KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Neurosciences

COMPLEX AND ADAPTIVE REPRESENTATIONS IN RAT AND MACAQUE VISUAL CORTEX

Kasper Vinken

Promoter: Prof. Dr. Rufin Vogels
Co-promoter: Prof. Dr. Hans Op de Beeck
Chair: Prof . Dr. Peter Janssen
Secretary: Prof. Dr. Wim Vanduffel
Jury members:
Prof. Dr. Valérie Goffaux
Prof. Dr. Richard Van Wezel
Prof. Dr. Wim Vanduffel
Prof. Dr. Patrick Dupont

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

December 2017

SAMENVATTING

De studie van visuele perceptie bestrijkt niet alleen een grote verscheidenheid aan disciplines, gaande van cognitieve wetenschappen tot computerwetenschappen tot neurowetenschappen, maar ook een waaier aan diermodellen, gaande van primaten tot knaagdieren tot zelfs insecten. Het onderzoek van deze thesis ligt op het raakvlak van cognitieve en neurowetenschappen, met enkele ideeën ontleend aan computerwetenschappen. Het is een verzameling van studies uitgevoerd met zowel ratten (Laboratorium voor Biologische Psychologie, KU Leuven) als apen (Laboratorium voor Neuro- en Psychofysiologie, KU Leuven), samengebracht onder de titel: *"Complexe en adaptieve representaties in de visuele cortex van de rat en makaak"*. In een informatieverwerkingssysteem zoals het visuele, verwijst de term *representatie* naar de informatie die door het systeem expliciet gemaakt is. Bijvoorbeeld, welke eigenschappen van de visuele omgeving drijven in een zeker stadium van verwerking de neurale activiteit? Over het algemeen stijgen deze eigenschappen in complexiteit doorheen de visuele stroom. Neurale activiteit wordt echter niet alleen door huidige visuele stimulatie gedreven, maar past zich ook aan bij voorafgaande stimulatie. Dus, het uiteindelijke product is een *complexe* en *adaptieve* representatie.

Voor onze eerste onderzoekslijn hebben we het visuele systeem van de rat onderzocht. Deel I van deze studies is gericht op objectherkenning en omvat hoofdstukken 1-3. In Hoofdstuk 1 gaan we visuele classificatievaardigheden van ratten met natuurlijke filmpjes na. Vervolgens rapporteren we in Hoofdstuk 2 veranderingen in neurale representaties van deze filmpjes doorheen een baan in de visuele cortex van de rat. Die baan wordt weleens voorgesteld als homoloog van de ventrale visuele stroom in primaten. Ten slotte onderzoeken we in Hoofdstuk 3 of deze neurale representaties de visuele classificatievaardigheden besproken in Hoofdstuk 1 zouden kunnen ondersteunen. Dit doen we door middel van een vergelijking van corticale representaties van de filmpjes met representaties in de lagen van een diep neuraal netwerk model. Over het geheel genomen, concluderen we dat de vermeende rat ventrale stroom resulteert in een relatief complexe representatie van visuele input: één die niet direct

categorie-gerelateerd is, maar die misschien toch generalisatie in een complexe classificatie taak zou kunnen ondersteunen.

Voor Deel II, wat enkel Hoofdstuk 4 inhoud, focussen we op dezelfde baan in de visuele cortex van de rat, om te onderzoeken hoe recente visuele stimulatie neurale antwoorden beïnvloedt. De resultaten verschillen opmerkelijk van die in de aap, en we speculeren dat deze baan in de rat misschien gespecialiseerd is in de detectie van veranderingen. Dit Hoofstuk geeft de overgang naar het onderwerp van neurale adaptatie aan.

In Deel III zetten we het onderzoek naar neurale adaptatie verder. Hier richten we ons op de relatie met de *predictive coding* theorie, die veronderstelt dat corticale antwoorden schendingen van eerdere verwachtingen signaleren. We verplaatsten de focus naar apen om meer geavanceerde cognitieve processen te bestuderen. In Hoofdstuk 5 onderzoeken we de interactie tussen perceptuele verwachtingen en neurale adaptatie in de inferieur temporale cortex van de makaak. We vinden specifiek dat noch aandacht, noch een erg relevante stimulus categorie zoals gezichten, voldoende zijn om die interactie te kunnen observeren. Ten slotte gebruiken we in Hoofdstuk 6 een model dat neurale adaptatie mechanismen simuleert. We tonen aan dat stimulus-gedreven effecten van neurale adaptatie de zogenaamde verwachtings-effecten van een recente studie kunnen verklaren. Samen toont dit aan dat er onvoldoende bewijs is voor een algemene rol van perceptuele verwachtingen in adaptatie van neuronen in de inferieur temporale cortex.

SUMMARY

The study of visual perception not only spans a wide variety of fields, ranging from cognitive science to computer science to neuroscience, but also covers an abundance of animal models, from primates to rodents to even insects. The work presented in this dissertation lies at the intersection between cognitive science and systems neuroscience, while borrowing some ideas from computer science. It is a collection of studies conducted with both rats (Laboratory for Biological Psychology, KU Leuven) and monkeys (Laboratory for Neuro- and Psychophysiology, KU Leuven), brought together under the title of: *"Complex and adaptive representations in rat and macaque visual cortex"*. In an information processing system such as the visual system, the term *representation* refers to the information that is made explicit by the system. For example, which features of the visual world drive neural activity in a certain stage of processing? These features typically increase in complexity across the visual stream. However, neural activity is not only driven by current visual input, but also adapts to previous stimulation. Therefore, the eventual product is a *complex* and *adaptive* representation.

For our first line of research, we investigate the rat visual system. Part I of these studies is focused on object recognition and encompasses chapters 1-3. In Chapter 1, we investigate visual classification abilities of rats with naturalistic movies. In Chapter 2, we report changes in neural representations of these movies across a pathway in the rat visual cortex. This pathway has been proposed to be a homologue to the primate ventral visual stream. Finally, we explore in Chapter 3 whether these neural representations might be able to support the visual classification abilities presented in Chapter 1. We do this by comparing cortical representations of the movies with representations in layers of a deep neural network model. Overall, we conclude that the putative rat ventral stream results in a relatively complex representation of visual input: one that is not directly category-related, yet might support generalization in complex classification tasks.

For Part II, which includes only Chapter 4, we continue with the same pathway in rat visual cortex to study how recent visual stimulus history affects neural responses. The

results differ markedly from those in monkeys, and we speculate that maybe this pathway is specialized in in change detection in rats. This chapter marks a transition to the topic of neural adaptation.

We continue our research on neural adaptation in Part III. Here, we focus on its relation to the predictive coding theory, which postulates that cortical responses signal violations of prior expectations. We move the focus to monkeys to be able to study more advanced cognitive processes. In Chapter 5, we investigate the interaction between perceptual expectations and neural adaptation in macaque inferior temporal cortex. Specifically, we find that neither attention, nor a highly relevant stimulus category such as faces, are sufficient for observing that interaction. Finally, in Chapter 6, we use a model that simulates neural adaptation mechanisms. We show that simple, stimulus-driven effects of neural adaptation can explain a recent study's proclaimed expectation effects. Taken together, the actual evidence does not support a general role of perceptual expectation in adaptation of inferior temporal cortex neurons.

ACKNOWLEDGEMENTS

Rufin and Hans both deserve my greatest gratitude. I was lucky to not only have one, but two exceptionally outstanding scientists as advisors. It is a privilege to be associated with their recognition in the scientific world, which has opened up significant opportunities for me (starting with a PhD fellowship). At the same time I have enjoyed an excellent training as a scientist, learning from Hans' astute scientific insight and Rufin's unparalleled rigour and skepticism. In addition, they have both always been prepared to make ample time for discussion or help whenever needed. In fact, rather than a co-promotor, Hans has always also been a promotor for me. I am truly honored that I have been able to spend my PhD years under Rufin and Hans' supervision.

Besides my supervisors, I would like to thank my jury members: Prof. Dr. Wim Vanduffel, Prof. Dr. Patrick Dupont, Prof. Dr. Valérie Goffaux, and Prof. Dr. Richard Van Wezel, for their helpful and particularly supportive comments on this dissertation.

To the (former) members of LBP: thanks to everybody for providing such a stimulating environment! There was always room for an interesting scientific discussion as well as a party, a mix that was brought to perfection at the retreat. In particular, I'd like to thank Gert and Ben, for training me early on. Christophe (Bossens), for our incessant, inspiring discussions about all things science related. Jessica, Lien, Nicky, Lore, and Christophe for the good times at work or after work. Finally, An, Nicky, and Leen, for their valuable help with administration or any other random issues.

To the (former) members of Neuro- and Psychophysiology: thanks to all the former and current colleagues for a very supportive and pleasant lab environment! Thanks especially to Christophe (Ulens) and Elsie for having helped me more than anybody with all sorts of problems. To my (former) office colleagues: Jess, Hamed, Ivo, Peter, Pradeep, Susheel, Satwant, Ioannis, Francesco, and Victor for the laughs, discussions and animal caretaking. In addition, the supporting staff in this lab is truly exceptional. So, naturally, my gratitude goes to Christophe, Inez, Astrid, Sara, Anne, Chantal, Wouter, Stijn, Marc, Piet, Gerrit, Jan, Ronny, and Kim. Thank you for always being so helpful!

I would like to thank all of my friends, whom I have always relied on for entertainment, distraction, and driving me around. Especially thanks to Michiel for the relentless Friday night distraction during my busiest weeks. In addition, thanks to Johan, Sara, and Jeroen (baard) for always being interested in my work more than anyone else, and to Wim, my ever so informative scientific career advisor!

I thank my family for their support and interest. Thanks Lore, Jelle, and in particular, mama, for providing infinite support with pretty much anything.

Finally, Saloni, my one and only source of bubbles! I cannot thank her enough for her support, for putting up with my crazy long working hours, for always believing in me… And, of course, for just being there.

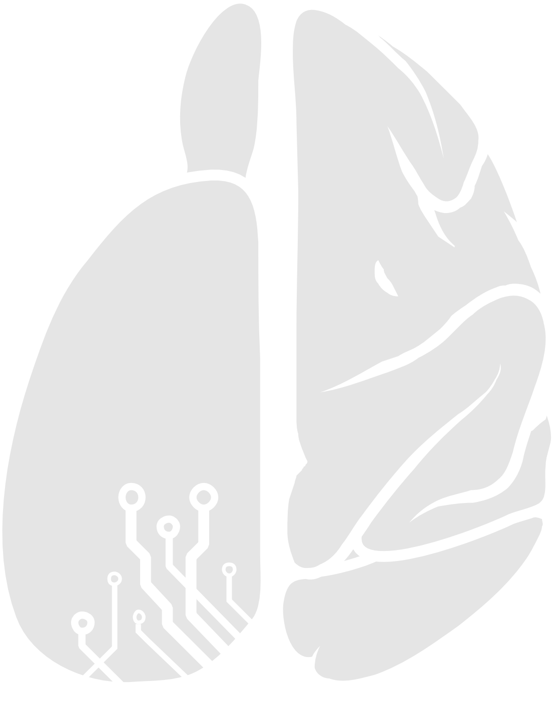
CONTENT

FIGURES AND TABLES

ABBREVIATIONS

2AFC	two-alternative forced choice
AB	alternation block
AF	anterior fundus face patch
AI	adaptation index
AIT, aIT	anterior inferior temporal cortex
AL	anterior lateral face patch
AM	anterior medial face patch
AT	alternation trial
BCa	bias-corrected accelerated bootstrap
BF	Bayes factor
CI	confidence interval
CIT, cIT	central inferior temporal cortex
DNN	deep neural network
EEG	electroencephalography
EI	equivalence index
FBNF1	Fisher 344 × Brown Norway hybrid
FFA	fusiform face area
FFT	fast Fourier transform
fMRI	functional magnetic resonance imaging
FSI	face selectivity index
HDI	highest density interval
IQR	interquartile range
IT	inferior temporal cortex
LFP	local field potential
LI	latero-intermediate cortex
LL	latero-lateral cortex
LM	lateromedial cortex
LOWESS	locally weighted scatterplot smoothing
M	mean
MDS	multidimensional scaling
MF	middle fundus face patch
MION	monocrystalline iron oxide nanoparticle
ML	middle lateral face patch
MMN	mismatch negativity
MRI	magnetic resonance imaging
MU	multi-unit
MUA	multi-unit spiking activity
OLS	ordinary least squares
PC	pixel change / principal component
PCA	principal component analysis
PER	perirhinal cortex
PIT, pIT	posterior inferior temporal cortex
PL	posterior lateral face patch
PSTH	peristimulus time histogram
RB	repetition block
RDM	representational dissimilarity matrix
RMS	root mean square (contrast)
ROI	region of interest
RT	repetition trial
SD	standard deviation
SU	single-unit
SVM	support vector machine
TO	lateral occipito-temporal cortex
V1	primary visual cortex
VEP	visually evoked potentials

INTRODUCTION



0.1 General introduction

Understanding the brain, and how it gives rise to behavior, is one of the most complicated yet important challenges in science today. It is hard to overstate the value of goals such as explaining the workings of brain disorders, or uncovering the mysteries of intelligence. Our brain is essentially who we are, everything we experience. Yet, we still do not understand this machine and how it processes information. That is, which computations transform sensory input to internal representations and eventually to an output, to behavior? Vision is a remarkable example of such information processing, which seems so effortless to us despite the tremendous computational challenge (Dicarlo et al., 2012). How does the brain construct a rich and meaningful representation of the outside world from the blizzard of photons entering the eye? This question, which has attracted scientists from many different fields over the centuries, is the backdrop of this dissertation. The study of this problem has branched out in a great variety of research topics, sometimes with very little overlap. Over the last few years I have worked on two of those topics, which I will briefly introduce here before going into more detail.

Historically, research on visual perception has mainly focused on humans – or, more generally, on primates. Like other primates, we are highly specialized in visual object recognition, which is the ability to recognize objects despite an enormous variation in the retinal projection. In order to achieve this, our visual system needs to construct an internal representation that is relatively invariant to changes in size, position, luminance, etc., but remains selective for object identity (Logothetis and Sheinberg, 1996; Tanaka, 1996; Dicarlo et al., 2012). We have long suspected that neural processing underlying this capability largely takes place in a visual pathway called the ventral visual stream (Mishkin et al., 1983), without much of an understanding of how this is actually implemented. After all, reaching an invariant representation while maintaining selectivity is a major computational challenge (Riesenhuber and Poggio, 2000; Rust and Dicarlo, 2010). A computational principle by which this can be achieved was introduced over 50 years ago in Hubel and Wiesel's (1962) seminal work on the primary visual cortex (V1). They proposed the idea that the output of several simple cells, with identical orientation preference but different retinal positions, is combined in complex cells with local tolerance to spatial shifts. Only recently, this principle has been generalized successfully in hierarchical neural network models that not only predict object selective neural responses surprisingly well, but also achieve unprecedented categorization performance in relatively confined object recognition tasks (Kriegeskorte, 2015).

Although much progress on vision has focused on primates, not much is known about other animals. In particular, we know very little about the rodent visual system. This is partly not surprising, given that they mainly rely on other senses such as touch, smell, and audition (Zoccolan, 2015). On the other hand, it is somewhat unexpected, because rats and mice are so widespread as a model in neuroscience. In recent years, however, there has been increased interest in their visual system, driven by the advances in scientific tools available to study their neural circuitry (Huberman and Niell, 2011). Still, there are reasons other than the available toolkit for investigating rodent vision. One argument, often advocated by Zoccolan (2009, 2015), is that a simpler system should be easier to understand, and that some principles might translate to a more complex system. Much like studying V1 allowed Hubel and Wiesel (1962) to discover a computational principle that may generalize well to more advanced stages of the visual system (and across species). On the other hand, the visual system of different animals is expected to be uniquely tailored to their functional requirements (Marr, 1982). Nevertheless, studying a diversity of systems with a related function can lead to the identification of universal computational principles (Carandini and Heeger, 2011). In addition, a comparative approach is one of the only ways to uncover how a complex systems and their specializations have evolved (Krubitzer, 2009). In Part I and II of this dissertation we focus on the rat visual system, and approach the subject by assessing to what extent it also expresses some general principles of the primate ventral stream.

While the success of deep neural network models of the ventral stream looks promising, they assume a fixed visual representation and ignore the temporal dynamics of the system. Yet, both object perception and neural responses change as a result of previous stimulation, even as a function of short-term stimulus history. These changes in neural responses are often referred to by the umbrella term *neural adaptation* (Kohn, 2007; Wark et al., 2007). Even though this phenomenon may be largely explained by bottom-up and local mechanisms (Solomon and Kohn, 2014; Vogels, 2016), some researchers have proposed a predictive coding account which emphasizes the role of top down influences of perceptual expectation (Friston, 2005; Summerfield and de Lange, 2014). The predictive coding framework is a theory of sensory processing that, if proven to be universal, has far reaching implications. In Part II and III of this dissertation, we focus on the relation between expectation and adaptation in the visual system.

In summary, in this dissertation I will present the work we have done on the topics of rat vision and on neural adaptation in the visual system. Before we proceed to the actual body of the dissertation, it seems appropriate to provide some background for both topics.

0.2 RAT VISION

There are several distinct advantages of studying neural circuits in rodents. The primary reason often cited is the development and widespread use in mice of molecular and genetic tools that allow, for example, for recording activity of a large number of individual neurons, or for reversibly silencing or activating specific cell-types (O'Connor et al., 2009; Huberman and Niell, 2011). In addition to that, thanks to the smaller overall size of rodent brains, it is possible to simultaneously monitor neural activity in a large part of their visual system (Andermann et al., 2011; Marshel et al., 2011; Garrett et al., 2014). Besides these scientific reasons, rats and mice are cheaper, easier to maintain and handle, and less subject to ethical concern than monkeys (Huberman and Niell, 2011; Baker, 2013).

However, rats and other rodents are not little monkeys. They have no fovea (Euler and Wässle, 1995) and a very low visual acuity of 1-1.5 cycles per degree for pigmented strains (Prusky et al., 2002). They are primarily nocturnal (Burn, 2008), and therefore rely extensively on whisker touch and smell when exploring or navigating the environment (Zoccolan et al., 2015). So, what do they use vision for? It has been suggested that their eye movements serve to keep a continuously overlapping (binocular) field for enhanced predator detection (Wallace et al., 2013). Basically, vision is the only useful modality for timely detection and avoidance of aerial predators (Yilmaz and Meister, 2013). A second function is that of spatial navigation. For example, primarily vision is used for guidance in laboratory water mazes (Burn, 2008). Thus, predator detection and spatial navigation seem important functions of the rat visual system, but is it capable of more high-level tasks such as object recognition?

0.2.1 VISUAL OBJECT RECOGNITION

Because invariant visual object recognition entails such a computational challenge, it has often been assumed to be a unique hallmark of the primate visual system (Dicarlo et al., 2012; Zoccolan, 2015). Nonetheless, over the past decade, several studies have investigated these capabilities in the rat. While there is a long history of preceding behavioral studies of rat vision, dating back to over a century (Zoccolan et al., 2015), I will restrict myself here to giving an overview of the recent line of research aimed specifically at visual object recognition.

An important issue when studying such advanced visual functions is that rats will tend to find the simplest strategy available to successfully do the task, even if their solution is inconspicuous to us. This pitfall was highlighted in a series of experiments by Minini and Jeffery (2006), showing that rats did not actually use shape in a shape discrimination task, but instead relied on a strategy based on local luminance differences in the lower hemifield. The tendency to use this strategy was later confirmed by Vermaercke and Op de Beeck (2012), who also revealed that the behavioral templates were context dependent, meaning that rats seem to adapt the complexity of their strategy when necessary. Thus, it seems like these animals need to be pushed to use more advanced visual functions.

Zoccolan et al. (2009) really pushed their rats with the goal of probing truly invariant visual object recognition, by using 3-D rendered objects under a wide range of transformations. Rats that were trained to discriminate these objects despite variations in size and viewpoint, could successfully generalize to novel transformations. The authors concluded that, given the substantial variation in object appearance, the rats could not have relied on low-level visual strategies. The evidence for tolerance in the rat visual system in this task was later strengthened by demonstrating visual priming across transformations (Tafazoli et al., 2012). An investigation of the strategies used by the rats suggested they relied on a combination of visual features that was relatively stable across object transformations (Alemi-Neissi et al., 2013).

In conclusion, rats do seem to be able to demonstrate capabilities reminiscent of invariant visual object recognition when they need to. However, it is still hard to tell what level of processing was required for them to perform the task successfully. What if they found another, more complex strategy that still does not require an actual shape recognition? Vermaercke and Op de Beeck (2012) hypothesized that rats are capable of using flexible mid-level strategies based on a combination of local contrast cues. While not ruling out this possibility, Zoccolan has argued that object recognition based on a specific spatial arrangement of contrast cues could be considered shape-based (Zoccolan, 2015). In any case, on their own the behavioral data do not tell us which information the rat visual system used and how. In order to approach this problem, it might help to look at the structural and functional organization of that system.

0.2.2 Two streams

In the last 30 years a decent amount of research has been done to investigate the parcellation and connectivity of the rodent visual cortex. The rat visual cortex consists of several distinct regions organized in a multilevel hierarchy around V1 (Espinoza and Thomas, 1983; Coogan and Burkhalter, 1993). Mice have a similar organization of their visual cortex (Wang et al., 2007), which has been shown to be structured in two clusters based on their connectivity (Wang et al., 2011, 2012). Similarly, the primate visual cortex is typically divided in the dorsal and ventral stream (Mishkin et al., 1983). These two primate pathways each serve distinct functions that are closely related to their connectivity: visually guided actions for the dorsal and object processing for the ventral stream (Kravitz et al., 2011, 2013). The ventral stream transforms visual input signals into an object- or category-related complex representation that is tolerant for a range of

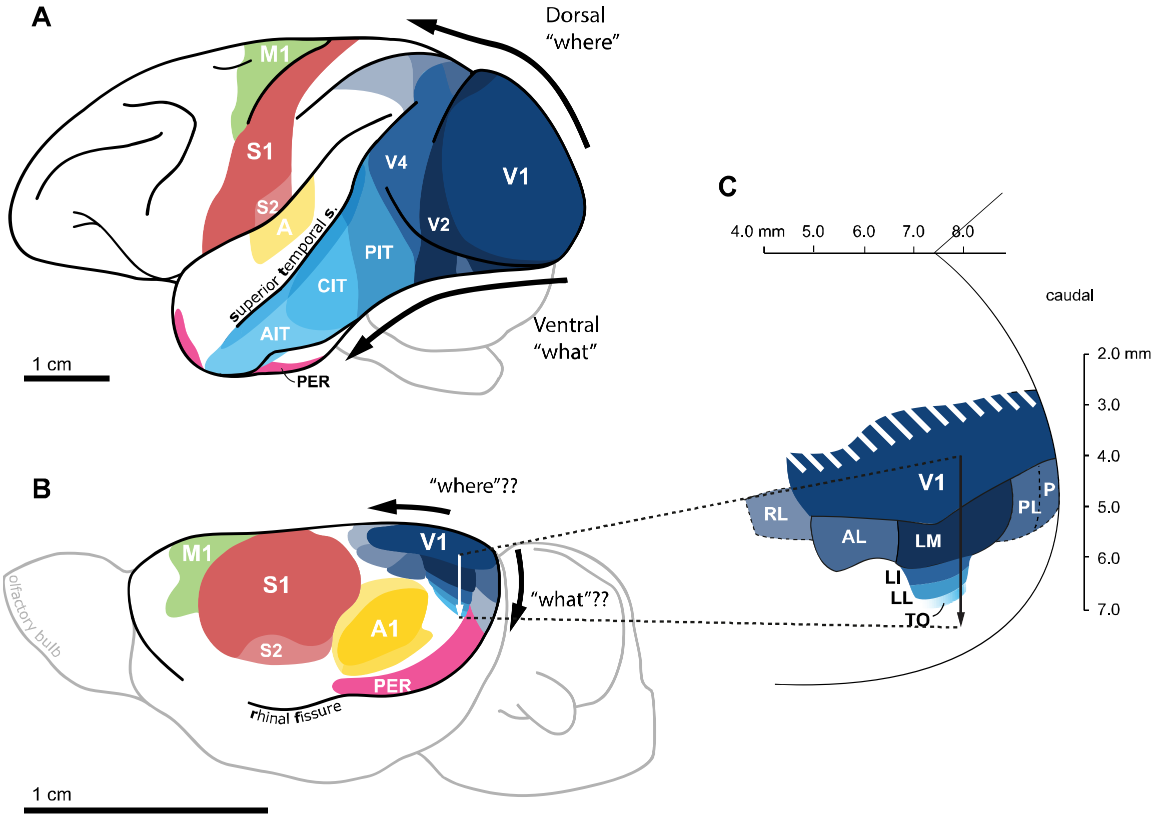


Figure 0.1. Two processing streams in macaque and rat visual cortex.
(A) Sensory cortices in the macaque brain. Area parcellation is based on Calabrese et al. (2015) and labeling on Felleman and Van Essen (1991). Arrows indicate the dorsal and ventral stream of visual cortical areas. Regions forming the ventral steam: visual area 1, 2, and 4 (V1, V2, and V4), and posterior, central, and anterior inferior temporal cortex (PIT, CIT, and AIT). Other regions: primary motor area M1, primary and secondary somatosensory areas S1 and S2, auditory cortex A, and perirhinal cortex PER. **(B)** Sensory cortices in the rat brain. Area parcellation in mainly based on Valdés-Hernández et al. (2011), Espinoza and Thomas (1983), Thomas and Espinoza (1987), and Vermaercke et al. (2014). Arrows indicate potential functional homologues of the primate dorsal and ventral visual stream. **(C)** Schematic top-view representation of V1 and lateral extrastriate regions in the rat, adapted from Vermaercke et al. (2014). The arrow illustrates how one electrode track can cross up to 5 different areas in a lateral pathway that could be a primate ventral stream homologue: V1, lateromedial (LM), latero-intermediate (LI), laterolateral (LL), and lateral occipito-temporal cortex (TO).

identity preserving transformations, such as changes in size, position, viewpoint, and illumination (Tanaka, 1996; Orban, 2008; Dicarlo et al., 2012).

Figure 0.1A shows a schematic representation of the sensory cortices in the macaque brain, with the two streams indicated for the visual cortex. **Figure 0.1**B shows an analogue schematic representation for the rat brain, which shows a similar constellation of sensory cortices that is typically shared amongst mammals (Krubitzer, 2009). Based on their location, the two aforementioned clusters have been proposed to be homologues of the two visual pathways in primates: a group of medial and parietal areas as a dorsal stream and a group of lateral temporal areas as the ventral stream (Wang et al., 2011, 2012). Indeed, posterior parietal lesions have been shown to selectively impair visuospatial functions in rats (Sánchez et al., 1997; Tees, 1999), while posterior temporal lesions lead to impaired object recognition and visual pattern discrimination (Wörtwein et al., 1994; Aggleton et al., 1997; Tees, 1999). Despite this evidence, surprisingly little is known of the functional properties of neurons in these regions.

In rat V1 the functional properties are very similar to those shared by cats and monkeys. The neurons are orientation selective with a distinction between simple and complex cells, but no organization in orientation columns (Girman et al., 1999). In mice, the posterior parietal cluster of extrastriate areas shows properties consistent with the idea of a dorsal stream for motion processing, such as increased direction selectivity relative to V1 (Marshel et al., 2011) and computation of global motion of complex patterns (Juavinett and Callaway, 2015). In rats, there have also been successful investigations into the functional properties of the putative ventral stream. Vermaercke et al. (2014) found, using simple shapes, that position tolerance increased along the pathway of the putative ventral stream. While an increase in position tolerance is typical for the primate ventral stream, they also reported an increased response to moving stimuli which is not typical. Recently, Tafazoli et al. (2017) greatly expanded upon those findings by showing that neural representations along the pathway increasingly support visual object discrimination despite changes in size, position, rotation, and illumination. However, this progression only became apparent when the authors extensively controlled for stimulus luminance. They conclude that "these findings strongly argue for the existence of a rat object-processing pathway" (Tafazoli et al., 2017).

Taken together, there is increasing anatomical and functional evidence in support of the two stream hypothesis of rat visual cortex that may to some extent be homologous to primates. On the other hand, from an ecological point of view it seems unlikely that object recognition is the main goal of the rat visual system. Furthermore, the connectivity in rodent visual cortex does differ in some notable ways: all visual areas receive input from V1, which additionally has extensive direct cross-modal connections (Wang et al., 2012; Laramée and Boire, 2015). This suggests that, as opposed to primates, rodent V1's outputs are directly integrated with other modalities. It is unknown to what extent these network differences indicate fundamentally different information processing. Thus, while the above results do sound very promising, we have yet to determine to what extent the mechanisms of visual object processing overlap with primates.

0.2.3 THE CAT MODEL

Rodents and non-human primates are not the only animal models commonly used for studying the visual cortex. Indeed, perhaps the most influential breakthrough for our understanding of the visual system was the early work of Hubel and Wiesel on receptive fields of single V1 neurons in anaesthetized cat (Hubel and Wiesel, 1959, 1962). Soon after, non-human primates became the dominant model to study visual processing. Hubel and Wiesel turned to monkeys to extend their work on V1, arguing that they are closer to humans in their visual capabilities (Hubel and Wiesel, 1968). At the same time there was an increased interest in studying higher visual functions, which required recordings in visual cortex of awake, behaving animals. Monkeys were the animal of choice for these recordings, because of their ability to perform complex tasks (Wurtz, 2009). Thus, studies of visual processing in cats have been more restricted to V1 recordings in anaesthetized animals. Like other primary (and secondary) sensory areas, V1 is considered to be evolutionary homologous across all mammals (Krubitzer and Hunt, 2007). However, parallels can be drawn even outside of V1: similar to primates and rodents, cat extrastriate visual cortex can be divided in two functionally distinct processing streams (Lomber et al., 1996). In addition, in cats there is evidence for early- to mid-level macaque visual area homologues (Payne, 1993). Still, while cats are more visual than rats and mice, they are phylogenetically even less related to primates (Krubitzer and Hunt, 2007). More importantly, however, cats lack the major advantages of rats and mice that we discussed earlier, such as the genetic toolkit, smaller brain size, low cost and ease of maintenance and handling.

0.3 NEURAL ADAPTATION

Neural responses in sensory cortex are not only dependent on the current input, but are also dependent on previous stimulation. A typical observation is that these responses are attenuated for repeated stimuli, a phenomenon called repetition suppression (Desimone, 1996). Repetition suppression refers specifically to a response reduction, while the more general term of neural adaptation can in principle also refer to response enhancement effects (Solomon and Kohn, 2014; Kaliukhovich and Vogels, 2016). However, it should be noted that in general these terms are often used interchangeably.

If adaptation for a repeated stimulus does not completely generalize to other stimuli, it is called stimulus-specific adaptation. This specificity is the basis of functional magnetic resonance imaging (fMRI) adaptation paradigms which are widely used to make inference about the functional properties of a neuronal population (Grill-Spector and Malach, 2001; Grill-Spector et al., 2006; Barron et al., 2016). The phenomenon is typically explained on the basis of relatively simple local and bottom-up mechanisms of neural fatigue. On the other hand, proponents of the predictive coding theory (Friston, 2005) have proposed that adaptation also involves top-down influences of perceptual expectations (Summerfield et al., 2008). In this section, I will provide a brief overview of the neural fatigue mechanisms and the perceptual expectation account of neural adaptation.

0.3.1 MECHANISMS OF NEURAL FATIGUE

In general, we can divide the fatigue related mechanisms into those acting on the level of a neuron and those acting on the level of a synapse. In addition, it is possible that a neuron’s response is suppressed indirectly, as a result of adaptation of other cells that provide input to said neuron.

Response fatigue

At the level of a neuron, response suppression can occur as a result of a hyperpolarization of the membrane potential (Carandini and, 1997; Sanchez-Vives et al., 2000a, 2000b). This hyperpolarization moves the state of the membrane potential away from the action potential threshold. The result is effectively a reduction of spiking probability in response to subsequent stimulation. Because of its dependence on the previous activity of the neuron, this mechanism is referred to as firing rate adaptation (Grill-Spector et al., 2006) or response fatigue (Vogels, 2016).

Synaptic depression

At the level of a synapse, repeated presynaptic activation can result in a reduced neurotransmitter release and thus reduced input for the postsynaptic neuron. Several known mechanisms can contribute to synaptic depression (Fioravante and Regehr, 2011), making it a possible source of neural adaptation. Nevertheless, a causal role in repetition suppression in the visual system has not yet been demonstrated (Vogels, 2016). Synaptic depression is a form of input fatigue and depends on activity of the presynaptic neuron (Vogels, 2016). Input fatigue can explain stimulus specific adaptation: if the first stimulus activates a different population of input neurons than the second, only input from that first population (that was activated by the first stimulus) will be adapted and the synaptic input for the second stimulus will not be affected.

Suppressed input neurons

Obviously, neurons are part of a network and should not just be considered in isolation. In a network, repetition suppression will be inherited from one neuron to the next, that is, when a presynaptic neuron’s activity is suppressed, the postsynaptic neuron receives less input (Vogels, 2016). Thus, for each neuron, adaptation can be a combination of response fatigue affecting the state of the actual neuron as well as input fatigue through adaptation inherited from input neurons or trough synaptic depression. In this framework, adaptation propagates through the pathway and dynamically changes the state of the network. When considered in the context of a sensory processing system, these relatively simple fatigue mechanisms can explain relatively complex phenomena resulting in both excitatory and suppressive signals (Solomon and Kohn, 2014).

0.3.2 Perceptual expectation and adaptation

Fatigue based mechanisms are very low-level: they basically emerge automatically from the hardware constraints of the network. They offer an explanation that is attractive for its simplicity and for the lack of unproven assumptions. Nevertheless, advocates of the predictive coding theory have proposed an alternative (or perhaps complementary) account of adaptation, emphasizing top-down mechanisms. Central to predictive coding is the view of the brain as a prediction machine that constructs prior expectations of the environment (Friston, 2005). Cortical responses are conceptualized as "prediction errors", meaning that sensory events that violate expectations elicit a stronger response.

However, perceptual expectation (unless triggered by a cue or similar) is often confounded by stimulus repetition: frequent, repeated stimuli become expected, while rare stimuli are unexpected. So how do we dissociate stimulus repetition and expectation? According to predictive coding theory, repetition suppression should occur when the sensory system expects a stimulus repetition (low prediction error) as opposed to when a repetition is unexpected (high prediction error). This hypothesis was supported by an influential study showing that fMRI adaptation in human fusiform face area (FFA) was modulated by the probability of a face repetition (Summerfield et al., 2008).

While several fMRI studies have later replicated such a repetition probability effect (Kovács et al., 2012, 2013; Larsson and Smith, 2012; Grotheer and Kovács, 2014; Ewbank et al., 2016), there are a number of inconsistencies. For example, while some studies could not find an effect of repetition probability for objects (Kaliukhovich and Vogels, 2011; Kovács et al., 2013), others did do so (Mayrhauser et al., 2014; Utzerath et al., 2017). Later studies have suggested that these effects are dependent on familiarity with the stimulus category (Grotheer and Kovács, 2014; Utzerath et al., 2017). In addition it is not clear at which level of processing these effects take place: some studies suggest they emerge as early as V1 (Larsson and Smith, 2012; Utzerath et al., 2017), but in other studies they emerge only after the object-selective lateral occipital cortex (Summerfield et al., 2008; Kovács et al., 2013). One study could not even find an effect of face repetition probability in any visual area (Olkkonen et al., 2017).

In conclusion, mechanisms of neural fatigue provide a relatively simple explanation of neural adaptation or repetition suppression. Considered in the context of a neural network, these mechanisms can lead to more complex effects, such as stimulus specificity. Conversely, evidence for a predictive coding account of neural adaptation is not always consistent, questioning the generality of the role of top-down mechanisms in adaptation. Nonetheless, under some conditions expectation and adaptation seem to interact. Most of the evidence for this interaction is based on human fMRI studies and the neural signature of these effects is unknown.

0.4 OBJECTIVES

Combining the two topics of rat vision and neural adaptation in monkeys has inescapably led to two distinct sets of research objectives (nevertheless united behind the common goal of understanding vision). In this section I will consider these objectives, broken down by chapter.

0.4.1 PART I: VISUAL OBJECT RECOGNITION IN RATS

At the time we started this project, rats and mice had become popular models in visual neuroscience. Our main objective was to further investigate the rat visual system, both on a behavioral and neurophysiological level. The reason why we chose to study rats instead of mice is that there is a lack of systematic behavioral studies that have investigated advanced visual abilities in mice. Even though the visual system of mice is very similar, it has been argued that it might simply be more difficult to test them in complex visual tasks (Zoccolan, 2015). This work was done at the Laboratory for Biological Psychology, KU Leuven.

Chapter 1: A behavioral investigation of rat visual abilities

Previously, behavioral tests of purely visual object recognition have only used simple shapes (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012) or rendered 3D shapes (Zoccolan et al., 2009). This leaves open the question of what these animals would do with visual stimuli that are less artificial and more like real-life visual input. *Can rats categorize naturalistic movies?*

Chapter 2: Natural stimulus representations in rat visual cortex

In our previous experiment, we had established that rats can be trained to discriminate categories of natural movies and generalize to novel exemplars. Meanwhile, research in our lab was showing promising evidence of position tolerance in their visual system (Vermaercke et al., 2014), a hallmark of object processing in primates. Would the rat visual system show other hallmarks of object processing? *Do we see a categorical representation emerge?*

Chapter 3: A bridge between behavior and neurons

While rats can categorize novel natural movies, we had not found evidence for a categorical representation in their visual system. Unfortunately, a direct comparison between our neural and behavioral data is not possible, because we have no neural responses for any of the movies that the animals had to generalize to. Meanwhile, deep neural network models (DNN) had been developed that predict neural responses and categorization performance on the same stimulus set in monkeys with unprecedented accuracy (Kriegeskorte, 2015). Quantifying our natural movies with a DNN allows us to ask new questions that connect the neural and behavioral data. A) *what level of processing of the DNN is required to support the categorization experiment?* B) *what level of processing of the DNN do the neural representations in rat visual cortex correspond to?*

0.4.2 Part II: Adaptation and expectation in rat visual cortex

Chapter 4: Change detection in rat visual cortex

All of our previous studies were aimed at investigating object recognition properties in the rat visual system. Here, we turn to a visual oddball paradigm in a study on adaptation and expectation. In human event-related potential studies, this paradigm is associated with a component called the mismatch negativity (MMN; Näätänen et al., 2007). This refers to a difference in response between frequent and rare events. In monkey IT cortex, this difference can be explained by repetition suppression for frequent stimuli and not by a surprise related enhancement for rare stimuli (Kaliukhovich and Vogels, 2014). In this final rat study, we use this paradigm to investigate adaptation and effects of expectation in the rat visual system. *Do we see a surprise-based response enhancement in the rat visual system?*

0.4.3 Part III: Adaptation and expectation in macaque visual cortex

The final rat study presented in Chapter 4 marks a transition from the topic of object recognition to the topic of neural adaptation and related expectation effects. For Part II, we moved on to macaques to investigate the theory of adaptation as a manifestation of perceptual expectation. These experiments could only work with monkeys, because of task complexity requirements and the importance of a face stimulus set. However, the topic is important, with implications for adaptation paradigms used in fMRI research (Grill-Spector and Malach, 2001) as well as for general theories of cortical responses (Friston, 2005). This work was done at the Laboratory for Neuro- and Psychophysiology, KU Leuven.

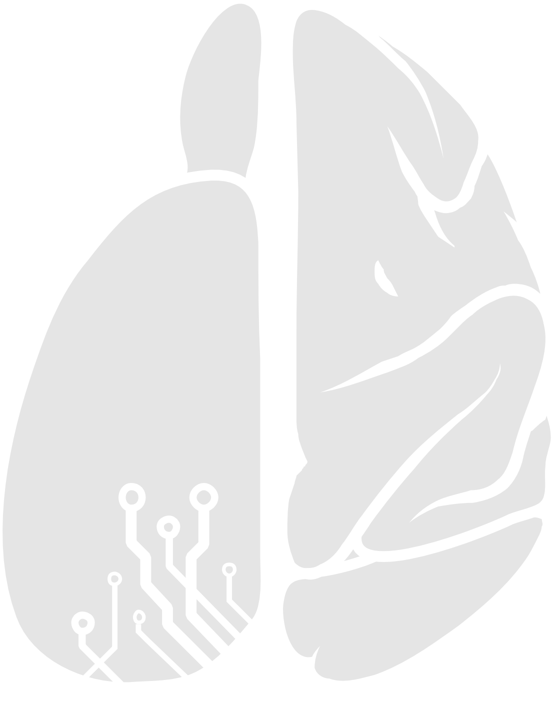
Chapter 5: The perceptual expectation account of neural adaptation

In contrast with fMRI studies, expectation effects on repetition suppression could not be replicated in neural responses in macaque IT (Kaliukhovich and Vogels, 2011). Subsequent fMRI studies pointed to the importance of attention (Larsson and Smith, 2012) or face specificity of the effect (Kovács et al., 2013). *Are these two conditions sufficient for observing expectation effects on repetition suppression in macaque IT?*

Chapter 6: Adaptation confounded as expectation

In Chapter 5, we did not find an effect of expectation on repetition suppression of face-responsive IT neurons. In contrast, Bell and colleagues reportedly found evidence for an expectation-based mechanism distinct from stimulus-driven adaptation (Bell et al., 2016). The authors used a design where stimulus repetition is confounded with expectation, but tried to control for repetition suppression with a linear regression approach. Using simulated neural responses, we investigate whether their method actually controls for the confound. *Could the analysis in Bell et al. lead to spurious effects of expectation?*

I VISUAL OBJECT RECOGNITION IN RATS



Chapter 1.

A BEHAVIORAL INVESTIGATION OF RAT VISUAL ABILITIES

Previously, behavioral tests of purely visual object recognition have only used simple shapes (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012) or rendered 3D shapes (Zoccolan et al., 2009). This leaves open the question of what these animals would do with visual stimuli that are less artificial and more like real-life visual input. *Can rats categorize naturalistic movies?*

Published as

Vinken K., Vermaercke B., Op de Beeck H. (2014). Visual categorization of natural movies by rats. *Journal of Neuroscience, 34* (32), 10645-10658.

VISUAL CATEGORIZATION OF NATURAL MOVIES BY RATS

Visual categorization of complex, natural stimuli has been studied for some time in human and non-human primates. Recent interest in the rodent as a model for visual perception, including higher-level functional specialization, leads to the question of how rodents would perform on a categorization task using natural stimuli. To answer this question, rats were trained in a two-alternative forced choice task to discriminate movies containing rats from movies containing other objects and from scrambled movies (ordinate-level categorization). Subsequently, transfer to novel, previously unseen stimuli was tested, followed by a series of control probes. The results show that the animals are capable of acquiring a decision rule by abstracting common features from natural movies in order to generalize categorization to new stimuli. Control probes demonstrate that they did not use single low-level features, such as motion energy or (local) luminance. Significant generalization was even present with stationary snapshots from untrained movies. The variability within and between training and test stimuli, the complexity of natural movies, and the control experiments and analyses all suggest that a more high-level rule based on more complex stimulus features than local luminance-based cues was used to classify the novel stimuli. In conclusion, natural stimuli can be used to probe ordinate-level categorization in rats.

1.1 INTRODUCTION

There is an increasing scientific interest in the visual perception of rodents. Several recent studies have focused upon the cortical organization in rodents, elucidating the extent of functional specialization in rodent extrastriate visual areas (Andermann et al., 2011; Marshel et al., 2011). However, the usefulness of this model depends on the visual capabilities of rodents. A number of studies have found behavioral evidence in rats for higher level visual processing (Zoccolan et al., 2009; Tafazoli et al., 2012; Vermaercke and Op de Beeck, 2012; Alemi-Neissi et al., 2013; Brooks et al., 2013). While these studies provide evidence for abilities reminiscent of higher level vision, none of them used very complex naturalistic stimuli. This leaves open the question of how rats would perform in more sophisticated visual tasks in which complex, dynamic stimuli are used for categorization and generalization to new stimuli.

The use of natural stimuli in visual neuroscience has been both defended and criticized. It has been argued that simple artificial stimuli are necessary for uncovering the specific response properties of neurons in each stage of visual processing (Rust and Movshon, 2005). Others have pointed towards evidence suggesting that visual processing cannot be entirely elucidated solely based on experiments with simple stimuli (Kayser et al., 2004; Felsen and Dan, 2005; Einhäuser and König, 2010). For example, humans are more efficient in classifying natural scenes compared to simplistic unnatural stimuli (Li et al., 2002). To find out the extent of the validity of rats as a model in vision research, it is very relevant to investigate to what extent experiments with complex stimuli can work.

In monkeys and humans, natural stimuli have been used effectively in highly demanding tasks for superordinate- and ordinate-level categorization (Thorpe et al., 1996; Fabre-Thorpe et al., 1998; Vogels, 1999a; Serre et al., 2007; Greene and Oliva, 2009; Peelen et al., 2009; Walther et al., 2009; Fize et al., 2011). Provided that the stimulus set contains sufficient variation, categorization of natural stimuli requires generalization relying on processing and extraction of category-specific features, invariant to the presence of other information. Therefore, successful performance of an animal on novel category exemplars provides information about the extent of the capabilities of their visual system.

In the present study, rats were trained to discriminate movies containing rats from movies containing other objects and from scrambled movies in a two-alternative forced choice (2AFC) task in a visual water maze (Prusky et al., 2000). After training, the animals were tested for generalization to unseen movies. Several tests were performed, starting with stimuli which could be considered as 'typical' exemplars, and gradually including more deviant movies, still images, and some controls to exclude the possibility that low-level cues would drive performance. The rats generalized well to new 'typical' movies, and generalization was still significant for slower movies, stationary snapshots, movies with differently colored rats, and movies controlling for local luminance cues.

Overall, the findings indicate that the rats were using relatively complex stimulus features to perform the categorization task.

1.2 MATERIALS AND METHODS

1.2.1 ANIMALS

Experiments were conducted with six male FBNF1 rats, aged 25 months at the start of the study. This specific breed was chosen for their relatively high visual acuity of 1.5 cycles per degree (Prusky et al., 2002). One subject was excluded from the data as a result of extreme response bias during training, preventing the rat from reaching above chance performance in over 1200 trials with the first stimulus pair. All rats had previously been used in discrimination experiments, but with unrelated stimuli: sinusoidal gratings of varying orientation and spatial frequency. Animals had ad libitum access to water and food pellets. Housing conditions and experimental procedures were approved by the KU Leuven Animal Ethics Committee.

