the random tripods (the *train* set) were used to fit the parameters of the model (i.e., CI_i, θ_0 and θ_1). The trained model was then used to predict the responses to the remaining 1/10 of the random tripods (the *test* set), and the goodness of such prediction was assessed by the logloss function. Critically, for each rat i, we also fitted five alternative models, using the same procedure described above, but with a key difference – rather than using the classification image CI_i (corresponding to the rat under consideration), we plugged into the equation the classification images CI_j (with $j \neq i$) of the other rats. This allowed testing whether the classification image obtained for a specific rat was able to predict its perceptual choices better than the classification images obtained for the other subjects (see *Results*, Figures 30 and 32)

Second, we used the same regression model (fitted to the full set of random tripods) to predict the rats' performances to all the distractors. Compared to the previous analysis, this was an even more stringent test of the ability of the classification images to predict rat perceptual choices, because the *train* and *test* image sets (i.e., the random tripods and the distractors) were not only different stimuli, but belonged to structurally distinct classes of visual objects (see *Results,* Figures 31, 32, 33). To be sure that the obtained predictions in the cross-rat comparison were not simply a result of overcompensation trough the bias parameter (θ_0) we preformed two additional analyses. We calculated how well the rats' classification accuracy correlates with the model based only on the classification images (see *Results,* Figure 35) showing that we can achieve a robust prediction without using any bias parameter. We also checked the matching between the observed and predicted values for the distractor evidence between the good and the poorer performing rats (see *Results,* Figure 34) in order to demonstrate that variability in rats' accuracy also can't be attributed to the bias parameter. Both analyses confirmed our conclusion that rats' perceptual strategy determines its proficiency level in the task.

Third, since using a fixed template matching still represents a low-level strategy, we used our model to predict rats' performance on the outline and "small size" (25º of visual angle) random tripods. Since our model's prediction drastically failed for the outline random tripods (see *Results*, Figure 36), we performed an additional analysis with the filled in versions of the outline random tripods under the assumption that these stimuli were perceived as solid bodies (see *Results*, Figure 38).

The statistical tests used to assess the significance of our findings are reported in the Results and in the legends of the figures. Here, we only describe in more details the binomial test used to evaluate the overall significance of the comparisons between *same-CI* models and *cross-CI* models in Figures 32 and 38. This test was carried out according to the following logic. We computed the probability of obtaining a number of successes in 6 Bernoulli trials that was equal or higher than the number of times a given model (e.g., the *same-CI* model) outperformed other alternative models (e.g., the *cross-CI* models). In the case of the *same-CI* model vs. *cross-CI* models comparisons (i.e., large colored dots vs. small colored dots in Figure 32 and large black dots and small colored cots vs. small black dots in Figure 38), the chance of success in each Bernoulli trial was set to 1/6. In the case of the *filled-in same-CI*^{o} model vs. *same-CI*^{o} model comparison (i.e., solid green dots vs. empty green dots in Figure 38), the chance of success in each trial was set to 0.5.

Chapter 4: Results and Discussion

4.1 Phase I

From the very early stages of training, it was noticeable that rats show a high level of difference in their performance, both in the recognition strategy and in the number of trials they perform per session. While some rats got accustomed to the (very artificial) surrounding of the operand box, others still had problems performing a sufficient number of trials on a day-to-day basis and varied very much in their recognition performance as well (Figure 22). For example, while the average number of trials for the rat No. 1 was close to 300 (294), it ranged from 123 trials per session to 482 trials per session, more then 400% difference. A similar trend was observed for rat No. 2 – average number of trials was 356, minimum 187, maximum 550. For rat No. 3, the average number of trials was significantly higher – 426, and much more stable but the minimum and maximum were still wide apart – 202 vs. 551. Rat No. 4 varied a little bit less – average 338, min. 205, max 447. Even though rat No. 5 had the best recognition performance, he still varied a lot in the number of trials on a daily basis (average 346, min. 169, max. 519). Rat No. 6 was the one with the biggest number of trials performed (average 530) and it also varied the least (min. 353, max 635). It is important to notice that minimums and maximums don't represent a slow rising trend but a zigzag line, meaning that rats would perform a big number of trials one day and a much lower one the next day. We can say that Phase I was as much about teaching the rats how to constantly perform a desired number of trials as training them to perform the object recognition task correctly.

Still, on average, just like with the performance, there was a steady rising in the number of trials performed, and (as expected) it became more and more stable as the training progressed. If we look at the recognition performance, one thing is apparent from the beginning – all the rats (with the exception of rat No. 5 in the very first sessions) achieved the best performance on the distractor #1 (which persisted throughout the experiment, even when the size transformation were included) reaching > 90% performance in just a few sessions. Qualitatively, this can be explained by distractor #1 having visual features that were easily discriminable from those of the reference tripod object (compare the overlap pictures in Materials and

Methods, Figure 12B). When it comes to the reference object, we can see that it was the hardest one to recognize, for most of the rats - only rats No. 5 and 6 were able to achieve a stable performance over 70% in phase I of the training. The other rats oscillated above

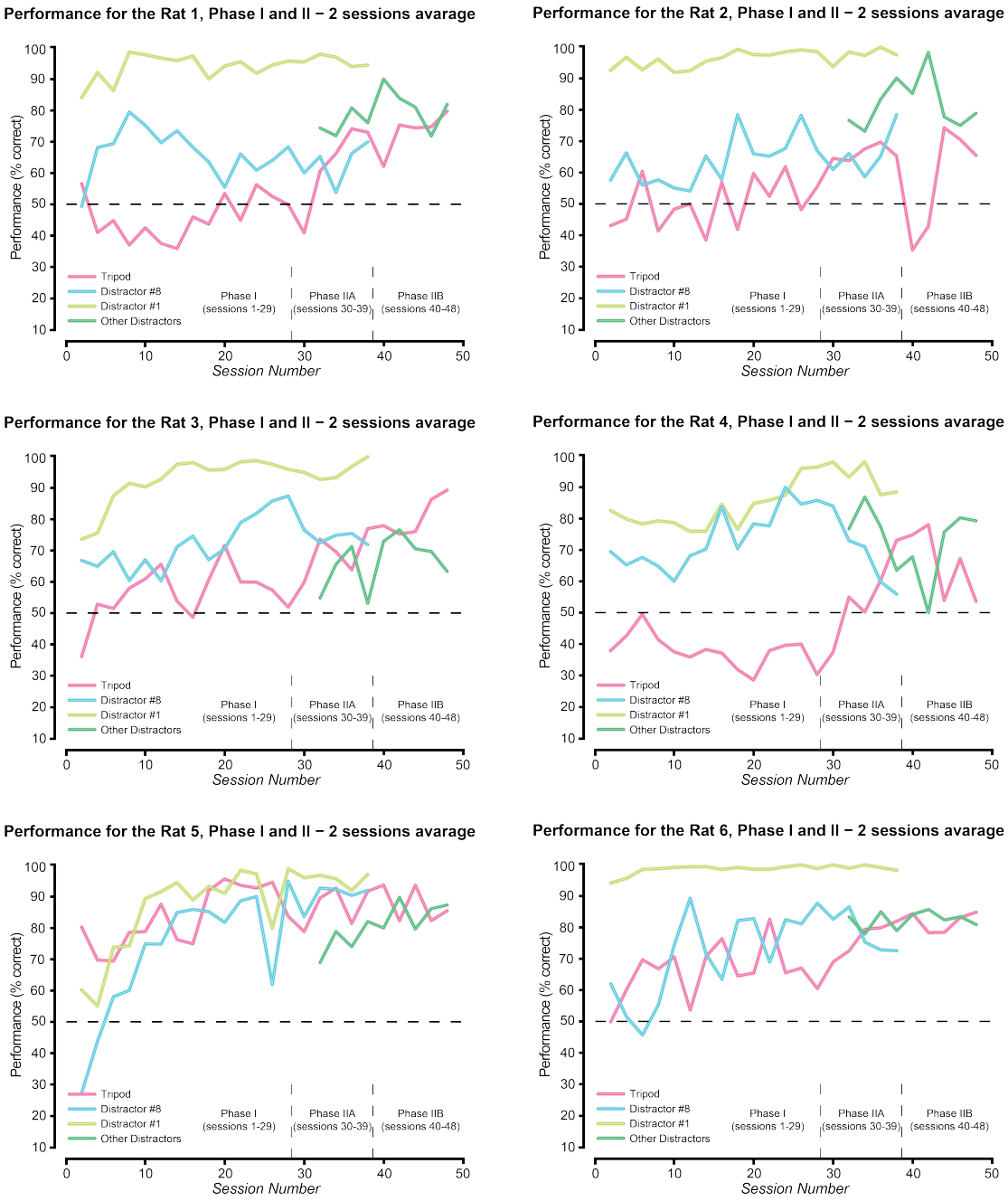

Figure 22. Recognition performance for all the rats, Phase I and II. Because of the variations on a day-to-day level, in order to smooth the curves and better represent the learning trend we plotted a two sessions average performance (instead of every session). For the Phase IIA distractors #1 and #8 are plotted separately from the other distractors and then for the Phase IIB we merged all of the distractors in a single line.

(rats No. 2 and 3) or below (rats No. 1 and 4) chance level, not reaching the desired recognition performance (>70%) until the end of Phase IIA. Since the task was designed from the beginning as "the tipod vs. everything else" and also as the result of the "shaping" procedure (see Materials and Methods), it makes sense that rats could decide to focus more on the "everything else" category (especially because they were so good at recognizing distractor #1), thus maximizing the amount of reward even if they recognized the tripod at the chance level. As we can see, the introduction of the additional distractors in the Phase II prevented this strategy and led to a fast increase in the performance with the reference tripod object. The performance for distractor #8 was always in between the performances for distractor #1 and the reference (except for the rat No. 5, who had the best performance for the reference from the beginning), usually varying around 70-80% for the rats.

4.2 Phase II

The introduction of the whole set of distractors forced the rats to start paying much more attention to the reference object, since the diversity of the distractors now made the recognition task much harder. As a result, by the end of Phase II, all rats reached a performance that was >70% correct on the reference.

The biggest improvement was achieved by the rat with the lowest performance on the tripod object (No. 4), which entered Phase II with just over 30% correct performance and managed to reach almost 80% correct by the end of phase II. Rat No. 3 raised its performance from chance level to 90% correct. Other two rats, oscillating around chance level (rats No. 1 and 2), raised their performance to 80% and 75% correct, respectively. And the rats with the best performance in Phase I (No. 5 and 6) continued to improve – rat No. 6 reached 85% correct, while rat No. 5 achieved an impressive 95% correct. The performance for distractor #1 was already almost perfect (for the most rats) in Phase I, so there was a little space for improvement – all the rats performed between 90-100% correct. For most rats, the performance for distractor #8 continued to rise, but for the rats No. 3 and 4, the introduction of all the distractors led to a drop in performance with this specific stimulus (although this drop was compensated by the rise of performance for the tripod object).

Similarly to the recognition performance, the number of performed trials continued to rise and became more and more stable on a day-to-day basis. The average number of trials for the rat No.1 increased to 375, with only one session below 300 trials. For the rat No. 2, the average number of trials rose to 378 (with only 3 sessions below 300 trials). Rat No. 5 performed on average 393 trials (one session below 300 trials). The other rats (No. 3, 4 and 6) all performed more than 400 trials on average (No. 3 – 426, No. 4 – 422, No. 6 – 513) with no sessions bellow 300 trials. Taken all together, at the end of Phase II, we were at the desired level of recognition performance and trial number for all the rats.

4.3 Phase III

Because of the reasons already mentioned in the Materials and Methods (distractors being presented only 50% of the time and size transformations including 99 different conditions for the distractors), Phase III ran for almost 100 sessions (96) and included tens of thousands of trials for each rat: rat No. 1 – 42613, rat No. 2 – 34793, rat No. 3 – 47173, rat No. 4 – 46121, rat No. 5 – 44720, rat No. 6 – 52486. Since we will provide a detailed presentation of the performance for all the stimuli across all the sizes, together with the classification images, when describing the results of Phase IV, here we will just show the time course of the performance (averaged across 10 sessions) for the reference tripod object (Figure 23) and the distractors averaged together (Figure 24) during the Phase III (including size transformations).

There are several important things to notice here: all rats continued to improve their performance on the reference object, with all the animals (at least at one point) achieving 90% performance at the default size (35º of visual angle) and most of them (excluding rat No. 2) almost reaching the perfect discrimination accuracy. The second thing is the stability of the performance. During almost 100 training sessions, the performance on the reference object varied (except for rat No. 2) by just a few percent. The third thing is the impact of the size transformations. After a very short period of only 10 sessions, in which the staircase was implemented (see Materials and Methods), rats generalized to all size transformations, but the smallest one (108 different conditions), and maintained a very high performance that was equal or slightly smaller than for the default size.

Regarding the smallest size (15º of visual angle), what can be readily observed from the performance plots is the substantially different performance between size 20° and size 15°. As we mentioned in the Introduction, probably the biggest limitation factor for rodents as experimental animals in the vision research is their low visual acuity. And 20° of visual angle seems to be the limit below which the performance drops drastically, at least in such complex shape discriminations as those tested here. Most rats performed bellow chance level at size 15° (except for the rat No. 6, the best performer for the tripod stimulus in this phase, which, even at size 15°, managed to reach 60%+, once again pointing out to inter-subject variability).

This drop in performance for the reference object at small sizes is relatively easy to explain, considering that the task we taught to the animals was “tripod vs. everything else”.

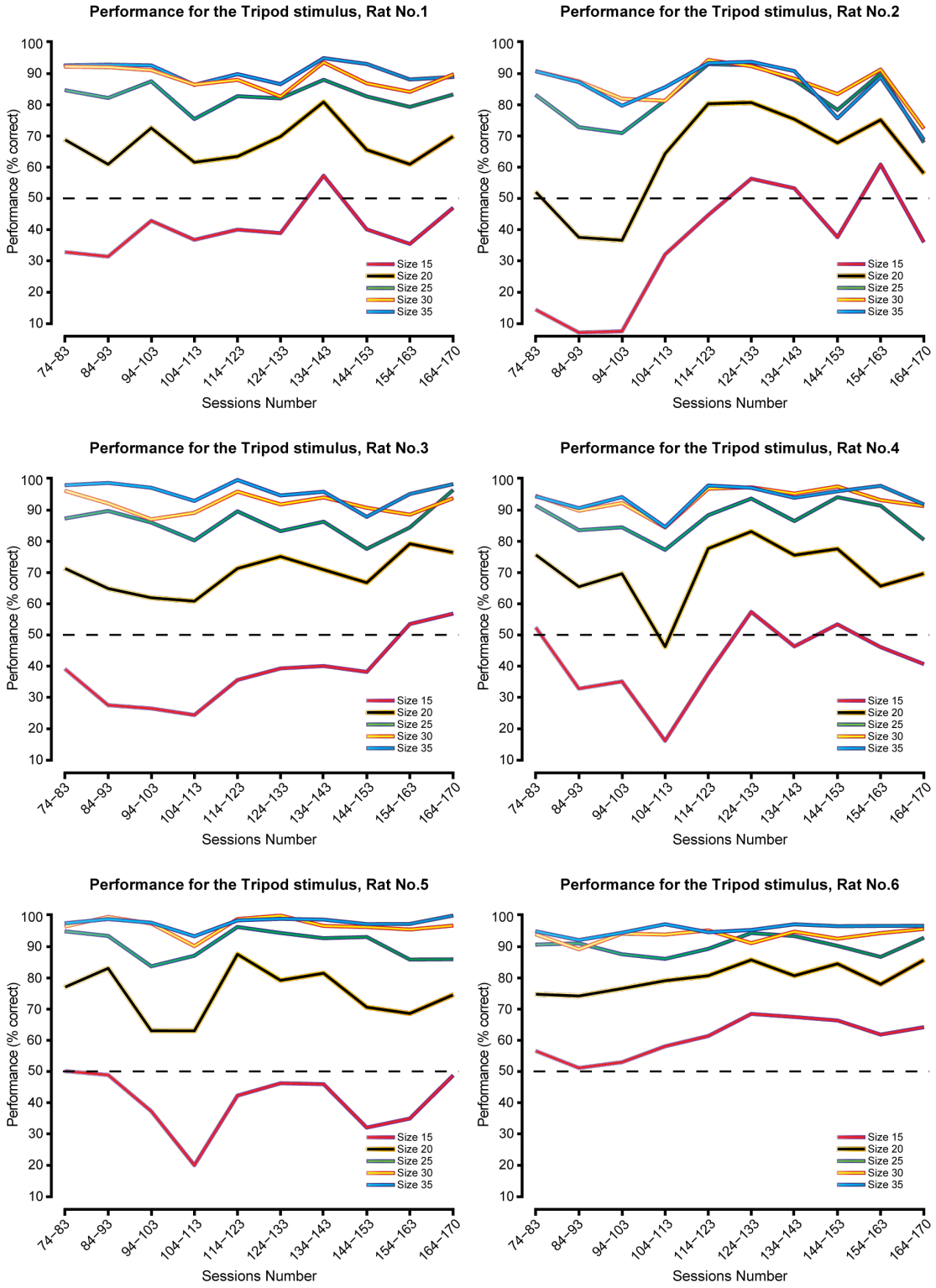

Figure 23. Recognition performance for the Tripod stimulus for all the rats, Phase III (sessions 74-170). Half sizes (.5) were omitted for clarity. Data points represent the average performance across 10 consecutive sessions, except for the last block (sessions 164-170).

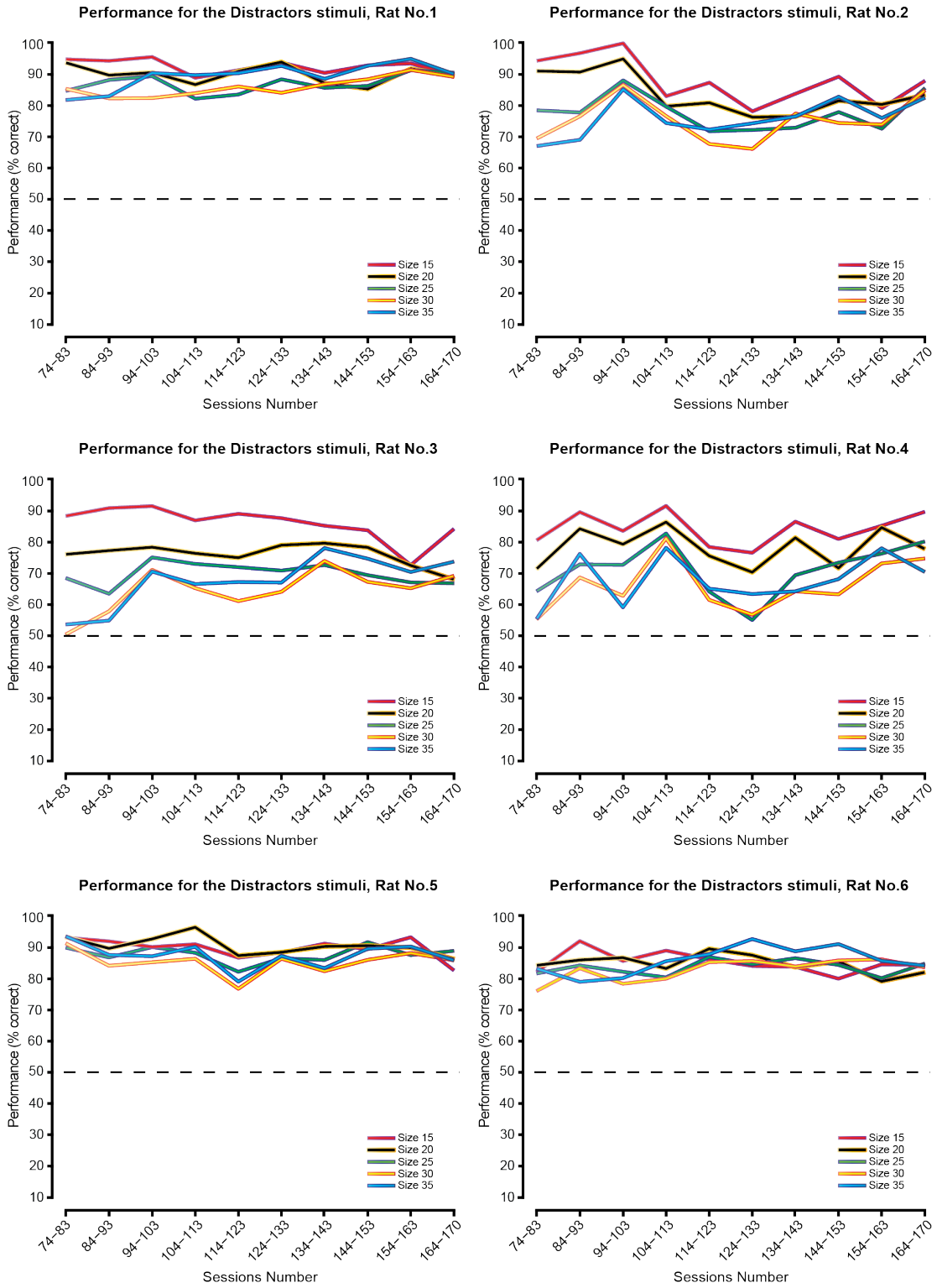

Figure 24. Recognition performance for the Distractors stimulus for all the rats, Phase III (sessions 74-170). Half sizes (.5) were omitted for clarity. Data points represent the average performance across 10 consecutive sessions, except for the last block (sessions 164-170). The performances of all the distractors at the same size were merger together.

Given the large number of presented conditions (99) that were not the reference tripod object, every time the rats were not sure about the identity of the presented stimulus, their tendency was to choose the "not tripod" response, and this obviously applied to

the cases where the tripod object was so blurred (because of its small size) to become hardly discernable. This is easily observable if we look at the average performance for the distractor stimuli. The plots for the distractors look like the mirror image of the plots for the reference object, when it comes to the ranking of the discrimination performances as a function of size (especially for the poorer performing rats 2, 3 and 4), which is exactly what would be expected: the more blurred was a distractor, the more its appearance fell in the "everything-else", "not-tripod" category. It is important to notice that, as the rats (with time) got better (in recognizing both the reference and the distractors), the performance with the distractors at the size 15° got slightly lower.

4.4 Phase IV

Phase IV represents the final and the most important part of our experiment. Although the previous phases yielded some important conclusions - the rats are able to achieve a high level of object recognition when exposed to multiple distractors instead of just one, and they are able to generalize and transfer this knowledge to the size transformed stimuli in a very short amount of time, the primary goal of our experiment was to discover the underlying recognition strategy and try to build a model, based on this strategy, that was able to predict rat perceptual choices.

First, let us take a more detailed look at the rats' performance for each stimulus independently. In Figures 25 and 26, we are presenting the recognition performance achieved during the sessions of Phase IV, when the data to obtain the classification images were collected (see Materials and Methods). This will enable us to make a direct comparison between the classification images and the performance. It could be a mistake to compare, for example, the classification images obtained in Phase IV with the recognition performance calculated for the sessions of Phase III, since these classification images (and therefore the underlying recognition strategy) could change in time (especially as a result of learning), but they would also not account for the effect that the presence of the random tripods in the stimulus pool might have had on rats' performance with the regular stimuli (i.e., the reference and distractor objects).

If we look at the performance for the reference object, we can see that the trend we already saw in the Phase III is completely preserved (there's a small drop in performance for the lower performing rats, probably because of the introduction of the random tripods which led to some confusion). The best performing rat (No. 6) had almost perfect performance and was able to achieve above 80% correct classification, even until 17.5° of visual angle. The other rats had a little bit lower performance, but everyone, except the worst performing rat (No. 4), remained above 70% until 20° of visual angle. As we already commented in the results for the Phase III, at 20° of visual angle rats visual acuity becomes a limiting factor (and the recognition performance starts to drop rapidly) leading some rats (No. 3 and 4) to fall below chance level. If we look at the plots for the performance on the distractors #1-#4, we will see that classification accuracy is close to perfect for all the rats across all the

sizes. Intuitively, these are the results that we expected. When we presented the results for the Phase I, we pointed out that distractor #1 is the "easiest" stimulus among rats and the reason is likely the number of

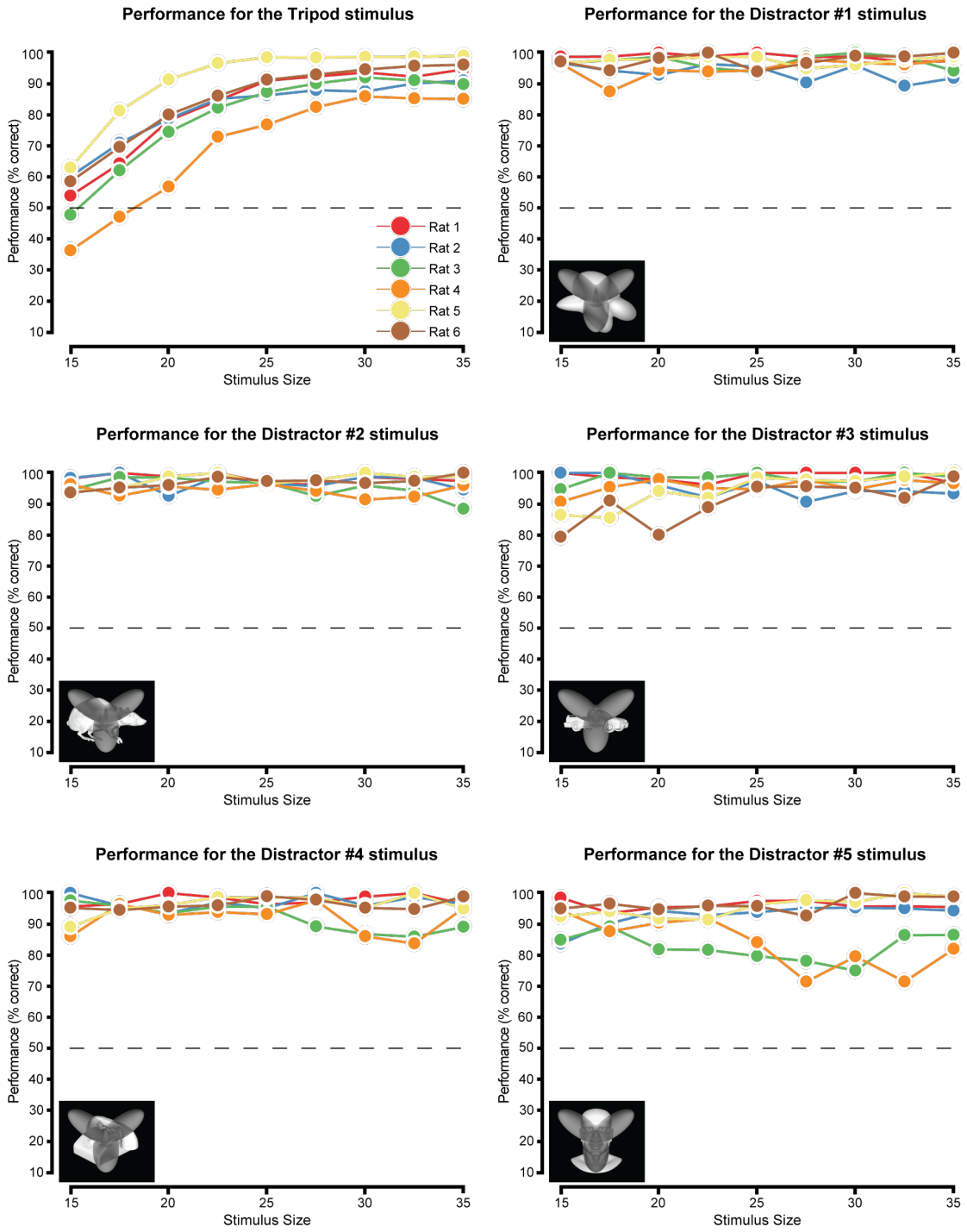

Figure 25. Recognition performance for the reference object and the Distractor stimuli #1-#5, Phase IV (sessions 220-252). Color code is included in the first plot. The rest of the plots contain the overlaps of the Tripod stimulus with the distractor for which the performance is shown.

of distinctive features the rats can use in order to discriminate it from the reference tripod object. But we also pointed to the low overlap with the tripod and to the two

horizontal features clearly sticking out for the distractor #1. If we look at the overlaps of distractors #2-#4 with the tripod, we can see the same pattern emerging – two distinctive horizontal features sticking out, overlapping minimally with the vertically-oriented and diagonally-oriented lobes of the tripod object.