1.2.2 STIMULI

Pairing of target and distractor stimuli

Natural movies were selected from our own database of 537 five-second movies created for the purpose of this experiment. They were recorded at 30 Hz (thus including 150 frames) and sized 384×384 pixels. For every stimulus, three vectors were calculated from the pixel intensities across columns $x = 1\ldots X$ and rows $y = 1\ldots Y$, but per frame $t = 1\ldots T$. Note that the monitors presenting the stimuli were gamma corrected to obtain a linear transfer function between pixel intensity values and luminance, thus using actual pixel values will not distort the metrics. The first vector contained the average pixel intensities as a function of time t:

$$\bar{I}(t) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} I_{xy}(t)\,;$$

the second the root mean squared contrasts:

$$RMS(t) = \sqrt{\frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_{xy}(t) - \bar{I}(t))^2} \;;$$

the third the average changes in pixel intensity (this time per frame transition for t = $1 \ldots T - 1$):

$$\overline{PC}(t) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}| I_{xy}(t+1) - I_{xy}(t) |.$$

Next, we reduced this information by taking the means and standard deviations across frames/frame transitions t to six features per stimulus: M($\bar{I}$), SD($\bar{I}$), M(RMS), SD(RMS), M($\overline{PC}$), and SD($\overline{PC}$). Doing this for every stimulus resulted in six feature-vectors summarizing our database of 537 stimuli in a relatively low-dimensional space. After taking Z-scores of each of these six feature-vectors (across all 537 movies), each rat movie was paired with a non-rat movie so that the Euclidean distance between them in this standardized space was less than one standard deviation. Without the constraint of one standard deviation, the average distance between all possible pairs of movies was 3.24 standard deviations with a 95% percentile interval of [1.19 6.60].

Training stimulus set

From these matched stimuli, a general set of five pairs was selected with variability of target and distractor in mind, along with three test sets (see **Figure 1.1**). Note that the previously described method of pairing stimuli with their most similar distractor could result in two rat movies being matched with the same distractor. This was the case for two rat movies in the test sets: one distractor is shared between a target movie of test set 1 and 2, and one between a target movie of test set 1 and 3. There was, however, complete separation of training and test sets. The target movies showed moving rats of the same strain, while three of the paired distractors contained a train, one a gloved hand moving in and out of the screen and one a moving stuffed sock. Both stimulus types had varying amounts of camera movement.

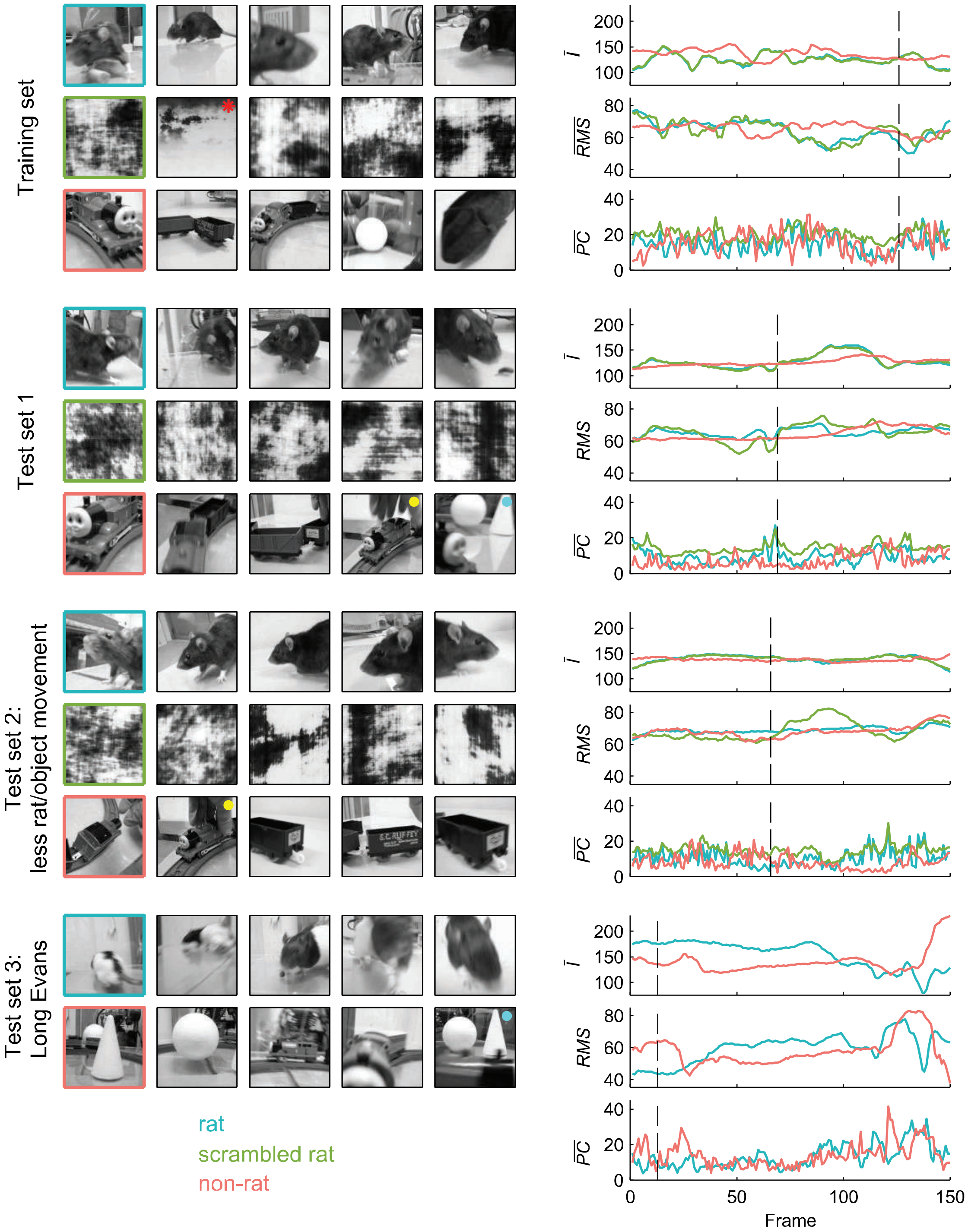


Figure 1.1. Stimulus sets.

The left side displays three rows of snapshots for each stimulus set, with the first row depicting the five rat movies, the second row the phase scrambled versions of these rat movies, and the third row the natural distractor chosen for each rat movie (for the last stimulus set the row with scrambled stimuli is omitted, because they were not used in the experiments). The snapshots of each target movie and its distractors are taken at the time point at which the frame of the target stimulus (i.e. rat movie) is most similar to the rest of the frames in that movie (i.e., minimal pixel-wise Euclidean distance). The red asterisk indicates the adjusted distractor (see materials and methods, stimuli). Yellow and blue dots indicate the two distractors that were each matched to two rat movies. The right side displays average pixel intensity ($\bar{I}$), root mean square contrast (RMS), and mean absolute pixel change ($\overline{PC}$).as a function of frame number for one target and its distractors of each set (the outline of the chosen movie is colored in the left panel). Dashed lines indicate the location in time of the frame displayed on the left.

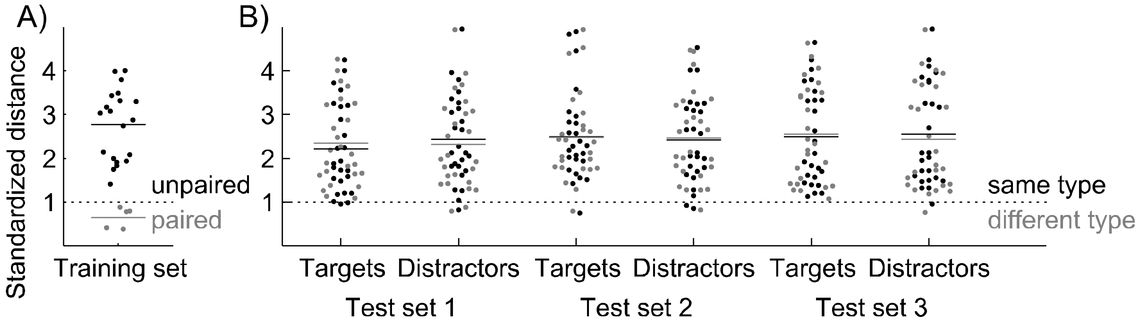


Figure 1.2. Standardized distances between individual stimuli.
Panel **(A)** shows a bee swarm plot of the distribution of all pairwise distances between target and distractor movies of the training set for paired (grey) and unpaired (black) target and distractor movies separately. Panel **(B)** shows bee swarm plots for all pairwise distances between either the targets (rat movies) or the distractors (non-rat movies) of a certain test set (e.g. the targets of test set 1) and either all training set movies of the same stimulus type (black), or all training set movies of the different stimulus type (grey). For example, for the targets of test set one, all pairwise distances to the rat movies of the training set are shown in black, while all pairwise distances to the non-rat movies of the training set are shown in gray.

The degree of variation is illustrated by the fact that the six-dimensional Euclidean distance between the target and distractor movies of the different pairs was relatively large (M = 2.76, calculated from all possible target-distractor combinations excluding the actual pairs), much larger than the distance between target and distractor movies from the same pair (M = 0.65; see **Figure 1.2**A for a plot of the distribution of these distances). Note that the magnitudes of these average distances are still in standardized space, expressed in units of SD across movies. In a later section we will show that a strategy based upon local luminance cues cannot explain generalization from this training set to test sets containing new stimuli.

Test stimulus sets

The first test set, used for generalization purposes, included movie pairs which were very different from the training pairs in terms of low-level properties (**Figure 1.2**B), but were judged to be relatively typical in terms of their high-level content by the experimenters (same strain of rats, similar motion properties, etc.). The second test set included movies in which the rats/objects were more stationary. In quantitative terms, the non-standardized M($\overline{PC}$) was on average 4.22 for the movies of this second test set, while it was 5.18 and 5.39 for the movies of the training set and first test set (note that the difference is rather small because there was still camera movement). For the third test set, the target movies included rats of a different strain (Long Evans) which are white/black spotted rather than uniformly dark. In each test set all of the natural distractors contained a (moving) train and some of them had objects (a ball, cone) present in them and/or a hand moving one of those objects. **Figure 1.2**B shows the distribution of all pairwise distances between either the targets (rat movies) or the distractors (non-rat movies) of a certain test set and either all training set movies of the same stimulus type, or all training set movies of the different stimulus type. It is clear that generalization cannot be explained by the six dimensions we used to match targets and distractors, because the distributions of target-target distances and distractor-distractor distances are not systematically lower than the distribution of target-distractor distances. The rat test set movies were not more similar to training set movies of the same type (targets) than to those of the other type (distractors), nor were the non-rat test set movies. **Figure 1.3** illustrates how all aforementioned stimulus sets compare to each other in the standardized six-dimensional stimulus space. **Figure 1.3**A and B clearly show that within-pair distances are much smaller than the between-pair variability. **Figure 1.3**C shows that on average test set 1 matches the training set best on all six dimensions. This plot also illustrates that on average test set 2 not only has the aforementioned smaller change of pixel values M($\overline{PC}$), but also less variability in average pixel intensity SD($\bar{I}$), contrast SD(RMS), and change of pixel values SD($\overline{PC}$), all of which is consistent with the more stationary rats/objects.

Scrambled distractors

For the training set and first two test sets, additional distractors were created by phase-scrambling the rat movies. On trials using scrambled distractors, a rat movie was only paired with its own scrambled version. The scrambled stimuli were created in three steps, as illustrated in **Figure 1.4**.

First the spatial amplitude spectrum $A(I_t)$ for each frame I_t in a rat-movie M was estimated by means of a two-dimensional fast Fourier transform (2D FFT). To each spatial frequency component, a random phase angle (drawn from a uniform distribution over the interval [-π, π]) was assigned, resulting in a new phase spectrum φ^*. A new movie M' was obtained by performing an inverse 2-D FFT on the combination of each

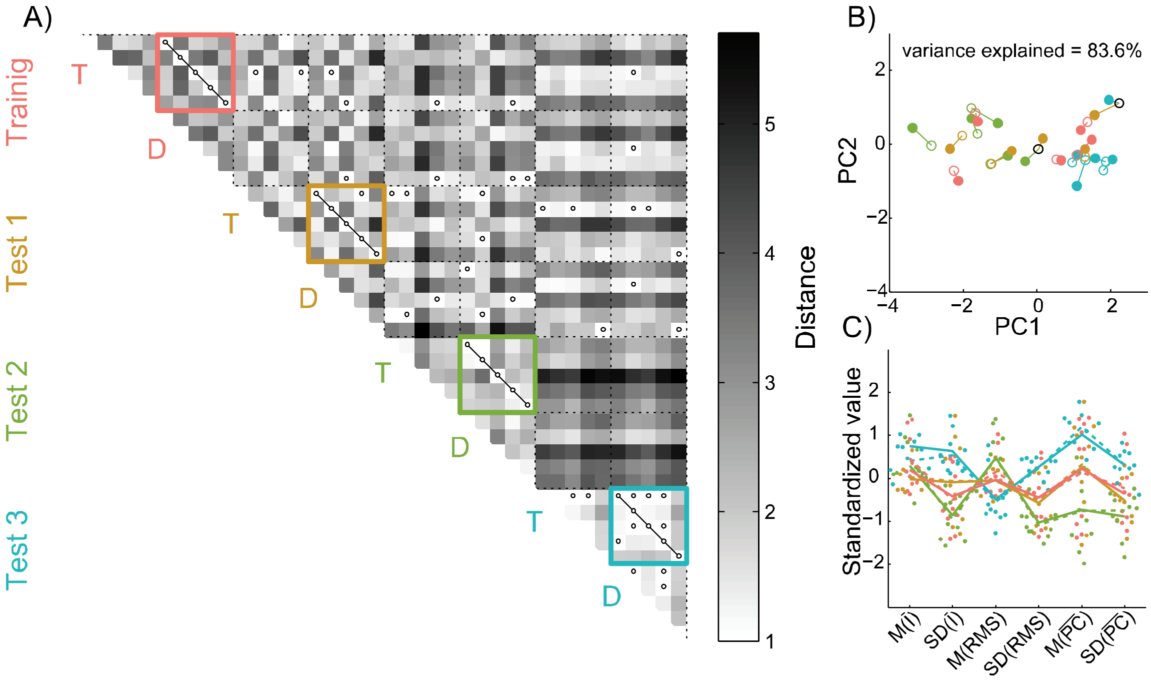


Figure 1.3. Stimulus dissimilarities.
Panel **(A)** depicts distance matrix for all natural movies used in the experiment (in SD units, see materials and methods, stimuli). Colored boxes highlight target (T) and natural distractor (D) combinations per stimulus set. Black circles indicate distances shorter than one (i.e. the criterion for target-distractor match). Panel **(B)** shows all stimuli in two-dimensional space after principal component analysis on the distance matrix from panel (A). Full markers indicate targets, while empty circles indicate distractors (color codes correspond to those in panel (A)). Black markers indicate distractors that are shared by two target movies. Variance explained by these first two principal components is 83.6%. Panel **(C)** is a parallel coordinates plot showing bee swarm plots of individual stimuli per dimension and the centroids (averages) of each stimulus set for targets (continuous lines) and distractors (dashed lines) separately in the standardized six-dimensional stimulus space (color codes correspond to those in panel (A)).

frame's amplitude spectrum with the new phase spectrum $A(I_t)e^{-i\varphi^*}$, for t = 1...T. This first step is equivalent to the scrambling method used by Schultz and Pilz (2009).

Note that this method uses the same phase spectrum φ^* for all frames of a particular movie (i.e. for every rat movie one set of random angles was generated and used), which results in excessive temporal correlation of pixel values in consecutive frames. Therefore, using a 3D FFT, the spatio*temporal* phase spectrum of this scrambled result $\varphi(M')$ and the spatiotemporal amplitude spectrum of the original movie $A(M)$ were taken and combined into a scrambled movie M'' by performing an inverse 3-D FFT on $A(M)e^{-i\varphi(M')}$. The result of this second step is a movie with consecutive frame correlations comparable to those of the original natural movie (see **Figure 1.4**).

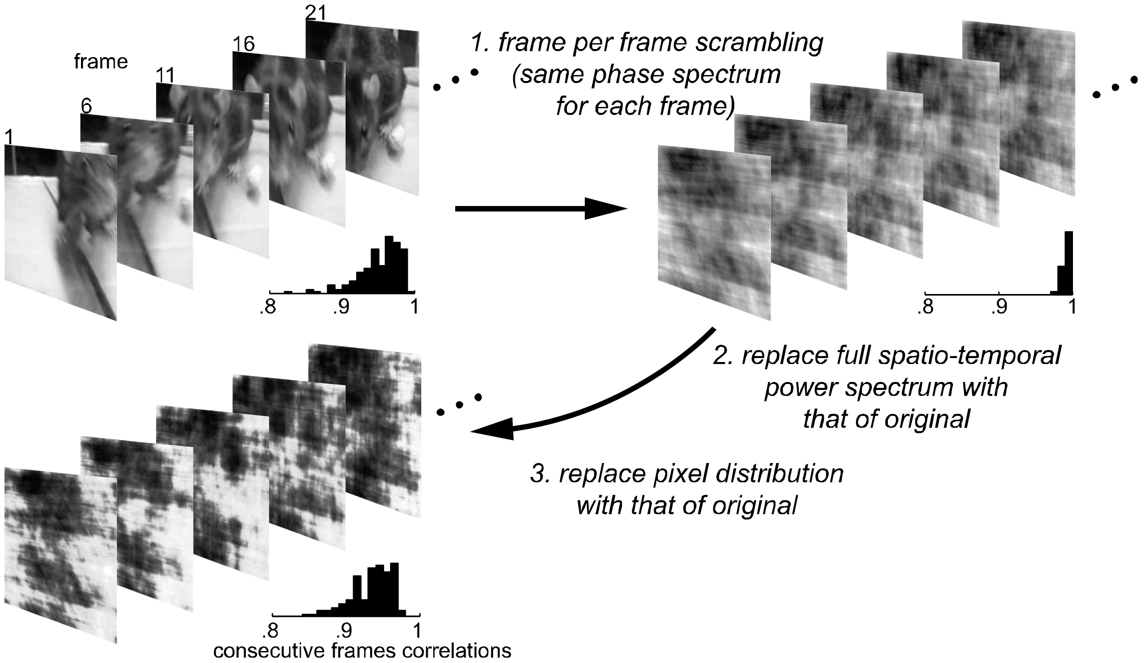

frame
1
6
11
16
21
1. frame per frame scrambling (same phase spectrum for each frame)
2. replace full spatio-temporal power spectrum with that of original
3. replace pixel distribution with that of original
.8
.9
1
consecutive frames correlations

Figure 1.4. Creation of scrambled stimuli.
One example rat movie is represented by five example frames taken in steps of five (the scrambled versions depicted here correspond to these exact frames). For each image sequence a histogram is inserted showing the distribution of Pearson correlations of pixel values belonging to consecutive frames.

Finally, to compensate for the expanded range of pixel values (i.e. values outside the range of 0 to 255), in the third step each scrambled movies′ M'' pixel distribution was replaced by that of the original movie M. The result was a temporally correlated image sequence with identical contrast, luminance and virtually identical spatiotemporal power spectrum, while not containing any recognizable content. The frame per frame contrast, luminance and spatial power spectra were highly similar.

The reason for using the trick in the first step instead of just generating random phase angles for each spatiotemporal frequency component is that the latter would result in a scrambled movie for which each frame is not matched as well with the corresponding frame of the original (a problem which is for example present in Fraedrich et al., 2010). For instance, for the example movie in Figure 1.4 the mean absolute deviation for frame per frame comparison between the original and 1000 scrambled samples generated using our method is on average 2.2 (95% percentile interval [1.5 3.1]) for pixel intensities $\overline{I}$, 3.5 (95% percentile interval [3.0 4.0]) for contrast values RMS, 4.9 (95% percentile interval [4.5 5.4]) for changes in pixel intensity $\overline{PC}$. The correlation between the spatial amplitude spectra for frequency components lower than one cycle per degree (excluding the zero-frequency component) of corresponding frames is on average .93 (95% percentile interval [.92 .93]). On the other hand, when these statistics are calculated for samples where scrambling is done by completely randomizing spatiotemporal phase, the values are 16.4 (95% percentile interval [10.7 21.1]) for pixel intensities $\bar{I}$, 5.1 (95% percentile interval [3.5 7.0]) for contrast values RMS, 5.9 (95% percentile interval [5.4 6.5]) for changes in pixel intensity $\overline{PC}$, and .87 (95% percentile interval [.84 .89]) for the spatial amplitude spectra.

The scrambled version of one rat movie in the training set was adjusted after nine sessions into the training, because we suspected that the rat which had started with this particular pair used the luminance difference in the lower part of the screens to achieve above chance performance suspiciously rapidly. To prevent this from happening the lower part of the frames of this one scrambled movie was made brighter by increasing the pixel values according to a linear gradient so that pixel values closer to the bottom of the movie frames were increased more. Specifically, the values of the gradient went from x at the bottom pixel row to zero at the top, where x was chosen so that the average (weighted by a linear gradient ranging from 100 to 0% from bottom to top pixel row) over all frames was equal to that of the original rat movie. Performance of this one animal dropped to chance immediately after this change (data not shown), hence for the continuation of the training this adjusted distractor was used.

1.2.3 EXPERIMENTAL SETUP AND TASK

Rats were trained to discriminate movies containing a rat versus movies without rat in a 2AFC task in a visual water maze (Prusky et al., 2000). Briefly, the setup consisted of a water filled V-shaped maze with two arms (**Figure 1.5**). At the end of each arm a stimulus was shown and a transparent platform was placed just below the water surface in front of the target stimulus. A trial started when a subject was placed in the water at the end opposite to the stimuli and ended when the rat reached the platform. For a quick escape, the animal had to choose the correct arm where the target movie was played. If the rat had entered the wrong arm, it had to swim back to the other arm and sit through a 20 second time interval before being rescued. In the case of an instant correct choice, this interval was 10 seconds to still ensure some stimulus exposure. Overall, the distribution of time to reach the platform had a mode of 4.7 seconds (with a 95% percentile interval of [3.8 22.5]) for correct trials and a mode of 8.6 seconds (with a 95% percentile interval of [5 27.7]) for incorrect trials. Note that the lower bound of 3.8 seconds is limited by swimming speed rather than animals waiting before responding, meaning that a mode of 4.7 seconds would be about one second extra.

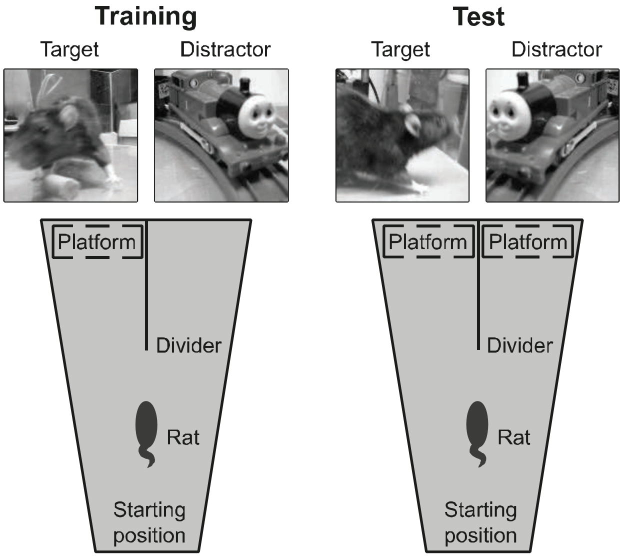


Figure 1.5. Schematic representation of the 2AFC setup as seen from the top.
In each trial, the rat had to find the hidden platform by swimming towards the side showing the target stimulus, while ignoring the distractor. During generalization sessions there was a platform present in front of each screen.

Between trials, the animals resided under a heat lamp. The water was kept at a temperature of 26-27° C. Stimuli were presented on two 768x1024 CRT screens at a width of 24 visual degrees as seen from the divider. Output to the monitors was linearized and the mean luminance was 53 cd/m^2 and 52 cd/m^2 for the left and right screen, respectively. The stimuli where played in an infinite loop alternating between forwards and backwards play, to always ensure a smooth transition. The target was always the stimulus containing a rat. The distractor was either the matched non-rat movie, or the scrambled version of the rat movie. Each animal was trained for two sessions per day. Each session included 12 trials carried out in an interleaved fashion (all rats were tested on trial n before any rat was tested on trial n+1). The side of the target stimulus was determined according to a standard sequence of LRLLRLRR (Prusky et al., 2000). Specifically, per rat and per session a random starting point in this sequence was chosen, with the only restriction that no rat could start with the target movie on the same side for two trials. Whenever the end of the sequence was reached, the procedure would jump to the beginning to fill in the remainder of target locations in order to get to 12 trials. We opted for adopting these stringent constraints used by Prusky et al. (2000) in an attempt to prevent development of response biases. However, this means that in theory a rat could predict the correct response if the target had been presented for two consecutive trials on the same side (i.e. after LL or RR). In addition, overall the probability that the next target will be on the opposite side is much higher than that it will be on the same side, meaning a strategy where a rat would switch sides would be relatively successful overall. Finally, because the sequence could not start with a repetition (i.e. LL or RR), the platform location on the second trial of each session could always be predicted from that on the first trial. However, the fact that the number of trials per session were limited and that the interleaved testing of animals resulted in a long inter trial interval of at least a few minutes argue against the hypothesis that a rat can pick up the regularities of this sequence.

Indeed, **Figure 1.6**A shows that the animals did not use any of these potential shortcuts: percentage correct does not fall to chance when trials are not predictable (e.g. the first), nor does it peak on trials that are predictable (the second trial or a trial following a repetition of the same target side). In addition, performance is well above chance even if the target side in a trial was a repetition of that in the previous one (i.e. a switch strategy would not be successful). Moreover, previous unsuccessful experiments at our lab using the same

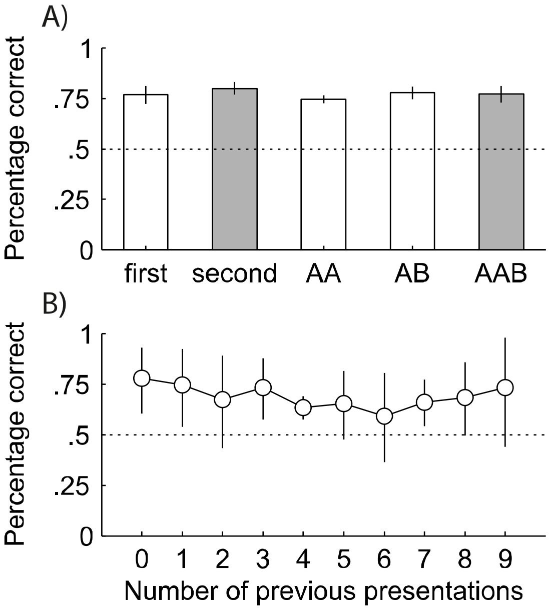


Figure 1.6. Control analyses to check for potential shortcuts related to the sequence for assigning stimuli to the left or right screen (A) and to check for learning of new pairs during two-platform trials (B).
Panel **(A)** shows percentage correct from all sessions where only one platform was used: for the first trial (*first*), the second trial (*second*), trials where the target was on the same side as on the previous trial (*AA*), trials where the target was on the other side as on the previous trial (*AB*), and trials following two consecutive trials in which the target was on the same side (*AAB*). The correct response for the cases indicated in grey were perfectly predictable in theory for a subject with full insight in the stimulus sequence, as opposed to all the other data shown here where the correct response was unpredictable. Panel **(B)** shows percentage correct averaged across animals from all trials with new stimulus pairs (and therefore using two platforms) as a function of how many times the animal had seen that pair before. Error bars in both panels (A) and (B) are the 95% confidence intervals obtained from a two-sided t-test on the arcsine of the square rooted proportions correct per animal (N = 5 per confidence interval).

protocol in tasks that turned out to be too challenging indicate that rats will not pick up any potential shortcut even after substantial training. Whenever a rat would reach a response bias of over 80% in one session, an anti-bias procedure was used in the subsequent session: on the first two trials the target was presented to the side opposite of the bias. If the bias persisted in the following session, the target was presented for 75% of the trials on the side opposite to the bias. For all but one animal, this procedure was sufficient to break any persistent response preference. This rat never learned any stimulus pair and thus the rat could not be included in the experiment. Note that bias correction trials were never used during any of the generalization sessions with two platforms. Only the data to test performance on all target-distractor combinations of the training stimuli include bias correction trials (see testing phase).

Training phase

At the start of training, the subjects were familiarized with the 2AFC task using a white screen as the target versus a black screen as the distractor. This shaping procedure was terminated when all rats had reached a performance of at least 80% correct on three consecutive sessions. The actual experiment consisted of two phases: a training phase, and a testing phase. During the former, rats were trained to discriminate the five rat movies of the training set from their distractors (see **Figure 1.1**). A rat would start the phase with one target movie and one distractor movie. Whenever performance would reach a criterion of at least 75% correct on four consecutive sessions, the same target movie would be presented with the other type of distractor (the two types being object movies and scrambled movies). Whenever a rat would fail to reach this criterion within a large number of trials (e.g., over 300, which would take about a month), the decision was made to move to the next pair to advance the training process. This happened a few times, because we had chosen some challenging combinations for the training set on purpose in order to push the animals: both the second training pair containing a movie of a rat relatively far away and high up the screen and the last training pair containing rat-like sock puppet as distractor proved to be difficult. Test set 1 did not include such challenging combinations. When the criterion was reached again (i.e., after the rat was trained with the two types of distractors for a certain target), a rehearsal intermezzo of the previous combinations started until performance for every pair (assessed on the 6 last trials per pair) was at least 75% again. Subsequently the rat moved to a new target movie with the distractor of the same type as that of the latest combination. Except for the first sessions, trials containing a new target or distractor were always mixed with trials containing the most recently learned combination. On every switch to a new movie, the new-old stimulus pair ratio was 1/2 and changed to 2/3 after a full session for which performance on the old pair was 75%. The order in which rats were trained on each target and their distractors was different for every animal. At the end of the training phase, final performance on the training stimuli was assessed by presenting all possible target and distractor combinations (thus no longer only including the original pairings of each target with its two distractors).

Testing phase

During the subsequent testing phase, generalization to the stimulus pairs of the three test sets was assessed. On these generalization trials, the protocol was changed to limit new learning: both arms contained a platform and the animals were rescued immediately upon reaching it. Thus, any response was rewarded and most importantly there was no negative reinforcement. **Figure 1.6**B shows that on average the percentage correct on a new stimulus pair did not increase as a function of the amount of times the animal has seen (any of the movies in) that pair. Rather, the figure suggests that if there was any learning at all during the test phase, it had a negative effect on the performance of the animals. Generalization trials were randomly mixed with trials using training movies and only one platform in order to keep the rats motivated to perform the task well. If a rat acquired a strong response bias during testing with two platforms (i.e., over 80% responses in one direction), the data for that particular session were removed from analysis and an anti-bias procedure (see first paragraph of experimental setup and task) was initiated using mixed target-distractor combinations with training stimuli and one platform only. The data obtained during these correction sessions were pooled with the data obtained at the end of the training phase using all possible target and distractor combinations to ensure a sufficiently large number of trials per target × distractor combination. After probing for generalization, specific hypotheses were examined by manipulating the stimuli of the first test set and assessing the effects on performance. It should be noted that not every rat underwent every testing condition, because of

Table 1.1. Number of trials used for data analysis per rat and per phase or test condition.

Phase	Type	Rat				
		1	2	3	4	5
Training		1668	1704	1692	2124	1896
Test training pairs		60	60	60	60	60
Test new combinations		204	180	119	180	240
Test generalization						
1: Typical	Natural distractor	48	48	48	48	48
	Scrambled distractor	48	48	48	.	48
	Reduced speed	48	48	48	48	48
	Single frame	48	48	42	24	.
	Changed luminance	48	48	48	.	.
	Single frame, changed luminance	12	48	48	54	.
2: Less rat/ object movement	Natural distractor	48	48	48	48	48
	Scrambled distractor	48	48	48	.	48
3: Long Evans	Natural distractor	48	.	48	.	.

Note. For the generalization data the numbers of trials are only taken from sessions without response bias (i.e., no more than 80% responses in one direction).

temporal constraints related to the fact that each animal finished training at a different time (see **Table 1.1**). In total, the experiment encompassed 251 behavioral sessions per rat (spread over about 6.5 months), with each containing 12 trials per rat and taking about one hour.

1.2.4 Data analysis

In some occasions we report the results from a classical one-tailed t-test based upon the across-rat variability (N=5) and using a significance threshold of α = .05. These tests are performed on the arcsine of the square root of the proportion correct trials (i.e., $y_j^{trans} = \sin^{-1}\sqrt{y_j/n_j}$, with y_j correct responses of rat j on n_j trials) in order to stabilize variance and approximate normality for the transformed proportions (Hogg and Craig, 1995). However, the t-tests do not take into account the number of trials on which the performance in each animal is based, in fact, the transformed numbers should not be treated as metric because then we ignore information about the number of trials n_j. The latter issue can be addressed by using a simple binomial test and pooling all trials over animals, but then the unmodeled dependencies can lead to meaningless results. On the other hand, logistic regression supersedes transformations for analyzing proportional data (Warton and Hui, 2011). In addition, a hierarchical model is the preferred method to approach dependencies between observations (Lazic, 2010; Aarts et al., 2014), since information on the uncertainty of the estimates on the within-subject level is not discarded, but used in the analysis on the population level.

Given the widespread familiarity with t-tests in the neuroscience community, we include t-tests for the results on each of the stimulus sets with data of more than three animals and binomial 95% confidence intervals per rat otherwise. However, given the disadvantages of such tests discussed in the preceding paragraph, in addition we turned to a more comprehensive hierarchical model which allowed us to take into account both the number of subjects and the number of trials per subject. Specifically, a within-subject logistic-binomial model was fit to the data to make inference on animal performances and on comparisons between the different stimulus sets:

$$\begin{gathered} y_{ij} \sim B\left(n_{ij}, p_{ij}\right) \\ p_{ij} = \operatorname{logit}^{-1}\left(\mu + \alpha_i SET_i + \beta_j RAT_j + \gamma_{ij}(RAT \times SET)_{ij} + \varepsilon_{ij}\right) \\ \beta_j \sim N\left(0, \sigma_\beta^2\right) \\ \gamma_{ij} \sim N\left(0, \sigma_\gamma^2\right) \\ \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \end{gathered} \tag{1}$$

In this model the observed number of correct trials $y_{ij}$of rat j on stimulus set i is assumed to have a binomial distribution with n_{ij} denoting the number of trials and p_{ij} denoting the probability of a correct trial. This probability is estimated by the logistic function $f(x) = 1 \div (1 + e^{-x})$ (which compresses values between zero and one) of a linear combination of predictors: one for stimulus set (SET, a nominal predictor with 13 levels: five stimulus sets with natural distractors, four with scrambled distractor, and four manipulations of test set 1), one for subject (RAT, a nominal predictor with five levels: one for each rat), and one for the interactions between subject and stimulus set ($RAT \times SET$, a nominal predictor covering all interactions between subjects and stimulus sets). The parameters α_i, β_j, and γ_{ij}(for all i = 1…13 and j = 1…5) are the estimated deflections from the central tendency μ for each stimulus set, rat, and combination of rat and stimulus set, respectively. These parameters are estimated on the log odds scale

($\text{logit}(p) = \log\big(p \div (1-p)\big)$ for percentage p), meaning that the increase or decrease in percentage correct corresponding to their value is not a constant but depends on the percentage correct from which the deflection is calculated (for a more detailed discussion of the interpretation of logistic regression coefficients, see Gelman and Hill, 2007, p 81-83). The residuals are assumed to be normally distributed with variance σ_ε^2. The regression weights for the subject predictor and subject interactions are also assumed to be normally distributed with variances σ_β^2 and σ_γ^2, respectively. All three variances, as well as the central tendency and all deflections, are estimated by the data. This model is formally equivalent to the example model of Gelman and Hill (2007 p.116–117). While only the effect of the nominal predictor SET is used for inference, all other parameters are necessary to model the dependencies present in the data (Lazic, 2010).

A slightly modified model was used for inference on different target and distractor combinations: stimulus pair was used as predictor instead of stimulus set, with a variance parameter for its regression weights ($\alpha_i \sim N(0, \sigma_\alpha^2)$ for i = 1…50). This parameter provides shrinkage towards μ on the performance estimates for stimulus pairs (i.e. regularizing the regression), which makes sense because they are estimated from a rather limited number of trials (between 14 and 33, *Mdn* = 30) while there is a large number of parameters (stimulus pair is a nominal predictor with 50 levels).

Estimation was done within the Bayesian framework by approximating the posterior distribution by means of Markov chain Monte Carlo sampling using JAGS (Plummer, 2003; an improved clone of BUGS, one of the most popular statistical modelling packages, Lunn et al., 2009) in R (R Core Team, 2015). JAGS uses Gibbs sampling, which is an algorithm that can draw samples from a joint probability distribution, given that all the conditional distributions (i.e. one for each parameter) can be expressed mathematically. The joint posterior distribution was approximated by generating 10000-20000 samples (using three chains to check for convergence). The joint posterior distribution quantifies the probability of each parameter value given the data, by combining a prior with the likelihood. Non-informative prior distributions were used as to let the data fully speak for themselves and not constrain the estimates in this respect. Specifically, priors for the regression weights were all normally distributed and centered around zero. Large standard deviations of magnitude 100 were chosen for parameters without hyperprior. Uniform priors ranging from 0 to 100 were chosen for the standard deviations that were estimated in the model (as in Gelman, 2006). Similar to confidence intervals, a 95% highest density interval (HDI), containing the 95% most probable parameter values, was used for inference, while the mode indicates the single most probable value, which will be called the point estimate from here on. A 95% HDI covers 95% of the posterior probability density (i.e. there is 95% certainty that the underlying population parameter that generated the data falls within the bounds of the interval) and, in addition, there is no value outside the interval that is more probable than the least probable value within the interval (the concept of HDI is explained in further detail in Kruschke, 2011 p.296–303). Values falling outside of the 95 % HDI are rejected based on low probability. Essentially this is a within subject ANOVA model. However, the logistic-binomial extension makes it appropriate for dichotomous predicted variables. In addition, by including the information about the magnitude of n_{ij}, one allows for appropriate inference taking into account the (often unbalanced) number of trials over subjects and conditions. In sum, the Bayesian framework permits us to choose the appropriate model for the experiment design and data type.

A different yet very similar model was used to check whether generalization to new stimuli could be explained by a strategy where rats use simple cues based on local luminance to achieve above chance performance:

$$\begin{gathered} y_{ij} \sim B\left(n_{ij}, p_{ij}\right) \\ p_{ij} = \operatorname{logit}^{-1}\left(\mu + \alpha_k X_{ik} + \beta_j RAT_j + \gamma_{jk} X_{jk} + \varepsilon_{ij}\right) \\ \alpha_k \sim t(0, \sigma_\alpha^2, \nu) \\ \beta_j \sim N\left(0, \sigma_\beta^2\right) \\ \gamma_{jk} \sim N\left(0, \sigma_\gamma^2\right) \\ \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \end{gathered} \tag{2}$$

This model is the same as the one described above, except that it uses metric predictors X_k. In a first test, X_{ik} for k = 1…36 denote the following predictors for each stimulus pair i: M($\bar{I}$) (local mean luminance) and M($\overline{PC}$) (variation in luminance) values of the target and distractor in nine locations of each screen (2 metrics × 9 locations × 2 stimuli equals 36 predictors). Indices i, j, and k denote stimulus pair, rat, and metric predictor, respectively. In a second test, for each stimulus pair *i*, X_{ik} for *k*=1…18 denote the M($\bar{I}$) and M($\overline{PC}$) for the distractor subtracted from the same metrics for the target for the corresponding nine locations of the screens. The t distribution on the regression coefficients α_k for our metric predictors provides regularization and avoids over-fitting to the data by only allowing strong predictors to have a substantial regression weight (Kruschke, 2011 p.463–467). A uniform prior ranging from 0 to .5 was used on the inverse of the degrees of freedom (1 ÷ ν) of the t distribution, allowing it to range from heavy tailed (e.g. $\nu = 2$) to more normal (i.e. ν becomes larger) depending on the data.

1.3 Results

1.3.1 Training

Rats were trained to categorize 5 rat movies versus 5 object movies and 5 scrambled movies. The training started with one pair of movies, and gradually other pairs were added. Subjects completed the training phase after a total of 139, 142, 141, 177, and 158 sessions, corresponding to 1668, 1704, 1692, 2124, and 1896 trials (M = 1816.8). Counting a rate of 40 training sessions per month (2 per working day), this is a training period of 3.5-4.5 months. **Table 1.1** contains the number of trials carried out per rat per condition for the training phase and all subsequent test phases.

Before the testing phase, 6 trials per training pair and per animal were conducted to assess performance at the end of training. Performance was significantly different from chance regardless of the distractor type: mean performance was 76.7% correct for the natural distractors (one tailed $t(4) = 8.76$, $p = .0005$, $d = 3.92$) and 83.3% correct for the scrambled distractors (one tailed $t(4) = 5.52$, $p = .0026$, $d = 2.47$). Subsequently, all training stimuli were presented in all previously unseen target-distractor combinations. Again, mean performance was higher than chance for both distractors: 75.9% correct for natural distractors (one tailed $t(4) = 10.2$, $p = .0003$, $d = 4.54$) and 80.3% correct for scrambled distractors (one tailed $t(4) = 8.24$, $p = .0006$, $d = 3.69$).

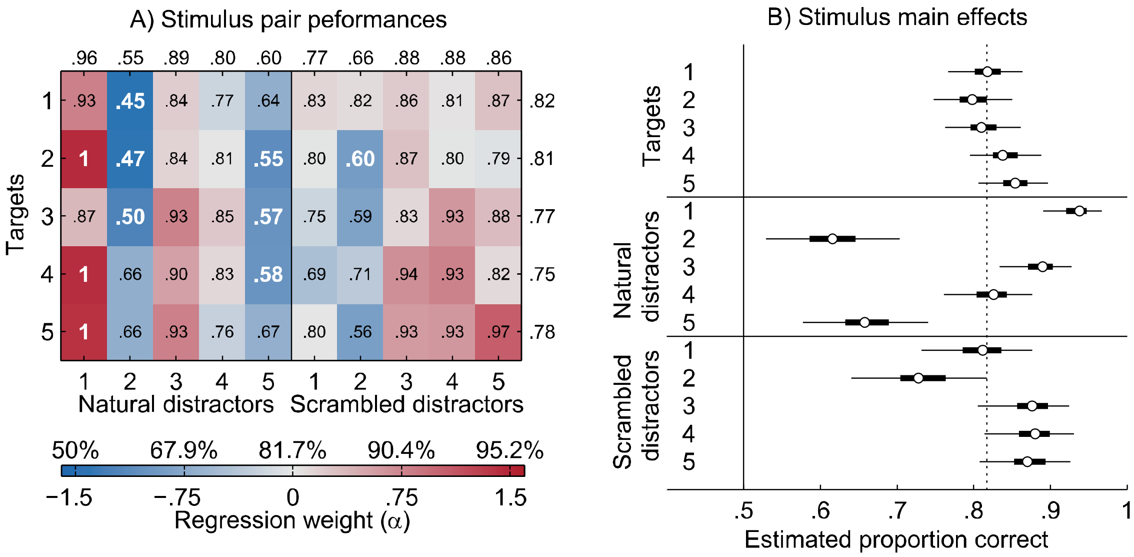


Figure 1.7. Performance on all target-distractor combinations of the training set.
Panel **(A)** displays a heat map of point estimates of the regression weights (α_k) for each different pair. Red indicates performance on this combination is estimated higher than the central tendency over all combinations (μ), while blue indicates the reverse (lower than the central tendency, which in most cases is still higher than chance performance). Percentages correct corresponding to regression weights (i.e. $100 \div \left(1 + e^{-(\mu+tickvalue)}\right)$) are indicated above the color bar. Numbers placed on the heat map are the proportion correct for each combination for all rats pooled together (with marginal proportions at the top and right side). White print indicates that the 95% HDI of the regression weight did not include zero, indicating high certainty (i.e. at least 95%) that performance on this pair was different from the central tendency. Panel **(B)** summarizes the marginal posterior distributions of estimated performances for targets, over distractors and vice versa. White dots indicate the mode, thick error bars the 50% HDI, and thin error bars the 95% HDI. The dashed line indicates the central tendency.

For more detailed inference the model of Equation 1 was fit to the data with a slight modification (see material and methods, data analysis). **Figure 1.7** displays the results of this analysis. While the numbers on top of the heat map in **Figure 1.7**A are measured proportions correct per target distractor combination, the colors represent a deflection (α) from the central tendency (i.e. overall performance, μ) for a specific level of the nominal predictor for stimulus pair. Recall that in logistic regression this deflection is on the log odds scale and an increase or decrease in percentage correct depends on the percentage correct from which the deflection is calculated. Specifically, a certain deflection on the log odds scale is compressed at the ends of the probability scale (or: the difference 55%-50% is not of the same magnitude as the difference 95%-90%). In using the parameter values α for the color scale, differences in color intensity correspond linearly to differences in performance on an unbound scale. For these data's central tendency of 81.7%, the percentages correct corresponding to different values of α are indicated on the color bar of the heat map.