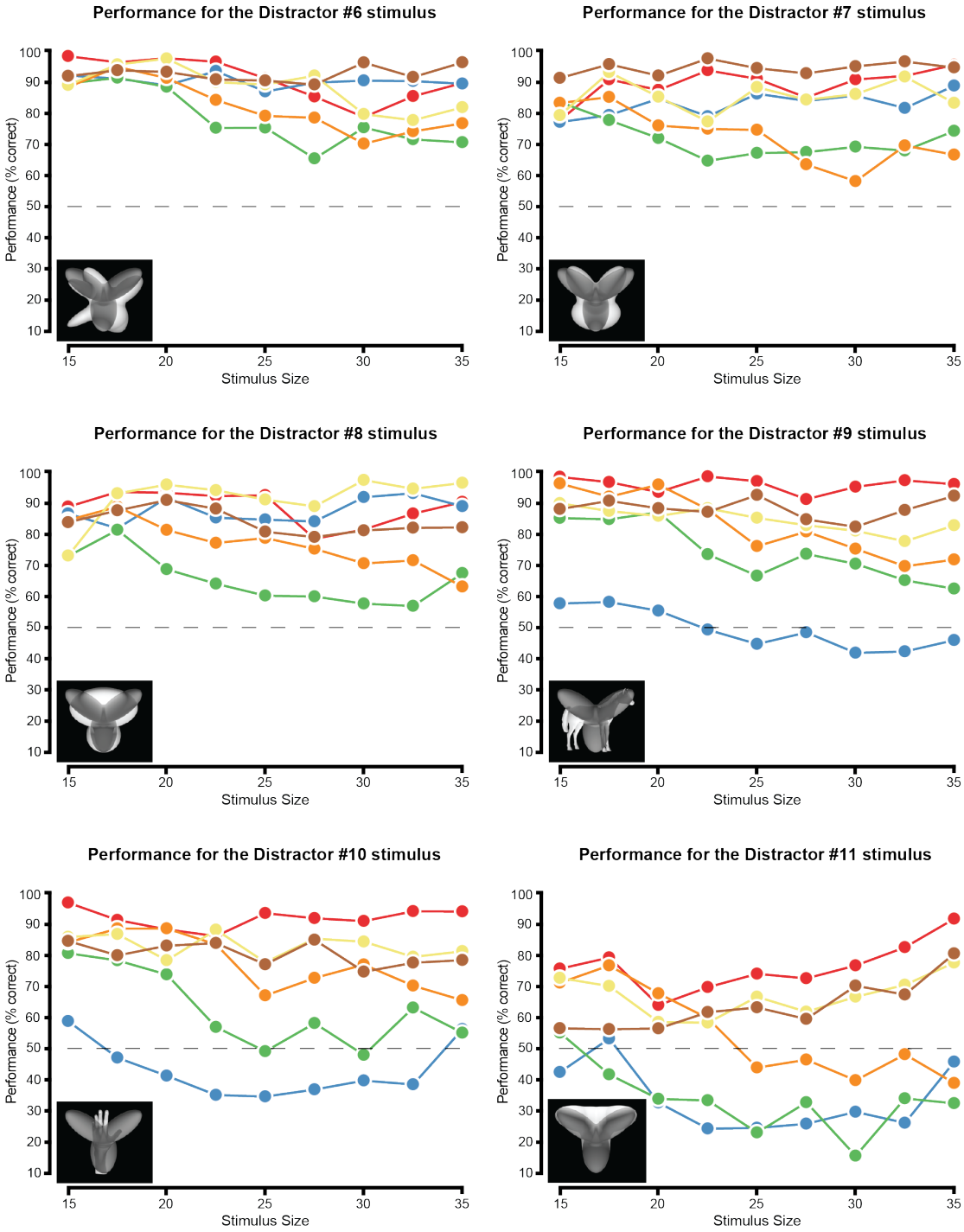

Figure 26. Recognition performance for the Distractor stimuli #6-#11, Phase IV (sessions 220-252). Color code is the same as in Figure 25. The plots contain the overlaps of the Tripod stimulus with the distractor for which the performance is shown.

Consistent with this, the performance is almost the same with these distractors (for all the rats across all the sizes) as with distractor #1.

But as the overlap with the tripod starts to increases (distractors #5-#11), we begin to see a bigger and bigger drop in performance and the "emergence" of two groups of rats - one really good (rats No. 1, 5 and 6) and the other one with, on average, poorer performance – (rats No. 2, 3 and 4). This is especially true for the three hardest distractors (#9-#11) where we can draw a clear line that separates these two groups. As was expected, distractor #11 (which almost completely overlaps with the reference object, except for the middle-top part) proved to be the hardest to successfully identify, with only half of the rats (No. 1, 5 and 6) managing to accomplish the recognition task. We will now take a look at the obtained classification images (Figure 27) and see how good of an explanation they provide for the rats recognition strategy.

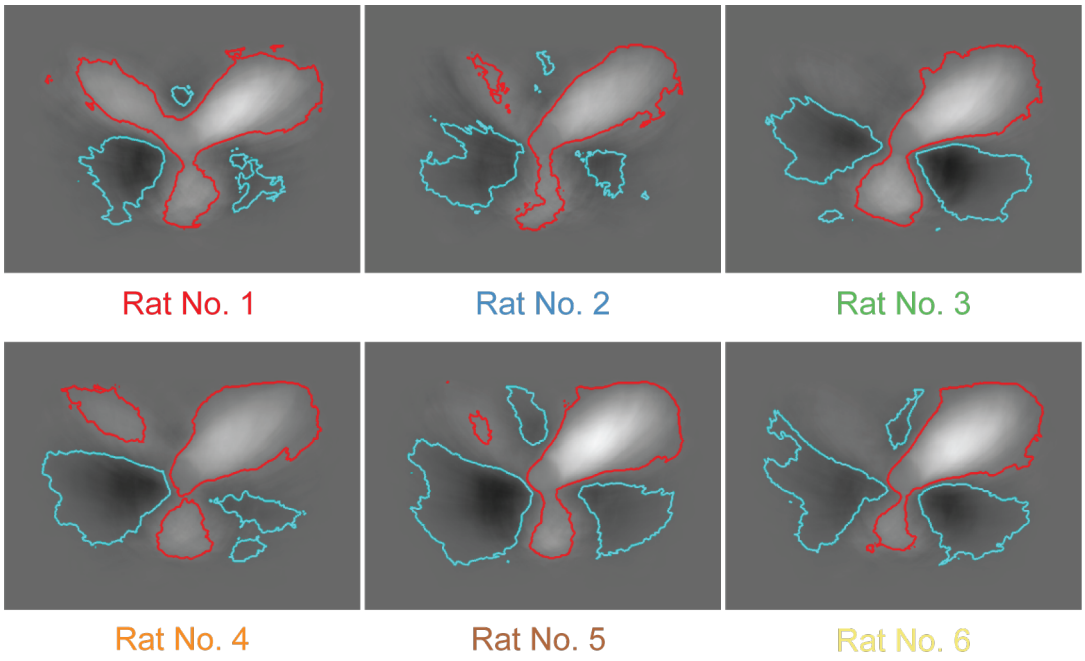

Figure 27. Classification images obtained using the random tripods at size 30° of visual angle, Phase IV (sessions 220-252). Color code is the same as in the Figure 25.

Visual inspection of the classification images reveals a variable pattern of salient and anti-salient features across the animals. All the rats used an elongated salient region, roughly oriented at 45º, matching the shape of the right lobe of the tripod. They also used a second salient feature, corresponding to the tripod's bottom lobe. Only some animals, however, relied on a third salient feature, corresponding to the tripod's left lobe (rats No. 2, 4, and 5) and, most noticeably, rat No. 1, whose

classification image featured three salient regions that were nearly equally prominent. All images also displayed two anti-salient regions, located to the left and right of the tripod, but their shape, extension and relative weight varied considerably among the animals. Finally, only some rats (rats No. 1, 2, 5 and 6) also relied upon a third, smaller anti-salient feature, located at the intersection of the tripod's top lobes. If we now look at the rats performance for the size 30° of visual angle (at which the random tripods were presented, so it's the most accurate comparison) together with the classification images (Figure 28), we can infer a tentative, qualitative explanation of how the different strategies can account for the variable proficiency of the rats with the distractors.

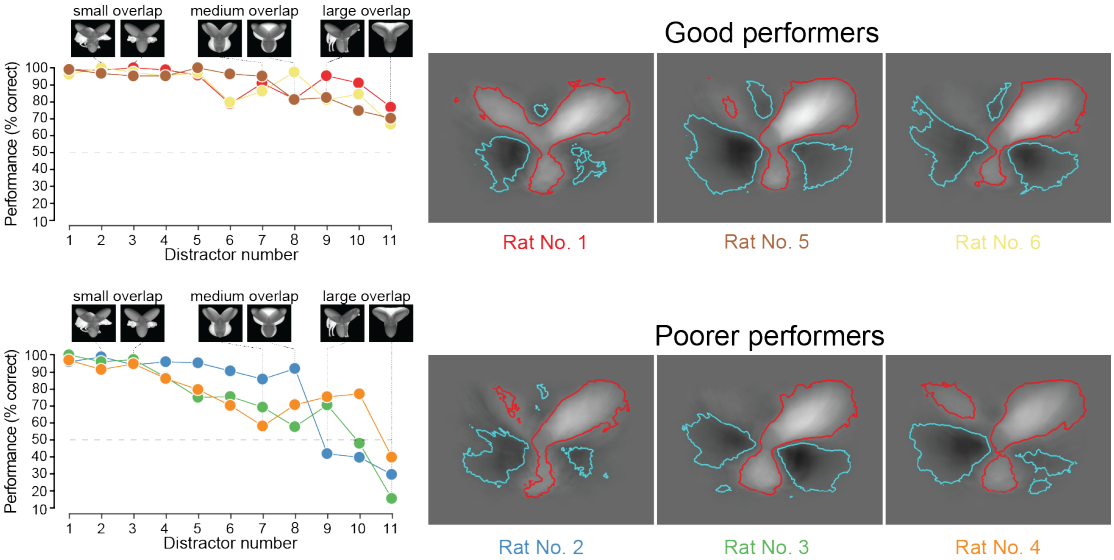

Figure 28. Recognition performance and the classification images (both for the size of 30° of visual angle) for the two groups of rats – very good performers (rats No. 1, 5 and 6) and poorer performers (rats No. 2, 3 and 4). Color code is the same as in the Figure 25.

Not surprisingly, the rat achieving the biggest discrimination performance across most of the distractors was rat No. 1. Its strategy was the one based on the bigger number of salient and anti-salient features, including the anti-salient patch, precisely located at the intersection of the top salient regions. Rats No. 5 and 6 displayed perceptual templates that were nearly as rich (also including the upper anti-salient patch) and reached similarly high performances. The remaining rats deployed a less refined perceptual strategies. Most noticeably, the top anti-salient feature was either absent from their classification images (rats No. 3 and 4) or shifted upward (rat No. 2). This seems to explain the inability of these rats to correctly classify the Distractor #11, whose lobes almost exactly matched those of the tripod object (if not for the different orientation of the top ones – horizontally, rather than diagonally

elongated in the case of Distractor #11). Perceiving such difference required assigning anti-tripod evidence to the region of the image located at the intersection of the tripod's upper lobes – the lack (or misplacement) of the top anti-salient feature likely prevented rats No. 2-4 from doing so. This seems to explain also the lower performance of these animals with other distractors that considerably overlapped with the tripod, such as distractors #7 and 8.

Interestingly, these objects were more challenging for rats No. 3 and 4 than for rat No. 2, whose accuracy (with many distractors) was not dissimilar from that of the good performers. This is consistent with the richer perceptual strategy displayed by rat No. 2, as compared to rats No. 3 and 4. The animal, in fact, relied on a combination of features that was similar to that of the most proficient rats, if not for the smaller and misplaced anti-salient patch on the top, and the greater prominence of the salient region corresponding to the tripod's right lobe – which likely accounts for the poor performance of the rat with the distractor #9. These observations, though still qualitative, suggest that rat accuracy in classifying a given distractor, far from being determined solely by the similarity with the reference object, is strongly affected by the complexity of the perceptual strategy adopted by each animal. In particular, rats building richer, more integrated perceptual templates (in terms of number and variety of diagnostic features) achieved higher success rates, also with very difficult distractors.

4.4.1 Building models of rat perceptual choices

While these qualitative observations provide a plausible explanation of rat performance, based on the recognition strategies inferred through the classification images, our main goal was to quantify the predictive power of such strategies. As we explain in the Materials and Methods, once we obtained the classification images, we trained our model using 9/10 of the trials the rats performed with the random tripods, in order to test how well the logistic regression model presented in the Materials and Methods is able to predict the responses of the animals to the remaining 1/10 of the trials. We used the logloss function (see Material and Methods) to measure the models' accuracy (Figure 29). The logloss results confirmed our previous observation that Rat No.1 classification image has the best prediction power (i.e., minimal

logloss), but it also confirmed the ordering we observed in the recognition performance and classification images complexity among the other rats (good performers – Rat No. 1, 5 and 6, poorer performers – Rat No. 2, 3 and 4, see Figure 28). This indicates that the rats relying on richer perceptual templates responded to the input images in a way that was more consistently based on the matching to the template (and, therefore, less prone to purely random choices), as compared to the animals using simpler perceptual strategies. But does the model confirm the uniqueness of the rats' strategies?

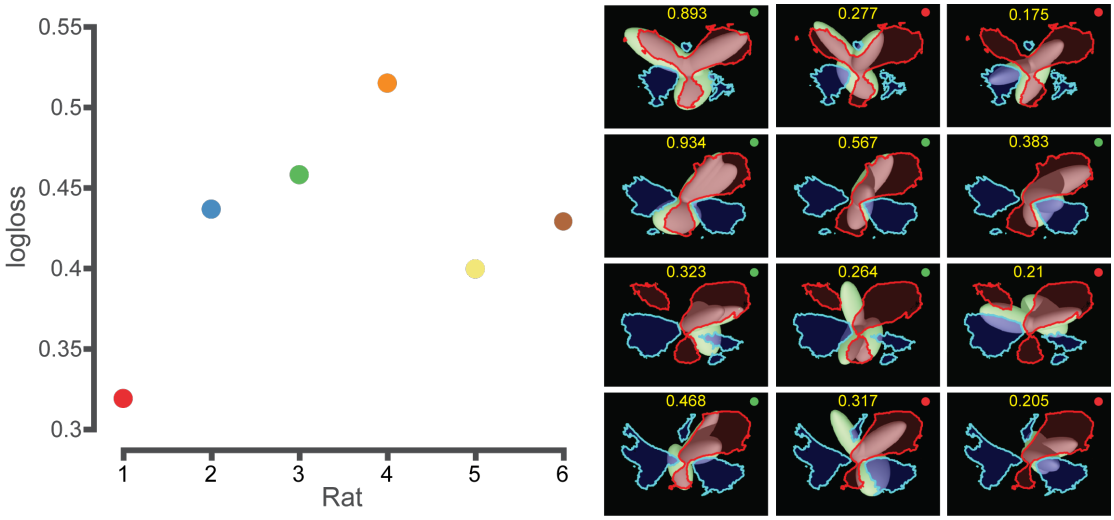

Figure 29. Logloss values of the model prediction for each rat (color code is the same as in the Figure 25) on the left and examples of the probability values for different random tripods and different rats (each row contains the classification image of a different rat – No.1, 3, 4 and 6 in that order

In order to test this we plugged into our model the classification images of the other rats and compared the obtained logloss values (Figure 32). For 5 out of 6 rats, we found that the *same*-CI model (large circles) did indeed yield logloss values that were lower than those returned by the *cross*-CI models (small circles), and this pattern was highly significant ($p < 0.0006$; binomial test). This implies that each rat relied on a distinctive perceptual strategy, whose specificity was well captured by the classification image, which further demonstrates the uniqueness of each rat object processing strategy.

To strengthen this conclusion, we carried out a second analysis, where, for each rat, we first fitted our model (see Materials and Methods) to the full set of random tripods, and then tested its ability to predict the responses to the 11 distractors, when presented at 30º of visual angle. Compared to the previous analysis, this was an even more stringent test of the ability of the classification images to predict rat perceptual choices, because the *train* and *test* image sets (i.e., the random

tripods and the distractors) were not only different stimuli, but belonged to structurally distinct classes of visual objects.

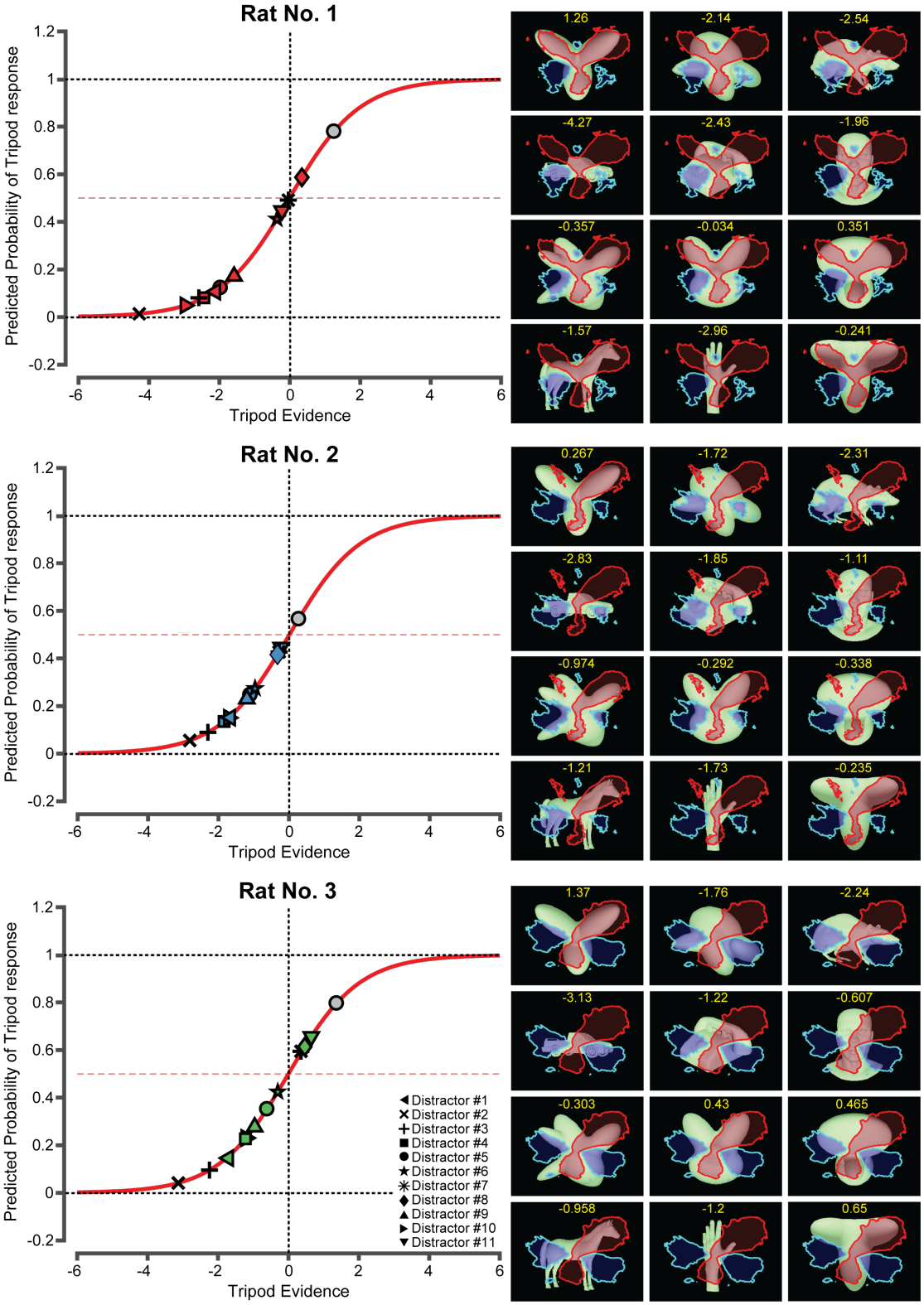

Figure 30. Predicted probabilities of a Tripod response for the rats No. 1, 2 and 3. Distractors are color coded for each rat (same as in the Figure 25), while the Tripod is marked with the gray color.

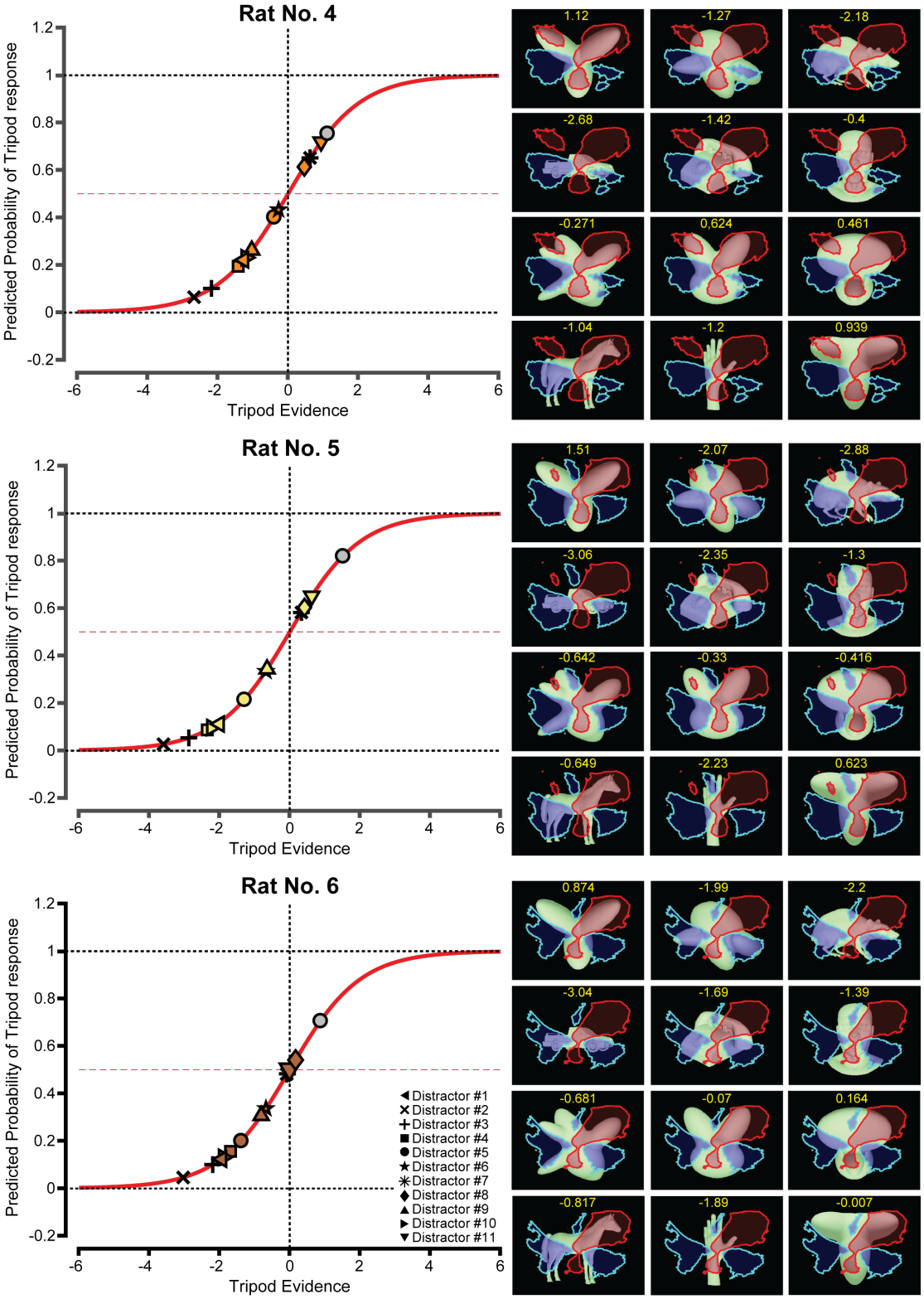

Figure 31. Predicted probabilities of a Tripod response for the rats No. 4, 5 and 6. Distractors are color coded for each rat (same as in the Figure 25), while the Tripod is marked with the gray color.

Figures 30 and 31 illustrate the way in which we calculated the predicted probabilities of the model. The distribution of values over the sigmoid for different rats can serve as additional visual aid demonstrating the individuality of each rats' strategy.

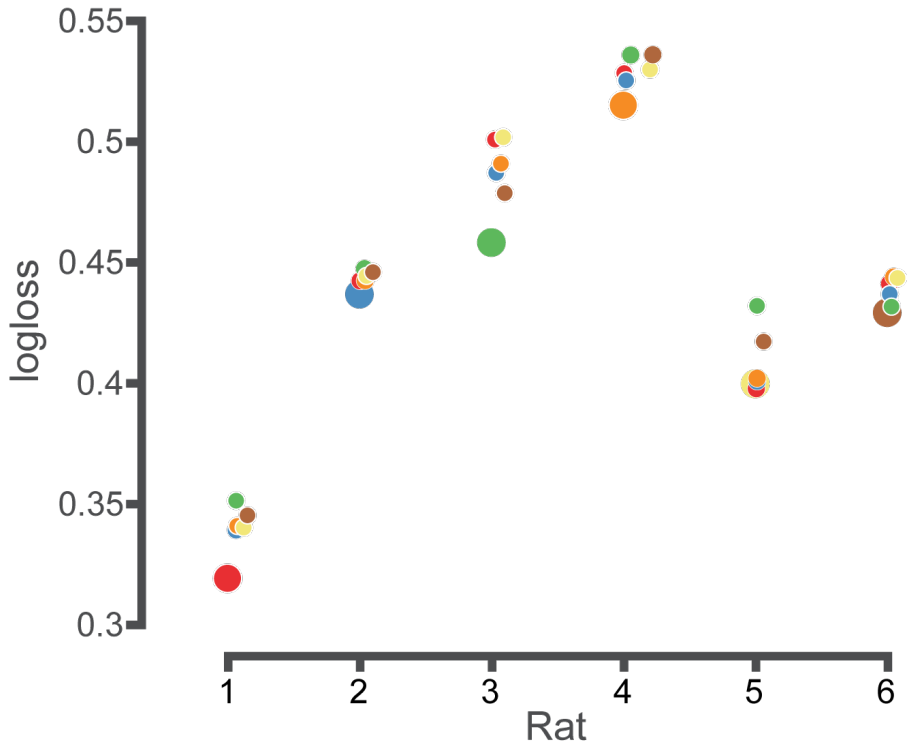

Figure 32. Logloss values for the model predictions depending on the input classification images. For each rat we compared model's logloss values obtained using its own classification image and the classification images of the other rats. Color code is the same as in the Figure 25.

We took the logit of the predicted and measured probabilities of correctly classifying the distractors, and we plotted them against each other. Figure 33 shows the resulting scatter plots for all the rats. The logit is the reverse of the logistic function (i.e., it yields the argument of σ in our model). As such, it measures the evidence that a rat has acquired about whether the input image is the tripod (or a distractor), before this evidence is translated into the probability of choosing the tripod (or distractor) response. Our model assumes that this evidence is linearly related to the similarity between input image and perceptual template. Therefore, comparing the logit of the observed performances to the logit of the predicted ones amounts to a direct test of the linearity of this relationship.

The two quantities were, in fact, strongly correlated ($r = 0.61$; $p < 0.0001$), with the objects that were more dissimilar from the tripod, such as the Distractor #2 and Distractor #3 yielding larger *distractor evidence* (both observed and predicted, symbols × and + in the Figure 33, left) and the objects that were more tripod-looking, such as Distractor #11 (symbol▼), providing lower distractor evidence. At the same time, the model managed to capture very well the variability of rats' responses to the

same distractor – e.g., rats No. 1, 5 and 6 achieved larger distractor evidence (both observed and predicted) than rats No. 2, 3 and 4 with the difficult Distractor #11.