Notice the presence of a pattern where color seems to vary predominantly across columns rather than rows. Since each column shows the data and estimates for a different distractor, this visual inspection already indicates that performance seems to be mostly modulated by the distractor. Target distractor combinations for which the 95% HDI of the regression weight did not include zero were indicated by printing the corresponding measured proportions correct in white bold font to further highlight those combinations for which performance deviates from the overall performance across all stimulus pairs. Recall that the 95% HDI indicates the range of values for which there is 95% certainty that the underlying population parameter that generated the data falls within the bounds of the interval. Values falling outside of the interval are rejected based on low probability. This comparison indicates that performances on several combinations with natural distractor 1 are higher than average, while performances on several combinations with natural distractors 2 and 5 and one combination with scrambled distractor 2 are lower than average (although still higher than chance performance which is 50%). On the two diagonals the proportions correct for the training pairs are located. Performance for three of these pairs is lower than the criterion of 75% correct which was upheld during training, because for some pairs we had to continue training without that criterion having been reached. To see the main effect per movie, we looked at the marginal posterior distributions for effects of targets and distractors separately. **Figure 1.7**B shows the estimated proportion correct (i.e. mode of the distribution) and its 95% HDI (indicated by thin error bars) for each target movie independent of the distractor, and for each distractor independent of the target. If the estimated proportion correct across all target distractor combinations (indicated by the dashed line) falls outside of the 95% HDI, meaning that this value is highly improbable for this target or distractor, we have strong evidence that this particular movie modulates performance independent of the movie it was paired with. Here we can clearly see that performance is substantially modulated by four natural distractors and one scrambled distractor only.

1.3.2 Generalization to new movies

Next, performance of the animals was tested on new stimuli. In order to limit learning effects, each arm of the maze contained a platform during the trials with new stimuli so there was no negative reinforcement. On the first test set the rats performed significantly higher than chance level. Mean performance was 78.3% correct for natural distractors (one tailed $t(4) = 4.91$, $p = .0040$, $d = 2.19$) and 76.6% correct for scrambled distractors (one tailed $t(3) = 7.69$, $p = .0023$, $d = 3.85$). The model of Equation 1 leads to the same conclusions for the natural and scrambled distractors of test set 1, since the 95% HDI did not include the chance level of 50% correct (95% HDI [72.0 85.4] and [68.5 84.3], for natural and scrambled distractors respectively; see **Figure 1.8**). Overall, we find that the animals were able to generalize to new movies to (a) categorize rat movies from scrambled movies, and to (b) categorize rat movies from movies containing another object.

After Test Set 1, a second set was presented, using target movies in which the rat was more stationary (as judged qualitatively by the experimenter). For the natural distractors mean performance was 58.8% correct, and estimated different from chance level (95% HDI [50.2 67.8], one tailed $t(4) = 2.42$, $p = .0366$, $d = 1.08$). Performance of 74.5% correct on the same targets versus scrambled distractors, is estimated to be substantially different from chance (95% HDI [66.8 83.5], one tailed $t(3) = 4.71$, $p = .0091$, $d = 2.35$). With natural distractors we find that performance on test set 2 was estimated lower than performance on test set 1 (non-overlapping 95% HDI). Thus, either the decreased amount of movement or another factor confounded with it makes generalization more difficult on test set 2. One potential confound might be that these movies were less similar to the trained movies (less ‘typical’) in more aspects than just the amount of motion. In a later section we will present specific manipulations of the motion in the movies of test set 1 which are meant to exclude such confounds.

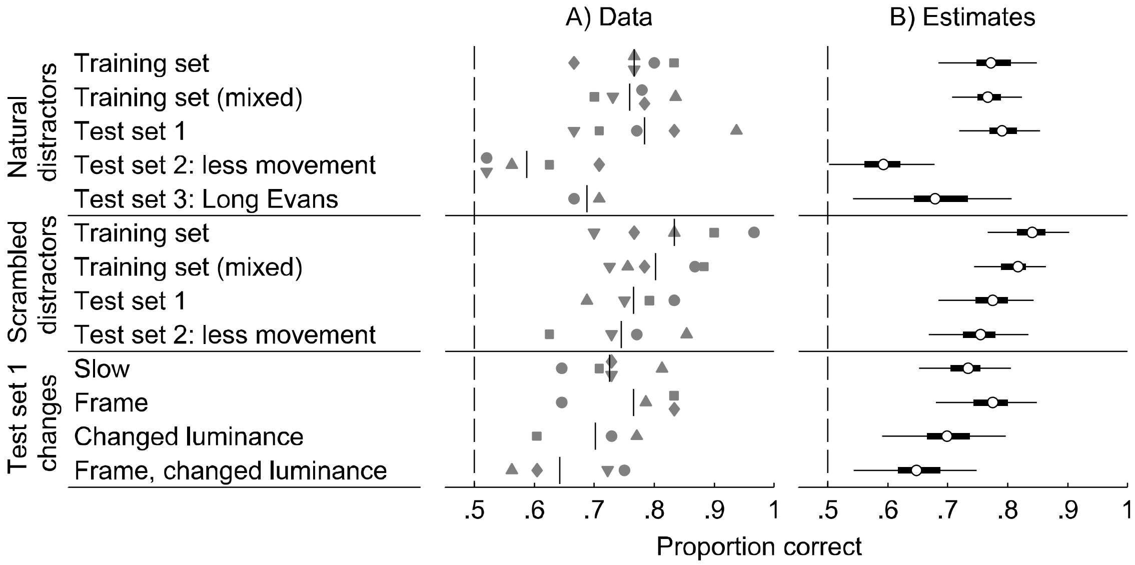


Figure 1.8. Performance per stimulus set as estimated by the model (Equation 1).
Panel **(A)**: raw performance data (each rat has its own marker), with the vertical lines signifying the mean. Panel **(B)**: posterior distributions are summarized, with white dots indicating the mode, thick error bars the 50% HDI, and thin error bars the 95% HDI. This plot indicates for each stimulus set which proportions correct are most probable given the data. Chance level (i.e., .5) is rejected when it lies outside the 95% HDI.

Combining the estimates for the training stimuli with those for test set 1 and 2, the proportion correct on stimulus pairs with a natural distractor is estimated lower than that on pairs with a scrambled distractor (95% HDI [-0.64 -0.03], on the log odds scale). Finally, performance of two rats was assessed for test set 3, which contained five target movies of a Long Evans rat, paired with natural distractors. To have a robust estimate in each rat, the rats performed each 48 trials with test set 3. Again, posterior distribution indicates performance to be higher than chance (95% HDI [54.2 80.6], binomial 95% confidence intervals for the two animals: [51.6 79.6] and [55.9 83.1]), based on an overall performance of 68.8% correct. So, the fact that the movie includes an animal which is no longer homogeneously dark, did not abolish generalization.

1.3.3 GENERALIZATION TO ALTERED VERSIONS OF THE MOVIES FROM TEST SET 1

Rats were also tested with several manipulations of test set 1. In all those manipulations the distractors were natural movies. The first two changes probed how the temporal variation of the movies affect performance, in order to have a more direct test of the effect of motion than provided by test set 2. First, we played the movies at 1/4th of their original speed and in a subsequent test only showed one static snapshot. The time point of the snapshot was that for which the frame of the target movie was most similar to all other frames in that movie (i.e., minimal pixel-wise Euclidean distance; see **Figure 1.1**). These tests were motivated by the observation of a lower performance on test set 2, where rats were more stationary in the target movies. Mean performance was 72.5% correct for the reduced playback speed and 76.5% correct for the snapshots. Both were estimated to be different from chance (95% HDI [65.3 80.5], one tailed $t(4) = 7.73$, $p = .0008$, $d = 3.46$, and 95% HDI [68 84.9], one tailed $t(3) = 5.77$, $p = .0052$, $d = 2.89$, for the speed reduction and static frame respectively). Comparisons with proportion correct on the unadjusted test set 1 do not indicate a decrease in performance (95% HDI [-0.26 0.87] and [-0.51 0.75], on the log odds scale, for the speed reduction and static frame respectively). Thus, most likely, the decrease in performance on test set 2 had to do with other confounding factors making the movies less typical. The amount of motion does not affect the ability to achieve above chance performance on test set 1 stimuli. The presence of motion is not necessary, and rats can differentiate between images containing a rat and other images based upon stationary cues.

Next, we tested whether generalization could be explained by local luminance differences. Indeed, previous studies have shown that whenever possible rats tend to use simple cues such as average luminance of the lower part of stimuli in visual discrimination tasks (Minini and Jeffery, 2006; Vermaercke and Op de Beeck, 2012). **Figure 1.9** shows that the average pixel value in the lower half of the target stimuli was consistently higher than that of the lower half of the corresponding natural distractors (note that this was also the case with four out of the five scrambled distractors; data not shown). Therefore, the lower part of target stimuli was made darker, while the reverse was done for the distractors. Specifically, pixel values where adjusted according to a linear gradient ranging from x to $–x$ from top to bottom pixel rows for target stimuli and from $–x$ to x for distractors, where x was chosen for each pair so that the difference in average lower half pixel values was just below zero (see **Figure 1.9**). In this way the global luminance was retained. Note that the average lower half pixel value was calculated across all frames.

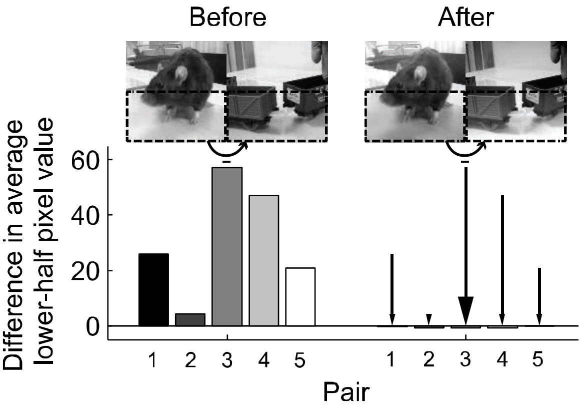


Figure 1.9. Difference in average lower-half pixel value between target and distractor before and after the adjustment.
Bar plots show the average (across width, height, and frames) pixel values in the lower half of the distractor movie subtracted from the same average of the target movie for each of the five pairs of test set 1 before (five bars on the left) and after luminance adjustment (five bars on the right; if a bar is not visible, its value is too close to zero). Positive values indicate that the lower part of the target movie is on average (across frames) lighter than the lower part of the distractor movie. Snapshots show the first frame of an example pair (corresponding to the third bar in each set of five, with the target on the left and the distractor on the right).

With full movies, the rats' mean performance was 70.1% correct, which is estimated to be different from chance (95% HDI [59.1 79.7], binomial 95% confidence intervals for each of the three animals: [58.2 84.7], [45.3 74.2], and [62.7 88.0]), and not different from the performance with the original test set 1 movies (95% HDI [-0.16 1.12], on the log odds scale). When the same luminance manipulation was applied to the static frame stimuli, mean performance was 64.2% correct, which is also estimated to be different from chance (95% HDI [54.4 74.9], one tailed $t(3) = 3.43$, $p = .0207$, $d = 1.72$), but in this case it was estimated substantially lower than the performance with the original test set 1 movies (95% HDI [0.06 1.13], on the log odds scale). Note that in this case there was only one frame, meaning that lower half pixel intensities were now equal simultaneously at all time. Overall, we still find significant generalization in both tests, showing that animals were not simply picking up a luminance difference in the lower half of the stimuli. While performance on the static pixel intensity adjusted frames did differ from that in test set 1 (without this being the case for the pixel intensity adjusted *movies*), the total picture is more complicated since there is no convincing evidence for a difference in performance between the static versus moving adjusted stimuli (95% HDI [-0.5 0.9], on the log odds scale). The lower observed performance can be explained by the fact that rats had been doing more trials with two platforms by then, which might decrease motivation as a result of mistakes being rewarded (see also **Figure 1.6**b). Another possible explanation is that the pixel intensity adjustment is more thorough in the case of one stationary frame, because it is now applied to the level of this individual frame. Most importantly, there is clear above chance performance for all test sets and for all included stimulus manipulations.

1.3.4 Do rats use a strategy based on local luminance

To determine whether rats used a strategy based on local luminance or pixel change a linear logistic-binomial regression model (see Equation 2) was fit on the rat performance scores with the following predictors: local mean luminance and local mean variation in luminance of the target and distractor in nine locations of each screen. The latter two statistics are the same as M($\bar{I}$) and M($\overline{PC}$) (defined in the method section), with the exception that they were calculated separately for different locations on the screen: each frame was divided in three by three equally sized (128×128 pixels) squares (which together cover the entire frame/screen). Concretely, this means that performance on each stimulus pair is estimated based on 36 metric predictors (2 metrics × 9 locations × 2 stimuli). The model was fit to the performance on all target-distractor combinations of the training stimuli (shown in **Figure 1.6**) in order to check whether a strategy that could be learned from these stimuli might allow rats to generalize to the test stimuli. **Figure 1.10**A depicts four templates based on the point estimates of the regression weights for the predictors (each of the nine squares of the templates corresponds to one of the nine locations on the stimulus). Note that the highest loading regression weights are for the properties of the distractor, not the target.

If rats use one or more of these stimulus properties in their generalization to new stimuli, the regression weights fit to the training data should accurately predict performance in the testing data. For example, if the regression weights represent a real strategy, then we would expect a distractor to be associated with better-than-average performance if it would have a higher-than-average luminance in the top right corner (the most positive regressor in the distractor pixel intensity template in **Figure 1.10**A and/or a lower-than-average luminance in the bottom right corner (the most negative regressor in this template). For all of the predictor regression weights the 95% HDI included zero (**Figure 1.10**B), so for none of these regression weights there was enough evidence to reject zero. Moreover, the proportion of variance explained by the model (1 - residual variance ÷

total variance) is .12 for the training data and .01 for the test data (which were not used to fit the model). Thus, the significant generalization of the rats and the variation in generalization performance among different targets and distractors cannot be explained by a strategy based upon local luminance cues.

Next, the same model was used but now with predictors referring to the local difference

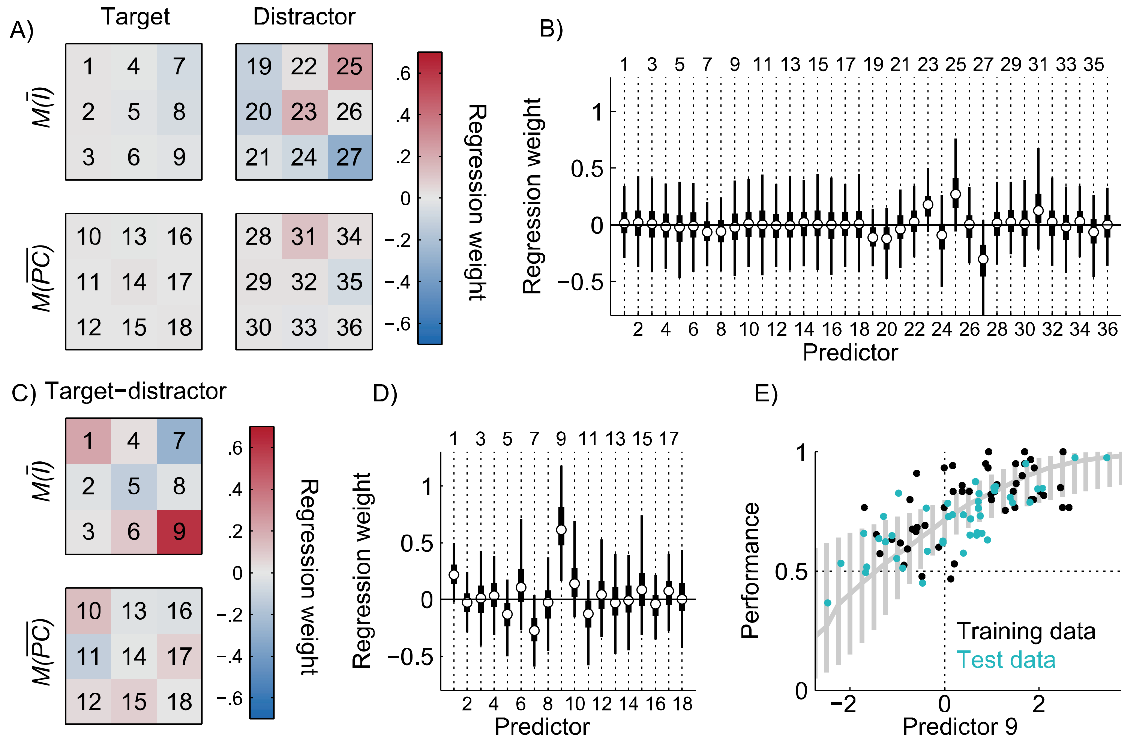


Figure 1.10. Local luminance cues and performance.
Panel **(A)** depicts templates based on modes of the posterior distributions of the regression weights for the 36 luminance predictors: average pixel values and mean absolute pixel change, each on nine locations of both stimuli (the nine squares in each template correspond to the nine locations on the square stimulus frames). Red indicates average luminance or pixel change in this area correlates positively with performance, while blue indicates the reverse. Panel **(B)** summarizes the posterior distributions of the regression weights for each of the 36 predictors, with dots indicating the mode, thick lines the 50% HDI and thin lines the 95% HDI. Panel **(C)** and **(D)** are analogue to panel (A) and (B), but show the results of a model based of the difference in each corresponding local luminance cue on the target and the distractor. Red indicates that a higher difference in average luminance or pixel change in this area for the target versus the distractor correlates positively with performance, while blue indicates the reverse. Panel **(E)** shows performance on each target-distractor combination (dot) as a function of the luminance difference in the lower right corner (predictor 9 of the difference template, which is estimated to be different from zero). Only the training data (indicated in black) were used to fit the model. The mode of the posterior distribution and 95% HDI's are indicated in grey as a function of predictor 9. This panel shows that this predictor cannot explain generalization because the model's intercept does not coincide with chance level. At the intercept, where there is no average difference in luminance in the lower right corner of the screens between target and distractor (i.e. predictor 9 is equal to zero), performance is well above chance (as indicated by the data points and the 95% HDI shown in grey).

in luminance between target and distractor (these new predictors correspond to a subtraction of the predictors of the previous model). This new model tests whether a strategy based upon differences between the target and distractor on corresponding local luminance or pixel change values could have allowed for successful generalization.

The templates (**Figure 1.10**C) and regression weights (**Figure 1.10**D) indicate that performance on the test set is positively correlated with a difference in luminance (for target minus distractor) on the lower right part of the screen. This difference template model has a proportion of variance explained of .12 for the training data and .13 for the test data. However, the predicted performance is still 71.5% correct (95% HDI [61.4 80.0]) for the intercept, which is the estimate for when there is no information in the difference template. Indeed, both the model as well as the training and test data shown in **Figure 1.10**E support the conclusion that while performance is modulated by a luminance difference in the lower right corner of the screen, it cannot explain generalization. Rats neither perform at chance when this predictor is zero, nor do they prefer the distractor when it is negative. Generalization performance is still around 70% even when there is no luminance difference.

1.4 Discussion

Five out of six subjects were able to complete the training phase. Mixing up the training pairs proved that the acquired decision rule(s) were not bound to these specific target-distractor combinations. In addition, these data with the training movies indicated that the variability in performance for different pairs can mainly be explained by the variability in natural distractors. Subsequently, the animals successfully generalized to a first typical test set, another test set with more stationary rats/objects, and one with a strain of differently colored rats. In general, performance with scrambled distractors was higher than with natural distractors.

Taken together, the results of the test phase show a successful generalization to a set of novel, unique stimulus pairs. This was the case for pairs with a natural as well as with a scrambled distractor. The latter are more different from the target movies in that they lack naturally occurring feature conjunctions. Even though performance was mainly modulated by the distractor, one cannot conclude that this means the animals used an avoid-distractor strategy. For example, this finding can be explained equally well by the simple fact that the content of the distractor movies was more variable than that of the targets.

1.4.1 Simple behavioral strategies which cannot explain the generalization to novel movies

We investigated several simple strategies which could underlie the main results. For instance, rats might have used general differences in motion energy or local luminance. Neither reducing the frame rate, nor presenting stationary frames, resulted in a substantial reduction of performance. This means that motion cues in the movies were not a critical factor. Likewise, there is no evidence that equalizing the luminance in the lower part of the target corrupted performance on test movies. The latter did affect performance on stationary frames, yet even in this case it remained well above chance.

Finally, we did a control analysis to see whether a more complex pattern of local luminance cues could explain generalization. The results show that these cues cannot explain above chance performance on the test sets. Therefore, we conclude that both simple local luminance and motion energy are insufficient to explain the achieved proportion correct on the test sets, which indicates that generalization relied on a more complex combination of features.

1.4.2 Behavioral strategies which might underlie the generalization to novel movies

Here we consider three non-trivial and interesting strategies. Although we discuss to what extent they might underlie performance in our experiments, further studies are needed to distinguish between these possibilities.

First, the rats might use contrast templates by comparing the luminance in different screen positions (instead of using the simple luminance cues which we ruled out). We recently suggested the use of such contrast strategies as an explanation of the behavioral templates in an invariant shape discrimination task (Vermaercke and Op de Beeck, 2012). Such contrast templates can be fairly complex by combining different contrast cues, as has been suggested in the context of face detection by the human and monkey visual system (Gilad et al., 2009; Ohayon et al., 2012). Nevertheless, these templates arise from low spatial frequencies and do not necessarily require orientation selectivity, edge detection, or curvature processing and are effectively used in computational face detection models (Viola and Jones, 2001; Viola et al., 2004).

This property sets the contrast template strategy aside from a second strategy based on shape cues such as edges/lines, corners, and curvature. Hierarchical computational models of object vision based upon the primate literature (Hummel and Biederman, 1992; Cadieu et al., 2007) aim to process the visual input in terms of such shape features which have been shown to drive neurons in inferior temporal cortex in monkeys (Kayaert et al., 2005; Connor et al., 2007). In rats we currently lack such neurophysiological evidence. Previous studies reporting the use of shape information by rats (Simpson and Gaffan, 1999; Alemi-Neissi et al., 2013) did not make this important distinction and therefore cannot exclude the use of contrast templates.

Third, we cannot exclude the possibility that rats would have a notion of rats as a 'semantic' category. However, we believe this possibility is very unlikely, at least when based on visual cues only. First, it takes quite some time to train them to categorize movies containing a congener from non-rat movies. If this category distinction would be salient to them, as it is for humans and other primates, we would expect that training with only one pair of movies would allow very good generalization to other pairs. In contrast, training was relatively slow, also for later movie pairs. This could be because they tend to use simpler cues first and/or because they do not make this distinction naturally. However, this study cannot say anything about a possible semantic representation of the category rats relying on one or more other modalities that are more ecologically relevant to rats as a species.

1.4.3 Comparison with categorization of natural stimuli in monkeys

At this point it is relevant to compare our findings to the two most similar studies in monkeys: Vogels (1999a) and Fabre-Thorpe (1998). Similar to both studies, rats could learn to successfully discriminate natural stimuli belonging to different categories and generalize to novel stimuli. Even though the training period (on average 151.4 sessions) might seem highly intensive, the average number of training trials (1816.8) is actually relatively low. Vogels (1999a) used probe stimuli to test whether a single low-level feature lead to generalization and concluded that at least feature combinations were required. Similarly, in the present study a number of probe tests were performed to exclude the simplest low-level strategies. Fabre-Thorpe (1998), on the other hand, focused on the speed of categorization during very brief presentations. The setup used in the current study neither allowed for such a short stimulus presentation, nor for fast response by the rat or for an accurate measurement of reaction times. A different setup using still images would be necessary to investigate that aspect of categorization in rats. Finally, the most obvious difference with both Vogels (1999a) and Fabre-Thorpe (1998) is that in the present study natural movies were used instead of still natural images. However, presenting snapshots of the movies did not disrupt generalization. Overall, there are interesting commonalities with previous findings in monkeys, but a more systematic comparison requires a study which tests both species on the same stimuli in the same task context.

1.4.4 Neural mechanisms

The swimming-based task used here was chosen for its relative ease to train rats and the very low error-rate the animals obtain with easy stimuli. This task cannot immediately be combined with experiments involving electrophysiological recordings. Of course, as mentioned in the introduction, uncovering the visual capabilities of rodents on a behavioral level to evaluate the validity of rodents as a model for vision (such as Tafazoli et al., 2012) is in itself relevant for the growing group of neuroscientist focusing on these animals. Furthermore, just as other swimming-based tasks used in neuroscience such as the Morris water maze, techniques such as lesioning, genetic or pharmacological manipulations, and activity-mapping with immediate early gene expression can be successfully applied in the context of our task. Finally, an extension with simultaneous neural recordings might use virtual navigation (Harvey et al., 2009) as a paradigm with similar behavioral responses (i.e. "running towards").

Which neural representations might underlie the categorization performance? In primates visual features that are encoded in the primary visual cortex (V1) are integrated into higher level representations in the extrastriate cortex (Orban, 2008). Traditionally these extrastriate areas are grouped into two anatomically and functionally distinct pathways: a ventral stream, providing the computations underlying object recognition, and a dorsal stream, mediating spatial perception and visually guided actions (Kravitz et al., 2011). Neurons in the ventral stream in monkeys display category specific responses which are tolerant to changes in various image transformations (e.g., Vogels, 1999b; Hung et al., 2005). The superordinate distinction between animals and non-animals has also been related to strong categorical responses in monkey and human ventral regions (Kiani et al., 2007; Kriegeskorte et al., 2008b).

Based on the high complexity and variability of the stimuli and on the evidence from the probe tests and the local luminance control analyses presented here, we suggest that a computation based on the integration of features encoded in V1 would have been necessary for generalization to novel stimuli. Consequently, extrastriate cortical regions might be involved as in primates. Previous research has suggested that the rodent visual cortex consists of two streams resembling the dorsal and ventral pathways in primates (Wang et al., 2012). It seems therefore natural to suspect that the putative ventral stream in rodents is involved in learning the categorical distinction between rat and non-rat movies. Indeed, one of these areas has been shown to respond to high spatial frequencies in mice, which might indicate a role in the analysis of structural detail and form (Marshel et al., 2011). However, for now this proposal remains very speculative given the many differences between rodents and monkeys and the lack of knowledge about rodent extrastriate cortex.

Chapter 2.

NATURAL STIMULUS REPRESENTATIONS IN RAT VISUAL CORTEX

In our previous experiment, we had established that rats can be trained to discriminate categories of natural movies and generalize to novel exemplars. Meanwhile, research in our lab was showing promising evidence of position tolerance in their visual system (Vermaercke et al., 2014), a hallmark of object processing in primates. Would the rat visual system show other hallmarks of object processing? *Do we see a categorical representation emerge?*

Published as

Vinken K., Van den Bergh G., Vermaercke B., Op de Beeck H. (2016). Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. *Cerebral Cortex, 26* (7), 3310-3322.

NEURAL REPRESENTATIONS OF NATURAL AND SCRAMBLED MOVIES PROGRESSIVELY CHANGE FROM RAT STRIATE TO TEMPORAL CORTEX.

In recent years the rodent has come forward as a candidate model for investigating higher level visual abilities such as object vision. This view has been backed up substantially by evidence from behavioral studies that show rats can be trained to express visual object recognition and categorization capabilities. However, almost no studies have investigated the functional properties of rodent extrastriate visual cortex using stimuli that target object vision, leaving a gap compared to the primate literature. Therefore, we recorded single-neuron responses along a proposed ventral pathway in rat visual cortex to investigate hallmarks of primate neural object representations such as preference for intact versus scrambled stimuli and category-selectivity. We presented natural movies containing a rat or no rat as well as their phase-scrambled versions. Population analyses showed increased dissociation in representations of natural versus scrambled stimuli along the targeted stream, but without a clear preference for natural stimuli. Along the measured cortical hierarchy, the neural response seemed to be driven increasingly by features that are not V1-like and destroyed by phase-scrambling. However, there was no evidence for category selectivity for the rat versus non-rat distinction. Together, these findings provide insights about differences and commonalities between rodent and primate visual cortex.

2.1 INTRODUCTION

Visual perception is the end product of a series of computations that start in the retina and culminate in several cortical areas. Although we can readily experience this end product effortlessly, decades of intensive research still have not yielded a full picture about the computations taking place beyond the point where visual information first arrives at the cortex, the primary visual area (V1). Until a few years ago the neural underpinnings of visual perception were mainly investigated in primates and cats. With the recent surge of rodent studies involving new techniques which have proven to be of high value to disentangle the mechanisms of visual processing (Huberman and Niell, 2011), questions concerning the functional properties and capabilities of areas in rodent extrastriate visual cortex have become highly relevant. Behavioral experiments have found evidence in rats for forms of higher level visual processing (Zoccolan et al., 2009; Tafazoli et al., 2012; Vermaercke and Op de Beeck, 2012; Alemi-Neissi et al., 2013; Brooks et al., 2013; Vinken et al., 2014; Rosselli et al., 2015; for review, see Zoccolan, 2015), fueling the idea that these animals might be useful as an alternative and experimentally more flexible model to tackle certain questions related to these complex visual capabilities.

In primates, extrastriate visual areas further integrate visual features that are encoded in V1 into more complex representations (Orban, 2008). These areas have traditionally been grouped into two anatomically and functionally distinct pathways: a dorsal stream and a ventral stream (Mishkin and Ungerleider, 1982; Kravitz et al., 2011). The latter is responsible for the transformations that eventually produce the ingredients necessary for extraordinary abilities such as object recognition, namely high selectivity distinguishing between objects, combined with tolerance for a range of identity preserving transformations, such as changes in size, position, viewpoint, illumination, etc. (DiCarlo and Cox, 2007; Dicarlo et al., 2012). The result is a high level representation that manifests itself in strong categorical responses in monkey and human ventral regions, with for example a high selectivity for the distinction between animal and non-animal pictures (Kiani et al., 2007; Kriegeskorte et al., 2008a). This category selectivity comes on top of a general preference in primate occipitotemporal cortex for natural, intact images compared to scrambled versions of these stimuli. Thus, in primates the computations along the ventral pathway introduce a bias in favor of coherent stimuli containing surfaces and objects over random texture patterns. This preference for intact coherent images was found higher up in this pathway through human functional magnetic resonance imaging (fMRI, Grill-Spector et al., 1998), monkey fMRI (Rainer et al., 2002), and monkey single-neuron physiology (Vogels, 1999c). This bias does not exist in lower levels of the pathway where sometimes even a preference for scrambled images is found (Rainer et al., 2002), potentially depending upon the exact scrambling procedure (Stojanoski and Cusack, 2014).

Can we find evidence for similar computations being performed in the rodent brain? Previous research has suggested that anatomically the rodent visual cortex consists of two streams resembling the dorsal and ventral pathways in primates (Niell, 2011; Wang et al., 2012). Already some steps have been taken to investigate the functional properties of rodent extrastriate cortex using drifting bars and gratings (Andermann et al., 2011; Marshel et al., 2011) and simple shapes (Vermaercke et al., 2014). Marshel et al. (2011) reported that mouse latero-intermediate area (LI) prefers high spatial frequencies, which might indicate a role in the analysis of structural detail and form. Vermaercke et al. (2014) report an increase in position tolerance, consistent with the primate ventral visual stream, along a progression of five cortical areas starting in V1 and culminating via LI in recently established lateral occipito-temporal area TO (Vermaercke et al., 2014). This increased position tolerance paralleled a gradual transformation of the selectivity for the simple shapes used in the study. However, these areas were hardly selective to stationary shapes and were more responsive to moving stimuli, which contrasts with the primate ventral visual stream. More complex stimuli such as natural movies have rarely been used in rodents, with two recent exceptions (Kampa et al., 2011; Froudarakis et al., 2014). In those studies, the focus was primarily on the population code in primary visual cortex. Kampa et al. (2011) measured responses of V1 layer 2/3 populations to dynamic stimuli (including natural movies), showing reliable stimulus-specific tuning and evidence for functional sub-networks (despite the lack of orientation columns in rodent V1). Froudarakis et al. (2014) found that natural scenes evoke a sparser population response compared to phase scrambled movies, leading to an improved scene discriminability that also depended on cortical state. Both studies focused on primary visual cortex and not explicitly on coding of movie content. Here we investigated whether the two most salient functional hallmarks of neural object representations in primates might also exist in rodents: preference for intact versus scrambled stimuli and category-selective responses. To achieve this, we recorded action potential activity in three areas belonging to this putative ventral stream in rats with the aim of systematically comparing how stimulus representations change across areas: V1, LI, and TO. LI is the most downstream area in the putative ventral visual pathway which has been identified in both mice and rats (Espinoza and Thomas, 1983; Wang and

Burkhalter, 2007); TO extends even further to rat temporal cortex and its responses to simple grating stimuli and shapes already suggested a higher-order processing compared to the other areas (Vermaercke et al., 2014). While recording neural responses, we presented natural movies belonging to different categories, as well as phase-scrambled versions of these movies. Based on the primate research, we would expect very different results in higher stages of the cortical processing hierarchy. First, a functional hierarchy would be supported by a systematic and gradual change in population representation of scrambled versus natural movies across areas. Second, a change culminating in a preference for natural movies would show that this functional hierarchy is comparable to the primate ventral visual stream in this respect, a notion that would even be more supported by a categorical representation towards the most downstream area TO.

2.2 MATERIALS AND METHODS

Much of the materials and methods have been described previously in detail (for descriptions of the apparatus, methodological details, and functional criteria, see Vermaercke et al., 2014; for a description of the stimuli, see Vinken et al., 2014). There was however no overlap and animals were completely naïve with respect to the stimulus set. Here we focus upon the details which are most important and most relevant in the context of the present study.

2.2.1 ANIMALS

Experiments were conducted with 7 male FBNF1 rats, aged 14 to 30 months (21 on average) at the start of the study. This specific breed was chosen for their relatively high visual acuity of 1.5 cycles per degree (Prusky et al., 2002). Surgery was performed to implant a head post and a recording chamber. The craniotomy was centered -7.9 mm anterioposterior and -2.5 mm lateral from bregma. This location allowed the electrode to pass through five different visual areas, including our three target areas, when entering at an angle of 45°: V1, latero-medial area, LI, latero-lateral area and TO (Vermaercke et al., 2014). As in Vermaercke et al. (2014), we performed histology in five out of the seven rat brains, confirming that the electrode tracks followed trajectories similar to that study (therefore also confirming a sampling bias towards upper layers in V1). After recovery the animals were water deprived and had ad libitum access to food pellets. Housing conditions and experimental procedures were approved by the KU Leuven Animal Ethics Committee.

2.2.2 Stimuli

The set of stimuli used in this experiment corresponds to the training set described in Vinken et al. (2014). The set consisted of 20 movies: 10 natural movies and the phase-scrambled versions of these movies. The natural movies had a duration of 5 seconds and were recorded at 30 Hz (thus including 150 frames) and sized 384×384 pixels. Five of them contained a rat, while the other five contained a moving object. For each rat movie, a non-rat movie was chosen from our own database of 537 five-second movies in order to match relatively well on pixel intensities, contrast and changes in pixel intensities (Vinken et al., 2014). The rat movies showed moving rats of the same strain as the subjects. Three of the paired non-rat movies contained a train, one a gloved hand moving in and out of the screen, and one a moving stuffed sock. For each movie a phase scrambled version was created according to the procedure described previously, which allows for a better frame-wise match according to statistics such as average pixel intensity, contrast, changes in pixel intensity across successive frames, as well as spatial power spectrum compared to standard methods (Vinken et al., 2014). See **Figure 2.1** for snapshots of each movie. The original movies and their scrambled versions were created at a size of 384×384 pixels (to reduce memory load), but in the electrophysiological experiment the movies were shown at a size of 768×768 pixels.

2.2.3 Electrophysiological recordings

As described at full length by Vermaercke et al. (2014), the rats were head-fixed and placed in front of a 24” LCD screen (1280x768 at 60Hz), which was gamma corrected to obtain a linear transfer function between pixel intensity values and luminance. The animal’s nose pointed at the left edge at an angle of 40° and a closest eye-to-screen distance of 20.5 cm. The movies were always presented at the full height of the screen

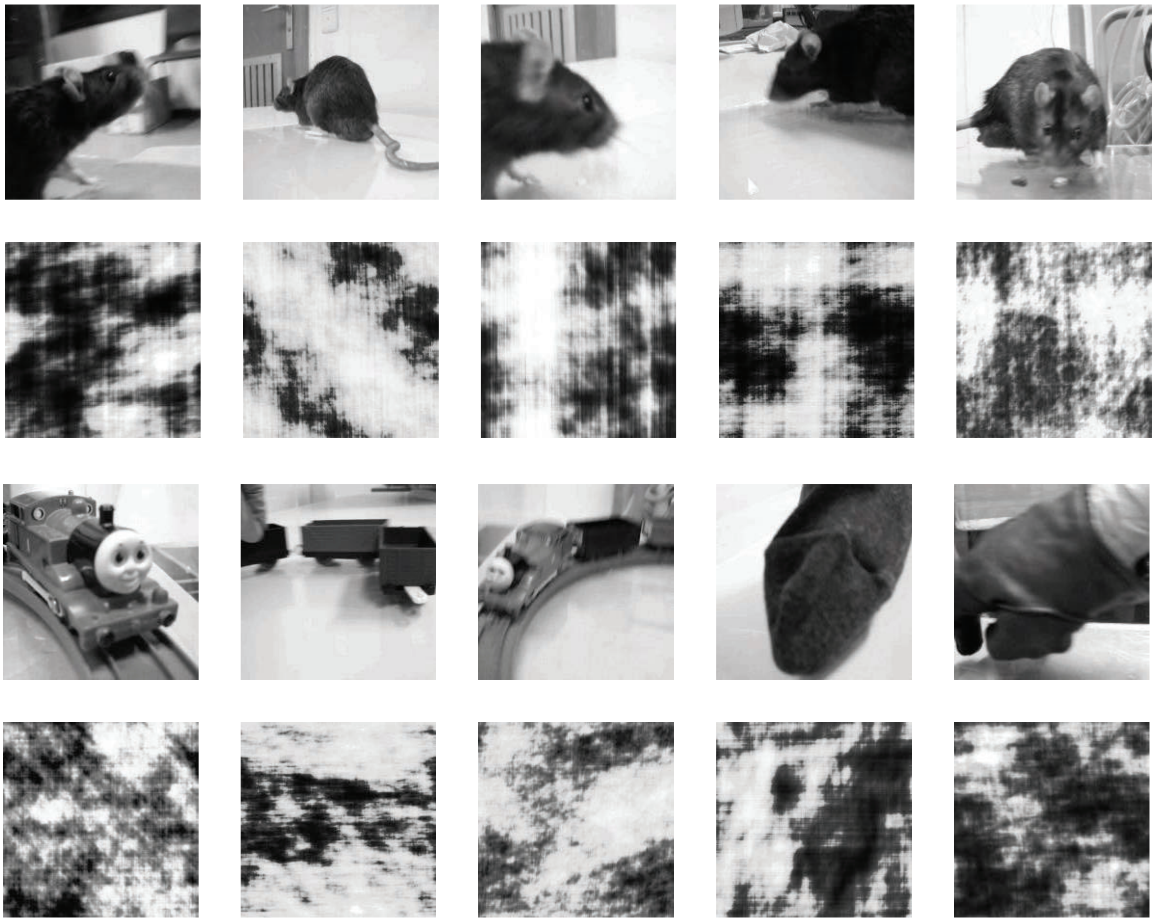


Figure 2.1. Representative snapshots of the movies used in the experiments.
First row depicts the original natural rat movies, with the corresponding scrambled versions represented on the second row. The third row depicts the natural stimuli belonging to the non-rat category, each matched to the rat movie displayed in the same column (see materials and methods, stimuli). From left to right: three movies of a toy train, one with a stuffed sock, and one with a gloved hand, with the corresponding scrambled versions represented on the fourth row. The full movies are available at:
http://ppw.kuleuven.be/home/english/research/lbp/downloads/ratMovies.

(768 pixels) and positioned on the horizontal axis according to the estimated receptive field location (see receptive field mapping below). This resulted in a stimulus width ranging from 50 to 74 visual degrees depending on the position (as the eye-to-stimulus distance varies according to position). During the experiments every fifth (movie experiment) or tenth (receptive field estimation) stimulus presentation a water reward was given.

Recordings were performed with a Biela Microdrive and single high-impedance electrodes (FHC, Bowdoin, ME; ordered with impedance 5 to 10 MΩ) in areas V1, LI, and TO. Spike detection was done using custom written code in Matlab (The MathWorks, Inc., Natick, MA), with the spike detection threshold set to detect spikes with a peak-to-peak amplitude of four times the standard deviation of the noise. Single-units (SU) were isolated based on cluster analysis of the properties of the recorded waveforms (the first n principal components, where n was optimized to the situation) using KlustaKwik 1.6, followed by a manual check in SpikeSort 3D 2.5.1. Spike waveforms that could not be separated into single units were pooled into one multi-unit (MU) cluster per recording site (each spike waveform was only used once, so there is no overlap between SU and MU). On average the peak-to-peak amplitude of the mean spike waveform was 12.4, 12.9, and 11.3 times the standard deviation of the noise for V1, LI, and TO units respectively. For all except two (one in V1 and one in TO) of the neurons included for analysis, this signal-noise-ratio was higher than the criterion value of 5 used by Issa and DiCarlo (2012). For multi-unit clusters these values were 4.9, 4.7, and 4.6 for V1, LI, and TO respectively (note that these values are limited in the lower end by the spike detection threshold of four times the standard deviation of the noise).

Receptive field mapping

Boundaries of the five aforementioned different areas were estimated based on changes in retinotopy as described previously (Vermaercke et al., 2014). A rough estimate of a site's population receptive field could be obtained manually by using continuously changing shapes or drifting gratings that could be moved across the screen. A quantitative estimate of receptive field size and location was achieved by flashing a hash symbol at 15 locations (3 rows by 5 columns) on the screen. Movies were translated along the horizontal screen axis in order to best cover the receptive field. We chose to record from V1, LI, and TO, and not the two additional intermediate areas LM and LL, because the elevation of the receptive fields encountered in V1, LI, and TO tends to be very similar (see Fig. 2C in Vermaercke et al., 2014). In contrast, the receptive fields encountered in LM and in particular LL show a very different elevation, which would make it difficult to compare results between the different areas (the receptive fields of the neuronal populations would then cover different parts of the movies).

Presentation of movies

In the main experiment, rats were passively viewing the 10 five-second natural movies and the phase scrambled versions of these movies. These were presented in random order intermitted by a two-second blank screen, with 10 repetitions per movie. The pixel intensity value of the blank screen and the part of the screen not covered by the movies was set equal to the average pixel intensity of all movies.

2.2.4 DATA ANALYSIS

We maintained two criteria to include units for analysis: units needed to be isolated for the full 10 presentations of each movie and have an average net response of more than 2 Hz for at least one movie.

Pre-processing

Before all analyses, peristimulus time histograms (PSTHs) with a bin width of 1 ms were made for each trial across the [-1999, 6000] ms interval (with stimulus onset at 0 ms). To estimate the response onset latency the PSTHs were averaged across trials and stimuli and smoothed with a Gaussian kernel (3 ms full width at half maximum). Response onset latency was defined per unit as the first time point after stimulus onset where the smoothed PSTH exceeded a threshold of the baseline activity (calculated from the]-1000, 0] ms window) plus three times the baseline activity standard deviation. For all further analyses only a 4800 ms time window after response onset was used, with the first 200 ms cut off to ignore the onset peak mainly for fitting the motion energy model. For consistency the same 4800 ms window was used for all other analyses, even though the inclusion of the window does not affect the results in any significant way.

Sparseness and reliability

Response sparseness for a certain neuron to a certain movie was quantified using the index defined by Vinje and Gallant (2000):

$$S = \left(1 - \frac{1}{n}\frac{(\sum_i r_i)^2}{\sum_i r_i^2}\right) \Big/ \left(1 - \frac{1}{n}\right)$$

where S is the sparseness index for a neuron with average (across trials) response r_i to frame i of a stimulus with n frames. Onset of the first bin is the estimated response latency plus 200 ms (see pre-processing) and the bin-width is ~33.3 ms, which corresponds with the frame rate. This sparseness index can vary between 0 and 1, with values close to 0 indicating a dense response, and values close to 1 indicating a sparse response. Response reliability for the time course of the response of a certain neuron to a certain movie was estimated using the Spearman-Brown correction as follows:

$$(nr_{xx'})/(1+(n-1)r_{xx'})$$

With the average correlation across time between two trials $r_{xx'}$ (across all combinations) and number of trials n. As before, the bin width to calculate the reliability was set to correspond with the frame rate of 30Hz.

Population representation

For each population of neurons (i.e. in V1, LI, or TO) pair-wise stimulus dissimilarities were calculated based on the correlation distance using the average responses to full movies. First, firing rates were averaged across trials and per stimulus across the entire 4800 ms interval, resulting in 20 responses per neuron. Second, responses of each neuron were transformed to Z-scores (across stimuli). Third, for each stimulus a response vector was created containing the transformed responses of each neuron to that particular stimulus. Finally, dissimilarity between a pair of stimuli is defined as 1 – r (Pearson correlation) between the response vectors of the two stimuli in question. **Figure 2.2** illustrates how we used this method to create dissimilarity matrices.

Spatiotemporal motion energy model

To simulate the relative response of V1-like cells we calculated the output of a spatiotemporal motion energy model (Adelson and Bergen, 1985). Specifically, the spatiotemporal receptive field of a modeled V1 neuron is based on a three-dimensional Gabor filter, with a certain frequency, orientation, and location relative to the stimulus. The output of the filter (calculated through linear multiplication with the stimulus) is then squared, and summed with the output of the quadrature pair to that filter which is 90 degrees out of phase. This squared and summed output gives a physiologically plausible measure of motion energy. The square root of this measure is our modeled response of a complex V1 cell (Nishimoto and Gallant, 2011). See **Figure 2.3** for a schematic representation of this process.

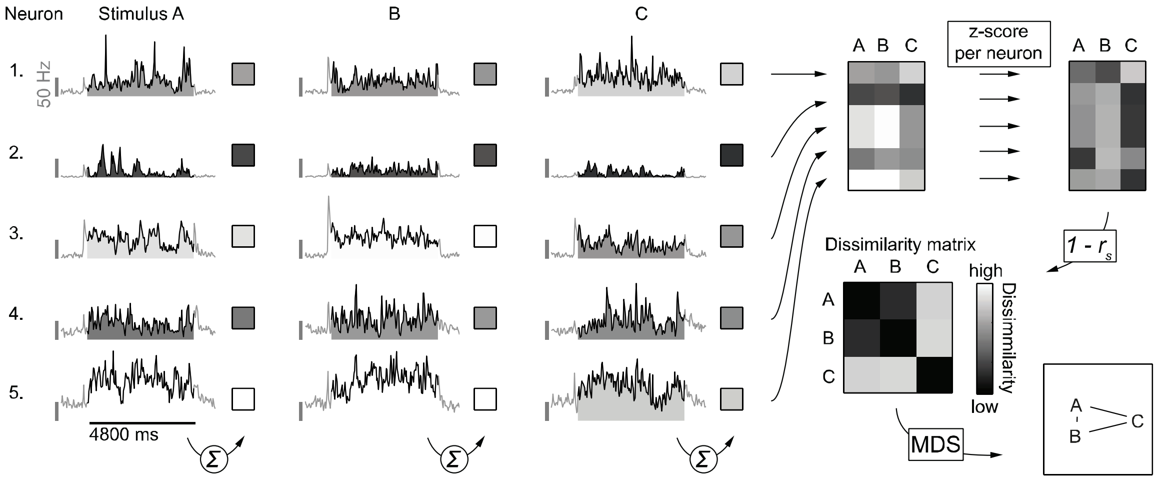


Figure 2.2. Schematic illustration of how dissimilarity matrices were calculated.
For the analysis of the population representation we started with the responses of single units, averaged across repeated stimulus presentations and summed across the 201 to 5000 ms window calculated from response onset (as indicated by the summation symbol). PSTH's in this example figure illustrate responses averaged across stimulus presentations of five single units to three movies (A, B, and C), all arbitrarily selected for the purpose of illustrating the methods. This was done for each stimulus (e. g. A, B, and C, in this example) to get raw response vectors. These raw responses were further standardized per neuron, by calculating the Z-scores across stimuli. Next the correlation matrix was calculated from these normalized response data by pair-wise correlation of the stimulus response vectors. For the stimulus dissimilarity matrix each value in this correlation matrix was subtracted from one, resulting in values between 0 and 2, where 0 indicates the lowest dissimilarity (i.e. an identical population response pattern) and 2 indicates the highest dissimilarity (i.e. a highly different population response pattern). To visualize the stimulus space as represented by the population of neurons, we performed multi-dimensional scaling on the dissimilarity matrix and present the stimuli using the first two dimensions. Stimuli plotted closer together (A and B in this example) have a more similar population response pattern than stimuli plotted further apart (A and C, and B and C in this example).

A wide range of Gabor filters spanning different frequencies, orientations, and locations is then used to model our set of V1 cells. Thus, we end up with a modeled V1 complex cell for each spatiotemporal frequency, orientation, and spatial location included in the model. The output of each modeled cell is then standardized by calculating the Z-score across all movie frames. The set of Gabor filters spanned eight different directions, six different spatial frequencies, and six different temporal frequencies. The spatial frequencies were log spaced between .04 and .15 cycles per degree and the temporal frequencies between 0 and 15 Hz, based on the optimal responses of rat V1 neurons reported by Girman et al. (1999). Each filter occurred at different spatial locations. Grid spacing was identical to what is reported by Nishimoto and Gallant (2011) and depended on spatial frequency: filters were separated by 2.2 standard deviations of the

Gaussian envelope, with one standard deviation set to half a cycle of the sine wave. Next, the output of these filters was used as predictors in a regularized linear regression model with an early stopping rule (David et al., 2007; Nishimoto and Gallant, 2011) fitted to the neural responses using code from the STRFlab toolkit (version 1.45, retrieved from http://strflab.berkeley.edu/). The model was estimated at five different latencies, ranging from 20 to 153 ms in steps of the duration of one frame. For each unit and latency, the model was fit 10 times, each time refraining two movies (i.e. a natural movie and its scrambled pair) from the fitting procedure for cross-validation and using the remaining 18 movies for training the model. Reported accuracies refer always to data that was not included in the training set.

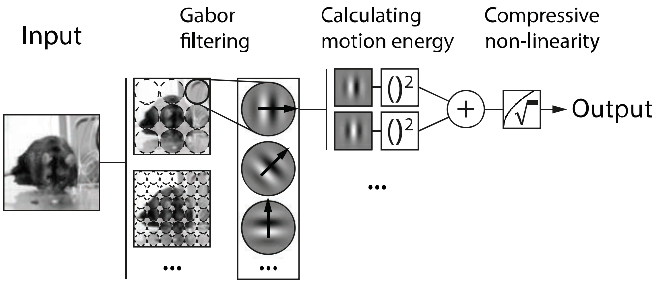


Figure 2.3. Schematic representation of the motion energy model
(based on Nishimoto et al., 2011), described under materials and methods, data analysis, spatiotemporal motion energy model. In short, input stimuli (movies) are run through a bank of quadrature pairs of Gabor filters, each with a certain spatiotemporal frequency and orientation and located on a grid covering the stimulus. The output of each pair is then squared and summed to give a physiologically plausible measure of motion energy. The end result is finally obtained by taking the square root to model a compressive nonlinearity. This final output is calculated for each of the spatiotemporal frequencies and locations covered by the bank of Gabor filters and standardized per filter across frames. In a next step the neural response is predicted as a linear combination of those standardized outputs.

Statistical analysis

For statistical inference we relied on the bias-corrected accelerated bootstrap (BCa; Efron, 1987) by random sampling with replacement (10000 iterations) of the neurons/units (unless indicated otherwise) to estimate the 95% confidence interval (CI) of the statistic in question. In addition, randomization tests (10000 iterations) are used to estimate the distribution of the test statistic in question under the null hypothesis in order to calculate p-values. In several places we report the slope of a linear regression to quantify gradual change across the three regions in the pathway under investigation for reasons of simplicity and interpretability, without the intention of making strong claims of linearity. However, we formally tested for a deviation of linearity by including categorical dummy variables for each region in the regression. In none of the cases where we report a slope did a categorical predictor show a significant effect, which would indicate that there would be a non-linear component. Thus, the simpler model with one linear trend is preferred.

2.3 Results

We recorded the activity of single neurons in three areas of awake rats, namely V1, LI, and TO (Vermaercke et al., 2014), while presenting natural movies containing a rat or not as well as scrambled movies. The recordings yielded 50 (out of 58, or 86%) responsive single units for V1, 53 (out of 88, or 60%) for LI, and 52 (out of 84, or 62%) for TO, as well as 25 (out of 25, or 100%), 33 (out of 35, or 94%), and 26 (out of 30, or 87%) responsive multi-unit sites for each area respectively (percentages indicate the proportion of units that passed the inclusion criteria for analysis, i.e. an average net response of more than 2 Hz for at least one movie). We tested for a) a change in representation of stimulus type (natural versus scrambled) across areas, supporting a functional hierarchy, b) the emergence of a categorical representation for the distinction between rat and non-rat movies, and c) the emergence of a preference for natural movies along these areas.

2.3.1 Stimulus representations in V1, LI, and TO

To get an idea of how the stimuli are represented by the populations of neurons, we investigated the neural stimulus dissimilarities in the *N*-dimensional representational space defined by the average response of *N* neurons to the individual stimuli (**Figure 2.2**). Two stimuli that elicit a very different response in the population of recorded neurons will result in a higher dissimilarity value. On the other hand, if a population of neurons shows the same response pattern to two different stimuli, the dissimilarity value will be zero. These dissimilarity values can be visualized in a dissimilarity matrix as in Figure4A (top row), where more yellow colors indicate higher pair-wise stimulus dissimilarity. Visual inspection of the dissimilarity matrices suggests that moving from V1 to TO, a pattern emerges which can be summarized as an increased structuring by quadrants in the matrix: between stimulus type (i.e. natural or scrambled) dissimilarities increase relative to within stimulus type dissimilarities, which leads to an increased dissociation of natural versus scrambled movies. This is also illustrated by the plots on the lower row of **Figure 2.4**A, where the similarity representations are visualized in two-dimensional space after performing non-metric multidimensional scaling (MDS) on each dissimilarity matrix (using the function mdscale in Matlab, The MathWorks, Inc., Natick, MA, with the number of dimensions set to 2 and criterion set to 'stress'). These plots show an increased separation between natural movies versus scrambled movies. This increased separation is also supported by further statistical analyses. The difference of average between stimulus type and average within stimulus type dissimilarities increases per area (**Figure 2.4**B; ordinary least squares, OLS, slope .16, 95% CI [.06 .26], p = .003), with a value of .20 (95% CI [.11 .32], p < .001) for V1 neurons, .39 (95% CI [.25 .54], p < .001) for LI neurons, and .53 (95% CI [.36 .70], p < .001) for TO neurons. Thus, the distinction between natural and scrambled movies becomes more dominant in the neural representation when we move up in the cortical hierarchy.

In order to relate single cell responses to this population effect we plotted the standardized (per neuron) responses of each neuron to each stimulus that were used to create the dissimilarity matrices (**Figure 2.4**D). Here we see that the curve showing average natural minus scrambled responses per neuron (to the right of each heatmap) is generally shifted to the right for LI compared to V1. This means that the distribution of a natural versus scrambled comparison shifts in favor of natural stimuli from V1 to LI causing more neurons to respond more to natural than to scrambled stimuli. For TO however, this curve has moved to the right nearly only for neurons responding more to natural stimuli. Thus, in TO the proportion of neurons responding more to natural stimuli is not necessarily different compared to LI, but the natural/scrambled difference is higher for those that do respond stronger to natural movies.

Is this increased sensitivity for natural versus scrambled movies accompanied by an increase in category selectivity, that is, a differentiation between movies that depict a rat versus movies without rat? This would result in a similar "structuring by quadrants" as described in the previous paragraph, but now within the left upper quadrant of the dissimilarity matrix.

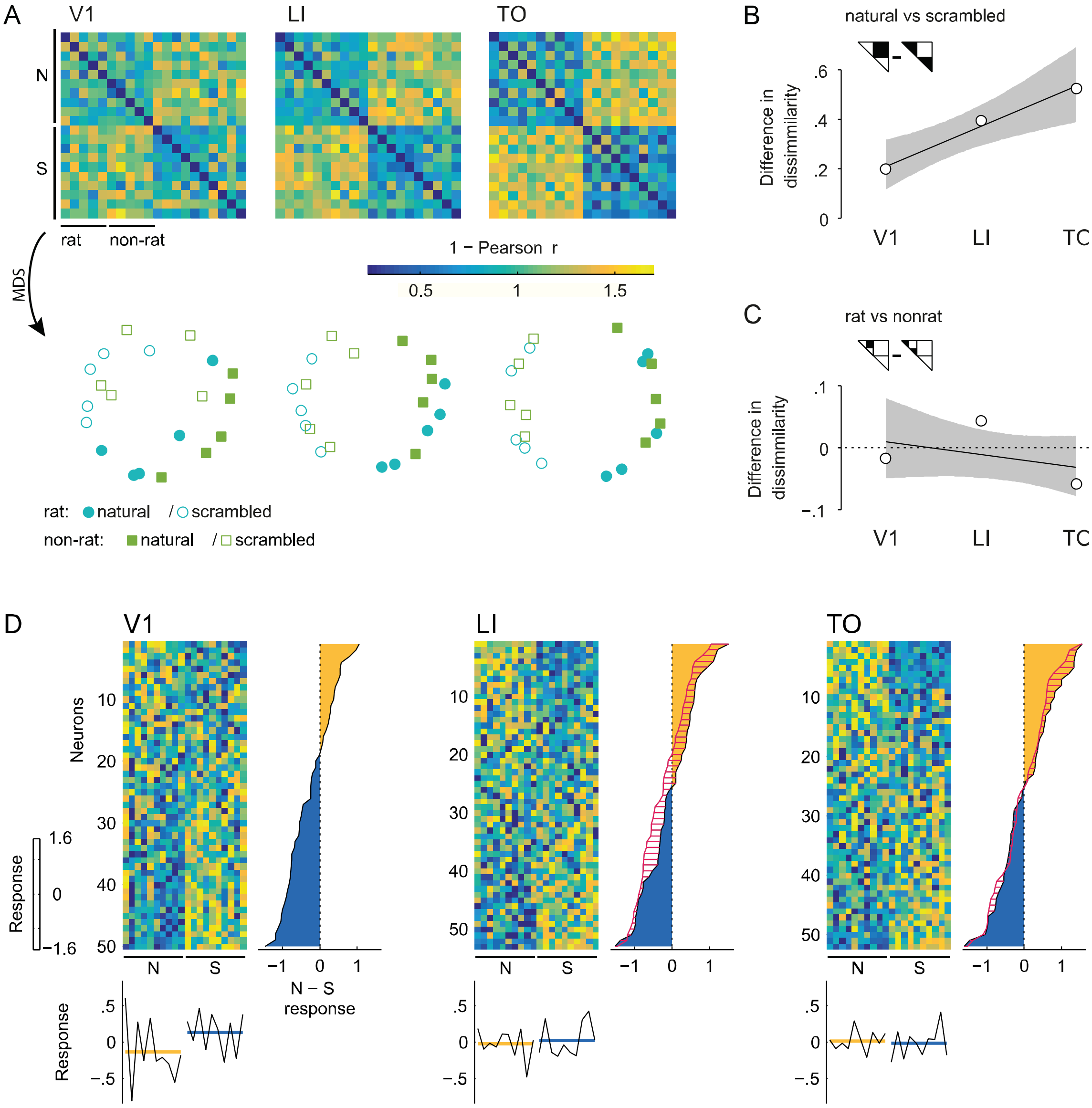


Figure 2.4. Stimulus representations based on responses averaged across movie durations. Panel **(A)** shows the stimulus dissimilarity matrices based on the correlation distance for each population of neurons recorded in V1, LI, and TO (upper row). Non-metric MDS is then used to represent the representational space in two dimensions (lower row). Panel **(B)** shows the difference between dissimilarities for stimulus pairs of the opposite stimulus type (i.e. natural versus scrambled) and dissimilarities for pairs of the same type in areas V1, LI, and TO. Grey area indicates 95% confidence bounds for OLS regression, calculated by BCa. Panel **(C)** is the same as panel B, but for the difference between dissimilarities for stimulus pairs of the opposite stimulus category (i.e. rat versus non-rat) and dissimilarities for pairs of the same category. Panel **(D)** contains heatmaps (one per area) displaying the average response to each movie standardized per neuron (Z-score across stimuli). Neurons (rows) are sorted in descending order according to the values of the average standardized response to natural movies minus that to their scrambled versions. To the right of each heatmap the average of this value used for sorting is plotted per neuron, with the yellow area indicating neurons that respond more to natural images than to their scrambled version and the blue area indicating the reverse. For LI and TO, red hatching indicates how this distribution changed from V1 and LI respectively. Stimuli (columns) are sorted in the same way as in the dissimilarity matrices: five natural rat movies, five natural non-rat movies, and their scrambled versions in the same order. Column averages are displayed below each heatmap (black lines), with the average across stimulus type (yellow for natural movies, blue for their scrambled version) indicated in color.

Visual inspection does not suggest that this pattern exists in the representational space for the populations of neurons recorded in either V1, LI, or TO. The MDS plots suggest an overlap in representations for rat versus non-rat movies without a clear separation. We performed further statistical analysis where we compared the average within stimulus category dissimilarities with average between stimulus category dissimilarities. The results do not show an emerging trend that would support an increased separation between representations of rat movies and those of non-rat movies (OLS slope -.02, 95% CI [-.06 .02], p = .432). Looking at each area separately, the difference in dissimilarity is -.02 (95% CI [-.08 .06], p = .696) for V1 neurons, .04 (95% CI [-.03 .13], p = .269) for LI neurons, and -.06 (95% CI [-.10 -.01], p = .148) for TO neurons. Positive values signify higher within category similarity than between category similarity, which is what one would expect in the case of a categorical representation. None of the areas shows such a categorical representation. To further strengthen these findings, we performed a potentially more sensitive population decoding analysis using support vector machines as a linear classifier. In agreement with the other analyses discussed above, the results of the linear classifier reveal no evidence for a categorical representation (Supplementary Material, Population Decoding Analysis).

As described previously (Vermaercke et al., 2014), moving from V1 to LI and to TO is characterized by systematic changes in retinotopic location and size of the receptive field. Furthermore, we might sample neurons with different receptive field properties in the three areas. This could mean that there are systematic changes in the area of the stimulus covered by our recorded samples of neurons across areas. However, a control analysis using average local stimulus statistics for each neuron’s receptive field shows that this confound cannot account for the increased natural/scrambled distinction from V1 to TO. In addition, there is no evidence that an emergence of a categorical rat/non-rat distinction could be hidden by such a confound (Supplementary Material, Receptive Field Confound).

2.3.2 Response statistics for natural and scrambled movies: Mean, sparseness, and reliability

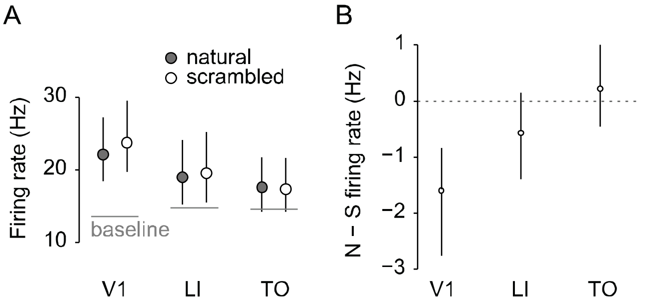


Figure 2.5. Firing rates per area.
Panel **(A)** shows average (across trials, stimuli, and neurons) response strength to natural (grey markers) and scrambled (white markers) movies per area, with the average baseline firing rate indicated by a horizontal line. Error bars indicate the 95% CI's calculated by BCa. Panel **(B)** shows the average difference in response strength to natural and scrambled movies per area. Negative values indicate a stronger response to scrambled versions of the movies. Error bars indicate the 95% CI's calculated by BCa.

Next, we looked at the average firing rates. **Figure 2.5**A shows the average firing rate for natural and scrambled movies (first averaged per neuron across movies for statistical analysis) for each area. Overall there is a statistically non-significant decrease in firing rate (OLS slope -2.7, 95% CI [-5.7 0.1], p = .083) moving from V1 to TO, while the average baseline firing rate does not seem to vary. The difference in response to natural movies versus scrambled versions is negative for V1 (see **Figure 2.5**B) and this difference disappears towards the other areas, or, quantitatively, decreases significantly (OLS slope .9, 95% CI [.4 1.5], p = .003). The average difference is -1.6 for V1 (95% CI [-2.8 -.8]), -.6 for LI (95% CI [-1.4 .2]), and .2 for TO (95% CI [-.4 1]). Similar results are obtained when the difference in firing rate is first divided by the average firing rate per neuron (OLS slope .04, 95% CI [.01 .08], p = .020), with an average difference of -.06 for V1 (95% CI [-.10 -.02]), .01 for LI (95% CI [-.03 .06]), and .02 for TO (95% CI [-.02 .09]).

Control analyses show that these differences in firing rate between natural movies and their scrambled versions cannot be explained by differences in location and size of receptive fields (Supplementary Material, Receptive Field Confound). The reason why V1 neurons would prefer scrambled movies is further explored in a later section. For inference per neuron, BCa 95% CIs on the average difference in response to natural movies versus their scrambled version were calculated for each unit by means of random sampling of natural/scrambled stimulus pairs. We decided a unit prefers natural stimuli when this 95% CI excludes zero. This criterion indicates that 4% (95% CI [0 14], based on the binomial distribution) of the units in V1 prefer natural stimuli, 23% (95% CI [12 36]) in LI, and 27% (95% CI [16 41]) in TO. Scrambled stimuli are preferred by 32% (95% CI [20 47]) of the units in V1, 19% (95% CI [9 32]) in LI, and 29% (95% CI [17 43]) in TO. We conclude that, while there is an increase in the percentage of units consistently responding more to natural stimuli from V1 to TO, no clear change is evident for the percentage of units consistently responding more to scrambled stimuli. Applying the same criterion, rat stimuli are preferred by 18% (95% CI [9 31]) of the units in V1, 6% (95% CI [1 16]) in LI, and 13% (95% CI [6 26]) in TO. Non-rat stimuli finally are preferred by 2% (95% CI [0 11]) of the units in V1, 11% (95% CI [4 23]) in LI, and 0% (95% CI [0 7]) in TO. In general, it seems that a higher percentage of units tend to consistently respond more to rat than to non-rat movies. However, since this is clearest for V1 neurons and since these percentages do not change progressively across areas, we conclude that this is most likely a result of lower level stimulus properties that V1 neurons typically respond to.

Next, we investigated the variation in responsiveness. Natural stimulation has been shown to increase the sparseness of the neural response (Vinje and Gallant, 2000). We looked at the sparseness (see materials and methods, data analysis, sparseness and reliability) of each neuron's response to natural movies and to their scrambled counterparts. In mouse V1, sparseness has been shown to be higher for responses to natural movies than to their scrambled counterparts (Froudarakis et al., 2014). We confirm this finding for single units in rat V1, with a difference in sparseness index of .035 (95% CI [.027 .047], $p < .001$; positive values mean higher response sparseness to natural movies; see **Figure 2.6**A). Also in LI and TO we find a higher sparseness for natural movies, with a difference of .028 (95% CI [.014 .042], $p < .001$) and .015 (95% CI [.004 .025], $p = .011$), respectively.

Importantly local luminance based stimulus sparseness calculated for each neuron's receptive field cannot explain the difference in response sparseness (Supplementary Material, Receptive Field Confound). However, if lower firing rates would tend to get higher sparseness index values and vice versa, some of these differences might be explained by the differences in firing rates shown before, in particular in area V1. Indeed, differences in firing rates are negatively correlated with differences in sparseness

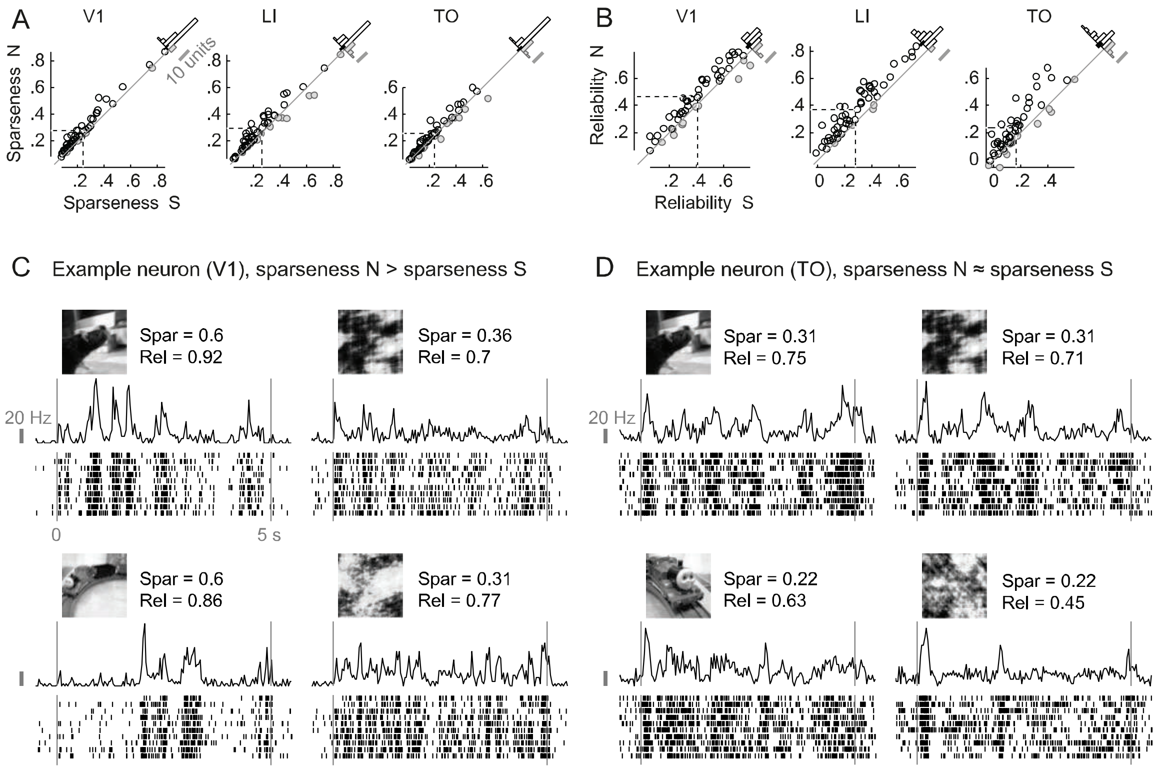


Figure 2.6. Sparseness and reliability of single neuron's responses.
Panel **(A)** contains scatterplots of the sparseness index for natural (N) compared to the same index for scrambled (S) movies for all neurons recorded in V1 (left), LI (middle), and TO (right). Neurons with a lower index for natural movies are greyed out. Dashed lines indicate the means. Histograms of the difference between natural and scrambled stimuli are shown in the top right corner of each plot, with the 95% CI (calculated by means of BCa) of the mean indicated by a black bar. Panel **(B)** contains the same figures, but for the reliability coefficient. Panel **(C)** shows raster plots for an example neuron with relatively high response reliability of responses to two natural movies and their scrambled versions. For this example, response sparseness is much higher for the original stimuli compared to their scrambled version. Panel **(D)** shows raster plots for another example neuron. In this case response sparseness indices are equal for the two stimulus types.

index for each area, with a Pearson correlation of -.28 in V1 (95% BC$_a$ interval [-.50 -.001], p = .050), -.54 in LI (95% CI [-.69 -.35], p < .001), and -.43 in TO (95% CI [-.67 -.02], p = .002). Thus, we controlled for differences in firing rates by taking for each cortical area the 30 units with a difference in firing rates evenly distributed around zero. For these units, the average difference in sparseness indices was .025 in V1 (95% CI [.017 .037], p < .001), .024 in LI (95% CI [.011 .037], p < .001), and .011 in TO (95% CI [.004 .019], p = .010). Thus, responses to natural movies show a higher sparseness than responses to scrambled movies, even when we control for overall responsiveness.

Finally, responses to natural movies are decisively more reliable in all three areas (**Figure 2.6**B). Reliability of V1 neural responses is on average .056 (95% CI [.037 .072], p < .001) higher to natural movies than to scrambled versions. For LI and TO neurons this difference is on average .090 (95% CI [.074 .108], p < .001) and .070 (95% CI [.047 .097], p < .001), respectively. In the case of reliability, we have relatively weak evidence of a possible influence of differences in response strength: the Pearson correlations between the difference in standardized reliability and the difference in standardized firing rates are .26 in V1 (95% CI [-.04 .47], p = .070), .12 in LI (95% CI [-.08 .35], p = .385), and .40 in TO (95% CI [-.040 .64], p = .003). The results for the 30 units selected to match for response strength (see paragraph above) of each cortical area are qualitatively similar to the results for the whole sample, with a difference of .059 in V1 (95% CI [.034 .081], p < .001), .095 in LI (95% CI [.072 .120], p < .001), and .052 in TO (95% CI [.029 .077], p < .001).

2.3.3 Using a V1 model to explain the preference for scrambled stimuli and to predict neural responses

To further investigate these findings, we used simulated V1 responses to see (a) if these simulated responses can predict observed responses and, if the response to (a) is affirmative, (b) if these simulated responses can explain the relative increase of the response for natural movies and the increased segregation between natural and scrambled movies. In the model spatiotemporal motion energy filters (Adelson and Bergen, 1985) are used as modeled V1 complex cells for a set of spatiotemporal frequencies, orientations, and spatial locations. The output of these filters can be used to estimate how strong responses in V1 would be to one stimulus relative to another one. Furthermore, the filters can be used in a model that is fitted to part of the data in order predict independent test data (Nishimoto and Gallant, 2011). Note that the V1-like filters are linearly combined to predict the neural responses, which is why the model might even capture the responses of neurons in higher visual areas to the extent that the complexity of their computations can be approximated by a linear combination of V1-like filters.

Predicting neural responses

The standardized output of each V1 filter was used as a predictor in a regularized linear regression model that was fit to the neural data, resulting in a spatiotemporal receptive field estimate consisting of a linear combination of these filters. This receptive field estimate can then be used to predict the response to a new stimulus. If this predicted response captures a certain amount of variability in the response to movies that were not used for fitting the model, then the fitted receptive field can explain some of the response properties of the neuron. For this test we also included the multi-unit data, since their firing rate is the sum (i.e. a linear combination) of the firing rates of the single units contained in the multi-unit cluster which can be accounted for because the model can use a linear combination of outputs of modeled single cells.

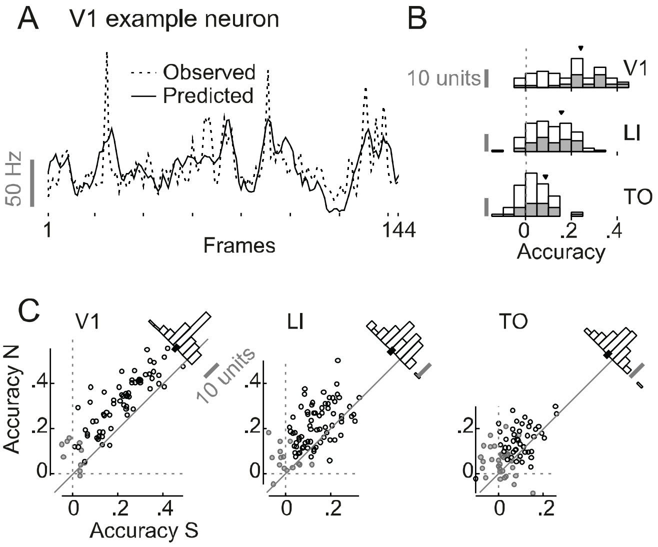


Figure 2.7. Performance of the V1-like motion energy model.
Panel **(A)** shows the observed (dashed line) and predicted (full line) response of a V1 example neuron to each frame of a natural movie (note that the number of frames is 144, because the first six were omitted to get rid of the onset peak). Panel **(B)** contains stacked histograms for the prediction accuracy (Pearson correlation) averaged across all movies for single neurons (white) and multi-unit clusters (grey) recorded in V1, LI, and TO. Panel **(C)** contains scatterplots of average (across movies) accuracy for natural versus scrambled movies for single and multi-units. Histograms of the difference between natural and scrambled stimuli are shown in the top right corner of each plot, with the 95% CI (calculated by means of BCa) of the mean indicated by a black bar. Greyed out markers represent units for which the 95% CI of the average accuracy as calculated by BCa (resampling all of the 20 stimuli) does include zero.

The observed and predicted responses of a V1 example neuron on one natural movie are illustrated in **Figure 2.7**A. Histograms with prediction accuracy (i.e. Pearson correlation between observed and predicted response) averaged across all stimuli are shown in **Figure 2.7**B. The model performs reasonably well for V1 units, with an average prediction accuracy of .24 (95% CI [.21 .27], randomization test for difference from zero p < .001). To put this number in perspective, this is lower than the prediction accuracy of .52 reported by Nishimoto and Gallant (2011) in monkeys. This was to be expected since the number of frames (data) we used to train the model (2736) is one order of magnitude lower (on average 27120 in their case) and since we used natural movies and not motion enhanced movies. Prediction accuracy is lower but still significantly different from zero for LI (.16, 95% CI [.14 .18], p < .001) and even lower but still significantly different from zero for TO (.09, 95% CI [.07 .10], p < .001). In general, accuracy decreases across areas (OLS slope -.08, 95% CI [-.09 -.06], p < .001). Even when including reliability as a covariate accuracy decreases across areas (OLS slope -.05, 95% CI [-.06 -.03], p < .001). Dividing prediction accuracy by reliability gives us an estimate of the proportion obtained prediction accuracy out of the total possible prediction accuracy. When we do this per neuron and movie, we get an average of .46 (95% CI [.41 .51], randomization test for difference from zero p < .001) for V1, .34 (95% CI [.30 .38], p < .001) for LI, and .20 (95% CI [.16 .24], p < .001) for TO. This amounts to a decrease across areas (OLS slope -.13, 95% CI [-.16 -.10], p < .001). Thus, the aforementioned decrease in response reliability cannot explain the decrease in the performance of the model from V1 to TO.

Focusing on the performance for single neurons and multi-unit clusters, prediction accuracy averaged across all 20 stimuli (natural and scrambled) is significantly different from zero (i.e. the 95% CI – calculated by resampling the stimulus labels – excludes zero), for 85.3% of the units in V1, 82.5% of the units in LI, and 56.4% of the units in TO.

To get an estimate of receptive field size, we estimated the percentage of pixels in the area covered by the movie that modulate the neural response. Specifically, we used the median regression weights of the V1-like filters (across the 20 training sets used for cross-validation) in order to estimate the spatial receptive field. Pixels that were estimated to modulate the response with a magnitude less than 50% of that of the pixel that maximally modulated the response were excluded, to ignore the pixels that contribute relatively nothing. Based on this approach, V1 neurons were estimated to be modulated on average by 8% of the movies' pixels (95% CI [7% 10%]), LI neurons by 14% (95% CI [12% 17%]), and TO neurons by 16% (95% CI [13% 20%]). This means that for LI neurons this movie frame coverage was 6% higher than for V1 neurons (95% CI [3% 9%], p < .001) and for TO neurons it was 8% higher than for V1 neurons (95% CI [5% 12%], p <

.001). The difference in coverage between TO and LI neurons was 2% (95% CI [-2 % 7%], p = .294).

In sum, the simulated V1 model allows us to predict neural responses in each of the investigated areas, and reveals several differences between the areas which can be expected given their position in the cortical hierarchy: the model works better for V1 than for the other areas, and the estimated receptive field size is smallest in V1. Thus, the model can be used to further investigate potential differences in how these three neuronal populations respond to natural and scrambled movies.

Preference for scrambled stimuli in V1 filters

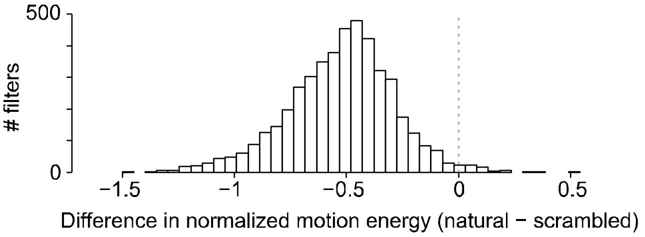


Figure 2.8. Stimulus type preference in output of the V1-like motion energy model.
Negative values indicate filter output for scrambled movies is on average higher than filter output for natural movies.

To investigate the preference for one stimulus type over another, we calculated the output of the V1-like spatiotemporal motion energy model, thus before combining the V1 filters into a spatiotemporal receptive field estimate. For the vast majority of modeled V1 filters, the overall (across time) response to natural movies is lower than that to their scrambled versions. **Figure 2.8** contains a histogram depicting the standardized response to scrambled movies subtracted from the response to their natural counterparts for all filters in the model (in this case 4616), with 98.6% of them preferring scrambled movies. This means that scrambled movies do seem to contain relatively more motion energy in virtually all spatiotemporal frequencies regardless of orientation and location. Other scrambling methods, such as segment/box scrambling (i.e. random repositioning of rectangular image segments), give qualitatively the same result (data not shown). This characteristic of scrambled stimuli has been reported before in the context of scrambling of still images using various methods including phase scrambling (Stojanoski and Cusack, 2014) and in an fMRI experiment using phase-scrambled movies (Fraedrich et al., 2010). The higher amount of motion energy in scrambled movies can explain the neural preference for scrambled stimuli especially prominent in the V1 data.

Prediction accuracy on natural versus scrambled movies

Scatter plots of average prediction accuracy on natural versus that on scrambled movies for each unit are shown in **Figure 2.7**C. For all three cortical areas, accuracy was higher for natural compared to scrambled movies: the difference in accuracy for natural minus that for scrambled was .10 for V1 (95% CI [.09 .12], p < .001), .07 for LI (95% CI [.06 .09], p < .001), and .05 for TO (95% CI [.03 .06], p <.001). Given the earlier finding that responses to natural movies are more reliable than responses to scrambled movies, this difference in prediction accuracy might be caused by the difference in reliability. Including reliability as a covariate still resulted in an estimated higher accuracy for natural movies of .07 for V1 (OLS estimate, 95% CI [.06 .09], p <.001), .05 for LI (OLS estimate, 95% CI [.03 .09], p = .003), and .05 for TO (OLS estimate, 95% CI [.03 .07], p <.001). Thus, the effect of scrambling on prediction accuracy does not seem to be the result of differences in response reliability.

2.4 Discussion

We investigated the neural responses to natural movies and their phase scrambled versions in rat V1 and two extrastriate visual areas LI and TO, which belong to a distinct pathway reminiscent of the primate ventral visual stream (Vermaercke et al., 2014).

First, we found an increased clustering of natural versus scrambled movie representations when progressing from V1 to TO. The increased dissociation of the two stimulus types correlates with a decreased overall preference for scrambled stimuli in spite of the stronger motion energy contained in scrambled movies. A closer look at single cell preferences suggests that the population effect is driven by the increase in the proportion of cells preferring natural stimuli and by an increase in strength of preference for those neurons that prefer natural stimuli.

Second, unlike what one would expect to see in an object representation pathway such as the primate ventral visual stream (Orban, 2008), the population representations of the stimulus set do not culminate into a higher level categorical representation in area LI or in the most downstream area TO. Of course, we are restricted in making strong claims about this by our small stimulus set and limited amount of neurons per area. However, the neurons used for our analyses were all responsive to at least one stimulus and did show selectivity, indicating that they did encode information. Furthermore, as far as we can judge from the available primate literature, the distinction between animate and non-animate stimuli in the monkey and human brain is very clear. For example, the matrices from monkey and human data as shown by Kriegeskorte et al. (2008b, Fig. 1) suggest that this animate/inanimate distinction in primates is at least as clear as the distinction between natural and scrambled in our rat data. Given that we can easily pick up the natural vs scrambled distinction in our rat data, we think we should be able to pick up a similarly sized effect of rat (animate) vs nonrat (inanimate). An effect of this size does not seem to be present for our stimuli in the recorded neuronal populations. Nonetheless, we cannot exclude that factors we did not control for, such as attention or rather the lack thereof, might have influenced such findings. Nor can we exclude the possibility that other areas in the rat brain would show such category selectivity. However, it is not very obvious which other areas would do so.

Finally, a V1-like model that has previously been used to model receptive field properties of neurons (Nishimoto and Gallant, 2011; Talebi and Baker, 2012) as well as voxels in human brain imaging (Nishimoto et al., 2011) could predict responses of V1 neurons reasonably well, especially when we take into consideration that our experiment was not optimized for fitting such a model. Similar to what is reported in humans when comparing primary visual cortex with extrastriate areas (Nishimoto et al., 2011), prediction accuracy was reduced in LI and even more so in TO, suggesting that such a model was progressively less able to capture response properties of areas further along this visual stream. Of course, this is only one model that will not capture all possible tuning properties of V1 neurons, therefore caution should be taken in drawing strong conclusions from this piece of evidence alone.

The sampling bias of upper cortical layers for the V1 recordings could explain (some of) the differences that we observe between recordings in V1 on the one hand and recordings in LI and TO on the other hand. A previous systematic comparison between responses in upper and lower layers of V1 did not show consistent differences (Vermaercke et al., 2014). Likewise, Froudarakis et al. (2014) report no differences between responses to natural movies in V1 layer 2/3 and V1 layer 4. Moreover, all the changes across the succession of areas that we do describe continue in the same direction from LI to TO, where there is no difference in layer sampling bias. Thus, we argue that a sampling bias of upper layers in V1 does not invalidate conclusions of gradual changes across the visual processing pathway under investigation.

Together, these findings support the idea of a functional hierarchy in these areas. The data suggest that an increasing number of neurons are driven by more complex stimulus features that are not captured by V1-like filters and destroyed by a phase-scrambling method. Indeed, a linear combination of V1-like receptive fields decreases in efficiency in predicting neural responses the further up this hierarchy. Nevertheless, the functional hierarchy does not seem to culminate in neural representations as found in primates.

2.4.1 COMPARISON WITH PREVIOUS RESEARCH ON RODENT EXTRASTRIATE VISUAL CORTEX

Previous research on the functional properties of rodent extrastriate areas LI and/or TO have only used drifting gratings (Marshel et al., 2011; Vermaercke et al., 2014) or simplistic artificial stimuli (Vermaercke et al., 2014).

The current study is the first that allows an investigation of selectivity in more downstream visual cortex for the complex features present during natural stimulation. The increased response to natural movies relative to scrambled movies and the decreased performance of a V1-like energy model significantly extend the earlier findings of position invariance and simple shape representations, and support the notion of an increasingly high-level stimulus representation when progressing from V1 to TO.

Nevertheless, we could not find any evidence for a category selective representation in rat extrastriate cortex and to the best of our knowledge there is no other neural data supporting this notion. In a recent study, rats could be trained in a two alternative forced choice task to discriminate rat movies from non-rat movies and could generalize to new previously unseen exemplars (Vinken et al., 2014). However, the training was difficult and took a substantial amount of time, which is consistent with the absence of a categorical representation in naïve animals.

2.4.2 Comparison with previous experiments using natural stimuli

Previous studies using stationary natural and scrambled images indicate a preference for natural images in responses of human lateral occipital complex (Grill-Spector et al., 1998) and monkey inferior temporal cortex (Vogels, 1999c; Rainer et al., 2002). In areas earlier in the ventral visual processing hierarchy responses have been reported to show a preference for scrambled images in V1 that disappears in extrastriate visual areas (Rainer et al., 2002). The current study included a method of scrambling which falls within the range of methods used previously. Recent studies zoomed in on this general difference between intact and scrambled images by including specific methods of scrambling and focusing upon particular characteristics of natural images. For example, Freeman and colleagues showed in monkeys and in humans that responses in V2 were stronger for naturalistic textures than for spectrally matched noise, while they were the same in magnitude in V1 (Freeman et al., 2013).

In the current study, the change in preference for natural and scrambled is similar to all this earlier work when the preference is expressed in relative terms: more preference for natural movies when moving away from V1. We find a stronger firing rate in response to scrambled movies compared to their original natural counterparts in V1 and this difference decreases in extrastriate area LI and ends in an equal firing rate for both stimulus types in TO. Similarly, an fMRI study in humans has reported stronger early visual cortex activity to spatiotemporally phase-scrambled movies relative to their original version (Fraedrich et al., 2010). This result is supported by the motion energy model we used to show an increased output for scrambled movies when passed through a bank of V1-like filters, as well as by previous modeling studies using still images (Stojanoski and Cusack, 2014). In the study by Freeman et al. (2013) the reported equal firing rate for natural and scrambled images in V1 might be the result of control stimuli that are more carefully matched in spectral properties than is allowed by our spatiotemporal phase scrambling of the movies. These previous reports related to phase-scrambling combined with our own modeling results indicate a parallel between our experimental results and the earlier findings in primates: namely, a preference for scrambled stimuli that disappears in extrastriate cortex (Rainer et al., 2002). Similar to the present study, Rainer et al. (2002) used a scrambling method that introduced distortions that the earlier visual system is sensitive to (Stojanoski and Cusack, 2014). However, in the present study this gradual change in stimulus type preference did not culminate in the higher response to natural stimuli that is observed in monkey inferior temporal cortex (Rainer et al., 2002) and human lateral occipital complex (Grill-Spector et al., 1998). This means that inasmuch as the succession of areas where we recorded can be compared with the primate ventral visual stream, we could not find support for a preference for natural stimuli typical of these primate higher visual areas, and find a resemblance with more mid-level areas, at best. Another recent study in rodents reports a stronger response to phase scrambled movies in mouse V1 when the animal was sitting still and not whisking (Froudarakis et al., 2014), which is consistent with our findings. On the other hand, their recordings when the animals were whisking and/or running as well as their recordings in anaesthetized animals showed an equal response to natural movies and their phase scrambled controls. These findings suggest that brain/behavioral state interacts with the effect of phase scrambling. Our rats were awake and passively viewing the stimuli during recordings, but we did not monitor behavioral cues such as whisking, so we cannot control for this. Overall, the animals tended to be sitting very still during the recordings. We speculate that behavioral state might increase the sensitivity for natural stimulation in rodent visual cortex overall, which then overcomes the difference in motion energy in V1. Several other comparisons in our report between natural and scrambled movies were also included in the investigation of V1 by Froudarakis et al. (Froudarakis et al., 2014), and for those indices the results tend to be consistent between the two studies. More specifically, we replicate a higher sparseness and more reliable responses for natural movies.

2.4.3 Implications for the rodent as a model for object vision

What are the implications of our findings on the idea of the rodent as a model for object vision? The rodent has become a popular model in the neuroscience community for tackling questions on the topic of higher-level and object vision (Glickfeld et al., 2014; Cooke and Bear, 2015). A large and consistent body of evidence exists in the behavioral literature revealing encouraging visual capabilities in rats (Zoccolan, 2015). However, while steps have been taken to show the existence of two anatomically and functionally distinct streams in mouse visual cortex (Andermann et al., 2011; Marshel et al., 2011; Wang et al., 2012), it remains an open question whether and to what extent the rodent ventral visual stream can be considered as homologous to that of the primate. Here we show that the proposed homologue of the rat ventral visual stream may not show certain properties to the same extent as the primate ventral visual stream, such as higher responses to natural images, and even lack defining properties like a categorical representation. This story parallels previous findings that show both typical (an increase in tolerance for stimulus position), as well as atypical (an increased response to moving stimuli) properties of the pathway (Vermaercke et al., 2014). Together, Vermaercke's (2014) and our results show that we should be cautious in assuming functional similarities in visual processing between rodents and primates. Perhaps we should reconsider the concept of a ventral visual stream tuned for object recognition in rats and mice. After all, these are non-foveal animals, with a very low visual acuity (Prusky et al., 2000), that might rely so much on their other senses for object recognition in natural situations that they lack the functional specialization in visual cortex. The situation is complicated further by the finer differentiation of this ventral stream in primates into multiple pathways (Kravitz et al., 2013), and it is unclear which pathway(s) might be present in rodents, if any.

2.4.4 Conclusion

We recorded neural responses in areas belonging to a proposed rodent homologue of the primate ventral visual stream in order to investigate two hallmarks of high-level representations in primates: preference for intact versus scrambled stimuli and category-selective responses. We found that our results parallel changes in response strength to natural versus scrambled stimuli from primate primary visual cortex to early extrastriate visual areas. However, unlike in primate ventral visual stream, in our results we failed to find a preference for natural stimuli in most temporal visual area TO, nor did the targeted pathway lead to category-selective representations.