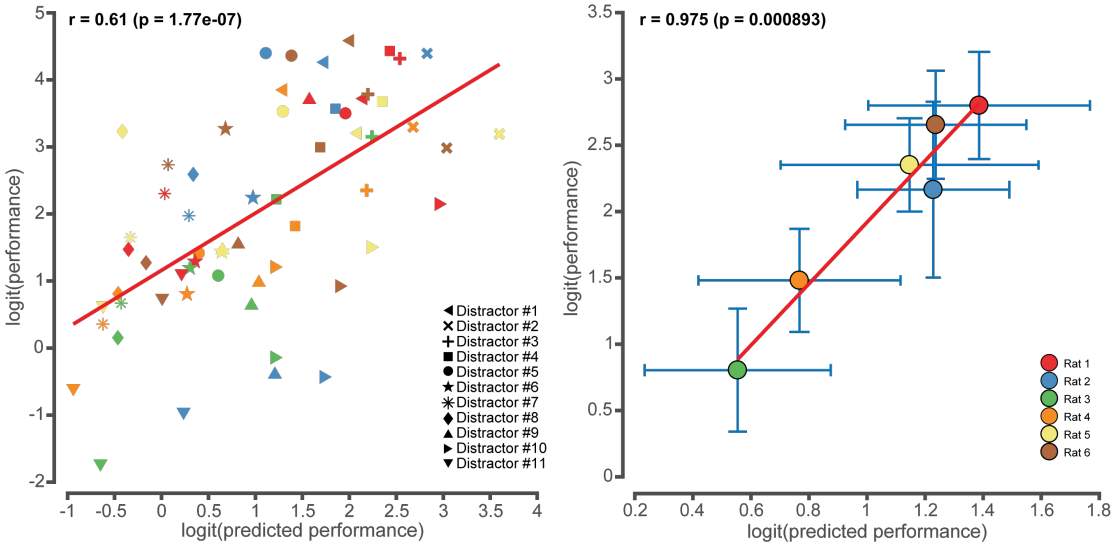

Figure 33. A scatter plot of predicted performance for all the distractors and all the rats (left). Average of predicted performance for individual rats (right). Color code is the same as in Figure 25.

To further test whether the model was capable of accounting for the variable proficiency of the rats in the discrimination task, we averaged, separately for each rat, the measured distractor evidences and the predicted distractor evidences, across all 11 distractors. The resulting scatter plot (Figure 33, right) showed a very clear linear relationship between the two evidences ($r = 0.98$, $p < 0.001$), with rat No. 1 achieving the largest values along both axes, followed by rats No. 5 and 6, then by rat No. 2, and finally by rats No. 3 and 4. Critically, this ranking matched very well with the one previously obtained for the predictability of rat choices (Figure 29) and, more importantly, the complexity level of rat perceptual strategies, as inferred by visual inspection of the classification images (Figure 27). Taken together, these trends indicate that each rat used a subject-specific perceptual strategy, whose complexity determined the animal's performance in the discrimination task.

This conclusion, however, required further validation, because, in our model, the probability of a correct choice depends not only on the perceptual template CI_i, but also on the offset θ_0, which can capture the overall propensity of a rat to choose the tripod (or distractor) category. Although this bias term did not influence the outcome of the within-rat analyses (Figure 32), it could potentially play a role in the between-rat comparisons (Figure 33, right) – i.e., the models obtained for the different animals may predict their fluency in the task, mainly because they accurately fit their

biases. Although this seemed unlikely, since the θ_0 obtained for the rats were very similar (Table 1), we carried out two control analyses.

	Bias ($\boldsymbol{\theta_0}$)	Scaling ($\boldsymbol{\theta_1}$), x10^{-5}
Rat 1	-4.08064	1.35654
Rat 2	-2.35565	1.14728
Rat 3	-2.38103	1.29263
Rat 4	-1.88381	1.27581
Rat 5	-2.82878	1.23493
Rat 6	-2.51752	1.10293

Table 1. Values of the parameters θ_0 (offset/bias) and θ_1 (gain/scaling) in our model for all the rats

The first relied on the fact that, if the variable accuracy of the rats with the distractors were determined by their different biases, this would affect also the discrimination of the reference object. That is, the best performing rats (No. 1, 5 and 6) would gather larger distractor evidence, as compared to the poorer performers (No. 2, 3 and 4), not only when presented with the distractors, but also when processing the tripod. As shown in Figure 34A, this was not the case, since the tripod yielded lower distractor evidence for the best performers. Crucially, model predictions followed the same trend (Figure 34B), thus showing that the difference among the θ_0 obtained for the rats was not at the root of the variable distractor evidence yielded by the models.

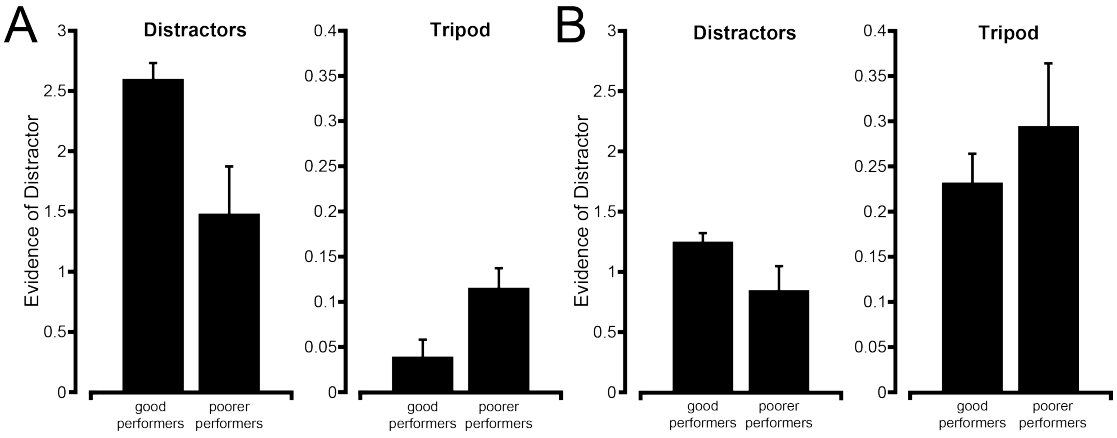

Figure 34. The variable fluency of the rats (and the models) in the discrimination task is not accounted for by differences in decision bias. A) Distractor evidence gathered by the good (rats #1, #5 and #6) and poorer (#2, #3 and #4) performers, when presented with either the distractors (left) or the reference tripod object (right). Note that the good performers achieved larger distractor evidence than the poorer ones, but only with the distractors objects. This indicates that the different performance of the rats with the distractors is not due to a difference in decision bias, but, rather, to a different ability of the animals to extract discriminatory shape information. B) Same as in A, but with distractor evidence yielded by the model. Note that the data follow the same qualitative trend shown in A. This means that the success of the models to account for the variable performance of the rats with the distractors (see Figure 33, right) is mainly rooted in the ability of the classification images to capture the differences, among the rats, in terms of shape processing strategies, rather than in the ability of the models to properly fit the animals' decision biases.

This conclusion was confirmed by a second analysis, in which the tripod evidence gathered by rat i, when presented with distractor x, was modeled as $\mathrm{CI}_i \cdot x - \mathrm{CI}_i \cdot t$, where t is the tripod stimulus. That is, the model measured the tripod evidence in terms of how close the match between distractor and template was to the match between tripod and template. Despite the lack of any bias term, this model yielded the same trends produced, in Figure 33, by the logistic regression (but with reversed sign, since tripod, rather than distractor evidence, was taken into account). Rat performance monotonically decreased as a function of tripod evidence (Figure 35, left) and a strong (negative) correlation was found between rat accuracy with the hardest distractors and tripod evidence (Figure 35, right). Overall, these analyses confirmed the conclusion that it was the perceptual strategy deployed by an animal to establish its proficiency level in the task.

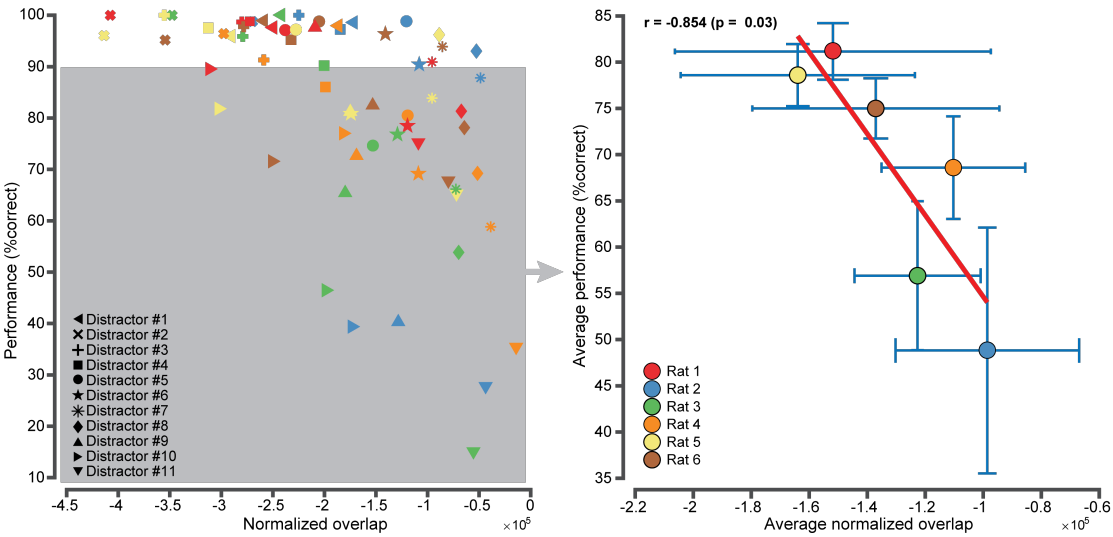

Figure 35. The perceptual choices of the rats are well predicted by models based exclusively on classification images. The left panel shows the relationship between the classification accuracy achieved by the rats with the distractors (ordinate axis) and the match between the perceptual templates inferred for the rats (i.e., the classification images) and the distractors (abscissa axis). The match was computed as described in the Results, and measured the similarity between each of the distractors and the perceptual template underlying the recognition of the reference tripod object. As such, it was a measure that modeled the tripod evidence collected by the rats. Therefore, it does have the opposite meaning of the quantity in the abscissa of Figure 33, which measured the distractor evidence. Because of this, the relationship shown here has a negative, rather than a positive slope. The gray area highlights the object conditions where this relationship was approximately linear. These conditions correspond to the distractors that were harder to recognize, yielding a performance lower than 90% correct discrimination. These are the conditions that were averaged, separately for each rat, to obtain the plot on the right, which shows a large and significant negative correlation ($p < 0.05$; one-tailed t-test) across the animals between classification performance and match to the perceptual template. As such, this plot confirms the ability of the classification images to explain the different proficiency of the rats in the discrimination task (as already shown in Figure 33, right), using models that do not include any bias terms. Color code is the same as in the Figure 25

4.4.2 Rat invariant recognition is not consistent with a low-level processing strategy

The success of the logistic regression model at predicting rat perceptual choices (Figure 33) does not imply that the animals processed every incoming stimulus using a fixed perceptual template. Such a rigid template-matching computation would amount to a low-level strategy, no matter how many features are integrated into the template, because it would prevent the animals from correctly classifying transformed versions of the objects – for instance, those producing global changes of luminosity, while leaving unaltered the shape, such as size variations. This can be appreciated by considering the argument of our model: the smaller (or dimmer) the object in the input image x becomes, the smaller its dot product with the template ($\mathrm{CI}_i \cdot x$) gets, and, as a consequence, the more likely is for the object to be classified as a distractor, regardless of its shape. This scenario is at odds with recent studies showing how rats are capable of recognizing visual objects in spite of identity-preserving transformations, such as size changes (Zoccolan et al, 2009; Alemi-Neissi et al, 2012; Rosselli et al, 2015) as well as our own results (see Figures 23-26). To directly show that a fixed template-matching strategy is not able to account for rats' invariant recognition, we measured how well our logistic regression model generalized to outline versions and scaled versions of the stimuli. We then compared the performance of the model to the actual performances of the rats with these transformed stimuli. As described in Materials and Methods we performed 32 sessions with outline versions of the same random tripods that we used to obtain the original classification images, as well as 32 sessions with scaled version of the random tripods (83% of their original size, corresponding to scaling the reference tripod object from 30º to 25º of visual angle, see Figure 20 and 36 for examples).

Reducing an object to its outline leaves the overall shape unchanged, while it substantially changes the luminance cues that define the object. As argued above, a template-matching strategy, developed to specifically process full-body stimuli, would show little cue invariance, bringing the rats to classify the outline random tripods as distractors, way more often than the animals did for their full-body counterparts. This can be appreciated by comparing, in Figure 36A, the two red bars, showing the probability of tripod responses to full-body and outline random tripods,

as predicted by the model (with CI_i, θ_0 and θ_1 fitted to rat responses to the full-body stimuli). The probability of tripod responses was half as large for the outlines, as compared to the full-bodies, and this difference was highly significant ($p < 0.001$, one-tailed, paired t-test). By contrast, rats displayed a fully cue-invariant behavior, with the fraction of tripod responses being virtually identical for outline vs. full-body random tripods (black bars; $p = 0.97$). As a result, the model significantly underestimated the probability of tripod responses for the outline stimuli ($p = 0.03$).