2.5 Supplemental information

2.5.1 Receptive field confound

Methods

The receptive field properties (i.e. size and location) are expected to change across areas (Vermaercke et al., 2014). Furthermore, the distribution of receptive field positions is not guaranteed to be the same in the sampled neurons from different areas. As a consequence, the part of the stimulus covered by receptive fields – and therefore the local stimulus properties – might change systematically. This could then cause changes in response properties that are not related to changes in actual functional properties across these areas. In order to assess the role of this confound, local stimulus statistics were calculated for each neuron, using only the area of the stimulus covered by the receptive field as estimated by a separate receptive field mapping experiment (see materials and methods, electrophysiological recordings, receptive field mapping). For this experiment, the screen area was divided in 3 × 5 square screen locations that could drive a neuron, out of which 3 × 3 locations overlap with the presented square stimulus. Thus, for each stimulus and per frame we calculated local stimulus statistics for each of those 9 locations: mean pixel value, root-mean-square contrast, mean absolute pixel change (between successive frames), skewness, and kurtosis. For each neuron, the corresponding stimulus properties were then defined as the average of the statistic in question across responsive locations as determined by the receptive field mapping experiment. Those statistics can then be used instead of the actual responses to calculate sparseness, dissimilarity matrices, etc. using exactly the same methods as for the neural data. As a criterion for responsiveness at a certain screen position, a Wilcoxon signed rank test was performed with a threshold of p = .05 (across hashtag stimulus presentations and per screen position).

Results

Stimulus representations in V1, LI, and TO.

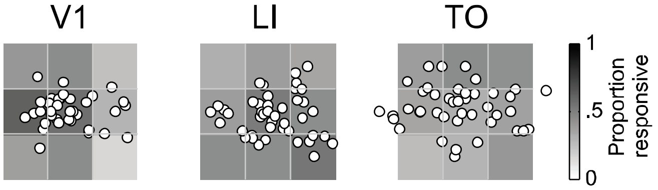


Figure 2.9. Stimulus coverage by receptive fields. White dots indicate the centers of gravity of the responsive screen positions as determined by the receptive field mapping experiment. Their position is calculated by taking the average responsive screen location (weighted by the net responses) and plotted relative to the square stimulus. Note that for four TO neurons that center of gravity fell just besides the stimulus, but the bigger receptive field size ensured overlap of the stimulus and receptive field. The greymaps indicate the proportion of neurons that have that part of the stimulus within their receptive field. Note that this stimulus coverage is more homogeneously distributed for LI and TO neurons.

Figure 2.9 illustrates the stimulus coverage of each neural population sample. A control analysis was performed to see whether systematic changes in receptive fields could underlie the change in population representation across areas. For this we used each neuron's average local stimulus statistics (i.e. the stimulus properties within the neurons receptive field) and used these to make dissimilarity matrices. Thus, the analysis is exactly the same, only now the values of the average local stimulus statistics are used for each neuron instead of firing rate. If a systematic change in receptive fields would underlie our finding of a systematic increase in natural versus scrambled clustering of representations, we should find a systematic increase in the local receptive field based stimulus statistics. This results in the following natural versus scrambled dissimilarity values based on: (a) pixel values; V1 .03 (95% CI [.00 .07], p = .144), LI .04 (95% CI [.01 .08], p = .024), TO .02 (95% CI [.00 .05], p = .243), (b) RMS contrast; V1 .07 (95% CI [.01 .13], p = .002), LI .11 (95% CI [.05 .16], p < .001), TO .10 (95% CI [.05 .17], p < .001), (c) skewness; V1 -.06 (95% CI [-.08 -.04], p = .005), LI -.06 (95% CI [-.07 -.04], p = .003), TO -.06 (95% CI [-.07 -.04], p = .006), (d) kurtosis; V1 .09 (95% CI [.05 .13], p < .001), LI .06 (95% CI [.03 .09], p = .001), TO .07 (95% CI [.03 .11], p < .001), (e) pixel change; V1 -.04 (95% CI [-.06 -.01], p = .051), LI -.04 (95% CI [-.06 -.02], p = .019), TO -.03 (95% CI [-.05 .01], p = .181).

In general, scrambled and natural movies are locally more different in terms of pixel values, RMS contrast, and kurtosis compared to within scrambled or natural comparisons. They are locally even more similar in terms of skewness and pixel change. However, there is no systematic change across areas in terms of local stimulus statistics such as we see from V1 to TO in the actual data (see Figure 4B). We conclude that changes in retinotopic location and size of receptive fields cannot account for the increasing distinction between natural and scrambled movies in neural representations from V1 to TO.

As we did before for the natural versus scrambled test, we report the same rat versus non-rat contrasts based on each neuron's average local stimulus statistics. Given the absence of categorical distinctions in the neural data, we should mainly check whether maybe this absence is due to differences in the local stimulus statistics in the other direction than expected, that is, differences that are counterproductive for finding categorical distinctions. More specifically, this would be the case if the local stimulus statistics would give negative values. This results in the following rat versus nonrat dissimilarity values based on: (a) pixel values; V1 -.16 (95% CI [-.19 -.11], p < .001), LI -.12 (95% CI [-.17 -.06], p = .002), TO .12 (95% CI [-.17 -.07], p = .004), (b) RMS contrast; V1 -.09 (95% CI [-.14 -.03], p = .025), LI -.08 (95% CI [-.12 -.03], p = .057), TO -.11 (95% CI [-.15 -.06], p = .008), (c) skewness; V1 -.06 (95% CI [-.11 .00], p = .139), LI -.09 (95% CI [-.14 -.02], p = .024), TO -.05 (95% CI [-.11 .02], p = .237), (d) kurtosis; V1 .02 (95% CI [-.05 .09], p = .710), LI -.02 (95% CI [-.08 .04], p = .565), TO .05 (95% CI [-.01 .13], p = .179), (e) pixel change; V1 -.19 (95% CI [-.21 -.16], p < .001), LI -.15 (95% CI [-.18 -.11], p < .001), TO -.15 (95% CI [-.19 -.10], p < .001). These values are negative for pixel values, RMS contrast, skewness and pixel change, indicating that differences on these local statistics might hide a categorical distinction. However, these contrasts are relatively small in absolute value (the minimum is -.19) compared to the size of the natural versus scrambled effect (.20, .39, and .53, for V1, LI, and TO, respectively). In addition, each local stimulus statistic is highly similar across regions, meaning that if a difference in local stimulus statistics would hide a categorical representation, it could equally affect the dissimilarity values of each region. This means we would then expect a change in dissimilarity values in our data across regions, since a categorical representation is expected to emerge along the pathway and would not be present in V1 responses. The absence of this change in our data means that we have no evidence that an emergence of a categorical distinction is hidden by local stimulus statistics.

Single unit response statistics for natural and scrambled movies: Mean and sparseness.

The difference in various local stimulus statistics (averaged across responsive stimulus parts and frames) for natural movies versus scrambled versions is not correlated with the difference in response to natural movies versus scrambled versions, with a Pearson correlation (across neurons) of .03 for pixel values (95% CI [-.17 .16], p = .731), -.01 for RMS contrast (95% CI [-.13 .22], p = .876), -.03 for skewness (95% CI [-.18 .16], p = .743), -.01 for kurtosis (95% CI [-.21 .12], p = .907), .01 for pixel change (95% CI [-.08 .25], p = .921). Thus, differences in local stimulus statistics cannot explain differences in firing rate between natural movies and their scrambled versions.

The differences in sparseness indices are not positively correlated with differences in sparseness indices calculated on the local pixel values, with a Pearson correlation of -.16 for V1 (95% CI [-.35 .10], p = .269), -.07 for LI (95% CI [-.33 .35], p = .727), and -.07 for TO (95% CI [-.44 .17], p = .651). Thus, we conclude that a difference in local luminance based stimulus sparseness does not underlie the difference in response sparseness.

2.5.2 Population decoding analysis

Methods

In addition to the dissimilarity matrices, we conducted a population decoding analysis to test whether a classifier trained on a few stimuli to discriminate stimuli of a different type would generalize to independent test stimuli. Specifically, we trained a linear classifier (support vector machine, linear kernel, least squares method, and a C-parameter of 1) to discriminate between movie types (natural/scrambled, rat/non-rat, and scrambled-rat/scrambled-non-rat). Generalization performance of the classifier was assessed by means of cross-validation where correct classification was tested separately for two stimuli (one from each movie type) using only all other relevant stimuli for training. The support vector machine analysis was performed exhaustively for each possible training and test set combination to obtain an average performance. For example, the classification of one rat movie was tested five times by training the classifier on the remaining four rat movies and four non-rat movies, each time leaving another non-rat movie out to have a balanced training set. This was then done for each of the ten rat or non-rat movies, resulting in 10 × 5 rat/non-rat classifications. Performance of the classifier was calculated as proportion of correctly classified test stimuli (out of 20 × 10 for natural/scrambled and out of 10 × 5 for both rat/non-rat and scrambled-rat/scrambled-non-rat classification), ensuring a complete separation of training and test stimuli. The distribution of the performance under the null hypothesis was estimated by randomly shuffling responses across neurons per stimulus in order to calculate p-values. The distribution of the difference between rat/non-rat performance and scrambled-rat/scrambled-non-rat performance under the null hypothesis was estimated by randomly flipping natural and scrambled labels across neurons in order to calculate p-values.

Results

In this additional analysis, we trained a linear classifier to discriminate between movie types (natural/scrambled, rat/non-rat, and scrambled-rat/scrambled-non-rat; see materials and methods, data analysis, population decoding analysis). In case of natural/scrambled classification the ceiling of 100% correct becomes quickly apparent, with a performance of .89 (p = .002) for V1, 1 (p < .001) for LI, and 1 (p < .001) for TO. For the categorical rat/non-rat test performance is .5 (p = 1) for V1, .9 (p = .007) for LI, and .4 (p = .528) for TO. This seems to suggest a possible categorical population representation in LI responses. However, if the classifier uses categorical information, which is not preserved by phase scrambling, it should perform better on the original stimuli than on their scrambled versions. Therefore, the appropriate baseline for comparison in this case is the performance on the scrambled versions of those stimuli. For the scrambled-rat/scrambled-non-rat test performance is .48 (p = .954) for V1, .66 (p = .303) for LI, and .28 (p = .188) for TO. A pair-wise (i.e. taking into account natural-scrambled stimulus pairs) comparison with this baseline shows no evidence of a difference in rat/non-rat versus scrambled-rat/scrambled-non-rat classification performance, with a difference of .02 (p = .930) for V1, .24 (p = .320) for LI, and .12 (p = .419) for TO. We conclude that in agreement with the other analyses, the linear classifier reveals no evidence for a categorical representation.

Chapter 3.

A BRIDGE BETWEEN BEHAVIOR AND NEURONS

While rats can categorize novel natural movies, we had not found evidence for a categorical representation in their visual system. Unfortunately, a direct comparison between our neural and behavioral data is not possible, because we have no neural responses for any of the movies that the animals had to generalize to. Meanwhile, deep neural network models (DNN) had been developed that predict neural responses and categorization performance on the same stimulus set in monkeys with unprecedented accuracy (Kriegeskorte, 2015). Quantifying our natural movies with a DNN allows us to ask new questions that connect the neural and behavioral data. A) *what level of processing of the DNN is required to support the categorization experiment?* B) *what level of processing of the DNN do the neural representations in rat visual cortex correspond to?*

3.0 BACKGROUND

The brain is essentially a large neural network. Brain functions such as complex information processing are achieved through interactions between neurons, the computational units of a neural network. Neural network models are computational models consisting of a collection of interconnected units for which the activation is a weighted sum of their incoming inputs, passed through a nonlinear activation function. DNNs are models in which the units are organized in multiple layers between input data and the final output layer. When such models are trained to solve complex problems, their layers learn complex representations of the data with different levels of abstraction (LeCun et al., 2015). Convolutional neural networks are a special case of (deep) neural networks that are inspired by the visual system and feature convolutional and pooling layers. In particular, each unit in a convolutional layer only processes a local patch of the data (i.e. receptive field). These units are organized in feature maps. Units within the same feature map tile the input space and share the same input weights (i.e. they respond to the same feature at a different location). Mathematically, this operation is equivalent to a discrete convolution, hence the name. Pooling layers then combine the outputs of a number of units from the same feature map with neighboring receptive fields (typically by taking the maximum). Convolutional and pooling operations are directly inspired by the idea that complex cells combine (pool) the output of several simple cells with identical orientation preference but different receptive field locations (Hubel and Wiesel, 1962). Several stages of convolutional and pooling layers are typically followed by fully connected layers in which each neuron receives input from all neurons in the previous layer.

A deep convolutional neural network that implements these principles in several layers can be trained to do natural image categorization with previously unprecedented accuracy (Krizhevsky et al., 2012). Interestingly, such models can match human level accuracy in certain object recognition tasks (Yamins et al., 2014) and even to some extent capture human shape sensitivity (Kubilius et al., 2016). The units in a trained convolutional network become feature detectors that encode increasingly complex features of the input image. Their activations in response to an input image can be taken to quantify the presence of each feature (i.e., feature extraction). Unit activations can also be used to predict actual neural responses in monkey IT and V4 (Cadieu et al., 2014; Yamins et al., 2014; Kalfas et al., 2017) and fMRI responses across the human ventral visual stream (Güçlü and van Gerven, 2015). Thus, deep neural networks provide an excellent framework for simultaneously predicting brain and behavioral responses (Kriegeskorte, 2015).

In the upcoming chapter, we revisit the behavioral categorization data of Chapter 1 and the neuronal data of Chapter 2. We refer to the respective chapters for methodological details of the experiments. Our understanding of these data was limited by the lack of a model that can quantify our natural movies in terms of complex visual features. Here, we attempt to address this issue by using a convolutional deep neural network to extract features of these movies. **Figure 3.1** shows the architecture of the DNN that we used in this paper. We extracted features at different layers of the network (shaded in blue), by

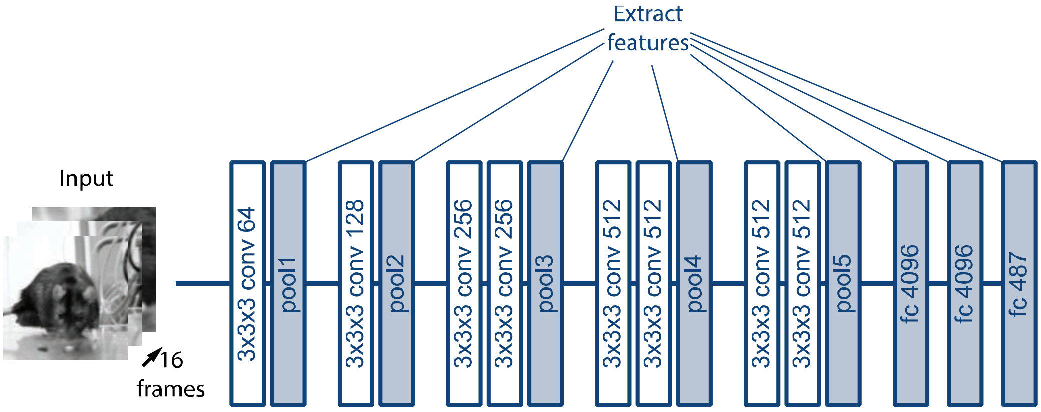


Figure 3.1. Schematic representation of the network architecture used for movie feature extraction.
The network is a 3D ConvNet (Tran et al., 2014) with a VGG-11 architecture (Simonyan and Zisserman, 2014). The figure should be read from left (input) to right (classification) and has 5 convolutional layer stacks followed by three fully connected layers. Convolution is done by 3 × 3 × 3 (height, width, and time) patches with 64, 128, 256, or 512 feature maps per convolutional layer. The fully connected layers have 4096 or 487 units. We only used the blue shaded layers for feature extraction.

taking their unit activations in response to our natural movies. Different layers represent different stages of processing, with units from lower to higher layers encoding increasingly complex visual features. For comparison with neural data, we applied an analysis analogue to the one described in detail for our neural data under “Population representation” in section 2.2.4, Data analysis. For comparison with behavioral data, we trained a linear classifier on features (activations) extracted from the training stimuli (for which we also have neuronal data) and tested for generalization on test stimuli.

Published online as conference paper

Vinken K., Op de Beeck H. (2017). Deep Neural Networks and Visual Processing in the Rat. *Annual Conference on Cognitive Computational Neuroscience*. New York City, NY, USA, http://ccneuro.org/conference-archives/2017-2/

DEEP NEURAL NETWORKS AND VISUAL PROCESSING IN THE RAT.

The increased use of rodents as a model for low- and higher-level visual functions has raised the question of how rodent visual processing compares to existing computational and primate models. Rodent visual cortex has two pathways, with one "lateral stream" anatomically resembling the primate ventral stream (Wang et al., 2012). This primate pathway is specialized in object recognition and the stages of processing are captured well by deep neural networks (DNNs; Güçlü and van Gerven, 2015). Here we compare the stages of processing of natural and scrambled movies in a 3D convolutional network (3D ConvNet) with three stages of the aforementioned rat lateral stream: primary visual (V1), laterointermediate (LI), and temporal occipital cortex (TO). As in rats (Vinken et al., 2016), a natural versus scrambled representation emerges in the convolutional layers of the DNN. The last of these layers can support generalization in a movie categorization task that rats could also learn (Vinken et al., 2014). The subsequent fully connected layers lead to a clear categorical representation not found in untrained rats (Vinken et al., 2016). This comparison reveals similarities between the rat lateral stream and a DNN that could explain relatively complex visual abilities.

3.1 RESULTS

We extracted spatio-temporal features from a movie stimulus set using a 3D ConvNet (Tran et al., 2014). Next, we compared DNN stimulus representations with neural representations in rat visual cortex (V1, LI, and TO). Finally, we assessed whether the DNN features allow a linear classifier to generalize in a movie categorization task that rats are able to learn.

3.1.1 DEEP NEURAL NETWORK FEATURE EXTRACTION

All stimuli are greyscale 384 x 384 movies of 5s (150 frames) each, with either a rat or a moving object. Frames were resized to 128 x 128 pixels and 112 x 112 center crops were taken. Features were extracted per window of 16 frames (9 windows spanning 144 frames) and averaged per movie (like for the neural data). We used a pre-trained (Sports

M-1) 3D ConvNet (Tran et al., 2014) with a VGG-11 architecture (Simonyan and Zisserman, 2014) that is extended to encompass the time dimension. This model is included in C3D-v1.0 (http://vlg.cs.dartmouth.edu/c3d/).

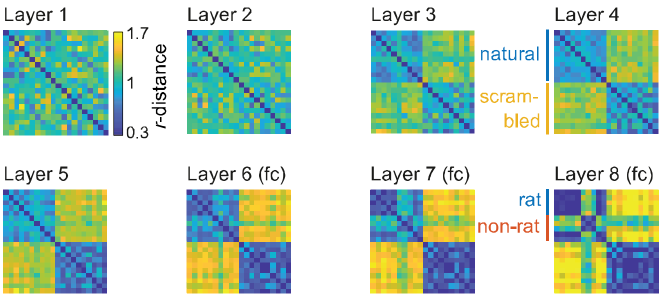


Figure 3.2. Deep Neural Network RDMs.
The first 5 are max-pool layers (each preceded by one or more convolutional layers). Layer 6-8 are fully connected.

We focus on a subset of 20 movies for which we have neural data: 5 rat movies, 5 non-rat movies, and their spatio-temporal phase scrambled version (Vinken et al., 2014). On the extracted features, we applied per layer principal component analysis (PCA) resulting in 19-dimensional vectors per stimulus. As for the neural data, these vectors were correlated pair-wise (Pearson *r*) in order to obtain representational dissimilarity matrices (RDMs) with distances 1 - *r*. Stimulus pairs that share a similar representation across features in a layer result in a lower dissimilarity. These matrices were calculated for 8 DNN layers (**Figure 3.2**). Across the max-pool layers (1-5) a natural versus scrambled movie pattern emerges: 4 large quadrants become visible. Later, in the fully connected layers 6-8 a categorical pattern emerges within the natural movies (i.e. a grouping within the left-upper quadrant).

3.1.2 Neural stimulus representations

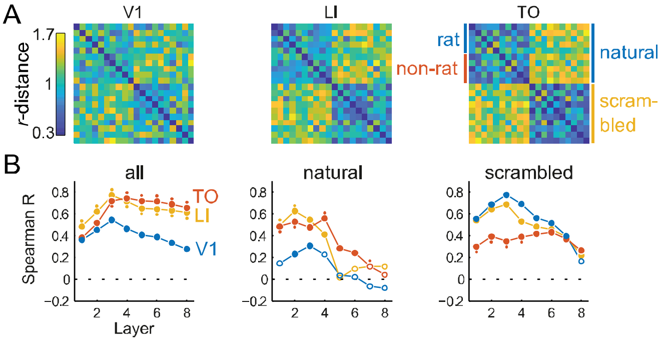


Figure 3.3. Neural RDMs per area (A) and their correspondence with DNN layers for all, natural or scrambled stimulus pairs (B).
Filled markers indicate the lower 95% CI bound was higher than zero. Layers for which the LI-V1 or TO-V1 95% CI excludes zero are marked with a dot.

Next, we compare the DNN RDMs with neural RDMs that previously revealed a natural versus scrambled dissociation but no categorical pattern in rats (Vinken et al., 2016). In short, per stimulus a neural response vector was obtained using each single and multi-unit's (SU and MU) normalized firing rate. This resulted in N-dimensional response vectors, with N = 50SU+25MU for V1, N = 53SU+33MU for LI, and N = 52SU+26MU for TO. Again, stimulus pairs that elicit a similar neural response result in a lower dissimilarity. These neural RDMs are shown in **Figure 3.3**A.

In **Figure 3.3**B we quantified the correspondence between neural and DNN RDMs by calculating the correlation (Spearman R) between off-diagonal upper halves of the matrices. In general, the correspondence increases up to layer 3 for V1/LI and up to layer 4 for TO. The maximum correlation is higher for LI and TO than for V1. For scrambled movies, TO corresponds less with earlier layers than V1. For natural movies, both LI and TO correspond more with earlier layers than V1. In particular, there is a decreased correlation for fully connected layers 6-8: the DNN representation of natural movies grows towards a categorical pattern that is absent in the neural representations.

3.1.3 Categorization performance

Next, features were extracted for a larger set of movies used previously in a behavioral experiment. Here, rats learned to classify rat movies from natural or scrambled distractors and could generalize to several new test sets (Vinken et al., 2014). To assess for each DNN layer whether its features would be able to support such a task, we trained a linear support vector machine (SVM) and tested for generalization on the test sets. PCA was used for feature reduction, only retaining the first N dimensions that explain at least 50% of the variance (more features generally lead to poor generalization).

Table 3.1. Rat versus non-rat generalization (% correct).

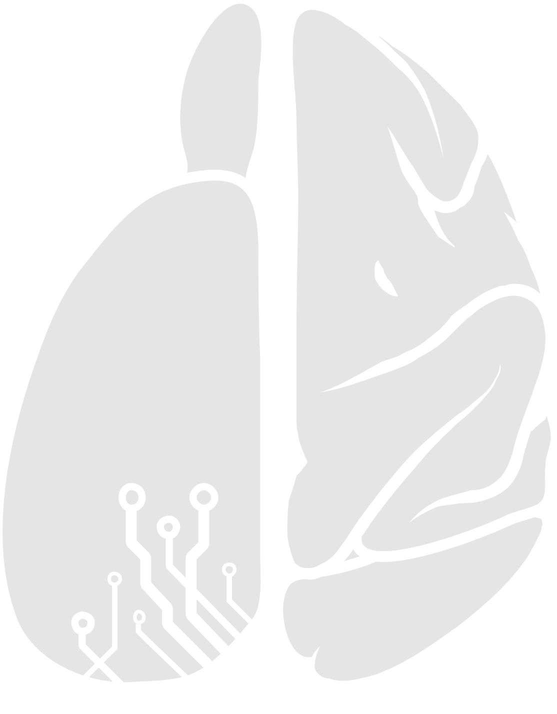
Test set	L1	L2	L3	L4	L5	L6	L7	L8
Natural	40	30	50	75	98	98	93	100
Natural slow	50	50	60	60	93	100	57	63
Scrambled	93	87	97	100	100	100	97	97

The SVM performance as a function of DNN layer is shown in **Table 3.1**. Representations in later convolutional layers (in particular at maxpool layer 5) can support successful generalization from training to test stimuli with natural distractors. Note that, as opposed to rats (Vinken et al., 2014), in fully convolutional layers the classifier fails to generalize to stationary or slow stimuli (labeled "natural slow").

3.2 CONCLUSION

In this work we compare stimulus representations in a DNN with those of the rat visual "lateral stream". We show a correspondence with convolutional layers that does not extend to the categorical representation of fully connected layers. In addition, later convolutional layers can explain visual categorization abilities in rats. Together, this suggests that rat neural responses and behavior relate to a mid-level representation in visual hierarchical processing.

II ADAPTATION AND EXPECTATION IN RAT VISUAL CORTEX



Chapter 4.

CHANGE DETECTION IN RAT VISUAL CORTEX

All of our previous studies were aimed at investigating object recognition properties in the rat visual system. Here, we turn to a visual oddball paradigm in a study on adaptation and expectation. In human event-related potential studies, this paradigm is associated with a component called the mismatch negativity (MMN; Näätänen et al., 2007). This refers to a difference in response between frequent and rare events. In monkey IT cortex, this difference can be explained by repetition suppression for frequent stimuli and not by a surprise related enhancement for rare stimuli (Kaliukhovich and Vogels, 2014). In this final rat study, we use this paradigm to investigate adaptation and effects of expectation in the rat visual system. *Do we see a surprise-based response enhancement in the rat visual system?*

Published as

Vinken K., Vogels R., Op de Beeck H. (2017). Recent Visual Experience Shapes Visual Processing in Rats Through Stimulus Specific Adaptation and Response Enhancement. *Current Biology*, *27* (6), 914-919.

RECENT VISUAL EXPERIENCE SHAPES VISUAL PROCESSING IN RATS THROUGH STIMULUS SPECIFIC ADAPTATION AND RESPONSE ENHANCEMENT.

From an ecological point of view it is generally suggested that the main goal of vision in rats and mice is navigation and (aerial) predator evasion (Wallace et al., 2013; Yilmaz and Meister, 2013; Zoccolan, 2015). The latter requires fast and accurate detection of a change in the visual environment. An outstanding question is whether there are mechanisms in the rodent visual system that would support and facilitate visual change detection. An experimental protocol frequently used to investigate change detection in humans is the oddball paradigm, where a rare unexpected stimulus is presented in a train of stimulus repetitions (Garrido et al., 2009). A popular “predictive coding” theory of cortical responses states that neural responses should decrease for expected sensory input and increase for unexpected input (Friston, 2005; Summerfield and de Lange, 2014). Despite evidence for response suppression and enhancement in noninvasive scalp recordings in humans with this paradigm (Jacobsen and Schröger, 2001; Czigler et al., 2002), it has proven challenging to observe both phenomena in invasive action potential recordings in other animals (Farley et al., 2010; Fishman and Steinschneider, 2012; Kaliukhovich and Vogels, 2014). During a visual oddball experiment, we recorded multi-unit spiking activity in rat primary visual cortex (V1) and latero-intermediate area (LI), which is a higher area of the rodent ventral visual stream. In rat V1 there was only evidence for response suppression related to stimulus-specific adaptation and not for response enhancement. Yet, higher up in area LI, spiking activity showed clear surprise-based response enhancement in addition to stimulus-specific adaptation. These results show that neural responses along the rat ventral visual stream become increasingly sensitive to changes in the visual environment, suggesting a system specialized in the detection of unexpected events.

4.1 RESULTS

We recorded the action potential activity of multi-unit sites in V1 and extrastriate area LI of awake rats, during a visual oddball paradigm with an equiprobable control condition (see **Figure 4.1** and supplemental information). The standard, deviant, and control

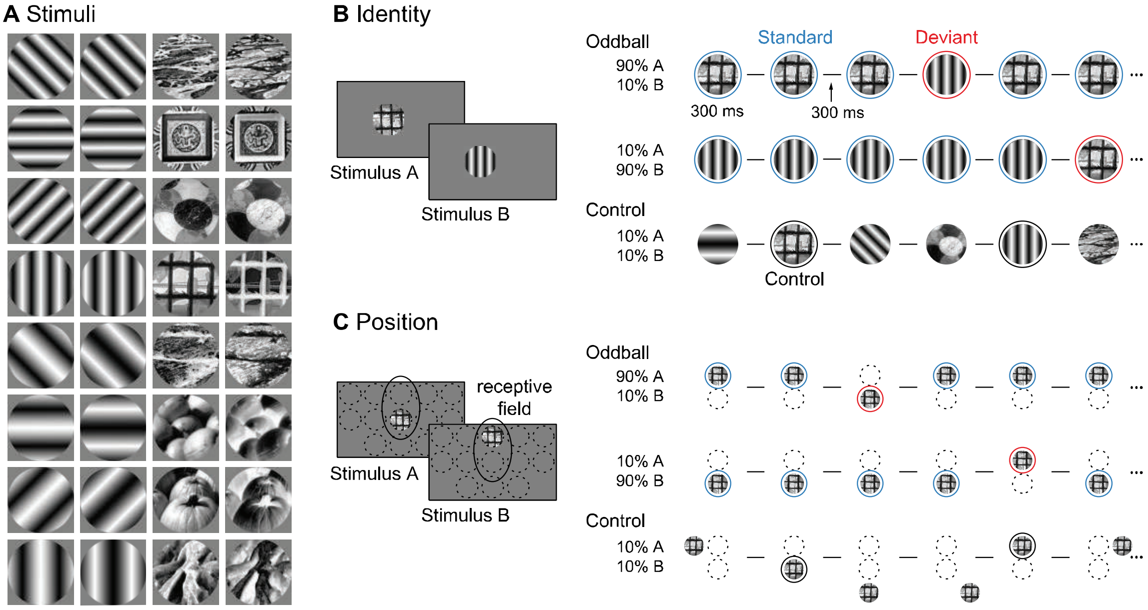


Figure 4.1. Stimulus set and visual oddball paradigm.
(A) Stimuli included sine wave gratings (4 orientations at 2 frequencies) and 8 textures. For each sine wave we included the quadrature phase shift and for each texture the negative version. Sine wave gratings typically drive neural responses in V1 well, but extrastriate area LI might be more sensitive to complex stimuli like textures (Vinken et al., 2016). **(B)** In the identity oddball experiment different stimuli were presented in different blocks of 100 randomized trials of 300 ms, separated by 300 ms. Oddball blocks consisted of two stimuli (A and B) shown at different probabilities: p(standard) = 0.9; p(deviant) = 0.1. Assignment of the probabilities to A and B was counterbalanced across blocks. In equiprobable control blocks, the probability of A, B, and 8 additional stimuli is 0.1, so that none of them stand out as a deviant. Responses to the standard, deviant, and control conditions were averaged across A and B for sites that had a positive net response to both (see supplemental information). **(C)** The position oddball experiment was identical, except that instead of presenting different stimuli, one stimulus was presented at different screen positions.

conditions allow us to identify two mechanisms: response suppression for highly probable stimuli (standard < control and < deviant) and response enhancement for unexpected stimuli (deviant > control).

4.1.1 Primary visual cortex: identity oddball experiment

In V1, we recorded multi-unit spiking activity (MUA) in 55 responsive sites (28 in Rat 1 and 27 in Rat 2, in 13 and 14 sessions, respectively) to sequences in which two stimuli were presented with different probabilities: p(standard) = 0.9; p(deviant) = 0.1 (**Figure 4.1**B; see supplemental information). In general, the V1 response to a stimulus was very transient, with a relatively low sustained response (**Figure 4.2**A; first 2 rows). Because of the transient nature of the response, we focused the analysis on its first 100 ms. Using the average net firing rates per condition, we calculated adaptation indices (*AI*) that indicate response suppression (*AI* < 0) and enhancement (*AI* > 0) for the standard (S) and deviant (D) relative to the response to the same stimulus in the equiprobable control condition (C):

$$AI_{SC} = \frac{(S - C)}{(|S| + |C|)}, AI_{DC} = \frac{(D - C)}{(|D| + |C|)}$$

The median *AI* demonstrated response suppression to the standard (AI_{SC} Rat 1: *Median* = -0.28, p < 0.0001, sign test, AI_{SC} Rat 2: *Median* = -0.31, p < 0.0001). However, we find no evidence for a change in response to the deviant, relative to the control (AI_{DC} Rat 1: *Median* = 0.01, p = 0.1849, AI_{DC} Rat 2: *Median* = -0.03, p = 1.0000; **Figure 4.2**B).

To account for the variability caused by differences amongst MUA sites across rats, we analyzed our data using a multi-level model (Lazic, 2010; Aarts et al., 2014; Vinken et al., 2014). We used a regression model where average raw firing rates per condition and per unit are modeled with a lognormal distribution. Responses of cortical neurons have been shown to follow a lognormal distribution (Buzsáki and Mizuseki, 2014) and this was confirmed in the present data. From this model we report parameter *δ*, which expresses the ratio of the net responses for deviant (δ_{DC}) and standard (δ_{SC}) conditions relative to those for the control condition (see supplemental information). The results (**Figure 4.2**C) indicated that the response to the standard was 57% of the response to the control (δ_{SC} = 0.57, **Figure 4.3**A), and the response to the deviant was 101% of the control (δ_{DC} = 1.01, **Figure 4.3**A). Both rats showed a lower response to the standard compared with those to the deviant – and lower than those to the control ($\delta_{DC} - \delta_{SC}$ = 0.44, **Figure 4.3**A), indicating stimulus specific adaptation.

4.1.2 Latero-intermediate area: identity oddball experiment

We performed the same experiment while recording MUA in 48 responsive sites in LI (29 in Rat 2 and 19 in Rat 3, in 11 and 8 sessions, respectively). The mean time course plots of firing rates (**Figure 4.2**A) indicate that a very small response was elicited to the standard compared to the control. The *AI*s (**Figure 4.2**B) showed strong reduction of the

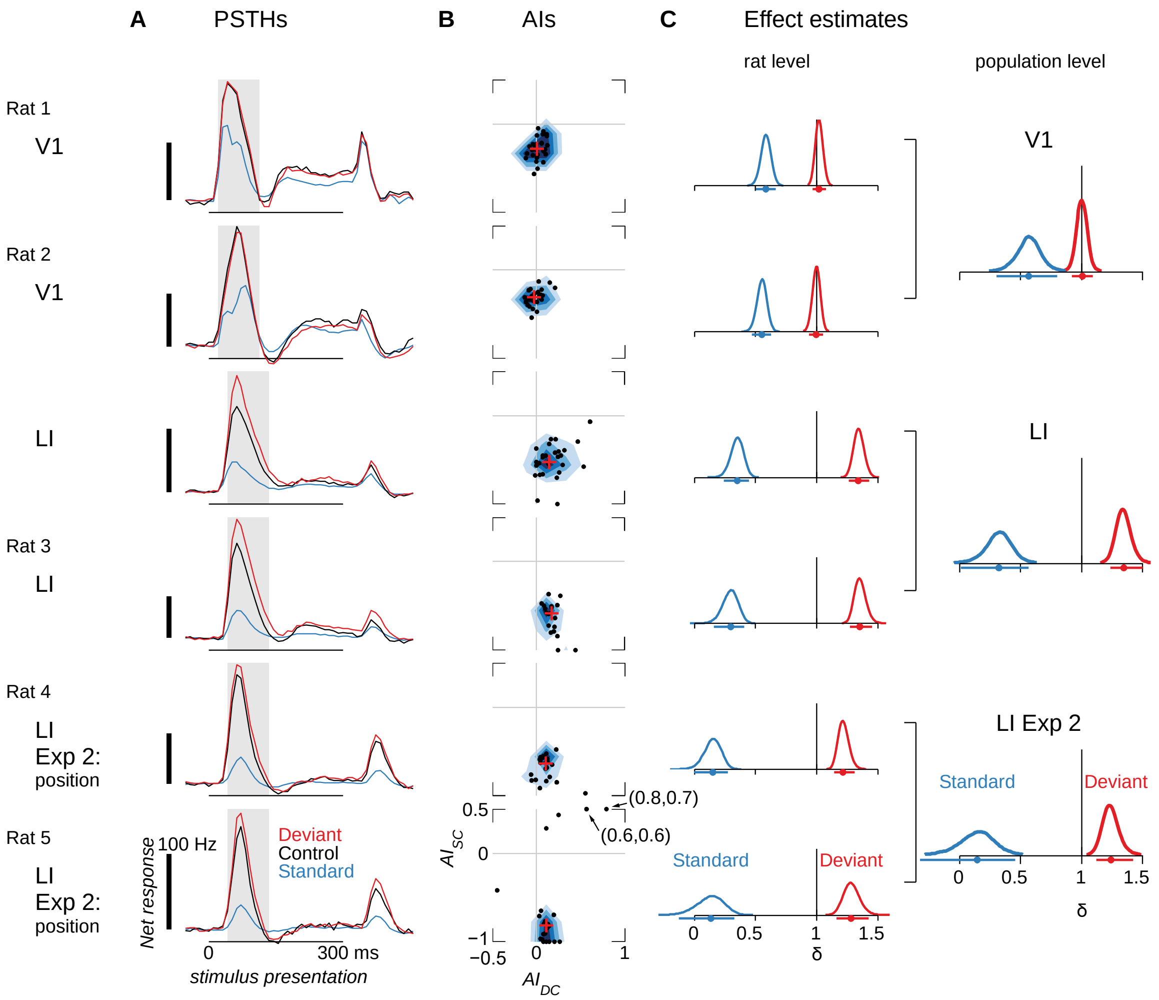


Figure 4.2. Multi-unit neural responses for deviant, standard, and control condition.
(A) The net (baseline subtracted) firing rate expressed in spikes per second is plotted across time (10 ms bins) averaged across deviant (red), standard (blue), or control (black) stimulus presentations. Vertical scale bars indicate a firing rate of 100 Hz. The horizontal line indicates the 300 ms stimulus presentation. The results are plotted separate for each rat × area combination. Further analyses were done on net spike counts of the first 100 ms onset of the response (area shaded in gray). **(B)** Scatterplots of adaptation indices of multi-unit sites for deviant (abscissa; AI_{DC}) versus for standard (ordinate; AI_{SC}) conditions, superimposed on a contour plot. Median values are indicated by a red cross. Points below the horizontal line ($AI_{SC} < 0$) indicate response suppression for the standard relative to the control condition. Points to the right of the vertical line ($AI_{DC} > 0$) indicate response enhancement for the deviant relative to the control condition. **(C)** Multi-level model effect estimates (δ) for standard and deviant condition on rat × area level and on population level. Values of δ indicate net responses for standard or deviant conditions as a proportion of net responses for the control condition. Estimated posterior distributions are plotted, which indicate the estimated probability density for each parameter value δ given the data: the higher the density, the more probable the underlying values. The effects' point estimates (posterior median) and 95% intervals are indicated below each distribution by a dot and a horizontal line, respectively (see also **Figure 4.3**). The value of 1 is indicated by a black vertical line. Estimates close to 1 indicate similar responses in the control relative to the responses for a standard or deviant. See also Figures S1, S2, and S3.

response to the standard (AI_{SC} Rat 2: *Median* = -0.53, $p < 0.0001$, AI_{SC} Rat 3: *Median* =-0.58, $p < 0.0001$).

Contrary to V1, the response to the deviant was stronger compared to the control (AI_{DC} Rat 2: *Median* = 0.14, p = 0.0001, AI_{DC} Rat 3: *Median* = 0.17, $p < 0.0001$). The latter suggests LI response enhancement to the unexpected deviant stimulus in LI. This difference between V1 and LI could not be explained by a sampling bias in cortical layers (see supplemental **Figure 4.4**). The multi-level model estimates (**Figure 4.2**C) indicated that the response to the standard was 32% of the response of the control (δ_{SC} = 0.32, **Figure 4.3**B) while the response to the deviant was 135% of the control (δ_{DC} = 1.35, **Figure 4.3**B). Logically, these results amounted to a strong difference in responses to standard and deviant ($\delta_{DC} - \delta_{SC}$ = 1.02, **Figure 4.3**B).

Comparing LI with V1, the deviant-standard response difference was estimated to be stronger in LI ($[\delta_{DC}^{LI} - \delta_{SC}^{LI}] - [\delta_{DC}^{V1} - \delta_{SC}^{V1}]$ = 0.59, **Figure 4.3**D). This difference between areas resulted from both an increase in stimulus-specific adaptation for the standard ($\delta_{SC}^{LI} - \delta_{SC}^{V1}$ = -0.25, **Figure 4.3**D), and a change in effect of the deviant ($\delta_{DC}^{LI} - \delta_{DC}^{V1}$ = 0.34, **Figure 4.3**D), with only the latter difference between the areas being significant.

We could compare the responses in V1 and LI within one animal (Rat 2). Like on the population level, the deviant-standard response difference was estimated to be stronger in LI than in V1 ($[\delta_{DC}^{LI} - \delta_{SC}^{LI}] - [\delta_{DC}^{V1} - \delta_{SC}^{V1}]$ = 0.55, **Figure 4.3**D). In addition, this animal showed in LI an increase in stimulus-specific adaptation for the standard ($\delta_{SC}^{LI} - \delta_{SC}^{V1}$ = -0.20, **Figure 4.3**D), in addition to a change in effect for the deviant ($\delta_{DC}^{LI} - \delta_{DC}^{V1}$ = 0.35, **Figure 4.3**D). Thus, the increase in the MUA difference between standard and deviant in Rat 2 LI compared to V1 resulted from both a stronger repetition suppression and stronger response enhancement.

4.1.3 Latero-intermediate area: position oddball experiment

A possible explanation for the deviant-control response difference in LI is that there was more cross-stimulus adaptation in the equiprobable blocks from the additional 8 stimuli. For example, if the neural site responds very well to most stimuli in equiprobable blocks, neural fatigue (Vogels, 2016) alone can cause reduced general responsiveness in these blocks. This can result in response suppression for the control condition which might explain a difference in neural response between deviant and control. Such suppression can also be caused by feature specific adaptation if the neural site responds to one or more features shared by the different stimuli in equiprobable blocks. We calculated a response equivalence index (*EI,* see supplemental information) that indicates whether the response to the additional stimuli is equivalent to (*EI* = 0) or lower (EI > 0) than the control. The *EI* was practically zero for the identity oddball recordings (Rat 1 V1 *EI*: *Median* = 0.03, *IQR* = 0.11; Rat 2 V1 *EI*: *Median* = 0.00, *IQR* = 0.10; Rat 2 LI *EI*: *Median* = 0.00, *IQR* = 0.14; Rat 3 *EI*: *Median* = 0.04, *IQR* = 0.08), indicating that cross-adaptation in the equiprobable blocks might indeed be present. This possibility was supported by a decrease of the responses with trial number in the control condition (see supplemental **Figure 4.5**C, D). We addressed this issue by performing an experiment where we manipulated stimulus position instead of stimulus identity (**Figure 4.1**C; see supplemental information). This allowed us to place stimuli for the control condition outside the receptive field, which should prevent cross-adaptation to the two positions used as standard and deviant in the oddball blocks. The data obtained in the position oddball experiment showed positive *EI*s for both rats (Rat 4 *EI*: *Median* = 0.65, *IQR* = 0.28; Rat 5 *EI*: *Median* = 0.33, *IQR* = 0.46), indicating that cross-adaptation in the equiprobable blocks should at least be reduced.

For this second experiment, we recorded MUA in 44 responsive sites in LI (22 in Rat 4 and 22 in Rat 5, in 6 and 5 sessions, respectively). Again, the *AI*s (**Figure 4.2**B) showed strong reduction of the response to the standard (AI_{SC} Rat 4: *Median* = -0.64, $p < 0.0001$, AI_{SC} Rat 5: *Median* = -0.81, $p = 0.0043$). In addition, the response to the deviant was still elevated compared to the control (AI_{DC} Rat 4: *Median* = 0.11, $p = 0.0009$, AI_{DC} Rat 5: *Median* = 0.11, $p < 0.0001$). The multi-level model estimates (**Figure 4.2**C) indicated that the response to the standard was 14% of the response to the control ($\delta_{SC} = 0.14$, **Figure 4.3**C). The response to the deviant was estimated as 124% of that to the control ($\delta_{DC} = 1.24$, **Figure 4.3**C). As before, this resulted in a strong difference in response between the standard and deviant conditions ($\delta_{DC} - \delta_{SC} = 1.09$, **Figure 4.3**C). Importantly, we no longer observed a decrease in response to the control as a function of trial number (see

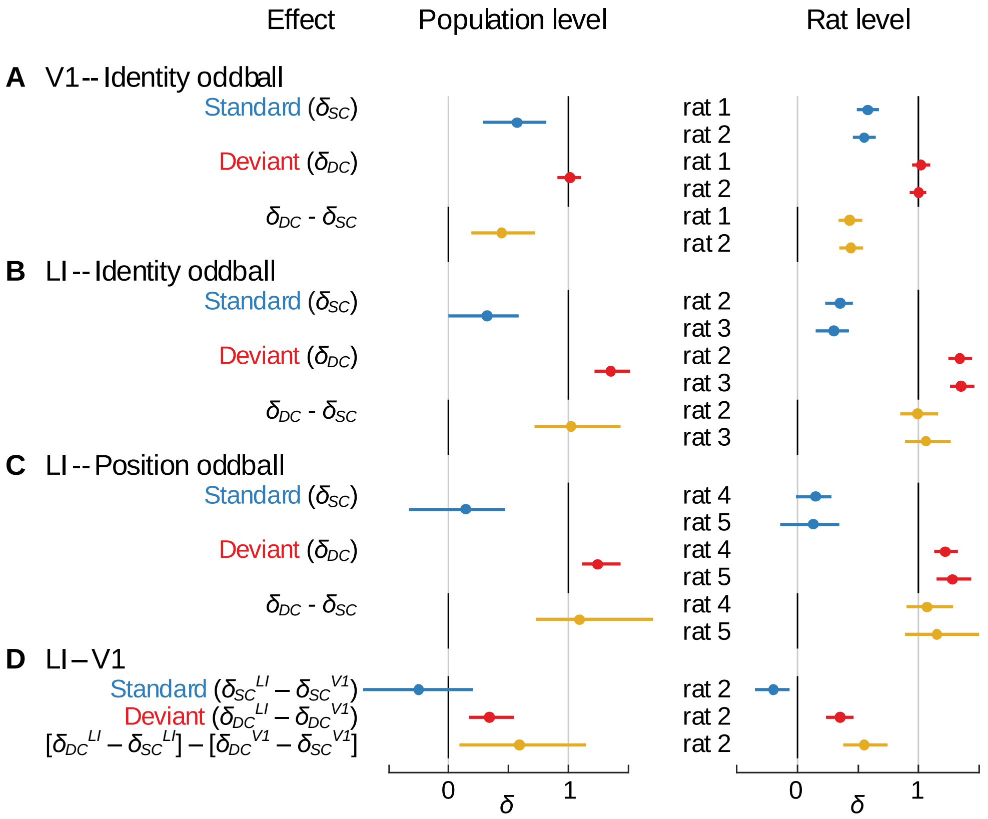


Figure 4.3. Multi-level model effect estimates on rat × area level and on population level. (A-D) Point estimates (posterior median) and 95% intervals (error bars) for each comparison. Relevant reference points (0 or 1) are indicated by black vertical lines. Note that effects for standard and deviant in panels (A)–(C) are the same as those presented in **Figure 4.2**.

supplemental **Figure 4.5**E, F), suggesting that cross-adaptation was successfully eliminated.