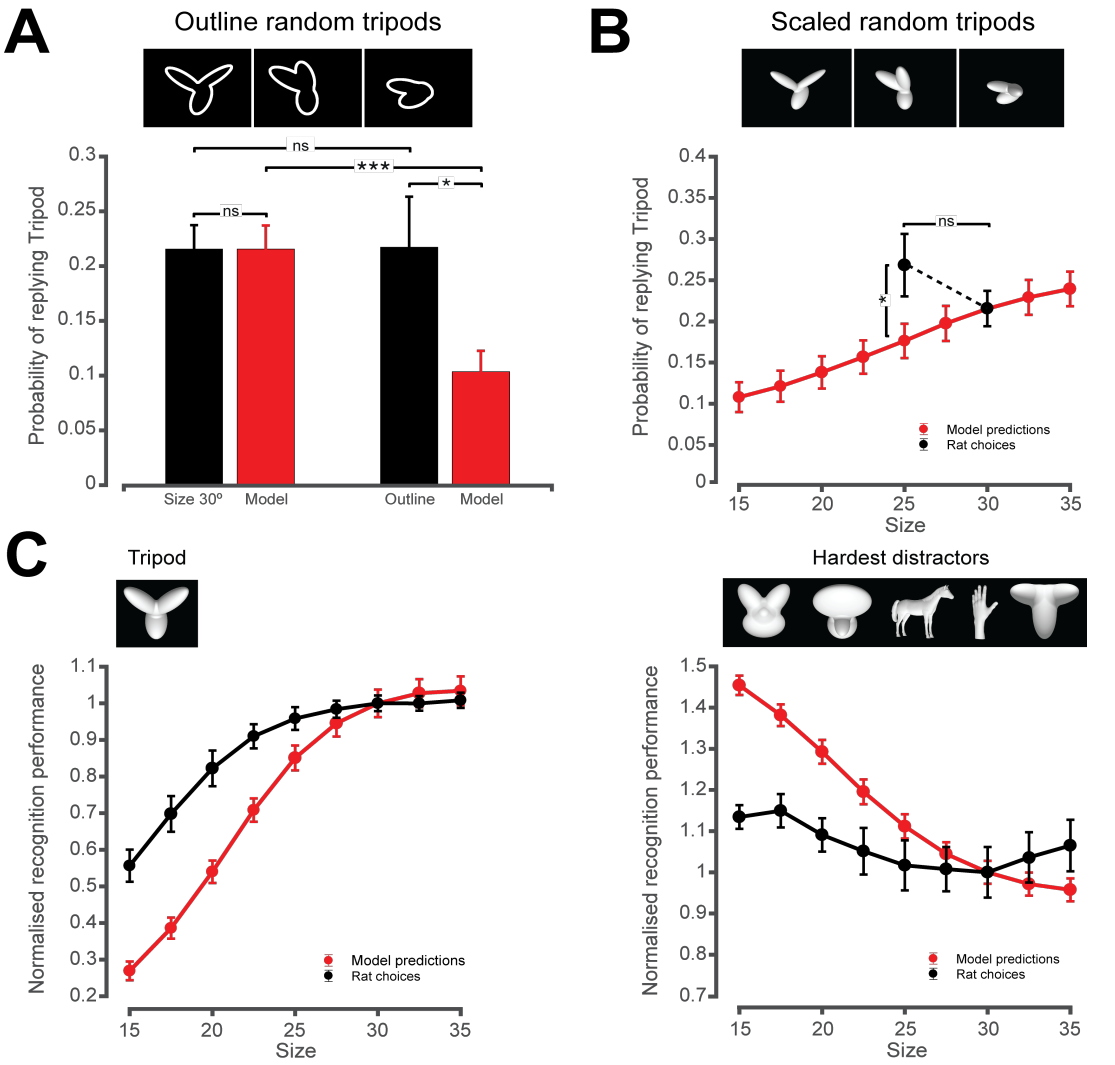

Figure 36. Inconsistency of rat invariant recognition with a simple template-matching strategy.
A) Fraction of random tripods classified as being the tripod by the rats (black) and by models (red) that were based on the classification images obtained from the full-body, regular size random tripods. Classification rates are reported for these full-body stimuli (left), as well as for their outlines (right; examples shown on the top). Stars indicate significant difference according to a one-tailed, paired t-test (* $p < 0.05$; *** $p < 0.001$). B) Same as above, but with classification rates referring to the random tripods presented at the default, regular size (30º) and additional sizes – the whole size range, in the case of model predictions (red curve), and size 25º, in the case of rat responses (black dots). Same statistical analysis as in A. C) Same as in A and B, but with classification rates referring to the reference tripod object (left) and to the five hardest distractors (right; stimuli shown on the top) presented across the whole size range. In all panels, bars/dots show mean classification rates (computed over 6 rats or 6 models) ± SEM.

A similar finding applied to size variations (Figure 36B). The logistic regression model, trained with the random tripods at size 30º, yielded a significant modulation of the probability of tripod responses as a function of the size of the random tripods ($p < 0.001$; $F_{8,40} = 113.488$, one-way ANOVA), with this probability steadily decreasing at smaller sizes (red curve). This trend did not match what observed for the rats, where the fraction of tripod responses to the smaller (25º) random tripods was actually slightly larger (although not significantly) than the fraction of tripod responses to their regular-size (30º) counterparts (black dots; $p = 0.12$; one-tailed, paired t-test). As a consequence, the model significantly underestimated the probability of tripod responses at size 25º ($p = 0.02$).

In the case of size variations, it was also possible to obtain model predictions for the responses to the reference object and to the distractors across the full range of sizes tested in our study. As already shown (Figure 25), rat recognition was very stable over a wide size span, with classification accuracy dropping substantially, for the tripod, only at the smallest sizes (< 20º of visual angle), while simultaneously increasing for some of the hardest distractors. These trends, although qualitatively consistent with a strategy based on matching the input images to a fixed template, were nevertheless much milder than predicted by such a strategy. This can be appreciated by comparing the black and red curves in Figure 36C, showing, respectively, rat group average accuracy (black) with the tripod and the five hardest distractors as a function of size, and the predictions of the logistic regression model (red). The curves (that were normalized to their values at 30º for a better comparison) all displayed some degree of modulation over the size axis, but this was much sharper for the model predictions than for rat performances. For the tripod (left), classification accuracy at size 15º dropped to just ~30% of its value at size 30º, in the case of the model, while, for the rats, it remained at ~60% of its value at 30º. The different trends obtained for rat performances and model predictions were confirmed by a two-way ANOVA with *size* and *observer* (i.e., either *rat* or *model*) as factors. The main effect of size was significant ($p < 0.001$, $F_{8,40} = 175.474$), as well as its interaction with observer ($p < 0.001$, $F_{8,40} = 22.387$), thus showing that the two performances dropped at a different pace along the size axis. With the distractors, the accuracy of the model increased sharply at the small sizes, being ~40% larger at size 15º than at size 30º (Figure 4C, right). By contrast, only a mild modulation was observed for rat performance, which, at size 15º, was just ~10% larger than at size 30º. Again, these

trends were assessed by a two-way ANOVA, yielding a significant main effect for *size* and its interaction with *observer* ($p < 0.001$, $F_{8,40} = 46.565$ and $F_{8,40} = 14.993$ respectively). This confirmed the failure of the single template-matching model to account for rat size-tolerant behavior.

4.4.3 Stability of rats' perceptual strategy under changes in object appearance

These findings are at once reassuring, with regard to the complexity of rats' recognition behavior, and far from unexpected, when our modeling approach is considered. In fact, building invariant representations of visual objects requires the same combination of diagnostic features (i.e., the same perceptual template) to be applied, in a filter-like fashion, across multiple scales, positions, etc., with the outputs of these filters being later merged through some non-linear computation (such as the *max pooling*, originally proposed in early models of visual cortex, Riesenhuber et al, 2000, and now successfully applied in deep convolutional neuronal networks, LeCun et al, 2015). On the other hand, computing a classification image is a strictly linear process, which returns a single perceptual template, and, as such, it cannot possibly capture the invariance of visual processing across identity-preserving changes. In this sense, the failure of our model to account for rats invariant recognition suggests that the rats' visual system, similarly to the primates' one and to the above-mentioned neural network models, processes the visual input through a bank of multiple perceptual templates that work in parallel. This hypothesis can be tested by iteratively computing classification images at different sizes, positions, orientations, etc., and then checking the consistency of the resulting perceptual templates, as well as their ability to account for rat recognition behavior. In our study, we applied this approach to obtain perceptual templates from the outline and the size 25° random tripods. The same procedure that was used to obtain the "original" (size 30°) classification images (see Materials and Methods) was also used to obtain the classification images for outline and size 25° versions.

In the case of the outline stimuli, the resulting classification images still displayed prominent salient and anti-salient features (Figure 37B, top), which, although thinner and more scattered, largely matched those found in the classification images derived from the size 30° random tripods (i.e., those shown in Figure 27 and

displayed again in Figure 37A for easier comparison). This means that, as already suggested by the bar plot in Figure 36A, rats responded to the outlines of the random tripods in a "meaningful" way, automatically classifying them according to the strategy they had developed in the *tripod vs. everything-else* discrimination, despite the lack of any explicit training with the outline stimuli. For each rat, the consistency between the perceptual strategies derived from the size 30° random tripods and their outlines can be appreciated by comparing matching images in Figures 37A and Figures 37B, top.

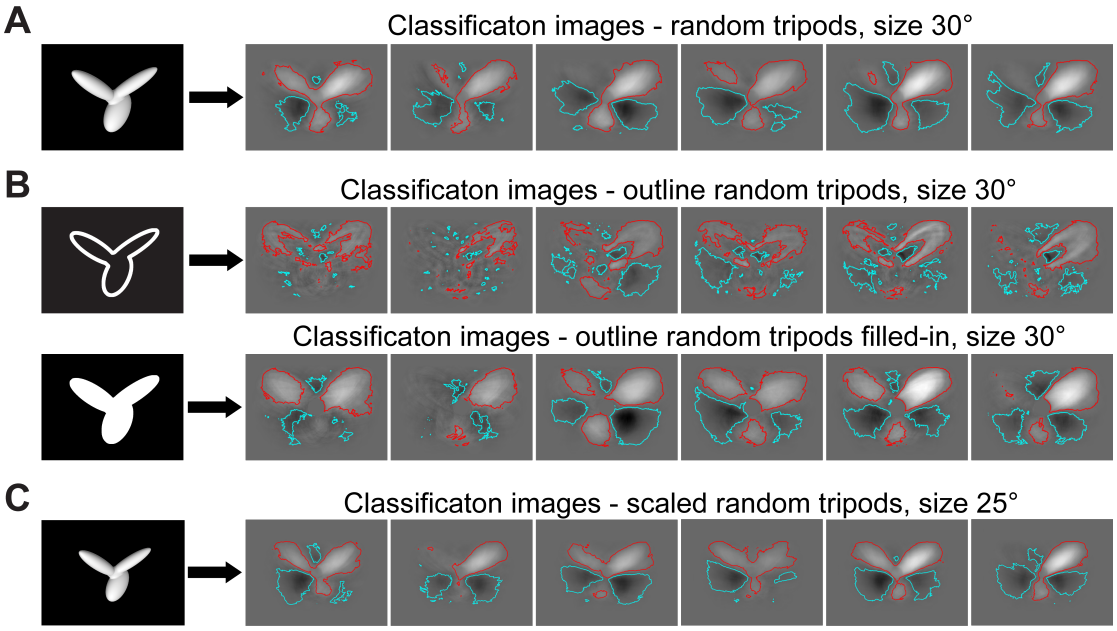

Figure 37. Transformation-tolerance of rat perceptual strategy. The classification images obtained, for the six rats, from: A) the regular-size (30º), full-body random tripods; B) their outlines (top) and the filled-in versions of their outlines (bottom); and C) the small-size (25º), full-body random tripods.

Such consistency becomes even more apparent by looking at the perceptual templates that were obtained by replacing the outline random tripods with their filled-in versions in the computation of the classification images (Figures 37B, bottom). The rationale of this analysis was to simulate a strategy in which a rat extracts diagnostic information about the identity of the outline random tripods not only from the features that are actually visible (the outlines), but also from the (empty) bodies of the stimuli. The resulting classification image allows inferring the strategy that a rat would deploy, if it was able to perceptually fill the constituent parts of the random tripods (the lobes), so as to process them as solid features.

To quantitatively assess whether the perceptual strategy of rat *i* was similar, when extracted from full-body or outline random tripods, we plugged the classification image CI_i^o obtained from the outlines into our model, and we fitted the

logistic regression model to the responses of the animal to the full-body stimuli. As previously explained when describing Figure 29, the fit was carried out using only 9/10 of the responses to the random tripods, so that we could measure how well the model predicted the responses to the remaining 1/10 of the stimuli. The performance of this *same-CI*^{o} model (measured by the logloss function) was compared to the performances of the *cross-CI* models, resulting from plugging into our model the classification images CI_j (with $j \neq i$) obtained, for the other animals, with the full-body random tripods (see previous description of Figure 32). Interestingly, when the *same-CI*^{o} model was built using the classification images derived from the actual outline stimuli (i.e., those shown in Figure 37B, top), it was outperformed by most of the *cross-CI* models (compare the empty green dots to the small black dots in Figure 36). By contrast, when the *same-CI*^{o} model was obtained using the classification images derived from the filled-in versions of the outline stimuli (i.e., those shown in Figure 37B, bottom), it yielded the best performance (i.e., the lowest logloss) for 3 out of 6 rats, and the second best performance for another rat (compare the solid green dots to the small black dots in Figure 38). Although the overall pattern was only marginally significant (the probability of obtaining at least 4 second best placements in 6 Bernoulli trials, with the chance of success in each trial being set to 2/6, is p = 0.06; binomial test), this results indicates that, for many rats, the perceptual strategy remained highly subject-specific and, therefore, largely tolerant under variation of the luminance cues that defined the visual stimuli. In addition, the superiority of the model based on the filled-in versions of the outlines, as compared to the one based on the actual outlines (a highly significant pattern, having been found for 5 out of 6 rats; p = 0.01; binomial test) strongly suggests that rats did indeed process the outline stimuli by perceptually filling their inside, so as to effectively treat them as solid-body objects.

A similar analysis was carried out for the classification images obtained from the small-sized random tripods (shown in Figure 37C). Again, these images were strikingly similar to those yielded by the size 30° random tripods (compare to Figure 37A). To quantitatively assess their consistency, we rescaled the classification images in Figure 37C, so as to match the size of those in Figure 37A, and, for each rat *i*, we used the rescaled version of the small-sized CI_i^s (size 25°) to carry out our cross-

validated model comparison. That is, the rescaled CI_i^s was plugged into our model and used to predict the responses of the animal to the size 30° random tripods.

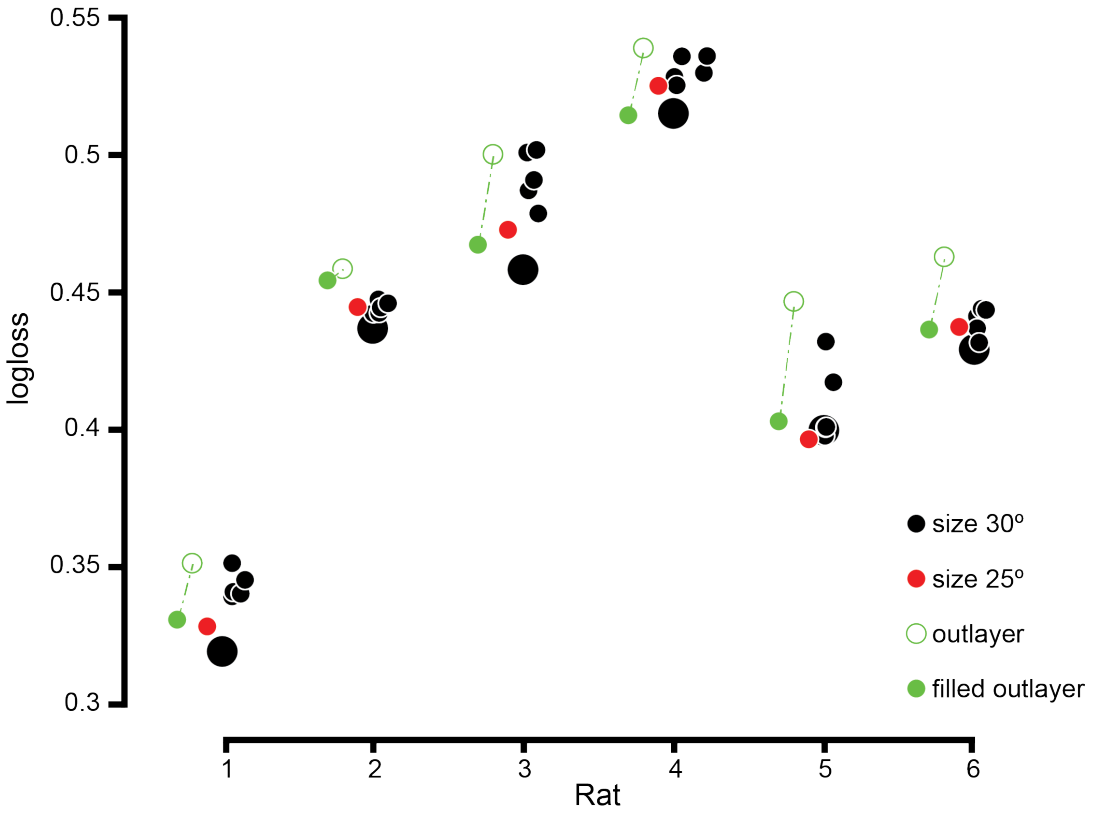

Figure 38. The comparison of the model accuracy for the classification images obtained for random tripods at size 30° (black dots), outline (empty dots) and filled outline (green dots, both at size 30°) as well as size 25° (red dots).

The performance of this *same-CI*^{s} model was then compared to the performances of the *cross-CI* models (same as described in the previous section). The *same-CI*^{s} model outperformed the *cross-CI* models for 5 out of 6 rats (compare the red dots to the small black dots in Figure 36) and this pattern was highly significant (p = 0.003; binomial test). Overall, this means that the perceptual template used by a rat to process the visual stimuli at a given size (25º) was able to predict its perceptual choices also at a larger size (30º), once properly scaled, better than any of the perceptual templates obtained, for the other animals, at the larger size itself. In other words, the perceptual strategy used by the rats remained highly subject-specific across size variations, and, therefore, largely size-tolerant.

For each rat i, additional evidence of size tolerance was obtained by measuring the ability of the model to predict the responses of the animal to the 11 distractors (same approach in Figure 33). As shown in Figure 39, no matter whether the model was based on CI_i (left panel) or CI_i^s (right panel), it predicted equally well rat

performances at both 25º and 30º of visual angle, in both the early sessions, when the regular-sized random bunnies were presented (black bars), or the late sessions, in which the small-sized random bunnies were shown (red bars). This indicates that the perceptual template obtained at a given size (e.g., 30º) was equally good at accounting for rat choices at that size and at a different one (e.g., 25º), once properly rescaled, thus confirming the tolerance of the object processing strategy to size changes.

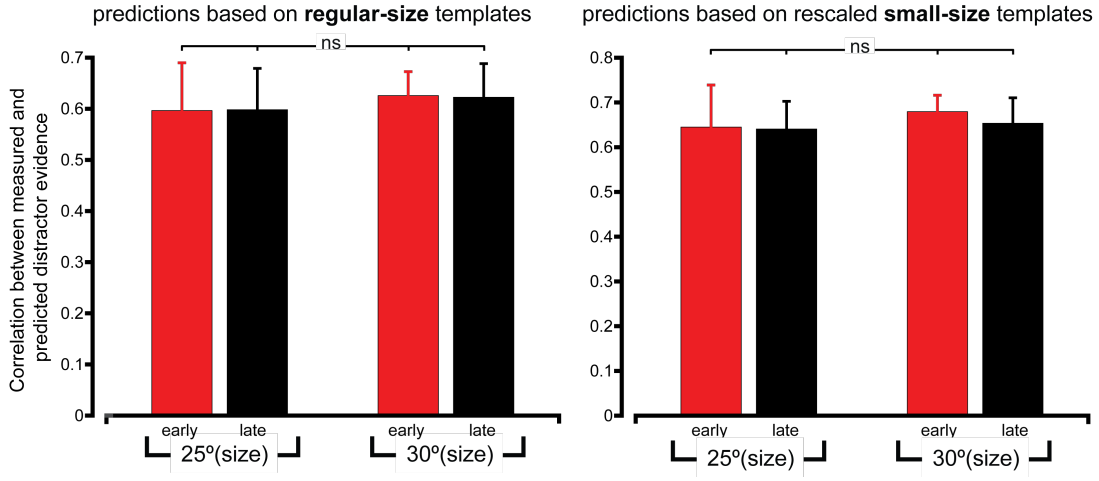

Figure 39. Correlation between measured and predicted distractor evidence, in the case of the distractor objects (i.e., analysis equivalent to Figure 33). The predictions were derived from models based either on the classification images obtained at the **regular size** (30º of visual angle) or on the rescaled versions of classification images obtained at the **small size** (25º of visual angle) – left and right panels, respectively. Correlations were obtained for the distractors shown both at size 25º and 30º, in both the early (red) and late (black) sessions, i.e., when the random tripods were shown, respectively, either at the regular or small size. No pairwise comparisons were significant according to a two-tailed, paired t-test

4.5 Discussion

Since we already discussed most of the results in the previous parts of this chapter (together with the results) we will use this section to summarize our findings and discuss the possible shortcomings.

Our study was conceived to overcome the limits of previous investigations of rat object vision (Zoccolan, 2015), including recent applications of the Bubbles method (Gosselin and Schyns, 2001; Gibson et al, 2005 and 2007) to uncover rat visual processing strategies (Alemi-Neissi et al, 2013; Rosselli et al, 2015; Vermaercke and Op de Beeck, 2012). To this aim, we designed a psychophysical test with three key, innovative features:

1) Rats were trained in a discrimination task that more closely matched the complexity encountered during natural vision, given that a reference object (the *tripod*) had to be distinguished from multiple distractors, encompassing a variety of shapes (Figure 12A). This was important for two reasons. First, we demonstrated that rats can effectively perform a two-alternative forced choice, image classification, task when exposed to more than just two stimuli. This finding can potentially be used in some future experiment concerning memory in rats. Second, and more important, our results show how fundamental is the choice of stimuli if we are set to discover the underlying recognition strategy in rats. If our study (for example) included only the first four distractors (Figure 12A), the resulting strategy would probably be very different then the one that we observed (compare the performances and the overlaps for the first four distractors, Figure 25). It seems very likely that, if not forced to use a more sophisticated strategy, rats can find a low level solution that would solve the task they face. This comes as a plausible explanation for the inconsistencies found in the other papers concerning the rats' recognition strategy (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012).

2) Rats' recognition strategy was inferred using a specially designed classification image method, where, rather than applying additive or multiplicative noise (Murray, 2011), altered versions of the reference (the *random tripods*) were produced by randomly varying the structure of the object itself (Figure 17). This allowed sampling a region of the shape space, centered over the reference, in a rather homogeneous way, yielding a perceptual template that was not confined within the boundaries of the

reference object, but spanned the whole image plane (Figure 18). As such, it was possible to incorporate it in a generalized linear model of rats' perceptual decisions (Figures 21 and 22).
3) And the key difference from previous studies – we went beyond the qualitative assessment of the perceptual strategy deployed by a rat, and we explicitly tested whether this strategy could predict the behavioral choices of the animal.

Our modeling approach yielded the following conclusions:
1) It showed that the visual processing strategies employed by the rats were subject-specific (Figures 32 and 38), thus confirming quantitatively what was inferred by inspecting the animals' perceptual templates (Figure 27).
2) The model's predictions for different animals were able to account for the overall pattern of performances observed across subjects and distractors, capturing the variable proficiency of the rats in the discrimination task (Figure 33). This indicates that the complexity of the perceptual strategy deployed by a rat played a key role in determining its performance with the distractors, as confirmed by the control analyses ruling out any major influence of the bias (Figures 34 and 35).
3) We found that rat successful discrimination of visual objects under variation of size and luminance cues could not be accounted by a low-level strategy, where input images are matched to a single perceptual template (Figure 36). Instead, we obtained evidence for a parallel processing of the visual input through multiple perceptual filters (Figures 38 and 39), whose shape was largely preserved across transformations (Figures 37A-C), but that were properly reformatted (e.g., rescaled) to deal with the specific transformations the objects underwent.

4.5.1 Possible shortcomings of our study

There are two issues that we didn't comment on directly but could potentially be seen as shortcomings of our study.

The first issue concerns the fact that all classification images for all the rats and all the variations (30º of visual angle, 25º of visual angle as well as filled in outlines) of random tripods produced the classification images with a very distinct salient upper right lobe which was especially prominent for the good performing rats (see Figure 37). Considering that our reference object is completely symmetrical, in

theory, there should be no reason for this type of preference. However, we believe that this is a result of the training process and the fact that rats had to lick the right sensor in order to make a reply and receive a reward for the correct identification of the reference object. Since the stimulus was presented at all times during the collection of the reward, rats spent much more time looking at the right part of the screen while the reference object was presented. It is also possible that rats apply some sort of decision tree by sequentially searching for salient and anti-salient features on the screen, but our study doesn't address this possibility since the evidence would have to include some sort of eye tracking.

The second issue is not so apparent but it deserves a comment nonetheless. It concerns the fact that both salient and anti-salient regions of the classification images contain within them a correlation that is a product of the way we produced our random tripods. Since the pixels within each single lobe are always correlated, we cannot be sure whether (for example) the tip of the lobe or the middle of the lobe carries a higher importance in the classification process. We addressed this problem by implementing the outline version of the random tripods, but a mathematically correct way to address this issue would be to apply a more stringent approach as the one called "decision images" developed by Macke and Wichmann (2010).

4.5.2 Conclusion

Overall, these findings account for the diversity of processing skills reported across and within studies of rat object vision (Zoccolan et al, 2009; Alemi-Neissi et al, 2012; Tafazoli et al, 2012; Rosselli et al, 2015; Vermaercke and Op de Beeck, 2012; Vinken et al, 2014; Bossens et al, 2016, Minini and Jeffery, 2006) establishing, in a quantitative way, the idiosyncratic nature of rat perceptual strategies. More importantly, our results provide the strongest behavioral evidence, to date, of advanced shape processing in a rodent species. In fact, the central role played by the featural complexity of rat perceptual templates, along with the tolerance of such templates to changes in object appearance, match very closely the two key computations implemented in the ventral stream of primates DiCarlo et al (2012) and in neural network models of the visual system (LeCun et al, 2015; Riesenhuber and Poggio, 2000; Kriegeskorte, 2015; Yamins and DiCarlo, 2016):

1) the construction of shape tuning, by combing multiple features into perceptual filters of increasing complexity; and
2) the construction of transformation tolerance, by iterating these filters across multiple scales, positions, etc.

Interestingly, our group have recently found evidence of similar processing mechanisms along the progression of extrastriate areas that, in the rat brain, run laterally to primary visual cortex (Tafazoli et al, 2017). This finding, together with other reports of high-level processing in these areas (Vermaercke et al, 2014 and 2015; Vinken et al. 2016 and 2017) argues for the existence of a rodent, ventral-like processing pathway. The data presented in our study add compelling behavioral evidence in support of this hypothesis, thus contributing to pave the way for the investigation of visual cortical processing using the powerful experimental approaches that rodent species afford.

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

Alemi-Neissi, A., Rosselli, F.B., Zoccolan, D. (2013) Multifeatural shape processing in rats engaged in invariant visual object recognition. J Neurosci 33:5939–56.

Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L., Reid, R.C. (2011). Functional Specialization of Mouse Higher Visual Cortical Areas. Neuron 72:1025–1039.

Anzai A., Peng X., Van Essen D.C (2007). Neurons in monkey visual area V2 encode combinations of orientations. Nature Neuroscience 10:1313 – 1321

Bishop, C.M. (2006), Pattern Recognition and Machine Learning, Springer

Baldassi, C., Alemi-Neissi, A. Pagan, M., DiCarlo, J.J., Zecchina, R., Zoccolan, D. (2016) Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. PLoS Comput Biol. 9(8): e1003167.

Bonin V., Histed M.H., Yurgenson S., Reid R.C. (2011) Local diversity and fine-scale organization of receptive fields in mouse visual cortex. J. Neurosci. 31:18506–21

Bossens, C., and Op de Beeck H.P. (2016). Linear and Non-Linear Visual Feature Learning in Rat and Humans. Front Behav Neurosci. 10:235.

Burkhalter. A., and Van Essen D.C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. J Neurosci. 6(8):2327-51.

Burn C.C. (2008). What is it like to be a rat? Rat sensory perception and its implications for experimental design and rat welfare. Appl. Anim. Behav. Sci. 112:1–32

Campbell F.W., and Gubisch R.W. (1966). Optical quality of the human eye. J. Physiol. 186:558–78

Campbell, B.A., and Messing, R.B., (1969). Aversion thresholds and aversion difference limens for white light in albino and hooded rats. J. Exp. Psychol. 82:353–359.

Chelazzi, L. et al. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. J. Neurophysiol. 80: 2918–2940

Coogan, T.A., and Burkhalter, A. (1990). Conserved patterns of cortico-cortical connections define areal hierarchy in rat visual cortex. Experimental Brain Research, 80:49-53.