4.2 Discussion

To summarize, we observed a clear difference between multi-unit responses to the deviant and standard stimuli in an oddball paradigm consistently across all rats and areas. This response difference was bigger in LI compared to V1. In LI it was the combined result of a strong response reduction for the standard and an enhancement for the deviant. V1 did not show such an enhancement and might have weaker stimulus-specific adaptation.

Stimulus specific adaptation has been documented to play an important role in modulating spiking activity in both auditory and visual cortices during modality appropriate oddball sequences (Ulanovsky et al., 2003; Farley et al., 2010; Fishman and Steinschneider, 2012; Kaliukhovich and Vogels, 2014), as well as recent sensory experience in general (Solomon and Kohn, 2014). Our results confirm this in both V1 and LI. In addition, this adaptation might be stronger in LI compared to V1. This agrees with the finding that stimulus-specific adaptation in auditory oddball sequences is stronger and faster outside primary auditory cortex (Nieto-Diego and Malmierca, 2016). In our data, the suppression for the standard was rapid and persistent, and seemed only to be relieved by the occurrence of a deviant in the immediately preceding trial (see supplemental **Figure 4.6**). Stimulus-specific adaptation also affected the response to the deviant/control, which was lower the more recent the previous deviant or the same stimulus was (see supplemental **Figure 4.6**).

Spiking activity in rat V1 did not show an enhanced response to the deviant. Rather, a response reduction for trial numbers later in the sequence points towards cross-adaptation from the standard (see supplemental **Figure 4.5**A, B). The absence of a surprise response in rat V1 corroborates multi-unit recordings in primary auditory cortex of both rat (Farley et al., 2010) and monkey (Fishman and Steinschneider, 2012). Recently, an increased response to a deviant was reported for neural responses in mouse V1 (Hamm and Yuste, 2016). However, the difference between standard and deviant stimuli (orthogonal orientations) was considerably greater than those between stimuli in their control condition. Thus, their effect can be explained by more cross-stimulus adaptation for the control than for the deviant (Farley et al., 2010). In contrast with V1, multi-unit responses in area LI did demonstrate a higher response to the deviant compared to the control. This enhancement was also present in an additional experiment which decreased cross-stimulus adaptation in the equiprobable sequences (i.e. the position oddball experiment). The absence of a decreased response to the control stimuli as a function of trial number (see supplemental **Figure 4.5**E, F) indicates that cross-adaptation was eliminated. Thus, we provide the first demonstration of a surprise response in spiking activity in an oddball paradigm when controlling for adaptation, which is an important prediction of the predictive coding framework (Friston, 2005). The timing of this effect is in the earliest phase of the response and thus may originate in LI itself. The high frequency of the standard stimulus could affect the processing of stimuli in the local LI circuit, giving rise to surprise responses. Perhaps NMDA receptor neurotransmission might be involved in deviant responses (Garrido et al., 2009) as opposed to stimulus specific adaptation (Farley et al., 2010). The relative contribution of local versus top-down processes (Gilbert and Li, 2013) requires further investigation. A trend for an enhanced response to the deviant was also present in single LI neurons, and this was significant when responses were not individually normalized for firing rate (see supplemental information). In addition, the results of simultaneously recorded local field potentials correspond to those of the MU activity (see supplemental **Figure 4.4**).

The results of our recordings in LI differ noticeably from those of monkey IT neurons. Various observations in monkeys show a sensitivity for statistical structure of visual information after weeks of exposure (Meyer and Olson, 2011; Kaposvári et al., 2016). Nevertheless, surprise-related enhancements in an oddball paradigm were not observed in primate visual cortical areas (Kaliukhovich and Vogels, 2014). We have to be careful when engaging in such species comparisons for various reasons, such as difficulties to know which areas and pathways correspond and differences in the details of experiments (stimulus size, behavioral tasks, reward schedules, etc.). Still, we stayed as close as possible to the experiment by Kaliukhovich and Vogels (Kaliukhovich and Vogels, 2014) and both rat and monkey were not actively engaged in a task with the stimuli. In addition , rodent LI belongs to a processing stream that has been suggested to be homologous to the primate ventral stream that culminates in IT (Wang et al., 2012). However, it remains an open question whether and to what extent they might be functionally similar. Other studies have reported unexpected properties of the proposed rat ventral pathway before, namely an increased response to moving stimuli (Vermaercke et al., 2014) as well as a lack of a categorical representation and lack of higher responses to natural stimuli (Vinken et al., 2016). Nevertheless, the same studies also reported commonalities with the primate ventral pathway, namely an increased tolerance for stimulus position (Vermaercke et al., 2014) and clustering of natural versus scrambled movie representations (Vinken et al., 2016).

A clear difference in neural response between regular and irregular stimuli is a necessary prerequisite for a system specialized in change detection. We observed this difference in all of our recordings and it increased between V1 and higher visual area LI. The fact that we see this transition might indicate that change detection is an important functional specialization of the processing stream that both areas belong to. This claim is compatible with previous reports emphasizing predator detection as one of the major ecologically valid functions of vision in rats and mice (Wallace et al., 2013; Yilmaz and Meister, 2013). Previous research has indeed shown that sensory adaptation facilitates perceptual detection of deviant stimuli by increasing the difference in neural responses (Musall et al., 2014). A response enhancement for the unexpected stimulus will only further increase this difference. Future studies are needed to study the behavioral relevance of this surprise-based response enhancement.

4.3 SUPPLEMENTAL INFORMATION

4.3.1 ANIMALS

Experiments were conducted with 5 male Long-Evans rats, aged between 6 and 25 months (12.8 on average) at the start of the study. The distribution of rats across experiments was as follows: V1 recordings for the identity oddball experiment in Rat 1 and Rat 2, LI recordings for the identity oddball experiment in Rat 2 and Rat 3, LI recordings for the position oddball experiment in Rat 4 and Rat 5. Surgical procedures were the same as previously reported (Vermaercke et al., 2014). Surgery was performed to implant a head post and a recording chamber. The craniotomy was centered on average 7.5 mm anterioposterior and 2.8 mm mediolateral. When entering at an angle of 45° this location allowed recordings in V1, as well as LI (Vermaercke et al., 2014). In one rat the craniotomy and recording chamber were placed at an angle of 90°, to allow sampling from different cortical depths in V1. After recovery the animals had ad libitum access to food pellets and had restricted access to water to train them to sit comfortably in the setup. After initial training the animals would remain still during recording even without water rewards, so water restriction was stopped. Housing conditions and experimental procedures were approved by the KU Leuven Animal Ethics Committee.

4.3.2 Stimuli

The stimulus set consisted of 16 sine wave gratings (quadrature phase pairs of 2 frequencies x 4 orientations) and 8 textures. While sine wave gratings are typically ideal to drive neural responses in V1, we included textures to ensure good responses in extrastriate area LI, which might be more sensitive to more natural visual stimulation (Vinken et al., 2016). The 8 textures where taken and modified from the MIT VisTex database (http://vismod.media.mit.edu/pub/VisTex/), and we included their negative version to balance the stimulus set with respect to the 8 quadrature phase pairs of gratings. All images where modified to have the same uniform pixel value distribution. The full stimulus set is displayed in **Figure 4.1**A.

4.3.3 Oddball paradigm

For the main experiment we based the design on the visual oddball paradigm as implemented by Kaliukhovich and Vogels (2014). In this protocol two stimuli (say A and B) are presented at different probabilities in oddball blocks of 100 stimulus presentations (trials) of 300 ms, each preceded by a 300 ms blank screen. One stimulus (the "standard") is presented for 90 trials, while the other (the "deviant") is presented for only 10 trials. The order of the standard and deviant trials is randomly shuffled. The standard and deviant are counterbalanced between blocks, meaning each stimulus is the standard in half of the oddball blocks and the deviant in the other half. As a control condition, the oddball blocks where interspersed with equiprobable sequences. In these sequences 10 different stimuli, including A and B, are presented each at equal probability (i.e. each for 10 trials). The trials of these stimuli are randomly shuffled. This leads to a total of 3 block types: two oddball blocks (counterbalanced for standard and deviant) and one equiprobable block. The order in which these three block types were presented was counterbalanced by means of a Latin square design (i.e. 1, 2, 3, 2, 3, 1, 3, 1, 2), with each block type randomly assigned to first, second, or third place. To minimize interference and adaptation between blocks, each block was preceded by serially presenting 50 random full screen stimuli (white noise of 10 by 6 squares) for 33.3 ms each. We used this oddball paradigm in two slightly different experiments.

Experiment 1: identity oddball experiment

Here we use a visual oddball paradigm where we manipulate the stimulus identity. Prior to starting the experiment, we completed two tests to determine (1) responsive receptive field positions, and (2) effective stimuli. In the first test, a hash shape was shown at 15 locations (on a 3 by 5 grid) on a black screen to determine the most responsive location(s) of the receptive field covered by the computer screen. At the center of the screen the shape diameter was 24 visual degrees and the shape centers were spaced 26° apart. The shape was presented for 500 ms, with an inter-trial interval of 500 ms plus up to 300 ms random jitter. Next, a responsive location was chosen for the stimulus selection protocol. In this stimulus selection protocol all of the 32 stimuli where presented in pseudorandom order for 7 to 23 times each (*Median* = 14). Each presentation lasted 300 ms and was separated from the next by an inter-trial interval of at least 300 ms. The stimulus diameter was about 30 visual degrees in the center of the screen. The background pixel value was 128, which is the same as during the oddball protocol. One grating and one texture stimulus that both elicited a clear response were then chosen as stimulus A and B for the oddball protocol. For the equiprobable sequences, 8 additional stimuli (4 textures and 4 gratings) where randomly chosen, with the restriction that none of them could be part of the quadrature phase pair of the selected grating (A) or the negative/original pair of the selected texture (B). Finally, the oddball paradigm described in the previous paragraph was run for a total of 18 to 28 blocks (*Median* = 24), with all stimuli presented at the screen location used in the stimulus selection protocol.

Experiment 2: position oddball experiment

Instead of manipulating stimulus identity, in this experiment we manipulate stimulus position in a visual oddball paradigm. The procedures were largely the same as those for the stimulus identity oddball experiment. The stimulus selection protocol was run with each of the 32 stimuli presented for 5 to 11 times (*Median* = 8). Since a stimulus would be presented in this experiment at different locations within the same blocks, their diameter was reduced to 20 visual degrees in the center of the screen in order to avoid any overlap. After the stimulus selection protocol, only one stimulus with a good response was chosen as stimulus A. Stimulus B, however, was now the same stimulus presented at a different location usually within the receptive field. Only 13 out of the 15 positions of the receptive field mapping test (see previous paragraph) where eligible as stimulus location, because the lower left and right were partially obstructed from view by photocells that serve to record onset and offset of the stimulus appearance on the screen. For the equiprobable sequences, eight additional locations where randomly chosen out of the remaining 11 locations. To sum up, experiment 2 was the same as experiment 1, except that now one stimulus was presented at different locations instead of multiple stimuli at only one location. This position oddball experiment was run for 23 to 30 blocks (*Median* = 24). See **Figure 4.1**B and C for an illustration of the oddball experiments 1 and 2.

4.3.4 Electrophysiological recordings

Recording procedures were identical to those previously reported (Vermaercke et al., 2014; Vinken et al., 2016). During recordings, rats were head-fixed, awake, and passively viewing stimulus presentations on a gamma corrected 24” LCD screen (1280x768 pixels at 60 Hz). Animals were positioned sideways next to the screen, with a closest eye-to-screen distance of 20.5 cm and a 40° angle formed by the screen and the rostrocaudal axis. We used single high-impedance tungsten electrodes (FHC, Bowdoin, ME; ordered with impedance 5 to 10 MΩ) fixed in a Biela Microdrive (Crist Instruments, Hagerstown, MD) for recordings of spiking activity in V1 and LI. The recording chambers were placed so that the electrode could enter the cortex orthogonally to record in different layers of V1 (Rat 1) or at an angle of 45° to enter in V1 and be able to reach LI (Rats 2, 3, 4, and 5). At each recording site of the identity oddball experiment, we tried to record large spikes of at least one cell for single-unit isolation. On-line single unit isolation was achieved by setting a threshold for either the peak or the trough of the spike waveform. The stimuli and their location of presentation were chosen for this isolated unit. In the position oddball experiment, where the intent was to record only multi-unit activity, the threshold was set low to get spikes from multiple neurons. Once off-line, all data still went through a procedure of spike detection and spike sorting using our own custom

Matlab (The MathWorks, Inc., Natick, MA) code, in order to retain a multi-unit cluster per recording site and one or more other single units if possible. Specifically, spikes with a peak-to-peak amplitude divided by the standard deviation of the noise (i.e. signal-to-noise-ratio (Issa and DiCarlo, 2012)) of at least 4 were detected and clustered with KlustaKwik 1.6 based on the first n principal components of their waveform (where n was optimized to the site). Automatic clustering was followed by a manual check based on properties such as spike waveforms, changes across time, and inter-spike interval histograms. At this stage all non-spike-waveforms were removed from the data and all remaining spikes were merged into one multi-unit cluster. Of all single units, the minimum signal-to-noise-ratio was 7.

4.3.5 Primary data analysis

For all analyses we used the spike count in the first 100 ms of the response and the spike count of the 100 ms of the baseline before stimulus onset. The first 100 ms response window was chosen for consistency across areas because LI neurons do not really show a longer sustained response. It should be noted however that including the sustained response of V1 neurons in their analysis does not affect the results in a qualitative way. The latency of the 100 ms response window was set at 20 ms after stimulus onset for V1 spiking activity, and 40 ms after stimulus onset for LI spiking activity. These values were chosen based on visual inspection of population peristimulus time histograms across all units and conditions per area. Note that also using a 40 ms latency window for analyzing V1 data did not affect the results in any notable way.

Based on Kaliukhovich and Vogels (2014) we included only those unit × stimulus combinations where the stimulus (A or B) evoked a positive net response (spike count in baseline subtracted from that in response window) in at least one of three conditions (standard, deviant, or control). Specifically, we required the p-value resulting from a trial-wise right-tailed Wilcoxon signed rank test to be lower than .05/3 for at least one condition. For each unit we then calculated condition averages from the spike counts, by averaging first across trials of the same stimulus and then across stimuli in case both were selected. If none of the two stimuli (A or B) were selected for a particular unit, then we did not include it in the analysis.

Response equivalence index

Our neurons or multi-unit recording sites can differ in the extent to which they respond more to the stimuli we used in the oddball blocks, compared to the 8 additional stimuli used to fill the equiprobable control blocks. To quantify this, we calculated a response equivalence index (*EI*) from the responses in the control blocks:

$$EI = \frac{R_{A,B} - R_{c,d,e\ldots}}{|R_{A,B}| + |R_{c,d,e\ldots}|},$$

with $R_{A,B}$ the average net response to the responsive stimuli that are used in the oddball blocks (A and/or B) and $R_{c,d,e\ldots}$ the average net response to the 8 additional stimuli (c, d, e …). This index's values can range from -1 (no response to the stimuli of interest: A and/or B) to 1 (no response to the 8 additional stimuli: c, d, e …). A value of zero would indicate an equal average response to both stimulus groups. Note that for calculating $R_{A,B}$ we only used those stimuli that were used to calculate responses for the standard and deviant conditions, i.e. those that evoked a positive net response according to the criterion explained in the beginning of this section. In short, a positive value indicates that the neuron or neural site responds stronger to the stimuli used for our standard and deviant conditions, compared to the 8 additional stimuli used for the equiprobable blocks.

Multi-level models

If we want to pool the data across rats for inference, we should take into account the dependencies between our observations. After all, these neural units are nested within rats, and the level of rats is nested within the areas we recorded in (except for rat 2, for which we recorded both in V1 and LI). Multi-level models take into account these different levels of variability in the data. We use a lognormal distribution to model the right-tailed, positive-only distribution of average raw firing rates.

Model 1. In the first model we compare standard, deviant and control conditions and include the data of all rats to allow for inference on the animal population level. Instead of treating the baseline firing rate as fixed and known, we model it as a random variable together with the response firing rate. The mean firing rate y_i (for condition and unit combination i = 1,…I, with I = 3 conditions × U units) is modelled as the linear combination of the predicted baseline b'_i (on the log scale) and 3 predictors per neural unit u (denoted as α_{cu}, for condition c = 1,2,3 and unit u = 1,…,U):

$$y_i \sim \text{lognormal}\big(b'_i + \alpha_{1u[i]} + \alpha_{2u[i]}X_{Si} + \alpha_{3u[i]}X_{Di}, \sigma\big),$$

$$\text{for } i = 1, \dots, I.$$

Index variable $u[i]$ codes unit membership for data point i. The unit intercepts are set to the control condition and are captured by regression weights α_{1u}. X_S and X_D are indicator variables (with a value of either 0 or 1) for the standard and deviant conditions, respectively. For each condition c we allow the regression weights α_{c*} to vary across neural units according to a t distribution with a standard deviation σ_c and ν_c degrees of freedom, which makes our model robust for outliers on the level of neural units (Lange et al., 1989). The average for these t distributions is a linear combination of predictors indicating the area, experiment, and rat × area in which the unit was recorded:

$$\alpha_{cu} \sim t\big(\beta_{V1c} + \beta_{LIc}X_{LIu} + \beta_{EXPc}X_{EXPu} + \gamma_{cr[u]}, \sigma_c, \nu_c\big),$$

$$\text{for condition } c = 1,2,3, \quad \text{unit } u = 1, \dots, U.$$

Index variable $r[u]$ codes rat × area membership for unit u. For each condition, the intercept is set at V1, with an indicator X_{LI} for LI recordings, and an indicator X_{EXP} for the position oddball experiment. Rat × area parameters γ_{j*} capture variability across rats (and the interaction with area for one rat) and are estimated by a normal distribution with standard deviation σ_{RATc} and a sum to zero constraint per condition c:

$$\gamma_{cr} \sim N(0, \sigma_{RATc}), \quad \sum_{r=1}^{R} \gamma_{cr} = 0,$$

$$\text{for condition } c = 1,2,3, \quad \text{rat × area } r = 1, \dots, R$$

Simultaneously, an almost identical model is fit for the baseline responses. The only difference is that for the baseline model we don't differentiate between standard and deviant (i.e. two conditions within the oddball block), because trials of these conditions are intermixed in the same blocks. For each baseline firing rate b_i the average of the lognormal distribution is a linear combination of 2 predictors per neural unit u (denoted as $\alpha^{base}{}_{bu}$ for block-type $b = 1{,}2$):

$$b_i \sim \text{lognormal}\big(\alpha^{base}_{1u[i]} + \alpha^{base}_{2u[i]} X_{ODDi}, \sigma\big),$$

$$\text{for } i = 1, \dots, I.$$

Index variable $u[i]$ codes unit membership for data point i. Again, the unit intercept is set to the control condition and is captured by regression weights $\alpha^{base}{}_{1u}$. X_{ODD} is an indicator for the oddball blocks (both standard and deviant conditions). This is the only difference with the response firing rate model, meaning that the modelling of baseline parameters $\alpha^{base}{}_{bu}$ is identical to that of response parameters α_{cu} explained earlier.

The model was fit by generating 30000 samples from the posterior distribution with the probabilistic programming language Stan (Stan Development Team, 2016a), using the RStan interface (Stan Development Team, 2016b) for R (R Core Team, 2015). We used uniform priors for all regression weights, and the standard deviation of the lognormal distribution. For all other standard deviation parameters (expressing variability across neural units and rats), we used a half-Cauchy distribution with location 0 and a uniform hyperprior for the scale parameter. This prior distribution is recommended for multi-level models in cases where the number of groups (in our case neural units and rats) is small (Gelman, 2006). For statistical inference we report the mean of a parameter's posterior distribution and its 95% interval (containing 95% of the posterior density) to express uncertainty.

The use of the lognormal distribution does affect our interpretation of regression coefficients with relation to the original scale (average firing rate). Instead of reporting parameters on the log scale, we report parameter δ that expresses predicted net responses on the original linear scale for deviant and standard conditions relative to those for the control condition:

$$\delta_{SC} = \frac{\left(S - S^{base}\right)}{\left(C - C^{base}\right)}, \delta_{DC} = \frac{\left(D - D^{base}\right)}{\left(C - C^{base}\right)},$$

with *S*, *D*, and *C* indicating the predicted gross responses on the linear scale to a stimulus presented as a standard, deviant, and control, respectively. The superscript *base* indicates the corresponding predicted baseline responses. For example, for V1 responses this relative predicted net response of the standard condition δ_{SC} is calculated as follows (with *S* indicating standard condition, *C* the control condition which also serves as intercept, and *base* indicating baseline model parameters):

$$\delta_{SC} = \frac{\exp\left(\beta_{V1C}^{base} + \beta_{V1S}^{base} + \beta_{V1C} + \beta_{V1S}\right) - \exp\left(\beta_{V1C}^{base} + \beta_{V1S}^{base}\right)}{\exp\left(\beta_{V1C}^{base} + \beta_{V1C}\right) - \exp\left(\beta_{V1C}^{base}\right)}$$

Thus, a δ_{SC} value of for example 0.5 would mean that the net response for the standard condition is 50% of the net response for the equiprobable control condition. For area LI, the position oddball experiment, or individual rats, the relevant coefficients (β_{LI^*}, β_{EXP^*}, or γ_{*r}, respectively) are included in numerator and denominator of the equation.

Model 2. In the second model we now compare within each rat the standard, deviant and control conditions at different time points in oddball or control blocks, or for different stimulus histories. This model is fit separately for each rat, so it does not allow for inference on the animal population level. As in Model 1, the response firing rate y_i is modeled as a linear combination of the predicted baseline b'_i (on the log scale), an intercept (see below) and *J* predictors per neural unit *u*. Again, the baseline firing rate b_i is modeled simultaneously and the only difference is that that the baseline model does not differentiate between standard and deviant conditions.

Contrary to Model 1, we now set the neural unit intercept to the presentation of the stimulus (A/B) at the first trial in a block. This should be independent of whether the stimulus will become a standard, deviant, or control in that block. Note that for the control blocks we only use responses to the stimuli that are used as standard or deviant in the other blocks. The *J* predictors are 3 × *N* indicators for the three conditions (standard, deviant, or control) at *N* different time points (first application of this model) or for *N* different preceding stimulus presentations (second application of this model). This means that now we model the mean firing rate y_i for each condition, unit *and* time-point/stimulus history combination $i = 1,\ldots I$, with $I = 3$ conditions × *U* units × *N*. Again, we make our model robust for outliers on the neural unit level, by allowing the regression weights α_{0*} and α_{j*} to vary according to a t distribution with averages π_0 and π_j, standard deviations σ_0 and $\sigma_{c[j]u}$, and ν_0 and ν degrees of freedom. The degrees of freedom parameter ν is set to be the same for all J predictors. The full model (excluding the baseline part) is the following:

$$y_i \sim \text{lognormal}\left(b_i' + \alpha_{0u[i]} + \sum_{j=1}^{J} \alpha_{ju[i]} X_{ji}\, , \sigma \right), \quad \text{for } i = 1, \dots, I, \quad \text{unit } u = 1, \dots, U.$$

$$\alpha_{0u} \sim t(\mu_0, \sigma_0, \nu_0),$$

$$\alpha_{ju} \sim t\left(\mu_j, \sigma_{c[j]u}, \nu\right),$$

$$\text{for predictor } j = 1, \dots, J, \quad \text{for condition } c = 1, \dots, 3, \quad \text{unit } u = 1, \dots, U.$$

Again, we used uniform priors for all regression weights, and the standard deviation of the lognormal distribution. For all other standard deviation parameters (expressing variability across neural units), we used a half-Cauchy distribution with location 0 and a scale of 1.

As we do for the results of Model 1, we report a parameter that expresses the predicted net responses on the original scale relative to the intercept. With the intercepts set at the occurrence of the stimulus at the first trial number in a block, this parameter (denoted as δ_1) expresses predicted net responses for deviant, standard, *and* control conditions relative to those to the stimulus in the first position of the block. The neural data and code for fitting these multi-level models are available at https://osf.io/mecg4/ .

4.3.6 FURTHER DATA ANALYSIS FOR SUPPLEMENTAL FIGURES

Cortical layer sampling bias

As reported before (Vermaercke et al., 2014; Vinken et al., 2016), there is a potential sampling bias of upper cortical layers for V1 recordings compared to LI recordings when entering the cortex (i.e. in V1) at an acute angle as we did in order to reach LI. This might explain differences between V1 and LI findings within Rat 2 (and between Rat 2 V1 and all other LI recordings). To eliminate this possible confound, we penetrated the cortex orthogonally for recordings across the entire width of V1 in Rat 1. The location of the

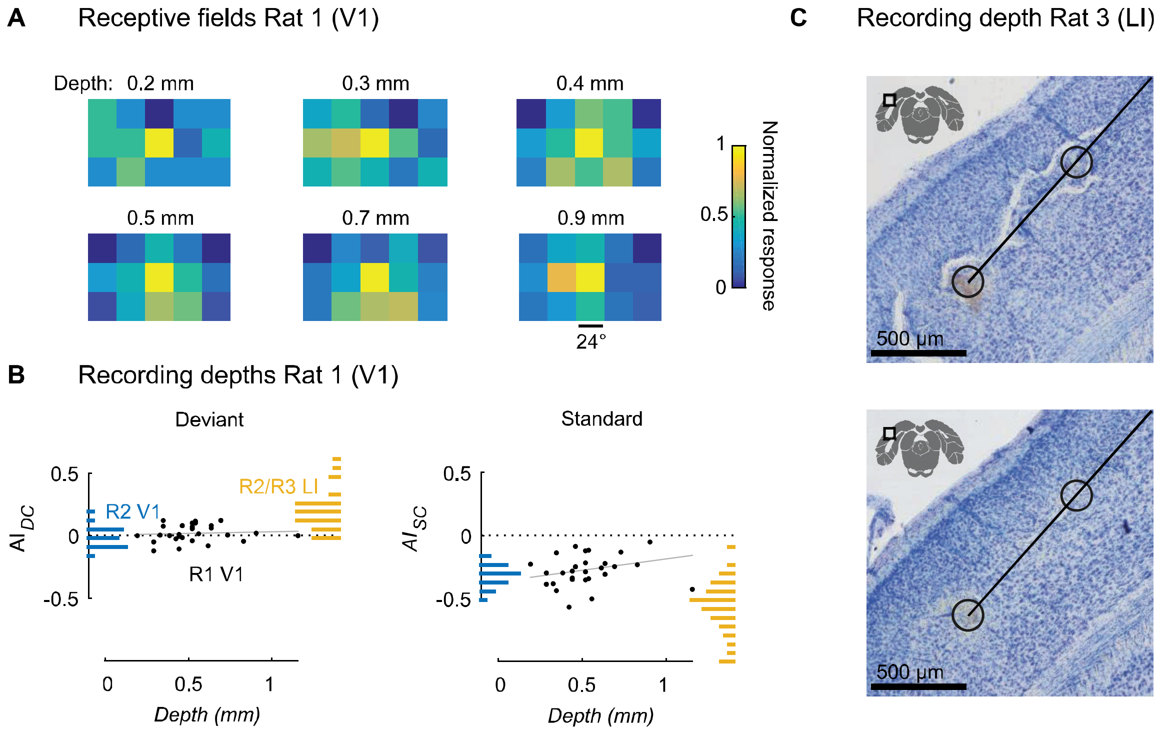


Figure 4.4. Recording depths in V1 (Rat 1) and LI (Rat 3).
Related to **Figure 4.2**. **(A)** The receptive field was mapped by presenting a hashtag symbol (diameter of 24 visual degrees) at 15 locations (see supplemental information, Oddball paradigm). The optimally responsive position on the computer screen does not change when advancing the electrode in Rat 1 (see supplemental information). This confirms that the penetration is orthogonal to the cortical surface (data obtained in one recording session). **(B)** Adaptation indices for deviant or standard stimuli for Rat 1 recordings are independent of the recording depths covering about 1 mm of cortex (grey lines indicate robust linear regression fit). The distributions of indices correspond with those from V1 recordings in the other Rat 2 (blue histograms), as opposed to those from LI recordings in Rat 2 and 3 (yellow histograms). **(C)** Histological sections (Nissl-stained) from Rat 3 with electrolytic lesions made along the electrode track indicate cortical recording depths fully covered by those of the V1 recordings in Rat 1.

receptive field remained constant across recording depth, demonstrating that the penetrations in Rat 1 were orthogonal indeed (**Figure 4.4**A).

It is clear from **Figure 4.2** that recordings in this rat agree with Rat 2 V1 results, as opposed to the LI results in all other rats. Furthermore, in Rat 1 there is no evidence for a correlation between recording depth and adaptation of the standard AI_{SC} (single-unit data: spearman correlation r_s = -0.16, p = 0.36; multi-unit data: r_s = 0.28, p = 0.15) or between depth and any effect for the deviant AI_{DC} (single-unit data: r_s = 0.07, p = 0.70; multi-unit data: r_s = 0.24, p = 0.21). The data behind these correlations are visualized in **Figure 4.4**B. The depths of the MU recording sites measured from the point of entry in the cortex ranged from about 200 µm up to about 1200 µm, covering the typical thickness of the rodent cortex of ~1 mm. Sampling was densest around a depth of 500 μm. It turns out that this distribution of depths overlaps substantially with the estimated distribution in our LI recordings. Evidence comes from histology in Rat 3. After the last recording session, we made two small electrolytic lesions at the functionally defined medial and lateral boundaries of LI in order to histologically determine the cortical depth of recording sites. Two nearly adjacent Nissl-stained sections, processed following previously described procedures (Vermaercke et al., 2014), show that we recorded LI MU activity at a cortical depth of about 500 μm in Rat 3 (**Figure 4.4**C), which overlapped with the V1 depths for Rat 1. Therefore, we conclude that a layer sampling bias cannot explain the marked and consistent differences in experimental results between V1 and LI.

Effect of trial number of stimulus presentations within blocks

In the main text we have quantified the effect for a standard and for a deviant stimulus in the visual oddball paradigm. The question remains whether these effects depend on the trial number at which the stimulus is presented within a block. Specifically, we expect the response to a deviant or a standard occurring in the beginning of a block to be different from when they occur later in a block. Typically a response reduction for both the standard and the deviant is observed compared with the first stimulus presentation (Kaliukhovich and Vogels, 2014; Nieto-Diego and Malmierca, 2016). The net response of each target stimulus (A or B) on the first trial in a block was averaged across all block types. The remaining 99 trial numbers of each block were grouped in 11 sets of 9 trials each to calculate an average net response per condition per set (first 9, second 9,…, eleventh 9 trials). Thus, for each stimulus × neural site combination, we have one value for the stimulus in the first trial of a block, and 11 values per condition: standard, deviant, and control.

Figure 4.5 summarizes the results per rat × area combination. The leftmost plot in each panel displays the averages across neural sites per trial number (set) and condition. The middle plot in each panel displays for each condition the multi-level model estimates for each of the 11 9-trial number sets, expressed proportional to the net response in the first trial (denoted as δ_1). The rightmost plot in each panel displays for each trial number set

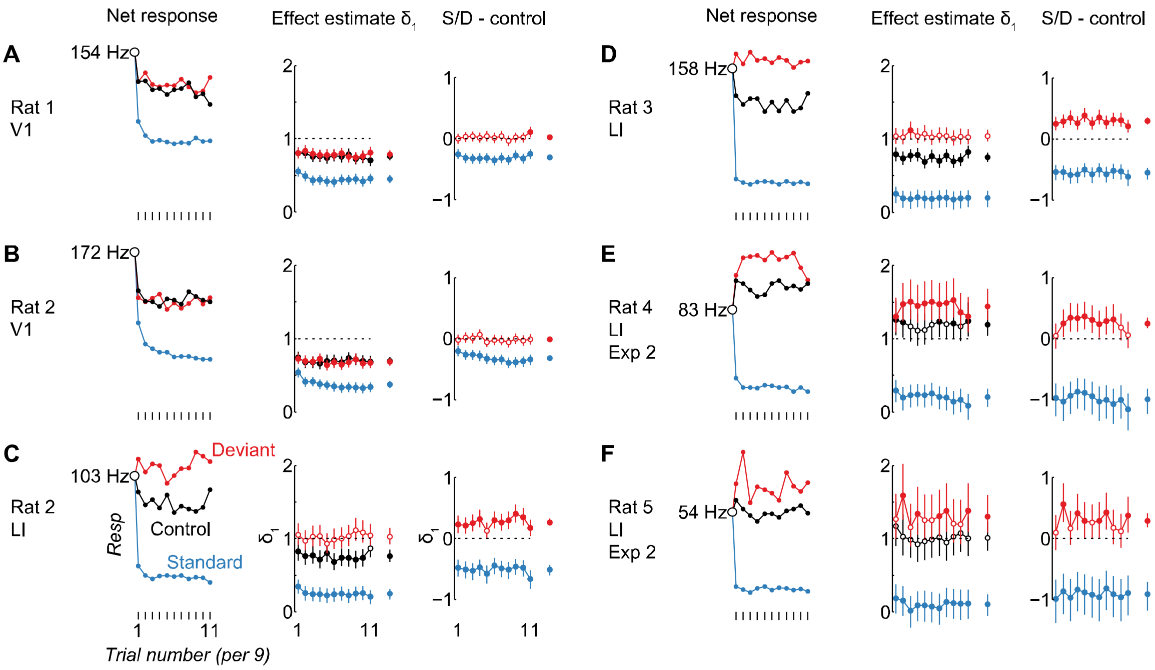


Figure 4.5. The effect of the trial number on multi-unit net responses in the visual oddball task.
Related to **Figure 4.2**. **(A-F)** For each rat × area combination: net multi-unit responses and effect estimates as a function of trial number. The left plot shows the average net response for the first trial (open black marker), with the firing rate indicated next to it. For the remaining 99 trials in a block, the average net response for each condition (blue for standard, red for deviant, and black for control) is plotted per set of 9 trial numbers. The middle plot shows for each of these trial number sets (and per condition) the estimated δ_1 effect, which is the estimated ratio of a trial number set's net response and the net response at trial 1. The right plot shows for each of these trial number sets (for both standard and deviant conditions) the difference with the control condition calculated from these multi-level model estimates. Marginal effects (average across the 11 trial number sets) are indicated in the right margin of each plot. All error bars in this plot indicate 95% intervals. Estimates for which the 95% interval excludes 1 or 0 (i.e. the standard, deviant or control's net response is increased or decreased relative to the first trial's or the control's net response, respectively), are indicated with filled markers.

the difference between these multi-level model estimates for the standard/deviant condition and those of the control condition.

For the V1 data (**Figure 4.5**A, B) it is clear that in both rats the average net response in trial numbers 2-10 (first set) is already well below the net response in trial 1 (95% intervals fall below 1). While this effect is strongest for the standard condition, it is also present for both the deviant and control conditions. Compared to the control condition, deviant responses do no really seem to differ (95% intervals are centered on 0), while responses for the standard are significantly lower and increasingly so across the first 3 to

5 trial number sets. In sum, we see increasing stimulus specific adaptation for the standard, as well as adaptation for the control and the deviant.

The results of the LI data from the same experiment (**Figure 4.5**C, D) paint a different picture. As for the V1 data, we see a response reduction for the standard and the control from the first set onwards (95% intervals fall below 1). In addition, deviant net responses across the entire block length are similar to the net response in the first trial (95% intervals are centered on 1). This means that the difference between deviant and control could be explained by stronger cross-stimulus adaptation in the equiprobable blocks.

In the next experiment we attempted to rule out the effect off cross-stimulus adaptation in the control condition, by using different stimulus positions that could be outside the receptive field instead of different stimuli. For these data (**Figure 4.5**E, F), there is no longer a reduction in response for the control condition across the block length (95% intervals are above or centered on 1), confirming the absence of cross-stimulus adaptation in the equiprobable blocks. In fact, in Rat 4, the average net response for the control condition is actually increased from the start of the block compared to the net response in trial 1. Importantly, deviant net responses are elevated compared to the control condition from the second trial number set onwards.

Effects of stimulus history

In addition to the trial number at which a stimulus is presented in a block, we expect the local stimulus history to affect the responses. For example, as was shown for monkey IT (Kaliukhovich and Vogels, 2014), the response for a deviant and even a standard stimulus can depend considerably on whether the previous stimulus was the same or different. A related question is how the enhanced response for the deviant condition is affected by the presentation of a deviant at different proximities in time. In particular, how many successive presentations of the standard do we need before we see an enhanced response for the deviant?

In order to answer these questions, we divided the responses into conditions based on the stimulus history of 5 trials back. Specifically, for the oddball blocks we look at whether and when a deviant occurred in one of the 5 previous trials. This results in 6 independent conditions which can be represented as follows: SSSSS ?, DSSSS ?, SDSSS ?, SSDSS ?, SSSDS ?, SSSSD ?, where 'S' indicates standard, 'D' deviant, and '?' the actual stimulus of interest, which could be both a standard or a deviant. For this analysis we omit the data where more than one deviant was presented in the previous 5 trials, since this was too rare to reliably estimate all the possible interactions. In addition, by definition we omit the first 5 stimulus trials from these conditions. However, the response for the first trial was included in the analysis in the form of the intercept of our multi-level model. Finally, for the equiprobable blocks we look at whether and when a same stimulus occurred in one of the 5 previous trials. Again, we only look at the data with only one same stimulus in the previous 5 trials, to have an appropriate reference for the 6 deviant conditions.

Figure 4.6 summarizes the results per rat × area combination. In V1 (**Figure 4.6**, A, B) response reduction relative to the first trial in a block is observed in all conditions (all 95% intervals fall well below 1 in both Rat 1 and 2). The responses for the deviant and control conditions are similarly reduced by the proximity of a recent presentation of the same stimulus (all but one of the 95% intervals in the deviant – control column for V1 recordings include 0). The response for the standard is relatively independent of local stimulus history: only if a deviant was presented immediately before the standard, we see a noticeably higher response compared to no deviant in the previous 5 positions (standard – history 6 95% intervals fall above 0 in both Rat 1 and Rat 2).

In LI recordings for the identity oddball experiment (**Figure 4.6**C, D), response reduction for the deviant relative to the first trial in a block is only observed when the last deviant was presented up to three trials back (95% intervals fall below 1 in both Rat 2 and 3). What is more, response enhancement for the deviant relative to the first trial in a block is observed when it was preceded by 5 or more consecutive presentations of the standard (95% intervals fall above 1 in both Rat 2 and 3). When comparing to the corresponding control conditions, the response for the deviant is enhanced when it is separated by 3 or more standards from the last deviant (deviant – control 95% intervals fall well above 0 in both Rat 2 and Rat 3). Again, the response to the standard only seems to be affected by a

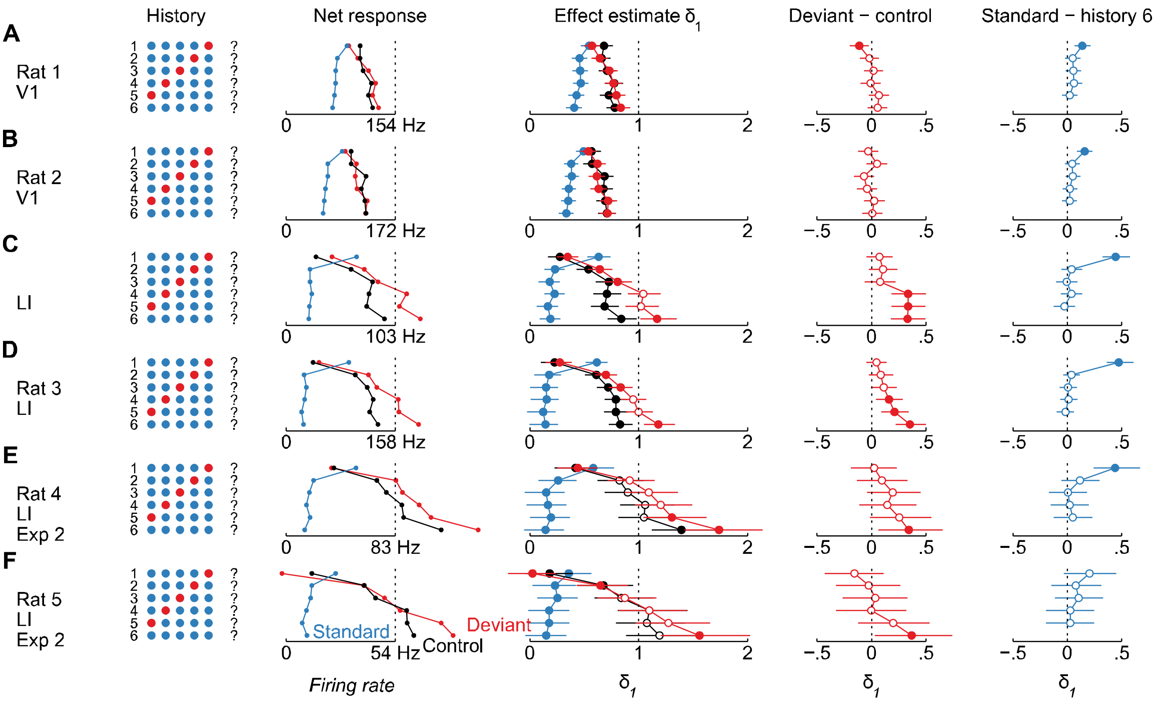


Figure 4.6. The effect of the local stimulus history on multi-unit net responses in the visual oddball task.
Related to **Figure 4.2**. For each rat × area combination: responses to a standard (blue), deviant (red), or control (black) condition separately for 6 local stimulus histories defined by the 5 previous stimulus presentations. The leftmost column indicates these 6 histories: 5 colored circles indicate the nature of the 5 stimuli that preceded the current stimulus (which is indicated by '?'), for which the actual response is plotted. For standard (blue) and deviant (red) conditions, blue and red circles indicate previous presentations of a standard and deviant stimulus, respectively. For the control (black) condition, a red circle indicates a presentation of the same stimulus as the current stimulus, and a blue circle the presentation of a different stimulus. The second column displays for each condition (standard, deviant, or control) the net response averaged across neural sites per local stimulus history. The third column shows for each condition the multi-level model estimates δ1 for each of the 6 local stimulus histories. These values reflect the net response expressed proportional to the net response in the first trial. The fourth column shows per condition the difference between the deviant and control condition calculated from these multi-level model estimates. The rightmost column shows for the standard condition the difference between the multi-level model estimates for each local stimulus history containing a deviant and that without a deviant. Error bars in the last three columns indicate 95% intervals. Filled markers indicate that this estimate's 95% interval excludes the value specified by the vertical dashed line.

deviant presented in the previous trial (standard – history 6 95% intervals fall well above 0 in both Rat 2 and Rat 3). This effect is very strong compared to the V1 recordings, with the reduction relative to the first trial in a block now being roughly half of what it is for other local stimulus histories.

The results of our LI recordings for the position oddball experiment (**Figure 4.6**E, F), tell a similar story. Response reduction for the deviant relative to the first trial in a block is only observed when the last deviant was only 1 or 2 trials back (95% intervals fall below 1 in Rat 4 and 5), while response enhancement is observed when it was more than 4 or 5 trials back (95% intervals fall above 1 in Rat 4 and 5). In comparison with the corresponding control conditions, the response enhancement for the deviant occurs when it is separated by 5 or more standards from the last deviant (deviant – control 95% intervals fall above 0 in Rat 4 and Rat 5). In contrast with all other rats, for Rat 5 (**Figure 4.6**F) the response to the standard does not seem to be clearly affected by a recent presentation of a deviant (no standard – history 6 95% interval falls completely above 0 in Rat 5). We purposefully refrain from trying to interpret the relatively small differences between rats in the identity and position oddball experiments, since we lack the data to discern experimental effect from inter rat variability here. Importantly, the enhancement of the response for the deviant is present and significant in the LI data of each of the 4 rats, regardless of the experiment.

4.3.7 Further data analysis on single-unit data and local field potentials

In the main text and the primary data analysis we focus upon multi-unit data. Here we describe further analysis on single-unit data and local field potentials.

Single-unit spiking activity

In this section we discuss the results of the single-unit data we collected for the identity oddball experiment in both V1 and LI. We isolated 72 responsive single units in V1 (36 in Rat 1 and 36 in Rat 2) and 63 in LI (41 in Rat 2 and 22 in Rat 3). The data for the *AI*s of the standard, relative to control, are consistent with those of the multi-unit activity: we see a clear response reduction for the standard in V1 (AI_{SC} Rat 1: *Median* = -0.24, p = 0.0012, AI_{SC} Rat 2: *Median* = -0.24, p < 0.0001), as well as LI (AI_{SC} Rat 2: *Median* = -0.28, p = 0.0115, AI_{SC} Rat 3: *Median* = -0.46, p = 0.0043). However, contrary to the multi-unit data, for V1 we find some evidence for cross-stimulus adaptation in the form of a response reduction of the deviant in Rat 1 (AI_{DC} Rat 1: *Median* = -0.10, p = 0.0039, AI_{DC} Rat 2: *Median* = -0.03, p = 0.2430). For LI, there was a nonsignificant trend for an enhanced response to the deviant (AI_{DC} Rat 2: *Median* = 0.06, p = 0.5327, AI_{DC} Rat 3: *Median* = 0.05, p = 0.1892).

This trend is smaller than the effect for our MU data, and might actually be attributable to cross-adaptation in the control blocks.