Coogan, T.A., and Burkhalter, A. (1993). Hierarchical organization of areas in rat visual cortex. Journal of Neuroscience, 13:3749-3772.

Cox, D. (2014). Do we understand high-level vision? Current Opinion in Neurobiology 2014, 25:187–193

Damasio, A, et al. (1980). Central achromatopsia: behavioral, anatomic, and physiologic aspects. Neurology. 30:1064–71

Dean P. (1981). Visual pathways and acuity in hooded rats. Behav. Brain. Res. 3:239–71

Deisseroth, K. (2011). Optogenetics. Nat. Methods 8:26–29.

Delorme, A., Richard., G. and Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. Vision Research, 40 (16): 2187–2200

Diamond M.E, von Heimendahl M., Knutsen P.M., Kleinfeld D., Ahissar E. (2008). “Where” and “what” in the whisker sensorimotor system. Nat. Rev. Neurosci. 9:601–12

DiCarlo, J.J., and Maunsell, J.H.R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J. Neurophysiol. 89:3264–3278

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

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How Does the Brain Solve Visual Object Recognition? Neuron *73*, 415–434

Downing P.E., Jiang Y., Shuman M., Kanwisher N. (2001). A cortical area selective for visual processing of the human body. Science. 293(5539):2470-3.

Dreher, B. (1972). Hypercoplex cells in the cat's striate cortex. Investigative Ophthalmology, 355-356.

Epstein R., and Kanwisher N. (1998): A cortical representation of the local visual environment. Nature 392:598–601.

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

Fabre-Thorpe, M. (2011). The Characteristics and Limits of Rapid Visual Categorization. Front Psychol. 2: 243

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

Fize D., Vanduffel W., Nelissen K., Denys K., Chef d'Hotel C., Faugeras O., Orban G.A. (2003). The Retinotopic Organization of Primate Dorsal V4 and Surrounding Areas: A Functional Magnetic Resonance Imaging Study in Awake Monkeys. J.Neurosci. 23(19):7395–7406

Gaffan, E.A. and Woolmore, A.L. (1996). Complex visual learning by rats. Learn Motiv 27:375–99

Gaffan, E.A., Eacott, M., Simpson, E. (2000). Perirhinal cortex ablation in rats selectively impairs object identification in a simultaneous visual comparison task. Behav Neurosci. 114:18–31.

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, 376–82.

Gibson, B.M., Lazareva, O.F., Gosselin, F., Schyns, P.G., and Wasserman, E.A. (2007). Nonaccidental Properties Underlie Shape Recognition in Mammalian and Nonmammalian Vision. Curr. Biol. 17, 336–340.

Glasser M.F., Coalson T.S., Robinson E.C., Hacker C.D, Harwell J., Yacoub E., Ugurbil K., Andersson J., Beckmann C.F., Jenkinson M., Smith S.M., Van Essen D.C. (2016). A multi-modal parcellation of human cerebral cortex. Nature. 536(7615):171-8.

Goddard E., Mannion D.J., McDonald J.C., Solomon S.G., Clifford C.W.G. (2011). Color responsiveness argues against a dorsal component of human V4. Journal of Vision. 11(4):3

Gosselin, F., and Schyns, P.G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. Vis. Res *41*, 2261–71.

Greenberg, D.S., Houweling, A.R., 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., Kourtzi Z., Kanwisher N. (2001). The lateral occipital complex and its role in object recognition. Vis. Res. 41:1409–1422.

Hasson, U., Harel, M., Levy, I., Malach, R. 2003. Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron 37:1027–41

Hirsch J., and Curcio C.A. (1989). The spatial resolution capacity of human foveal retina. Vision Res. 29:1095–101

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J Physiol. 148(3):574–591.

Hughes, A. (1977). The refractive state of the rat eye. Vision Res. 17:927–939.

Hung, C.P. et al. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science 310: 863–866

Ito, M. et al. (1995). Size and positionin variance of neuronal responses in monkey inferotemporal cortex. J. Neurophysiol. 73: 218–226

Jacobs, G. H., Fenwick, J. A., Williams, G. A. (2001). Cone-based vision of rats for ultraviolet and visible lights. J. Exp. Biol. 204:2439-2446.

Krechevsky, I. (1938a). An experimental investigation of the principle of proxim- ity in the visual perception of the rat. J Exp Psychol. 22:497–523.

Krechevsky, I. (1938b). A note on the perception of linear Gestalten in the rat. Pedagog Semin J Genet Psychol. 52:241–6.

Krieg, W.J.S. (1946). Connections of the cerebral cortex; the albino rat; topography of the cortical areas. J. Comp. Neurol. 84, 221–275. & 277–323

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu. Rev. Vis. Sci. 1, 417–446.

Langley, G. (2006). Next of Kin: A Report on the Use of Primates in Experiments, British Union for the Abolition of Vivisection, p. 12

Lashley K.S. (1930a). The mechanisms of vision: III. The comparative visual acuity of pigmented and albino rats. J. Genet. Psychol. 37: 481–4

Lashley, K.S. (1930b). The mechanisms of vision I – A method for rapid analysis of pattern vision in the rat. J Genet Psychol 37:453–60.

Lashley, K.S. (1938). The mechanisms of vision: XV. Preliminary studies of the rat's capacity for detail vision. J Gen Psychol 18:123–93.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444

Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R. 2001. Center-periphery organization of human object areas. Nat. Neurosci. 4:533–39

Lewis, J.W., and Van Essen, D.C. (2000) Architectonic parcellation of parieto-occipital cortex and interconnected cortical regions in the Macaque monkey. J. Comp. Neurol. 428:79-111.

Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. Annu. Rev. Neurosci. 19:577–621.

López-Aranda M.F., López-Téllez J.F., Navarro-Lobato I., Masmudi-Martín M., Gutiérrez A., Khan Z.U. (2009). Role of layer 6 of V2 visual cortex in object-recognition memory. Science. 325(5936):87-9.

Lyon, D.C., and Kaas J.H. (2002). Evidence for a modified V3 with dorsal and ventral halves in macaque monkeys. Neuron. 33(3):453- 61

Malach, R. (1989). Patterns of connections in rat visual cortex. J Neurosci 9:3741-3752

Marion R., Li K., Purushothaman G., Jiang Y., Casagrande V.A (2013). Morphological and neurochemical comparisons between pulvinar and V1 projections to V2. J Comp Neurol. 521(4):813-32

Marshel, J.H., Garrett, M.E., Nauhaus, I., Callaway, E.M. (2011). Functional Specialization of Seven Mouse Visual Cortical Areas. Neuron 72:1040–1054.

McDonald, A.J., and Mascagni, F. (1996). Cortico-cortical and cortico-amygdaloid projections of the rat occipital cortex: a Phaseolus vulgaris leucoagglutinin study. Neuroscience 71:37–54.

Merigan W.H., and Katz L.M. (1990). Spatial resolution across the macaque retina. Vision Res. 30:985–91

Miller, M.W., and Vogt, B.A. (1984). Direct connections of rat visual cortex with sensory, motor, and association cortices. J Comp Neurol 226:184-202

Minini, L., and Jeffery, K.J. (2006). Do rats use shape to solve "shape discriminations"? Learn. Mem. Cold Spring Harb. 13:287–297.

Mishkin M., Ungerleider L.G., Macko K.A. (1983). Object vision and spatial vision: two cortical pathways. Trends in Neurosci. 6:414–417.

Montero, V. M., Bravo, H., Fernandez, V. (1973). Striate-peristriate cortico-cortical connections in the albino and gray rat. Brain Res., 53:202–207.

Montero, V.M. (1993). Retinotopy of cortical connections between the striate cortex and extrastriate visual areas in the rat. Experimental Brain Research, 94:1-15.

Moran, J., and Desimone, R. (1985). Selective Attention Gates Visual Processing in the Extrastriate Cortex. Science. 229 (4715): 782–4.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5), 2-2.

Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P., Reid, R.C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature 433, 597–603.

Olavarria, J., and Montero, V.M. (1981). Reciprocal connections between the striate cortex and extrastriate cortical visual areas in the rat. Brain research, 217:358-363.

Olavarria, J., and Montero, V.M. (1984). Relation of callosal and striate-extrastriate cortical connections in the rat: morphological definition of extrastriate visual areas. Experimental Brain Research, 54:240-252.

Op De Beeck, H. and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. J. Comp. Neurol: 426, 505–518

Polack, P.O., and Contreras, D. (2012). Long-Range Parallel Processing and Local Recurrent Activity in the Visual Cortex of the Mouse. J. Neurosci. 32:11120–11131.

Prusky, G.T., West, P.W., Douglas, R.M. (2000) Behavioral assessment of visual acuity in mice and rats. Vision Res. 40:2201–2209

Prusky, G.T., Harker, K.T., Douglas, R.M., Whishaw, I.Q. (2002). Variation in visual acuity within pigmented, and between pigmented and albino rat strains. Behavioral Brain Research. 136:339-348

Qiu, F.T., and von der Heydt, R. (2005). Figure and ground in the visual cortex: v2 combines stereoscopic cues with gestalt rules. Neuron. 47(1):155-66.

Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. Nat Neurosci *3 Suppl*, 1199–204.

Rolls, E.T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron. 27:205–218

Rose, M. (1929). Cytoarchitektonischer atlas der grobhirnrinde der maus. J. Psychol. Neurol. 40:1–51..

Rosselli, F.B., Alemi, A., Ansuini, A. Zoccolan, D. (2015). Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. Frontiers in neural circuits, 9.

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

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of 58 inferior temporal neurons in awake macaques. Exp. Brain Res. 77: 23–30

Sefton, A.J., Dreher, B., Harvey, A. (2004). Chapter 32 – Visual system. The Rat Nervous System (Third Edition):1083–1165

Sereno M.I., Dale A.M., Reppas J.B., Kwong K.K., Belliveau J.W., Brady T.J., Rosen B.R., Tootell R.B.H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science. 268:889–893

Sergent, J., Ohta, S., MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain. 115 (1): 15–36

Shapley, R., and Hawken, M. J. (2011). Color in the Cortex: single- and double-opponent cells. Vision Research, 51(7), 701-717.

Sokoloff L. (1982). The Radioactive Deoxyglucose Method. Advances in Neurochemistry. 4:1-82.

Sutherland, N.S. (1961). Visual discrimination of horizontal and vertical rectan- gles by rats on a new discrimination training apparatus. Q J Exp Psychol. 13:117–21

Sutherland, N.S., Carr, A.E., Mackintosh, J.A. (1962a). Visual discrimination of open and closed shapes by Rats. I. Training. Q J Exp Psychol. 14:129–39

Sutherland, N.S. and Carr, A.E. (1962b). Visual discrimination of open and closed shapes by Rats. II. Transfer tests. Q J Exp Psychol. 14:140–56

Szel, A. and Rohlich, P. (1992). Two cone types of rat retina detected by anti-visual pigment antibodies. Exp. Eye Res. 55:47–52.

Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F.B., Vanzella, W., Riggi, M., Buffolo, F., Panzeri, S. and Zoccolan, D. (2017). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. eLife 6:e22794.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. 19:109–139.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature. 381: 520–522.

Tootell R.B.H., and Hadjikhani N. (2001). Where is 'Dorsal V4' in Human Visual Cortex? Retinotopic, Topographic and Functional Evidence. Cereb. Cortex 11 (4):298-311.

Tsunoda, K. et al. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat. Neurosci. 4: 832–838

Uchida N., Mainen Z.F. (2003) Speed and accuracy of olfactory discrimination in the rat. Nat Neurosci 6:1224–9

Ungerleider, L.G., and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. J. Comp. Neurol. 248:190-22.

Ungerleider L.G., Galkin T.W., Desimone R., Gattass R. (2008). Cortical connections of area V4 in the macaque. Cereb. Cortex. 18(3):477-99.

Van den Bergh, G., Zhang, B., Arckens, L., Chino, Y.M. (2010). Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. J. Comp. Neurol. 518:2051–2070.

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.

Vermaercke, B., Gerich, F.J., Ytebrouck, E., Arckens, L., Op de Beeck, H.P, Van den Bergh, G. (2014). Functional specialization in rat occipital and temporal visual cortex. J. Neurophysiol. 112:1963–1983.

Vermaercke, B., Van den Bergh, G., Gerich, F., and Op de Beeck, H. (2015). Neural discriminability in rat lateral extrastriate cortex and deep but not superficial primary visual cortex correlates with shape discriminability. Front. Neural Circuits 9, 24.

Vinken, K., Vermaercke, B., and Op de Beeck, H.P. (2014). Visual Categorization of Natural Movies by Rats. J. Neurosci. *34*, 10645–10658

Vinken, K., Van den Bergh, G., Vermaercke, B., Beeck, O. de, and P, H. (2016). Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. Cereb. Cortex 26, 3310–3322

Vinken, K., Vogels, R., and Op de Beeck, H. (2017). Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. Curr. Biol. 27, 914–919.

Wagor, E., Mangini, N.J., Pearlman, A.L. (1980). Retinotopic organization of striate and extrastriate visual cortex in the mouse. Journal of Comparative Neurology, 193(1):187-202.

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.

Wang, Q. and Burkhalter, A. (2007). Area map of mouse visual cortex. Journal of Comparative Neurology, 502(3):339-357.

Wang, Q., Gao, E., Burkhalter, A. (2011). Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. J. Neurosci. 31:1905 –1918.

Wang, Q., Sporns, O., Burkhalter, A. (2012). Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. J. Neurosci. 32:4386–4399.

Wang, Y. et al. (2002). Contribution of GABAergic inhibition to receptive field structures of monkey inferior temporal neurons. Cereb. Cortex 12:62–74

Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356–365.

Zeki, S.M. (1969). Representations of central visual fields in prestriate cortex of monkey. Brain Res. 14:271-291

Zeki S.M, (1990). A century of cerebral achromatopsia. Brain. 113 (Pt.6): 1721–77

Zoccolan, D., Oertelt, N., DiCarlo, J.J., Cox, D.D. (2009). A rodent model for the study of invariant visual object recognition. Proc Natl Acad Sci. 106:8748–53.

Zoccolan, D. (2015). Invariant visual object recognition and shape processing in rats. Behavioural Brain Research. 285:10–33