The absence of a significant enhancement for the deviant in the LI single units suggests that the data from our single-unit samples might not be very representative of our multi-unit data. One possible explanation is that of a selection bias in the type of single neurons that we recorded from, leading to results that are not representative of the general neural population. This is a known issue with extracellular recordings (Towe and Harding, 1970). In addition, multi-unit activity cannot be normalized for each individual neuron's response strength. Therefore, just like with un-normalized single-unit data, neuronal types such as interneurons with a high firing rate will contribute relatively more to the measured multi-unit firing rates. We tested this hypothesis by summing the firing rates of single neurons and calculating the *AI*s from these summed responses. For statistical inference, we calculated bias-corrected accelerated bootstrap confidence intervals (Efron, 1987) based on 10000 random neuron samples. These adaptation indices did not differ from the Median values reported in the previous paragraph, but the deviant enhancement for LI is now significant as a result of a narrower confidence interval (V1: Rat 1: AI_{SC} = -0.23, CI [-0.33 -0.14], AI_{DC} = -0.09, CI [-0.15 -0.05], Rat 2: AI_{SC} = -0.23, CI [-0.28 -0.19] , AI_{DC} = -0.05, CI [-0.10 -0.02]; LI: Rat 2: AI_{SC} = -0.45, CI [-0.53 -0.34], AI_{DC} = 0.06, CI [0.01 0.12], Rat 3: AI_{SC} = -0.49, CI [-0.61 -0.39] , AI_{DC} = 0.08, CI [0.01 0.12]). The reason is that this approach is not affected by erratic index values that can result from calculating indices from single neurons with low unreliable firing rates.

Local field potentials

Simultaneously with spikes, we recorded local field potentials (LFPs) which were sampled at 1 kHz and band-passed between 1.66 Hz (stimulus + inter-stimulus interval presentation frequency) and 170 Hz. Line noise was removed by means of a 50Hz notch filter. Spectral analysis was based on a time-frequency Morlet wavelet decomposition as described previously (Kaliukhovich and Vogels, 2014), using Fieldtrip Toolbox (F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, the Netherlands; http://fieldtrip. fcdonders.nl). Frequencies below 15 Hz were excluded from the wavelet analysis to avoid wavelets overlapping multiple stimulus presentations. Visually evoked potentials (VEPs) were computed by stimulus-locked averaging of the LFPs per condition. Trials for which the LFP signal exceeded the 5-95% window of the total input range were removed.

In total, we recorded the LFP signal during the identity oddball experiment in 47 V1 sites (19 in Rat 1, 28 in Rat 2) and in 52 LI sites (30 in Rat 2, 22 in Rat 3) and during the position oddball experiment in 50 LI sites (26 in Rat 4, 24 in Rat 5). Visual inspection of the LFP power spectra (**Figure 4.7**A, B) supports a stronger suppression for the standard in LI compared to V1, as well as an enhanced response to the deviant in LI. More detailed analysis of frequency bands per rat × area combination shows that, in particular for gamma frequencies, power in the first 100 ms of the response is significantly enhanced for the deviant in LI only (**Figure 4.7**C-H, left plots). Analysis of the VEPs per rat × area show the same deviant enhancement consistently for the first peak in LI recordings (**Figure 4.7**C-H, right plots).

Summarized, LFPs confirm all the major findings observed in MUA: (1) a difference in responses to the deviant and standard that (2) was bigger in LI compared to V1, with (3) a surprise-based response enhancement in LI.

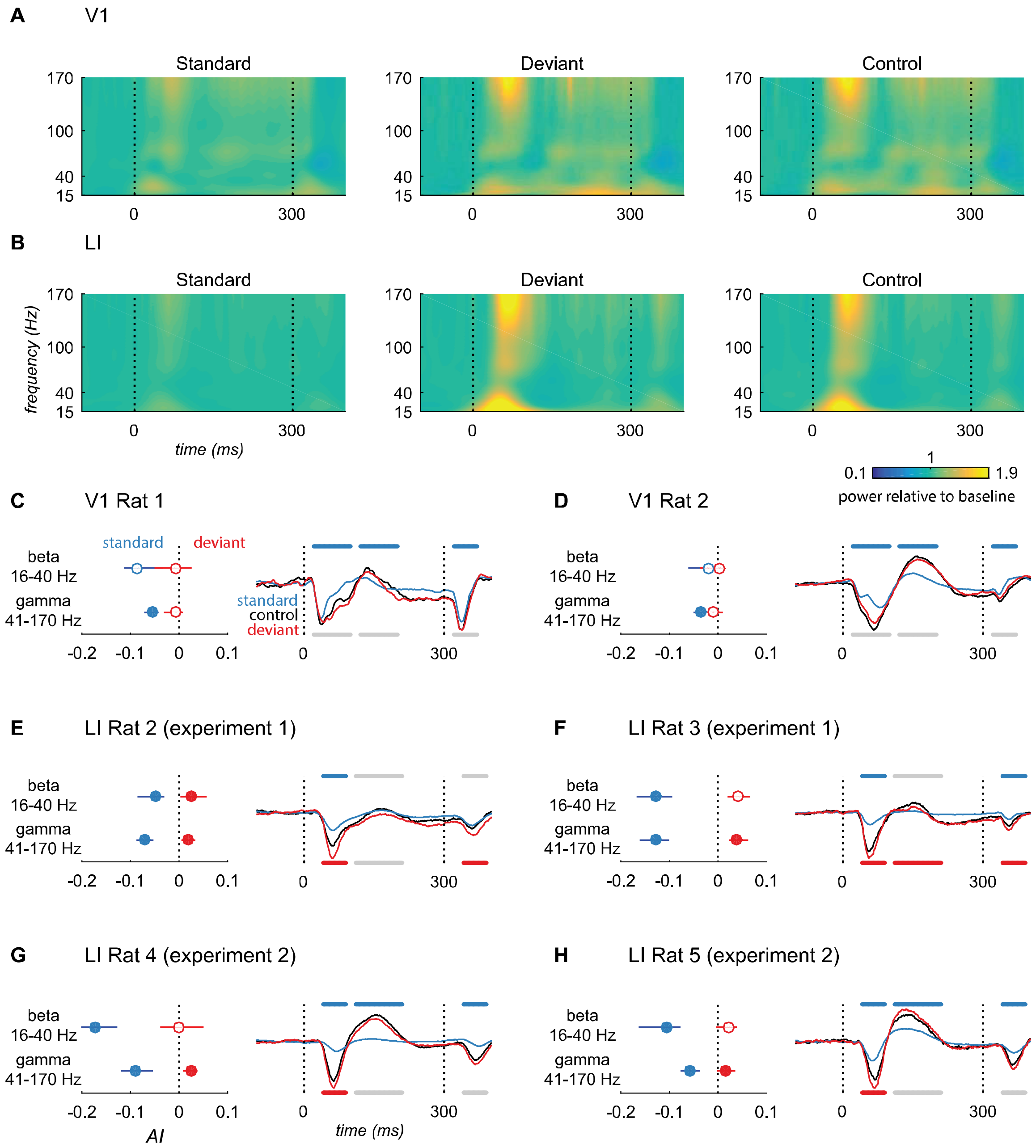
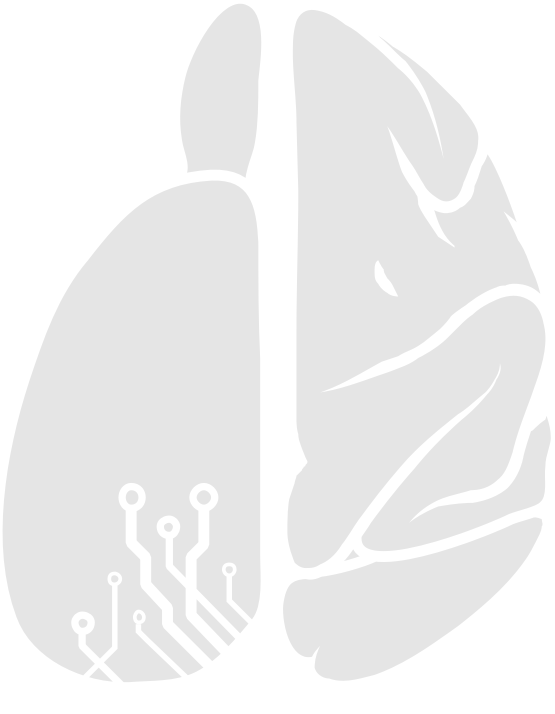


Figure 4.7. Analysis of LFP data.

Related to **Figure 4.2**. **(A, B)** average (across sites, then across animals) LFP power spectra for V1 and LI recordings, respectively. **(C-H)** Left: Median adaptation indices for standard and deviant stimuli per frequency band (beta: 16-40 Hz, gamma: 41-170 Hz). Filled makers indicate a p-value <.05 for a two-tailed sign test against a Median of 0. Error bars indicate 95% bias-corrected accelerated bootstrap confidence intervals (Efron, 1987) based on 10000 random recording site samples. Right: average VEPs (first normalized by absolute peak response for the control condition) for standard (blue), deviant (red), and control (black) stimuli. Adaptation indices were calculated for standard and deviant stimuli from the average VEP signal of the three peaks in V1 (20 ms latency + 1-80 ms, 101-180 ms, and 301-350 ms) and in LI (40 ms latency + 1-50 ms, 71-170ms, and 301-350 ms). Blue colored bars at the top of each graph indicate a p-value <.05 for a two-tailed sign test of the Median adaptation index for the standard against a Median of 0. Red colored bars at the bottom of each graph indicate the same for the Median adaptation index for the deviant.

III ADAPTATION AND EXPECTATION IN MACAQUE VISUAL CORTEX



Chapter 5.

THE PERCEPTUAL EXPECTATION ACCOUNT OF NEURAL ADAPTATION

In contrast with fMRI studies, expectation effects on repetition suppression could not be replicated in neural responses in macaque IT (Kaliukhovich and Vogels, 2011). Subsequent fMRI studies pointed to the importance of attention (Larsson and Smith, 2012) or face specificity of the effect (Kovács et al., 2013). *Are these two conditions sufficient for observing expectation effects on repetition suppression in macaque IT?*

FACE REPETITION PROBABILITY DOES NOT AFFECT REPETITION SUPPRESSION IN MACAQUE VISUAL CORTEX

Repetition suppression, which refers to reduced neural activity for repeated stimuli, is typically explained by relatively simple adaptation mechanisms (Vogels, 2016). However, recent theories have emphasized the role of top-down processes, suggesting that this response reduction reflects the fulfillment of perceptual expectations. To support this, an influential functional magnetic resonance imaging (fMRI) study (Summerfield et al., 2008) showed that the magnitude of suppression is modulated by the probability of a repetition. No such effect was found in macaque inferior temporal (IT) cortex (Kaliukhovich and Vogels, 2011), calling into question the generality of the role of top-down mechanisms. Here, we combined three measures of brain activity in search for expectation effects: spiking activity, local field potentials (LFPs), and fMRI. Specifically, we investigated two conditions that might be necessary: using face stimuli (Kovács et al., 2013) and a stimulus related task (Larsson and Smith, 2012). In an experiment similar to Summerfield et al. (2008), we simultaneously recorded spiking activity and LFPs in middle lateral face patch (ML) of one monkey (male), and a face-responsive region of another (female). While we observed clear repetition suppression, there were no effects of repetition probability, even when in a second experiment repetitions were task-relevant. Next, we performed a separate fMRI study with the same animals. Here, we did find effects of repetition probability which were inconsistent with direct measures of neural activity and in opposite directions for each monkey. In conclusion, even with face stimuli and a stimulus related task, we failed to replicate the Summerfield et al. (2008) results in macaque (face-selective) visual cortex. This further challenges a general perceptual expectation account of neural adaptation.

5.1 INTRODUCTION

Sensory processing in the brain does not only depend on the current input from the senses, but is also affected by previous sensory experience. A well-known example is the reduced neural activity when stimuli are repeated, called repetition suppression (Desimone, 1996). Research on this phenomenon is not only important for understanding its role in sensory processing, but also because repetition suppression paradigms are widely used in functional magnetic resonance imaging (fMRI) research (Grill-Spector and Malach, 2001; Grill-Spector et al., 2006; Barron et al., 2016). Several simple bottom-up or local adaptation mechanisms are thought to underlie these changes in neural responses (Vogels, 2016). However, it has also been suggested that repetition suppression can be explained by a reduction of responses that encode a prediction error, through a mechanism involving top-down influences of perceptual expectation (Friston, 2005; Summerfield et al., 2008).

In support of this theory, several studies have found evidence for a stronger fMRI suppression in blocks of trials were a repetition is more frequent, compared to those where a repetition is infrequent (Summerfield et al., 2008; Larsson and Smith, 2012; Kovács et al., 2013; Grotheer and Kovács, 2014). These repetition probability effects were originally reported for the fusiform face area (FFA; Summerfield et al., 2008) and later generalized to other upstream visual areas (Kovács et al., 2012; Larsson and Smith, 2012), but not in every study (Kovács et al., 2013). On the other hand, a single-cell study (Kaliukhovich and Vogels, 2011) found no evidence for an effect of repetition probability on repetition suppression in macaque inferior temporal (IT) cortex. What is more, a recent fMRI study did not support such an effect either (Olkkonen et al., 2017). These results imply that the relation between adaptation and expectation remains controversial. Therefore, an important question is how general the reported effect of repetition probability is.

Several studies have contributed to this question by narrowing down the conditions under which the effect was replicable. First, there is one study suggesting that attention is necessary for the expectation effect to be measurable (Larsson and Smith, 2012). Second, it has been implied that the expectation effects are specific for certain stimulus categories such as faces (Kovács et al., 2013), because of a dependence on prior experience (Grotheer and Kovács, 2014). Both constraints could potentially explain the absence of an effect of repetition probability in the study by Kaliukhovich and Vogels (2011): (a) the monkeys were passively fixating and perhaps paying little to no attention towards the content of the stimuli; (b) fractal patterns or various object images were used instead of faces or another stimulus category that monkeys are familiar with.

Here, we tried to address both issues in an experiment using a paradigm almost identical to that of Summerfield et al. (2008), where we (a) only use face stimuli and (b) make the monkeys perform a stimulus related task that was orthogonal to the manipulation of face repetitions (i.e. the task was unrelated to face repetitions or alternations). In a second experiment, we made face repetitions task relevant in case the orthogonal task would not be sufficient. During these experiments we recorded spiking activity and LFPs in the macaque middle lateral face patch (ML), an area that typically shows face category selective activity (Tsao et al., 2006; Aparicio et al., 2016). ML might be homologous to human FFA based on its location on the occipito-temporal axis (Tsao et al., 2008), but not according to every view (Yovel and Freiwald, 2013).

In a final experiment, we recorded fMRI responses to investigate the possibility that measurable expectation effects are restricted to neuroimaging signals. After all, there is evidence that such signals can contain task-related components that are poorly related to local spiking activity or LFPs (Cardoso et al., 2012; Lima et al., 2014).

Together, these experiments allowed us to investigate several conditions under which repetition suppression might be affected by repetition probability in different brain signals in macaque visual cortex. Concretely, we made the following predictions. First, in all experiments, independent of the task (repetitions relevant or not) or brain signal (spiking activity, LFP, fMRI), we expect stimulus-specific adaptation: more suppression for repetition trials than for alternation trials. Second, we expect an effect of repetition probability on repetition suppression in all experiments if face stimuli, a stimulus-related task, and/or face area specificity are sufficient conditions. Third, if repetitions need to be task relevant, we only expect an effect of repetition probability during such a task and not during the orthogonal task. Finally, if repetition probability effects are restricted to LFPs or fMRI, we only expect to observe the effects in these signals.

5.2 Materials and methods

5.2.1 Subjects

Experiments were conducted with two rhesus macaques (Macaca mulatta; 1 male G and 1 female D). Surgical procedures for implant placement were the same as previously reported (Kaliukhovich and Vogels, 2011). Surgeries were performed for the placement of a head post and recording chamber. The location of the latter was guided with a preoperative anatomical magnetic resonance imaging (MRI) scan. Animal care and experimental procedures were approved by the KU Leuven Animal Ethics Committee and in accordance with the national and European guidelines.

5.2.2 Face patch localization

In each monkey, face-selective patches were localized using a functional MRI (fMRI) experiment described previously in detail (Taubert et al., 2015). Briefly, we used 80 naturalistic greyscale stimuli originally used in Tsao et al. (2003) of 5 categories (16 images each): human faces, human (headless) bodies, fruits, manmade objects, and hands. The images were presented at a visual angle of 8° on a grey background with a red fixation dot in a categorical block design during continuous fixation. The 5 categorical blocks were presented in pseudo-random order and each time all had to be presented before they could be repeated again. Each block duration was 16 s: 16 images presented in shuffled order for 1 s each, no inter-stimulus interval. A fixation block (16 s) was presented after every 5th categorical block presentation. Each block was presented 5 times per run of 490 s.

Imaging data were acquired with a 3 Tesla full-body scanner (MAGNETOM Prisma, Siemens), using a custom-made 8 channel phased-array receive coil and radial transmit-only surface coil (Ekstrom et al., 2008). We used a gradient-echo T2*-weighted echo-planar imaging sequence of 34 horizontal slices (Monkey G; voxel size = 1.5 mm isotropic, TR = 2 s, TE = 15 ms, flip angle = 90°) or 40 horizontal slices (Monkey D; voxel size = 1.25 mm isotropic, TR = 2 s, TE = 18 ms, flip angle = 90°). Signal-to-noise ratio was enhanced with a MION contrast agent (monocrystalline iron oxide nanoparticle, Rienso: Takeda, 8-11 mg/kg) injected intravenously before scanning (Vanduffel et al., 2001).

The functional images were preprocessed separately per day using SPM12 for slice-time correction and spatial realignment to the first volume of the first run. Next, the mean of the realigned functional scans was used to calculate transformation parameters for co-registration with a skull-stripped anatomical MRI of the subject (JIP Toolkit v3.1). After co-registration, the images (resliced at 1 mm isotropic voxel size) were spatially smoothed with an isotropic 3D Gaussian kernel (2 mm full width at half maximum; SPM12).

For statistical analysis we used SPM12 to fit a general linear model to the functional images, estimating regression coefficients per run. Regressors were convolved with the MION response function (Vanduffel et al., 2001) and included one for each block type (image category) as well as motion and eye movement regressors of no interest. Face patches were defined with xjView (v9.0) using a threshold of T = 5 (positive activations only) on the contrast faces versus all other categories.

5.2.3 Repetition probability experiment

For the main experiments, we generated 50,000 images of unique human faces seen from the same frontal perspective (FaceGen Modeler, v 3.5, https://facegen.com/). They were presented in trials of two stimulus presentations of 250 ms separated by 500 ms. In a repetition trial the same image was repeated, while two different faces were shown in an alternation trial (**Figure 5.1**B). A trial was initiated by 500 ms of maintained fixation and was interrupted whenever fixation was broken. For maintained fixation the monkey's gaze had to stay within an area (fixation window) of about 2 by 2 visual degrees centered on the fixation dot. Each face was practically trial unique as a result of the large number of faces and the restriction that all images had to be used once before they could be used again. Compared to some earlier studies (e.g. Summerfield et al., 2008; Larsson and Smith, 2012), our computer-generated face stimuli were relatively homogeneous (viewpoint, lack of hair,…). To make sure that the faces presented in alternation trials were visually distinct, we predetermined face pairs as follows. First, we down-sampled the images to 32 by 32 pixels and unfolded the resulting image matrices into image vectors. Next, we performed principal component analysis on the full set of 50,000 image vectors, and retained only the first 50 principal components. Then, we calculated all pair-wise Euclidean distances in 50D space. Finally, we sequentially selected 25,000 face pairs by each time taking the two images with the maximum pair-wise distance (face dissimilarity) from the remaining pool of images. As a result, the average difference between faces in alternation trials was substantially larger than it would be in the case of random pairings. See **Figure 5.1**A for example face pairs.

The alternation and repetition trials were both presented in blocks of 40, 100, or 120 trials (the number was changed between sessions). A block had either a high or low repetition probability: repetition blocks (75% repetition trials) and alternation blocks (25% repetition trials). The first 5 trials of a block were always of the same type (i.e. repetition or alternation) as the block type. Both block types were presented alternatingly and the type of the first block in a recording session or run was randomly determined. Between

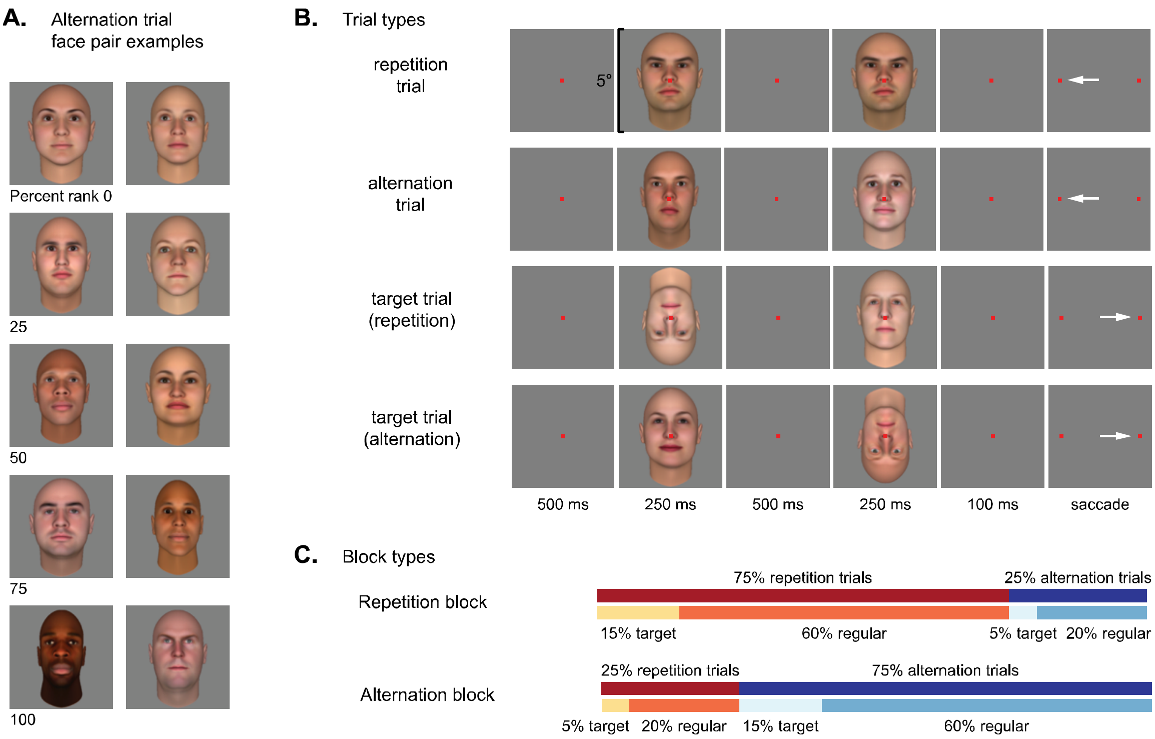


Figure 5.1. Stimuli and experimental paradigm.
(A) Example face pairs that we used for alternation trials, selected according to percent rank number of face dissimilarity (Euclidean distance in 50D PC space; see materials and methods): from the most similar pair with rank number 0, to the most dissimilar pair with rank number 100. Thus, the former is the most similar pair that we used and the latter the most dissimilar. **(B)** The different types of trial sequences. Subjects were required to fixate throughout the entire sequence and give the correct saccade response in order to receive a fluid rewards. Each trial new stimuli were selected until all 50,000 were used (after which the cycle restarted). For target trials either the first or the second face could be inverted. Note that there were no target trials in Experiment 2, where the monkey had to indicate repetition (left) versus alternation (right). **(C)** Composition of repetition and alternation blocks in terms of repetition and alternation trials and regular and target trials.

blocks, there were 5 trials with 300 ms presentations of a full screen color (blue, yellow, green, orange, or purple) during maintained fixation. Except for the task, these procedures were identical to the ones previously described in detail in Kaliukhovich et al. (2011).

Experiment 1: orthogonal task

Each block contained a number of target trials (20%) where either the first or second face was inverted (i.e. presented upside down; **Figure 5.1**B). At 100 ms after the end of the presentation of the second face, the subject was required to indicate whether it had been a regular or target trial by making an eye movement to the left (regular) or right (target). The monkey received a fluid reward after maintaining fixation throughout the trial sequence and giving the correct response. A higher reward for target trials was required to keep the subjects motivated to do the task, because a left response on all trials would already result in 80% correct. Note that we only analyzed responses to regular trials, which all had the same reward for a correct response regardless of being alternation or repetition. See **Figure 5.1**C for an illustration of the composition of blocks in this experiment.

Experiment 2: task relevant repetitions

Experiment 2 was identical to Experiment 1, except for the task: the monkey had to indicate whether a trial was a repetition (left) or an alternation (right). This task makes the face repetitions task-relevant. There were no inverted face trials in this experiment and all trials received the same reward for a correct response. Note that we did not do this experiment in the scanner.

5.2.4 Electrophysiological recordings

We recorded LFPs simultaneously with single or multi-unit spiking activity using Epoxylite-insulated tungsten microelectrodes (FHC Inc., impedance of around 1 MΩ in situ). For every recording session, a single electrode was lowered into the brain with a Narishige microdrive through a stainless steel guide tube that was fixed in a Crist grid. Spikes of single neurons were isolated online using a window discriminator. In addition, when no single neuron could be isolated, spikes of multiple neurons were thresholded online to record multi-unit activity. Stimuli were displayed on a CRT monitor (1024x768 pixels at 75 Hz; Philips Brilliance 202P4) at an eye-distance of about 57 cm. The point of gaze was continuously tracked by means of a video-based eye tracker using one eye (SR Research EyeLink; sampling rate 1 kHz).

Spiking activity

While advancing the electrode in search for responsive units, we presented 16 human face images and 16 non-face images in a category selectivity experiment as described previously (Taubert et al., 2015). The images were taken from the image set used in the fMRI localizer, but with the noise background removed. The set of 16 non-face images consisted of 4 images per category (bodies, fruits, manmade objects, and hands). To initiate a trial, the subject had to fixate (2 by 2 visual degree fixation window) for 300 ms, followed by 300 ms of stimulus presentation, and an additional 300 ms fixation period before receiving a fluid reward. The lower bound of the inter-trial interval was set to 500 ms, but it could be longer based on the behavior of the monkey as they were required to initiate each trial. All 32 stimuli were presented in random order, with the restriction that all images had to be presented before one could be repeated again. For each stimulus presentation we calculated the net response using the firing rate in the 300 ms window starting 50 ms after stimulus onset, minus that in the 50 ms window before stimulus onset. For each neuron or multi-unit site, spiking activity was recorded during at least 3 presentations of each image in order to assess the category selectivity. Specifically, we quantified the face category selectivity using the following index (Tsao et al., 2006; Taubert et al., 2015):

$$FSI = \frac{R_{face} - R_{nonface}}{|R_{face}| + |R_{nonface}|},$$

with R_{face} the mean net response to the 16 faces and $R_{nonface}$ the mean net response to the 16 non-face images. This face-selectivity index is > 0 for neurons or multi-unit sites that respond more to faces than non-faces (i.e. they are face category selective).

After the category selectivity experiment, we ran the repetition probability experiment for at least 4 blocks (2 repetition and 2 alternation) per neuron or neural site. For the analyses of the data recorded in this experiment, we calculated for each stimulus presentation the firing rate in the 250 ms window starting 50 ms after stimulus onset. The data for target trials and the first 5 unaborted trials of each block were excluded from analysis. The gross firing rates for the first (S1) and second (S2) face in each unaborted trial were used to calculate an adaptation index as follows:

$$AI = \frac{S1 - S2}{S1}.$$

This number expresses the proportional difference in response strength between the first and second stimulus and is > 0 if the response to the second stimulus is lower (e.g. repetition suppression), < 0 if it is higher, and = 0 if responses are equal. In order to have stimulus specific adaptation, the suppression for repetition trials needs to be stronger than for alternation trials. Thus, the AI should be positive and larger for repetition trials than alternation trials.

Local field potentials

At most spiking activity recording site, we also recorded LFPs sampled at 1 kHz. Offline, the signal was band-passed between .2 and 170 Hz and line noise was removed using a 50 Hz notch filter (48–52 Hz). We used time-frequency Morlet wavelet decomposition for spectral analysis as described previously (Kaliukhovich and Vogels, 2011), using FieldTrip (Oostenveld et al., 2011). Frequencies below 10 Hz were excluded from the wavelet analysis to avoid wavelets overlapping adapter and test stimulus presentations. At each frequency we normalized power by division by the average baseline power (200 ms window before stimulus onset). For LFP power responses to S1 and S2 we used the average normalized power of the 250 ms window starting 50 ms after stimulus onset over 4 frequency bands: 12-25 Hz, 26-60 Hz, 61-100 Hz (i.e. the 3 windows used by Kaliukhovich and Vogels, 2011), and 101-170 Hz. Like we did for spiking activity, these LFP power responses were then used to calculate AIs.

5.2.5 fMRI

Scanning details were almost identical to the ones described under Face patch localization, except for the following. For one subject (Monkey G), we used higher spatial resolutions for the first 5 scanning days (voxel size = 1.2 mm isotropic, 40 horizontal slices, TR = 2 s, TE = 18 ms, flip angle = 90°), which we decreased for the last 3 days (voxel size = 1.5 mm isotropic, 34 horizontal slices, TR = 2 s, TE = 15 ms, flip angle = 90°). After pre-processing (see face patch localization), which upscales the voxel size to 1 mm isotropic, these scanning days were combined for data analysis. For Monkey D the latter resolution was used throughout the 9 scanning days.

For the repetition probability experiment, we changed the minimum inter-trial interval to an average of 3 s (uniform distribution between 2 and 4 s) in accordance with Summerfield et al. (Summerfield et al., 2008). However, as for the electrophysiology experiment, the interval could be longer based on the behavior of the monkey. The length of a run was 800 s and after a few sessions was increased to 820 s. During this time, the subjects could usually finish 4 blocks (2 repetition and 2 alternation). The block length was always 40 trials, and the task was always to detect an inverted face. For data analysis we only used the data of completed blocks and discarded the imaging data collected during the last unfinished block. For the general linear model, we used 8 regressors in addition to motion and eye movement regressors: 1) repetition trials (excluding the first 5 trials of a block) in repetition blocks, 2) alternation trials in repetition blocks, 3) repetition trials in alternation blocks, 4) alternation trials (excluding the first 5) in alternation blocks, 5) repetition trials and 6) alternation trials in the first 5 trials of a block, 7) target trials, and, 8) full screen color presentations between blocks. Regressors were convolved with the MION response function (Vanduffel et al., 2001). Trials were modeled as 1 s events in SPM12 (0 s for color presentations), and because of the low temporal resolution of fMRI each face pair was treated as a compound trial (Summerfield et al., 2008).

The regions of interest (ROI) for analysis of the beta values were the face patches defined by the functional localizer, as well as several anatomically defined regions of interest. The latter were defined by an intersection between a contrast indicating voxels responsive to our stimuli (T > 5 on average across regressors 1-7 of the repetition suppression experiment) and the following anatomical regions: visual area 1 (V1), visual area 2 (V2), visual area 4 (V4), dorsal and ventral posterior IT (pIT), dorsal and ventral central IT (cIT), and dorsal and ventral anterior IT (aIT). Anatomical areas were based on the parcellation of Felleman and Van Essen (1991), included in Caret (v 5.61) software, coregistered to each monkey's native space.

5.2.6 Statistical analysis

For statistical inference we generally relied on bias-corrected accelerated bootstrap confidence intervals (Efron, 1987) and randomization tests, unless indicated otherwise. The bootstrap estimates are based on random sampling with replacement (10,000 iterations) of the neurons, recording sites (for multi-unit activity), or runs (fMRI). P-values (uncorrected for multiple comparisons) were calculated using randomization tests (10,000 iterations) to estimate the distribution of the test statistic under the null hypothesis. SPM T-maps are based on parametric t-tests and visualized in FslView (v 4.0.1) for **Figure 5.2** (t value threshold = 5 for faces versus bodies, fruits, manmade objects, and hands contrast and 12 for faces versus fixation contrast). In addition, on several occasions we use the JZS Bayes factor with the default $\sqrt{2}/2$ scale parameter to quantify evidence for the null hypothesis of one sample *t* tests (Rouder et al., 2009).

5.3 Results

We recorded spiking activity and fMRI signals in one male (Monkey G) and one female monkey (Monkey D) during trials where either two different faces were shown (alternation trial) or the same face was repeated (repetition trial). In separate blocks, we manipulated the probability of a repetition trial: 75% repetition trials and 25% alternation trials for repetition blocks and vice versa for alternation blocks. If repetition suppression for faces reflects fulfillment of perceptual expectations (Summerfield et al., 2008), it should be stronger in repetition blocks where a repetition is expected, compared to alternation blocks where it is unexpected.

5.3.1 Face category selectivity

We localized face-selective regions using an fMRI block design with images of 5 categories: faces, bodies, fruits, manmade objects, and hands (Tsao et al., 2003). We collected 13 runs for Monkey G and 31 for Monkey D. The results show the 6 prototypical face patches (Tsao et al., 2008) in IT cortex of Monkey G: posterior lateral (PL), middle lateral (ML), middle fundus (MF), and anterior lateral (AL) bilaterally, anterior fundus (AF) only in the right hemisphere, and anterior medial (AM) only in the left (**Figure 5.2**A). For IT cortex of Monkey D, there was only one face patch (bilaterally), which we identified as AL based on its location. There were no other face-selective patches defined by the contrast faces versus all other categories in this monkey. However, the contrast faces versus fixation did peak around the expected location of ML, suggesting that this area responds strongly to faces, albeit not selectively on the level of voxels. We will call this region putative ML from here on (**Figure 5.2**B).

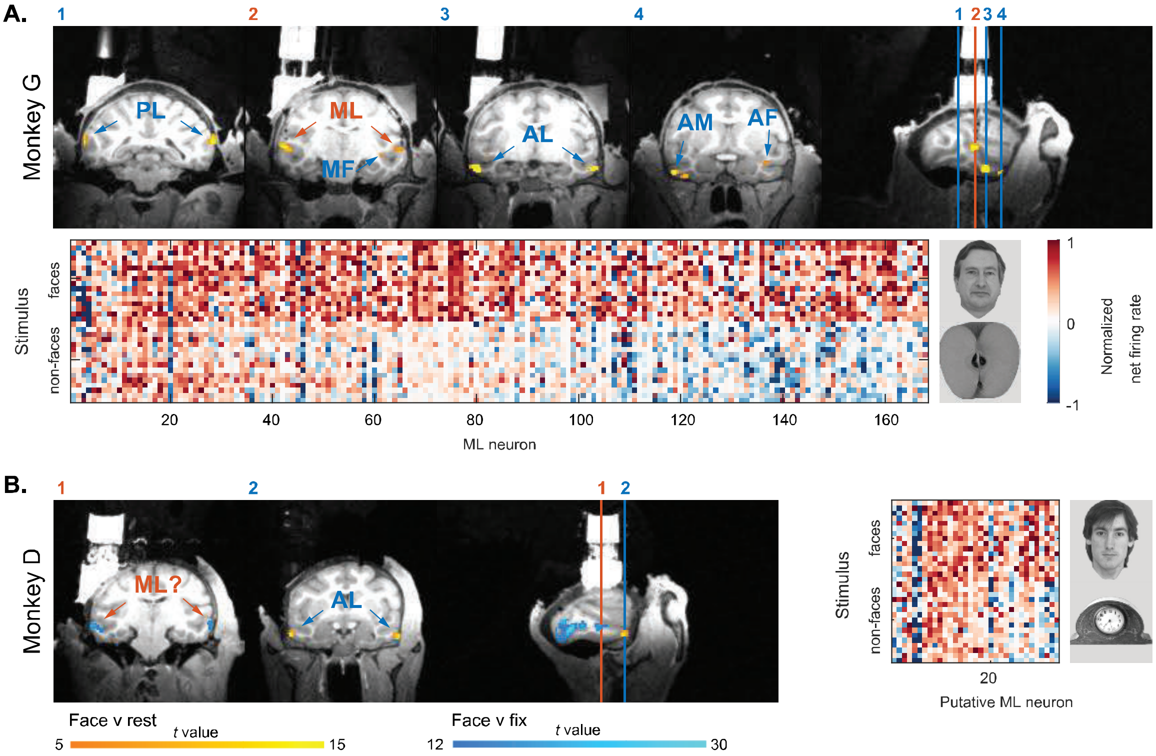


Figure 5.2. fMRI localized face patches and single cell face category selectivity.
(A) For Monkey G we were able to identify 6 face-selective patches: PL, ML, MF, and AL bilaterally; AM and AF unilaterally (faces versus bodies, fruits, manmade objects, and hands contrast; t value threshold = 5). The locations are indicated on 4 coronal slices (slices 1-4 selected along the posterior-anterior axis as indicated on the sagittal view). The heat-map below the images shows the face category selectivity profile of spiking activity in ML. Each row represents one image (16 faces and 16 non-faces) and each column represents one neuron (168 cells sorted by FSI). Values are net responses normalized by the maximum. To the right we show the preferred face and non-face. **(B)** For Monkey D we were able to identify only 1 face-selective patch: AL. Responses to faces (face versus fixation contrast; t value threshold = 12) did peak at the anatomically expected location of ML. We call this region putative ML. The locations are indicated on 2 coronal slices. The heat-map to the right of the images shows the face category selectivity profile in putative ML (34 cells, same conventions as in A).

Both fMRI localized ML of Monkey G and putative ML of Monkey D are the areas we targeted for our recordings of spiking activity during our repetition probability experiments. First, we ran a face category selectivity experiment to validate the results of the fMRI localizer. Spiking activity recorded in ML (Monkey G) showed face-selectivity for both single neurons (mean FSI = .69, 95% CI [.62 .74], SD = .38, 168 neurons; **Figure 5.2**A) and multi-unit sites (mean FSI = .71, 95% CI [.66 .76], SD = .35, 219 sites). Spiking activity recorded in putative ML (Monkey D) did also show face-selectivity for both single neurons (mean FSI = .35, 95% CI [.19 .49], SD = .44, 33 neurons; **Figure 5.2**B) and multi-unit sites (mean FSI = .45, 95% CI [.28 .62], SD = .38, 18 sites). However, it should be noted that our recordings were biased towards higher FSI values because we did not record from neurons that showed no or little response to faces (since faces were the only stimuli in our main experiment).

5.3.2 The effect of repetition probability on spiking activity and LFP signals

Here we assessed the effect of repetition probability on the adaptation of electrophysiological signals. In general, we expected stimulus specific adaptation: AI repetition trials > 0 and > AI alternation trials. If there is an expectation effect, this difference (AI repetition trials > AI alternation trials) should be larger for repetition blocks compared to alternation blocks. In a first experiment the task was orthogonal to the manipulation of repetition trials. In a second experiment, face repetitions were task-relevant. Spiking activity was recorded simultaneously with LFPs in fMRI localized regions: ML of Monkey G (experiment 1: 97 single, 110 multi-units, 68 LFP sites; experiment 2: 20 single, 60 multi-units, and 80 LFP sites), and putative ML for Monkey D (experiment 1: 34 single, 18 multi-units, and 52 LFP sites).

Orthogonal task

In our first experiment we implemented the task used in the main experiment of Summerfield et al. (2008): the monkey had to detect inverted faces, occurring in 20% of all trials. In these target trials, either the first or the second face could be inverted. After each trial, the monkey had to indicate with saccades whether it had been a target trial (left), or not (right). This task requires the monkey to attend the stimuli that are being presented, but is unrelated to face repetitions or alternations. In our initial recording sessions we used a block length of 40 trials following previous studies (Kaliukhovich and Vogels, 2011; Kovács et al., 2013), which is already two times the 20 trials initially reported by Summerfield et al. (Summerfield et al., 2008). Later this was increased to 120 trials, because longer blocks increase the chances of finding an expectation effect: more trials provide more information about the trial probabilities. Since there was no indication/evidence of an effect of block length, we pooled the data of the different block lengths. Monkey G's performance (proportion correct) for target and non-target trials was: .993, 95% CI [.991 .994] (non-target) and .990, 95% CI [.987 .993] (target) for repetition blocks, and .993, 95% CI [.991 .995] (non-target) and .991, 95% CI [.987 .993] (target) for alternation blocks. Monkey D's performance was: .996, 95% CI [.985 .998] (non-target) and .968, 95% CI [.950 .980] (target) for repetition blocks, and .996, 95% CI [.988 .999] (non-target) and .966, 95% CI [.948 .978] (target) for alternation blocks.

Spiking activity. For the single unit data recorded in Monkey G (N = 97), we observed stronger suppression for a face repetition than for an alternation, without any block effect. This is clear from both the peristimulus time histogram (PSTH) as well as the AIs (**Figure 5.3**A). The response reduction for a repeated stimulus was about 18%, compared to 8% for an alternation. This translates to a stimulus specific reduction of about 10% for each block (Repetition Block (RB): M = .10, SD = .24, p < .001; Alternation Block (AB): M = .10, SD = .23, p < .001), with no evidence for a difference between blocks (M = .00, SD = .26, p = .95). Assuming a normal distribution, the Bayes factor (Rouder et al., 2009) in favor of no block effect (BF_0) is 8.9. Thus, it is about 9 times more likely that there is no effect of repetition probability in the population average, given the data. We have previously reported a discrepancy between adaptation related effects from single versus multi-unit data, perhaps as a result of a single cell sampling bias (Vinken et al., 2017). Thus, we examined adaptation in multi-unit activity for potential probability effects. The results for multi-unit data recorded in Monkey G (N = 110, **Figure 5.3**B) are very similar to the single unit data, with a stimulus specific reduction of about 9-11% (RB: M = .11, SD = .13, p < .001; AB: M = .09, SD = .11, p < .001), with no evidence for a difference between blocks (M = .02, SD = .17, p = .22, BF_0 = 4.5). In Monkey D, for single unit data (N = 34, **Figure 5.3**C) the stimulus specific reduction was 5-8% (RB: M = .05, SD = .22, p = .25; AB: M = .08, SD = .15, p = .004) and 8-10% for multi-unit data (N = 18, **Figure 5.3**D; RB: M =

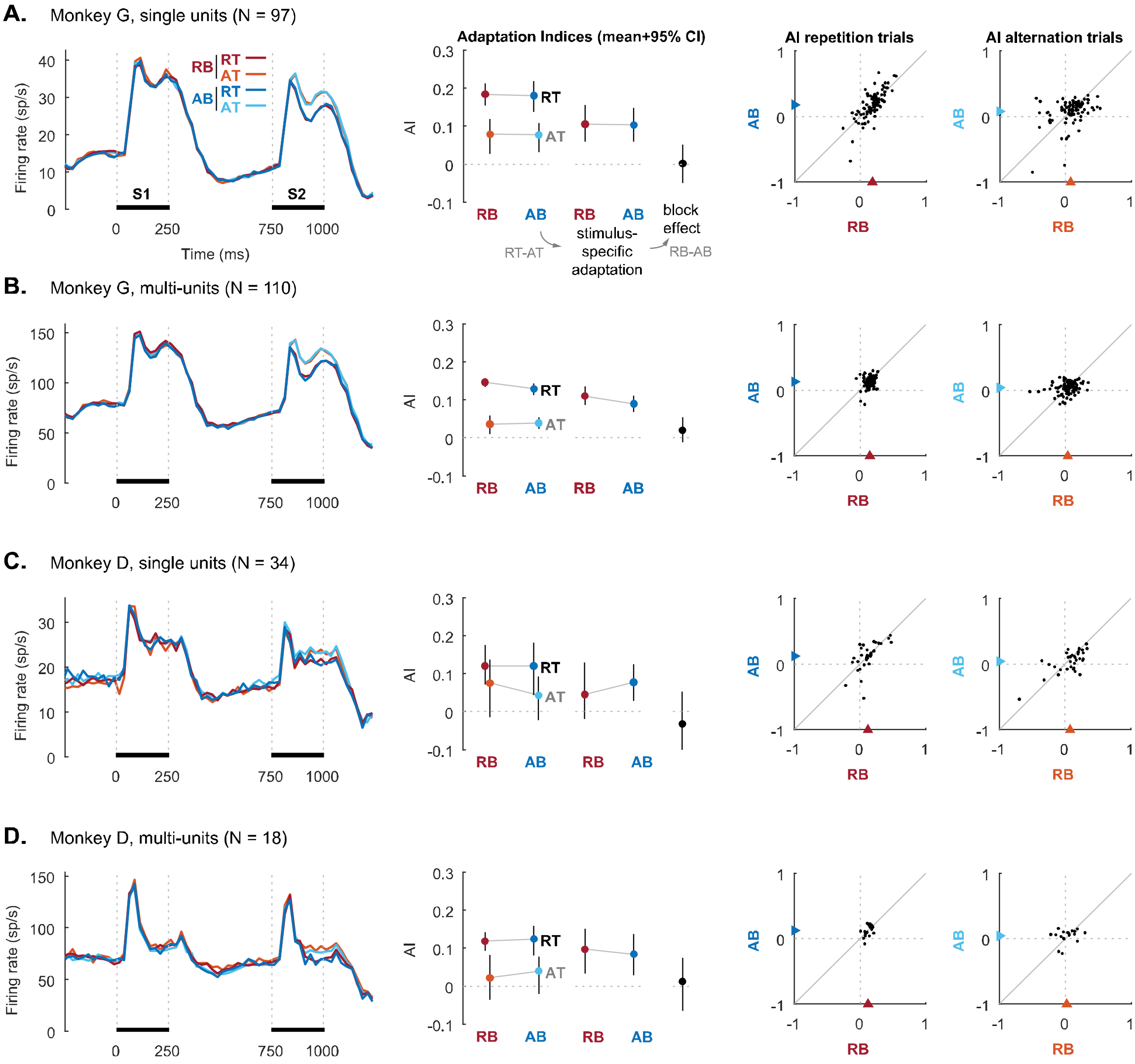


Figure 5.3. Spiking activity recorded during the orthogonal task.
(A-D) First column: population PSTHs showing the firing rate during the two stimulus presentations (S1 and S2) of repetition trials in repetition blocks (red) and alternation blocks (blue), as well as alternation trials in repetition blocks (orange) and alternation blocks (light blue). Second column: AIs for each trial x block combination (positive values mean suppression for S2), stimulus specific effects (AI repetition trials - AI alternation trials), and block effects (difference in stimulus specific effect: repetition block - adaptation block). Third column: scatter plot of AIs for repetition trials in repetition blocks (abscissa) and alternation blocks (ordinate). Triangles on axes indicate mean values. Fourth column: scatter plot of AIs for alternations trials (see third column). **(A)** Monkey G single and **(B)** multi-unit results. **(C)** Monkey D single and **(D)** multi-unit results.

.10, SD = .13, p = .009; AB: M = .08, SD = .12, p = .009). There was no evidence for a difference between blocks for either (single unit: M = -.03, SD = .23, p = .43, BF_0 = 4.0; multi-unit: M = .01, SD = .15, p = .73, BF_0 = 3.9).

LFP signals. The relationship between fMRI signals and spiking activity is complex. Depending on task conditions, the two signals can correlate less good (Logothetis et al., 2001; Maier et al., 2008) and sometimes better (Lima et al., 2014) than the correlation between fMRI and other measures such as LFP. To exclude the possibility that electrophysiological expectation effects are restricted to LFPs, we analyzed LFP data recorded together with spiking activity. **Figure 5.4**A shows baseline normalized time-frequency power maps per trial x block combination for Monkey G (N = 76 sites). Consistent with previous reports (De Baene and Vogels, 2010; Kaliukhovich and Vogels, 2011), these maps indicate that there is a stimulus-specific adaptation effect only for frequencies of about 70 Hz and higher. This is confirmed by the AIs for the 101-170 Hz window, which show a stimulus specific power reduction of 10% (RB: M = .10, SD = .14, p < .001; AB: M = .10, SD = .11, p < .001). As with spiking activity, there is no evidence for a difference between blocks (M = .01, SD = .18, p = .70, BF_0 = 7.3). For the other frequency bands, there was either a stimulus unspecific suppression (26-100 Hz) or enhancement (10-25 Hz) for S2, without evidence for a block effect (10-25 Hz: BF_0 = 7.9, 26-60 Hz: BF_0 = 6.3, 61-100 Hz: BF_0 = 7.2). The results for Monkey D (N = 52 sites, **Figure 5.4**B) indicate a stimulus-specific reduction for both the 61-100 Hz and 101-170 Hz band: 4% for the former (RB: M = .04, SD = .09, p = .001; AB: M = .04, SD = .08, p = .003) and 5-6% for the latter (RB: M = .06, SD = .09, p < .001; AB: M = .05, SD = .09, p < .001). Neither showed evidence for an effect of block (61-100 Hz: M = .01, SD = .11, p = .72, BF_0 = 6.2; 101-170 Hz: M = .01, SD = .12, p = .63, BF_0 = 5.9). In addition, the lowest frequency band of 10-25 Hz showed a stimulus-specific enhancement of 6-9% (RB: M = -.06, SD = .21, p = .039; AB: M = -.09, SD = .20, p = .001), with no evidence for a difference between blocks (M = .03, SD = .31, p = .44, BF_0 = 4.9).

Together, these data suggest that neither the use of face stimuli, nor an orthogonal stimulus-related task are sufficient conditions for an effect of repetition probability on the adaptation of spiking activity or LFPs. Therefore, in the next experiment we made repetition probability task-relevant.

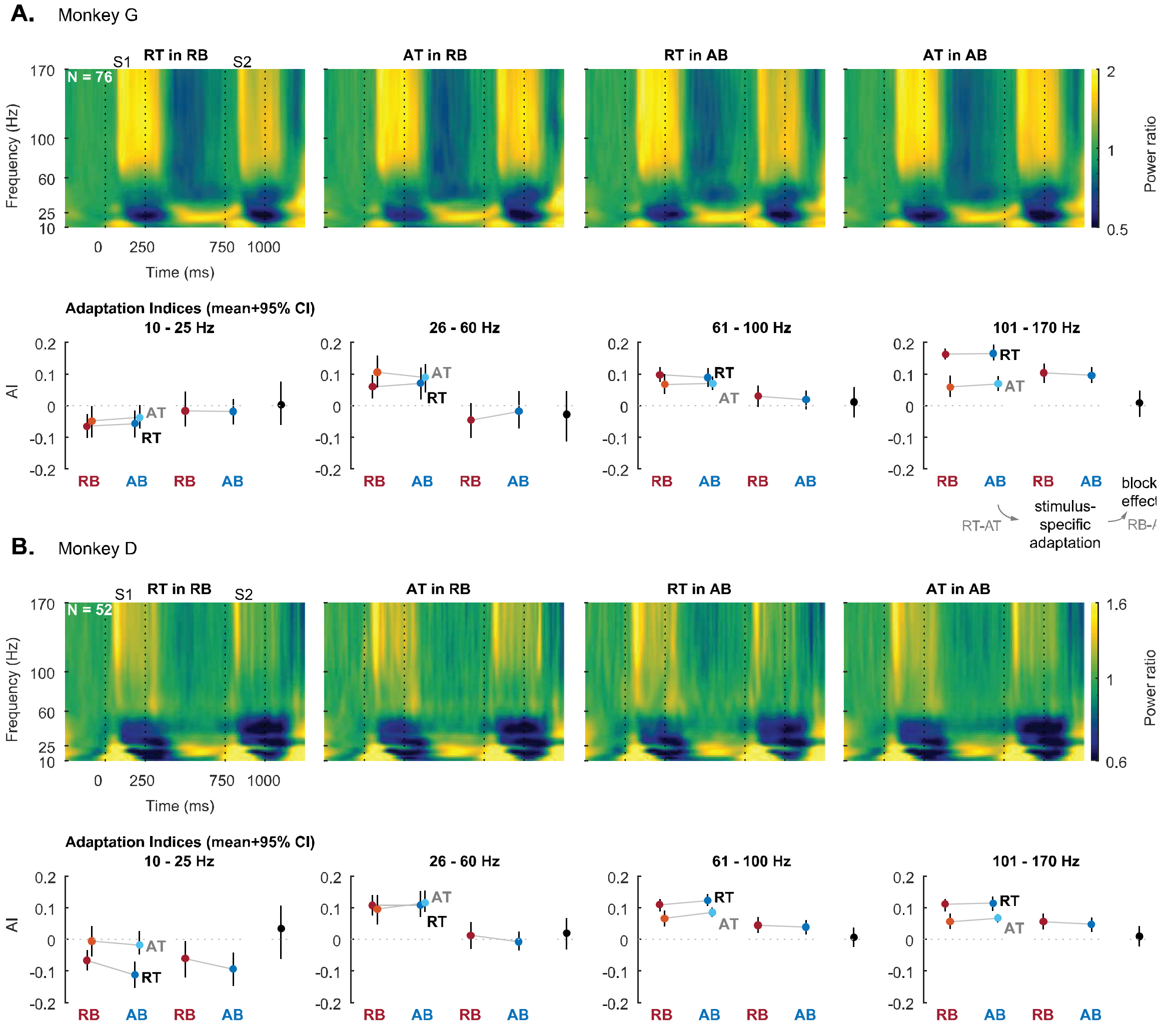


Figure 5.4. Time-frequency power spectra of LFP signals recorded during the orthogonal task. **(A)** Results for Monkey G. First row: time-frequency maps of power relative to baseline (-200 - 0 ms). We used a base 10 logarithmic color scale to give power suppression (values < 1) equal contrast as enhancement (values > 1). Second row: AIs calculated for separate frequency bands (same conventions as in **Figure 5.3**). **(B)** Results for Monkey D (same conventions as in A).

Making repetition probability task-relevant

In this experiment, there were no inverted target trials. Instead, after each trial, the monkey had to indicate whether it had been a repetition (left saccade) or alternation trial (right). In this way, the manipulated probability of a repetition became directly relevant for the task. We conducted this experiment with Monkey G, using block lengths of 40 and 100 trials. An important advantage is that we can now assess the effect of repetition probability on behavior in addition to neural activity. There was a clear interaction between block type and trial type: percentage correct for repetition trials was higher for repetition blocks compared to alternation blocks (RB: .94, 95% CI [.94 .95], AB:

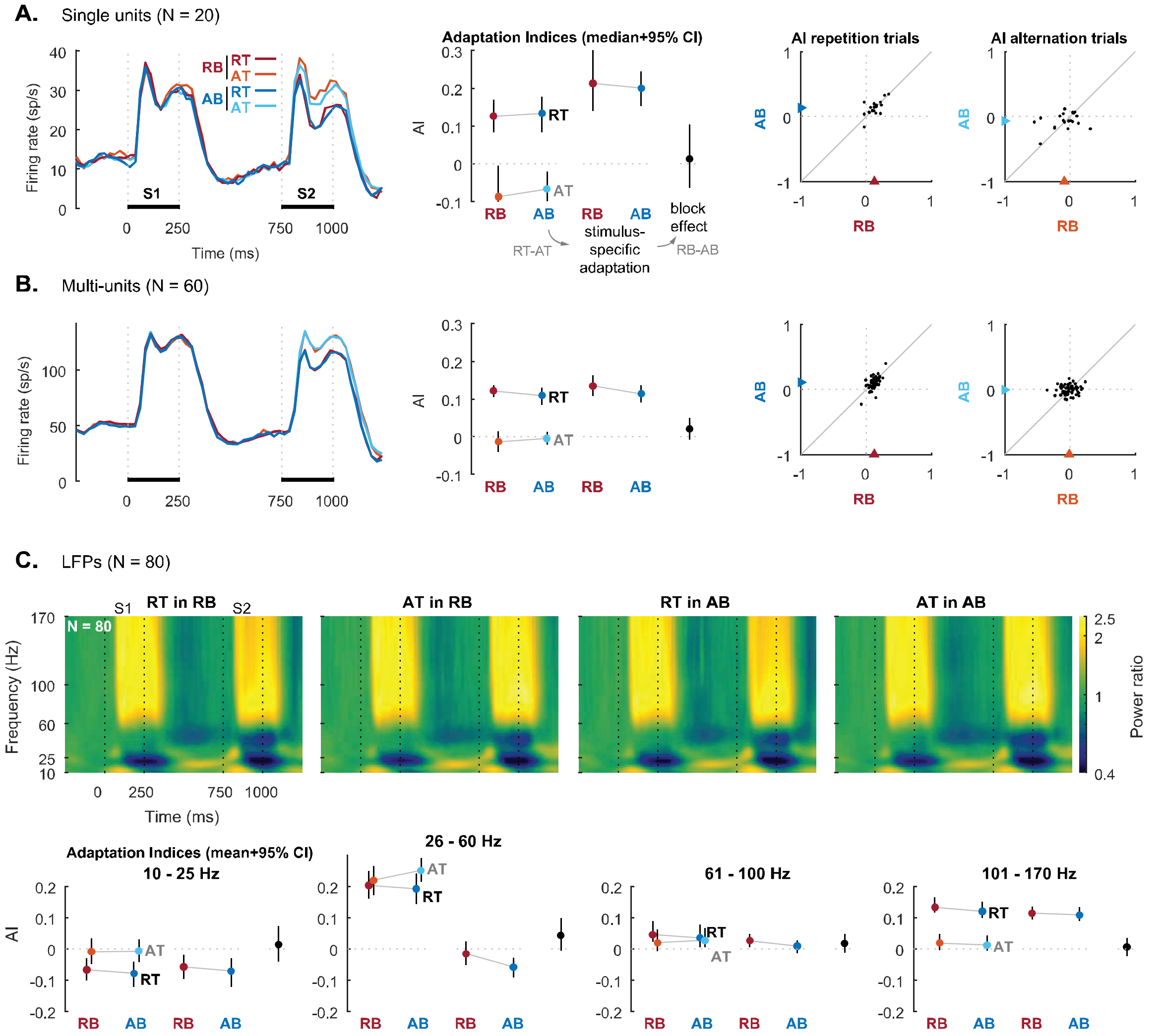


Figure 5.5. Spiking activity and LFPs recorded in Monkey G during a repetition versus alternation task.

Same conventions as **Figure 5.3** for single **(A)** and multi-unit **(B)** results and as **Figure 5.4** for LFP results **(C)**.

.83, 95% CI [.81 .85], N = 80 sessions), while the reverse was true for alternation trials (RB: .81, 95% CI [.79 .83], AB: .91, 95% CI [.90 .92]).

Despite this behavioral effect, repetition probability has no effect on repetition suppression of neural activity in ML. Spiking activity shows a stimulus specific reduction of 20-21% for single unit data (N = 20, **Figure 5.5**A; RB: M = .21, SD = .19, p < .001; AB: M = .20, SD = .11, p < .001) and 11-13% for multi-unit data (N = 60, **Figure 5.5**B; RB: M = .13, SD = .11, p < .001; AB: M = .11, SD = .09, p < .001). There was no evidence for a difference between blocks for either (single unit: M = .01, SD = .20, p = .78, BF_0 = 4.1; multi-unit: M = .02, SD = .11, p = .15, BF_0 = 2.6). In addition, there was no positive correlation across sessions between the behavioral interaction and block effect (spearman r; single units: r = -.34, p = .15; multi-units: r = -.07, p = .62). LFP power (N = 80, **Figure 5.5**C) shows a clear stimulus specific reduction of 11% for the 101-170 Hz band (RB: M = .11, SD = .09, p < .001; AB: M = .11, SD = .10, p < .001) as well as an enhancement of 6-7% for the 10-25 Hz band (RB: M = -.06, SD = .18, p = .004; AB: M = -.07, SD = .21, p = .002), with no evidence for a difference between blocks (10-25 Hz: M = .01, SD = .25, p = .63, BF$_0$ = 7.3; 101-170 Hz: M = .01, SD = .13, p = .69, BF$_0$ = 7.5). In conclusion, even when repetition probability was task relevant and modulated task performance, it did not affect the adaptation of spiking activity or LFP signals.

5.3.3 THE EFFECT OF REPETITION PROBABILITY ON FMRI SIGNALS

Finally, we ran the repetition probability experiment with the orthogonal task while recording fMRI responses. For our analysis, we look at the face patches defined by the functional localizer, as well as several anatomically defined ROIs (see materials and methods). We collected 90 runs for Monkey G and 100 for Monkey D. For Monkey G, in ML (**Figure 5.6**A) we observed lower responses for repetition trials than for alternation trials in alternation blocks (M = 1.2, SD = 5.3, p = .035, BF$_0$ = 0.97), but not in repetition blocks (M = -.45, SD = 5.5, p = .45, BF$_0$ = 6.4). As a result, there was evidence for a difference between blocks (M = -1.6, SD = 7.3, p = .034, BF$_0$ = 0.99), but in the direction inconsistent with predictive coding: repetition suppression was stronger in alternation blocks, where a repetition should be surprising. For none of the other ROIs in this monkey there was a block effect (**Figure 5.6**C).

For Monkey D, in putative ML (**Figure 5.6**B) we observed stronger responses for alternation trials than for repetition trials in repetition blocks (M = 1.5, SD = 6.4, p = .026, BF$_0$ = 0.76), but not in alternation blocks (M = .04, SD = 5.8, p = .94, BF$_0$ = 9). This resulted at best in very weak evidence for a block effect (M = 1.4, SD = 8.6, p = .10, BF$_0$ = 2.4) that would be consistent with Summerfield et al. (2008), i.e. that repetition suppression is stronger when repetitions are expected. For several other ROIs there was evidence for such a block effect, which was significant for V2, V4, cIT, and aIT (**Figure 5.6**D).

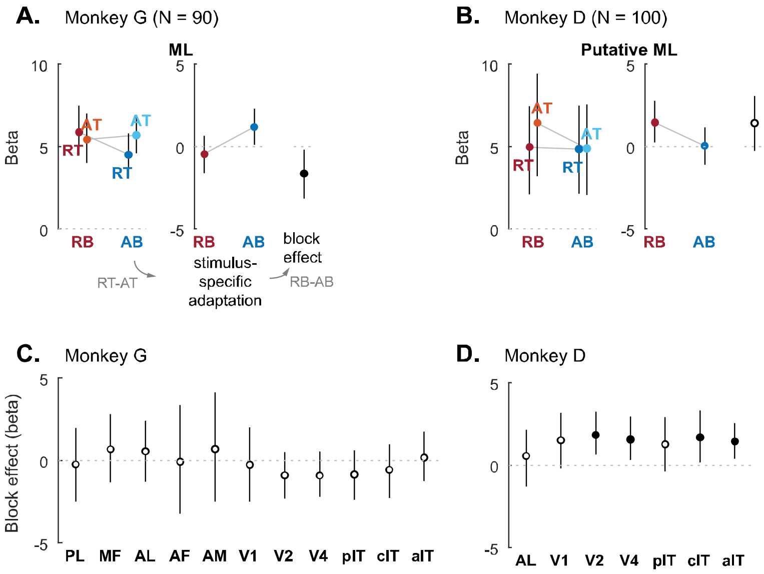


Figure 5.6. fMRI activity measured during the orthogonal task.
(A) Beta values (mean + 95% CI) for each trial × block combination in ML of Monkey G, stimulus specific adaptation (beta value repetition trials – beta value alternation trials), and block effects (difference in stimulus specific effect: repetition block - adaptation block). A filled marker for the block effect indicates p < .05. **(B)** Beta values in putative ML of Monkey D (same conventions as (A)). **(C)** Block effects (mean + 95% CI) for all other ROIs in Monkey G (face patches PL, MF, AL, AF, and AM; anatomical areas V1, V2, V4, pIT, cIT, and aIT). Filled markers indicate p < .05. **(D)** Block effects for all other ROIs in Monkey D (same conventions as (C)).

To sum up, even though we did find weak evidence for effects of repetition probability on repetition suppression in fMRI signals, it was inconsistent with the hypothesized direction (Monkey G), inconsistent across ROIs (Monkey G), and inconsistent across monkeys (Monkey G versus Monkey D).

5.4 Discussion

We investigated the effect of face repetition probability on adaptation of spiking activity, LFPs, and fMRI responses in (putative) ML of macaque IT cortex. In none of the electrophysiological recordings there was evidence in favor of an effect of repetition probability on adaptation and this was consistent in 2 monkeys. In contrast, the fMRI results showed weak effects of repetition probability but in opposite directions for each monkey.

The results of our electrophysiological data were consistent with the previous investigation by Kaliukhovich and Vogels (2011). Yet, here we had improved the paradigm by including two supposedly crucial conditions for observing the repetition probability effect: a task directing attention towards the stimuli (and later the repetition) (Larsson and Smith, 2012) and the use of faces, which might form a special stimulus category based on prior experience (Grotheer and Kovács, 2014). In addition to that, we recorded responses in face selective IT patch ML, which might be homologous to human FFA (Tsao et al., 2008), where the originally reported effects were clearest (Summerfield et al., 2008). Thus, neither attention, nor face stimuli were sufficient conditions for observing a perceptual expectation effect in electrophysiological responses in face patch ML.

Similarly, previous work by Kaliukhovich and Vogels (2014) demonstrates that IT neurons show no surprise response to deviants in visual oddball sequences. In contrast, a recent study did report effects of perceptual expectations on responses of mostly face selective macaque IT neurons (Bell et al., 2016). The authors claimed their effects were distinct from low-level sensory adaptation, based on a multivariate regression analysis where they tried to control for repetition suppression. They argued that expectation effects might have been absent in previous studies of spiking activity because of a lack of attentional requirements of the task. However, recently (Vinken and Vogels, 2017) we have shown that the analysis of Bell et al. (2016) did not properly control for adaptation. Indeed, we get the same 'expectation' effects in simulated neurons that only include mechanisms of low-level sensory adaptation. Hence, we argue that their results are also in line with simple bottom-up and local mechanisms of adaptation and do not require a perceptual expectation account.

Compared to electrophysiological data, our fMRI results were much less consistent. For starters, in ML of Monkey G we observed a block effect that had the opposite sign to what the perceptual expectation hypothesis predicts. Thus, instead of being larger, there was no suppression for expected repetitions. This effect was not present in any of the other (face selective) regions. The lack of stimulus specific fMRI adaptation in repetition blocks is also inconsistent with the clear stimulus specific adaptation for electrophysiological responses in these blocks. Of course, this block effect for ML might be a false positive. On the other hand, several regions in Monkey D did exhibit a block effect consistent with perceptual expectation: suppression for expected repetitions was stronger. This effect was the result of an absence of stimulus specific fMRI adaptation in alternation blocks. This absence was also the case for putative ML, which is inconsistent our electrophysiological recordings. However, for putative ML the block effect was not statistically significant. Interestingly, the only face selective patch as identified by our fMRI localizer, i.e. AL, did not show this block effect. Thus, the fMRI results were inconsistent and contradicted direct measures of neural activity.

Recently, Olkkonen et al. (2017) failed to replicate previous studies of an effect of expectation on fMRI adaptation in the FFA (or any visual area), showing that in humans the effect is not consistently found. Olkkonen et al. (2017) contemplate the possibility that their computer-generated stimuli (for which they also used FaceGen) did not attract enough attention compared to real faces during an orthogonal task. They do find a clear behavioral effect of repetition probability, but this was only assessed with a different task during a separate experiment without fMRI recordings. In our second experiment we also used a task that allowed us to observe a clear behavioral effect. Despite this, we did not find any evidence for an expectation effect in our simultaneously recorded electrophysiological responses. In any case, both Olkkonen et al. (2017) and our results show that adaptation is independent of expectation and that a role of such higher level processes is not very general. Indeed, even in the presence of attention (Larsson and Smith, 2012) and with an experience-based stimulus category like faces (Kovács et al., 2013; Grotheer and Kovács, 2014), we could not replicate the effect.

Finally, there is the question of the lack of face patches in Monkey D. Typically, six or more patches of face selective cortex can be defined in the temporal cortex using an fMRI localizer (Tsao et al., 2008). These patches form a hierarchical system for the processing and perception of faces (Moeller et al., 2008, 2017; Freiwald and Tsao, 2010). However, in Monkey D we could only find one anterior patch which we presume to be AL based on its location. It has been shown that the formation of the face domains requires exposure to faces during development (Arcaro et al., 2017). Yet, Monkey D was not visually deprived during development or reared in any unusual way. In addition, given the hierarchical nature of the face processing system it is a puzzle why this monkey would develop an anterior face patch while missing the earlier stages. One possible explanation is that there was no clustering of face selective neurons at earlier stages of processing in the functional hierarchy (at least not at a voxel level). For our experiments, we did record neural responses in a face responsive patch at the expected location of ML (i.e., putative ML) in Monkey D. The results for our main experiment were consistent with the results from the actual face patch ML of Monkey G, both in terms of clear stimulus-specific adaptation and the absence of an expectation effect.

In sum, we investigated whether repetition probability affects repetition suppression of single neurons in monkey IT under two supposedly necessary conditions: (a) the use of face stimuli, and (b) a task that requires attention for the stimuli (or repetitions). Even under these specific circumstances, we did not find any effect of repetition probability in any of the two monkeys. These results were confirmed by recordings of LFPs and multi-unit spiking activity. In an independent fMRI experiment, we did find evidence for effects of repetition probability that went in opposite directions for each monkey – and thus in the 'wrong' direction for one monkey. Importantly, both effects were inconsistent with the three electrophysiological measures recorded in the same region. We conclude that while fMRI recordings showed inconsistent results, direct measures of neural activity consistently suggested that there was no effect of face repetition probability on repetition suppression of face-responsive IT neurons. These results further call into question the importance of repetition-induced top-down mechanisms in neural adaptation.

Chapter 6.

ADAPTATION CONFOUNDED AS EXPECTATION

In Chapter 5, we did not find an effect of expectation on repetition suppression of face-responsive IT neurons. In contrast, Bell and colleagues reportedly found evidence for an expectation-based mechanism distinct from stimulus-driven adaptation (Bell et al., 2016). The authors used a design where stimulus repetition is confounded with expectation, but tried to control for repetition suppression with a linear regression approach. Using simulated neural responses, we investigate whether their method actually controls for the confound. *Could the analysis in Bell et al. lead to spurious effects of expectation?*

In press as

Vinken K., Vogels R. (2017). Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex. *Current Biology, in press.*

Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex.

In predictive coding theory, the brain is conceptualized as a prediction machine that constantly constructs and updates expectations of the sensory environment (Friston, 2005). In the context of this theory, Bell et al. (2016) recently studied the effect of the probability of task-relevant stimuli on the activity of macaque inferior temporal (IT) neurons and observed a reduced population response to expected faces in face-selective neurons. They concluded that "IT neurons encode long-term, latent probabilistic information about stimulus occurrence", supporting predictive coding. They manipulated expectation by the frequency of face versus fruit stimuli in blocks of trials. In such design, stimulus repetition is confounded with expectation. Since previous studies showed that IT neurons decrease their response with repetition (Vogels, 2016), such adaptation (or repetition suppression), instead of expectation suppression, could explain their effects. The authors attempted to control for this alternative interpretation with a multiple regression approach. Here we show by using simulation that adaptation can still masquerade as expectation effects reported in Bell et al. (2016). In addition, the results from the regression model used for most analyses cannot be trusted, because the model is not uniquely defined.

6.1 Results

We simulated three 1000 neuron populations (see supplemental information). The response levels roughly matched the mean responses of Bell et al. (2016) for the third population (simulation C). In simulation A, no adaptation was present. We fitted the same regression models to the simulated responses as Bell et al. (2016) did to the actual data: Model 1 where the expected probability of a face, p(face), is estimated using a Bayesian model (their equation 1), and Model 2 (their equation 4) where p(face) is estimated using a reinforcement learning model. Model 2 is problematic, since prediction error Δp(face) is computed as a linear combination of other predictors: stimulus – p(face). Since the coefficients are not uniquely defined (singular design matrix), we employed the Moore-Penrose pseudoinverse of the design matrix as Bell et al. (2016) (C. Summerfield, personal communication). Bell et al. reported evidence for expectation responses for two predictors: a negative value for Model 1's β3 (stimulus × p(face)) and a positive value for Model 2's β3 (Δp(face), or prediction error). For simulation A, Model 1 correctly showed only the stimulus selectivity (**Figure 6.1**A).

However, Model 2 showed an effect of Δp(face) and p(face), although the simulated neurons were sensitive for neither. Further examination showed that a positive value of the prediction error coefficient β3 required relatively higher responses for faces (**Figure 6.2**A), which was the case in the neurons of Bell et al. (2016). We believe that these spurious expectation effects resulted from the collinearity between the model's predictors. This means that the majority of further analyses by Bell et al. (2016) are based

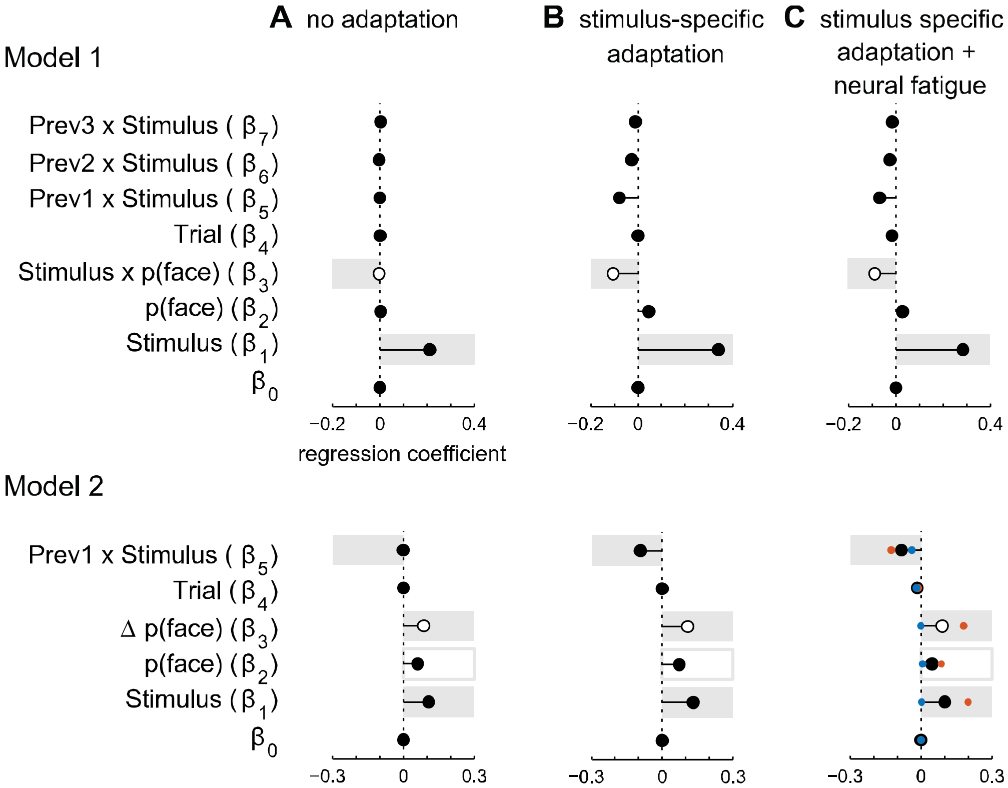


Figure 6.1. Results based on simulated neural responses for the Bell et al. study.
The two rows show the mean estimated regression weights for linear regression Models 1 and 2, respectively. In each subplot, the relevant regression coefficient that demonstrated putative expectation effects in Bell et al. (2016) is indicated by a white marker. The gray bars indicate the direction of significant regression coefficients observed in the neural data of Bell et al. (2016). The p(face) coefficient of Model 2 showed a non-significant positive trend which is indicated by the open bar. 95% confidence intervals were smaller than the symbol diameter. **(A-C)** We fitted each model for three different populations of simulated neurons: simulation A with no adaptation effects, simulation B with only a stimulus-specific adaptation effect, and simulation C with a firing-rate dependent response fatigue and recovery effect in addition to stimulus-specific adaptation. The colored dots for the Model 2 plot indicate the mean regression weights for neurons for which the response difference between faces and fruit is above (red) or below (blue) the median. See also **Figure 6.2**.

on coefficients of a regression model that cannot be used as evidence for expectation effects.

For simulation B we included stimulus-specific adaptation (Vogels, 2016). Here we used a simple resource decay model (Mill, 2014), where the response to a face/fruit is proportional to its corresponding input resources. These stimulus-specific resources decrease with each presentation and recover between them, reflecting synaptic depression in the input population (Fioravante and Regehr, 2011). The interaction between depletion and recovery allows suppression to build up, reach a stable state, or recover over time depending on the interval between repetitions. This simulates the finding that repetition suppression in IT increases with number of repetitions, even with intervening stimuli, and decreases with the inter-stimulus interval (Sawamura et al., 2006). For average adaptation rate γ and the lower bound δ parameters, the stimulus-specific suppression for a repetition was between 0% and 25%. When including adaptation in the simulation (**Figure 6.1**B), Model 1 showed a negative effect for stimulus × p(face). The magnitude of this effect was related to γ and δ (**Figure 6.2**B). Model 2 showed again the two expectation effects.

Bell et al. did not observe any putative expectation effects for fruits (Bell et al., 2016). We think this lack of an effect can in principle be explained by response fatigue: a reduction in excitability proportional to the previous response (Vogels, 2016). Specifically, high responses to face choice stimuli will cause more fatigue when p(face) is high, while low responses to fruit will result in more recovery from fatigue when p(face) is low. Thus, for simulation C, the firing rate for subsequent stimuli was reduced proportional to the previous response, reflecting a mechanism like after-hyperpolarization of the membrane potential (Sanchez-Vives et al., 2000a). We simulated firing rate recovery taking place in the interstimulus interval, which increased with decreasing response strength to the previous stimulus. For example, if the normalized response to a particular trial's choice equals 1, the response to the next cue is reduced by 12%, but recovers by 4%, resulting in a net reduction by 8% (see supplemental information). The regression analyses (**Figure 6.1**C) again showed 'prediction error' effects, with little to no effect on average fruit responses (**Figure 6.2**C). In an additional analysis, we show the results of a regression that separates the contributions of cue and choice (supplemental information).

6.2 Discussion

In summary, we replicated the results reported by Bell et al. (2016) using simulations that included only stimulus- and response-dependent processes that are thought to underlie adaptation in visual cortex and IT (Vogels, 2016). Although the simulated neurons were not sensitive to expectation-related signals, applying the multiple regression models of Bell et al. (2016) resulted in spurious effects of expectation and prediction error. Specifically, a higher response to faces is a sufficient condition for prediction error effects in regression Model 2, while the effect in Model 1 additionally requires stimulus-specific adaptation. By no means do we claim that this simple model captures everything IT neurons do in the design of Bell et al. (2016). Indeed, adaptation cannot explain the decoding of the forthcoming cue identity from baseline activity in "expectation" trials (Bell et al., 2016): faces in high p(face) blocks versus fruits in low p(face) blocks. However, this is essentially a decoding of block membership from baseline activity, which could have resulted from temporally correlated fluctuations in baseline activity unrelated to p(face) (supplemental information). We could replicate the major regression effects that were interpreted as "expectation suppression" rather than "repetition suppression" (Bell et al., 2016) in our simplified simulation exercise which only used basic passive adaptation mechanisms. This shows that their regression method did not control sufficiently for repetition suppression. Note that adaptation can be dissociated from expectation with a proper experimental design (Todorovic and de Lange, 2012) and adaptation and expectation are not just different semantic labels of the same mechanisms. In conclusion, expectation effects in the regression coefficients reported by Bell et al. (2016) can be explained in principle by adaptation and do not provide unequivocal evidence for encoding of probabilistic information in IT.

6.3 Supplemental Information

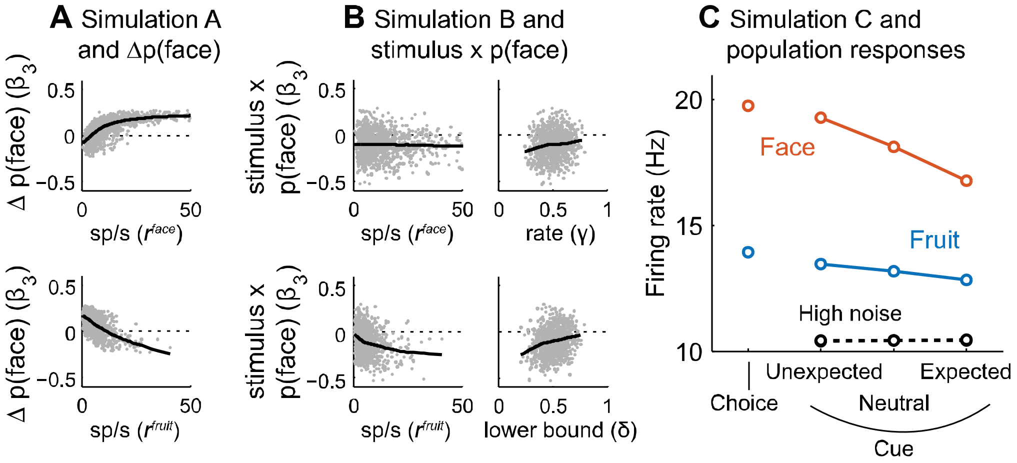


Figure 6.2. Supplemental results based on simulated neural responses.
Related to **Figure 6.1**. **(A)** Values of β3 (prediction error) for Model 2 for each neuron of simulation A, as a function of their response to faces and fruits. The black line shows LOWESS (locally weighted scatterplot smoothing) fits. **(B)** Values of β3 (stimulus x p(face)) for Model 1 for each neuron of simulation B, as a function of their response to faces and fruits, and stimulus specific adaptation parameters (γ and δ). **(C)** The population responses from simulation C are comparable to the response levels reported by Bell et al. (2016) for all conditions. Importantly, they show an effect of the stimulus probability (or expectation) for low noise face cues only.

6.3.1 Simulations

We simulated responses as a Poisson process of which the rate for a particular stimulus presentation r_t (i.e. a cue or choice stimulus) is determined by a simple model incorporating firing rate dependent fatigue, firing rate dependent recovery, and stimulus-specific adaptation. Specifically, r_t is a combination of a fixed rate for the baseline r^{base} and the unadapted rate associated with the cue stimulus r^{stim} (referring to either a low noise face r^{face}, a low noise fruit r^{fruit}, or a high noise cue response r^{noise}), which can be suppressed according to a stimulus-specific adaptation variable A_t that captures the state of adaptation for that stimulus at that time (0 = complete suppression, 1 = no adaptation). In addition, a fatigue variable F_t captures the state of neural fatigue at a particular trial. Note that we also enforced a stronger response to choice stimuli by multiplying r^{stim} by a factor of 1.2.

$$r_t = r^{base} + A_t F_t r^{stim}.$$

A_t is determined by stimulus-specific resource variables R_t^{face} or R_t^{fruit} and a lower bound parameter δ. For example, in case of a face stimulus:

$$A_t = \delta + R_t^{face}(1-\delta)$$

Both R_t^{face} and R_t^{fruit} start at a value of 1 and are updated after every stimulus presentation following a simple resource decay model (Mill, 2014). That is, after a presentation of a low noise face cue R^{face} decreases and R^{fruit} increases (or vice versa after the presentation of a fruit cue):

$$R_{t+1}^{face} = \gamma R_t^{face}$$

$$R_{t+1}^{fruit} = R_t^{fruit} + \gamma\left(1 - R_t^{fruit}\right)$$

Similarly, R_t^{face} and R_t^{fruit} are updated after the presentation of a choice stimulus. Since there were no stimulus specific effects for high noise cue stimuli in the data reported by Bell et al., we considered them as neither face, nor fruit and let both R^{face} and R^{fruit} increase. The interstimulus interval between a choice stimulus and the next cue (mean = 1350 ms, not counting time to initiate a new trial) is longer than that between the cue and choice within a trial (mean = 350 ms, not counting response time). We account for this difference by attenuating the decay of the resource variable after a choice stimulus by a factor of 3. For example, for a face choice that would be:

$$R_{t+1}^{face} = \left(1 - \frac{1-\gamma}{3}\right) R_t^{face}$$

To update the fatigue variable F after every stimulus presentation t, we used the net rate for that stimulus, normalized by the maximum unadapted net rate for the face and fruit choice stimuli (r^{fruit} and r^{face} multiplied by a factor of 1.2):

$$r_t^{norm} = \frac{r_t - r^{base}}{\max(r^{facechoice}, r^{fruitchoice})}.$$

F is then updated with the fatigue parameter α and recovery parameter β:

$$F_{t+1} = F_t - \alpha r_t^{norm} + \frac{\beta}{1 + r_t^{norm}}.$$

The value of F is constrained between 0 and 1. If the net firing rate to the previous choice stimulus decreases, the amount of recovery will eventually become higher than the amount of response fatigue (if it is 0, F will increase with β). We accounted for the difference in interstimulus interval mentioned earlier by attenuating the amount of recovery after a cue stimulus by a factor of 3 (i.e. using β/3).

Populations of neurons were simulated by generating firing rate values r^{base}, r^{face}, r^{fruit}, and r^{noise} from lognormal distributions. The average baseline firing rate was 10 Hz (SD = 5) and the average net response was 15 Hz (SD = 15) to faces, 6 Hz (SD = 6) to fruit, and .5 Hz (SD = .5) to high noise stimuli. This produced for simulation C response levels similar to those shown in Figure 3A of Bell et al. (2016). Fatigue and recovery parameters α (mean = .12, SD = .15) and β (mean = .08, SD = .1) and stimulus specific adaptation parameters γ (mean γ = .5, SD = .1) and δ (mean δ = .5, SD = .1) were generated using beta distributions. The experimental design was as described by Bell et al. (2016), except that a block had a fixed number of 50 trials (the mean block length in Bell et al. (2016)). For each simulated neuron, each of the five block-types (0%, 25%, 50%, 75%, 100% faces) was repeated three times in random order, resulting in a total of 750 trials. Behavioral responses were randomly generated to approximate the mean behavioral performances per p(face) as reported in Figure 1B of Bell et al. (2016).

6.3.2 Contributions of cue and choice stimuli

To examine whether the effect is driven by the probability of a face cue or that of a face choice, one could use the following regression model (C. Summerfield, personal communication):

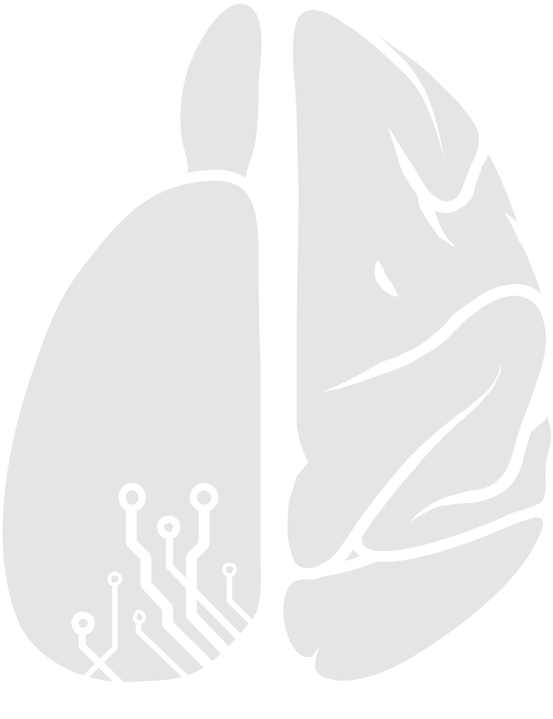
$$\begin{aligned} y = \ & \beta_0 + \beta_1 stimulus + \beta_2 stimulus \times p(face_{cue}) + \beta_3 stimulus \times p(face_{choice}) \\ & + \beta_4 stimulus \times prevchoice_1 + \beta_5 stimulus \times prevchoice_2 \\ & + \beta_6 stimulus \times prevchoice_3 + \beta_7 trial, \end{aligned}$$

where p($face_{cue}$) is identical to the p(face) of Model 1, and p($face_{choice}$) is based on the history of choices instead of cues. For our simulated data, the effect for the cue is clearly much stronger than for the choice (β_2 = -0.05, versus β_3 = -0.01). This could lead one to conclude that it is indeed the expectation of a face cue that drives the effect (which is false because we did not simulate expectation).

6.3.3 Decoding forthcoming cue identity

Bell et al. (2016) could decode the upcoming cue from baseline activity on trials were expectation and the stimulus were congruent: faces in high p(face) blocks versus fruits in low p(face) blocks. While they argued that “neural expectation signals carried information about its likely identity”, it cannot be excluded that these decoding results arose from temporally correlated slow fluctuations in baseline activity, which were unlikely to be equally spread across high p(face) and low p(face) blocks within a session. We confirmed this possibility by first arbitrarily dividing real monkey spiking data in blocks, and then decoding block membership from baseline activity, leading to an above chance accuracy similar in magnitude to theirs (data not shown).

DISCUSSION



In Part I-III, I have presented in detail all the work that we have done for this dissertation in the Laboratories for Biological Psychology and for Neuro- and Psychophysiology at KU Leuven. Here, I will try to provide a synthesis that reflects both direct implications, more indirect speculations, as well as my personal opinion as shaped by this work. I will start with an overview of the main results and implications per study. Next, I will discuss and speculate about what these results might mean more generally for each of the two research topics. Finally, I will end with some concluding remarks that reflect how this work has shaped my view on (visual) neuroscience in general.

7.1 SUMMARY OF THE MAIN RESULTS

7.1.1 PART I: VISUAL OBJECT RECOGNITION IN RATS

In Part I, we presented three studies related to visual object recognition in rats. Our angle was to use stimuli that are closer to a realistic visual experience for these animals, at least compared to artificial stimuli such as gratings, 2D shapes, or 3D shapes. The idea was that perhaps such natural stimuli were necessary and sufficient in order to easily elicit presumed visual object recognition capacities. For these experiments we used videos of rats and of inanimate moving objects. Our reasoning was simple: in their cages these animals are constantly surrounded by their cage mates, surely the visual image of other rats should be familiar to them?

First, we needed to investigate to which extent these animals can process our videos. With this goal in mind, we performed the experiment presented in **Chapter 1**, where we successfully trained five rats to go for target rat movies paired with distractor non-rat movies (moving objects or phase-scrambled version of the rat movies) in a two-alternative forced choice task. The crucial part is that these animals were able to generalize this classification to new pairs they had never seen before. They were able to do this in the face of considerable variation of stimuli both within and between training and test sets. In an additional test we showed that they did not rely on motion cues for successful classification. A control analysis showed that, while target-distractor differences in local screen luminance do explain some variability in classification accuracy, they do not explain the overall accuracy. We suggested that a further integration of features encoded in V1 might be required for successful generalization in this task.

In **Chapter 2**, we explored candidate areas where this further integration might take place. We used the same movies that had been used to train the rats in our categorization experiment and recorded neural responses in the visual cortex. In particular, we targeted neurons along a latero-temporal pathway (V1, LI, and TO) that is a putative homologue of the primate ventral stream, and looked for two hallmarks of the latter: preference for intact versus scrambled stimuli (Vogels, 1999c) and a category-related representation (Kiani et al., 2007). We found neither an overall preference for intact stimuli, nor evidence of a category-related representation. However, there was an increasingly different response pattern for natural versus phase-scrambled stimuli driven by an increased proportion of neurons preferring natural stimuli, perhaps paralleling changes from primate V1 to early extrastriate visual areas.

Unfortunately, we only have neural data for the training set videos, which makes it impossible to directly relate generalization performance to neural representations. One way to tackle such a problem, is to quantify and map the stimulus set to a feature space. We can then not only compare video representations in feature space with those in rat visual cortex, but also assess whether those features could support generalization in the behavioral task. In **Chapter 3**, we extracted features from several layers of a deep neural network (Tran et al., 2014). These features typically change across layers from V1-like to category-related (Güçlü and van Gerven, 2015). We found that stimulus representations in rat extrastriate visual cortex (LI and TO) corresponded best to up to mid-level representations in the neural network (late convolutional layers) and that generalization in the behavioral task could be supported by these mid-level representations.

Taken together, these studies suggest that the putative rat ventral stream results in a relatively complex representation of visual input: one that is not directly category-related, yet might support generalization in complex classification tasks.

7.1.2 Part II: Adaptation and expectation in rat visual cortex

After several studies on visual object recognition in the rat, in Part II we moved on to a second topic: neural adaptation and effects of expectation. The paradigm we focused on was a visual oddball paradigm, which is used for investigating pre-attentive processes of change detection in human electroencephalography (EEG) studies (Stefanics et al., 2014). Our motivation was a recent study that used this paradigm to characterize its effects in monkey IT neurons (Kaliukhovich and Vogels, 2014) and thus provided a frame of reference. They had only found repetition suppression for frequent standard stimuli, but no "surprise" response enhancements for rare deviant stimuli.

We performed a very similar experiment in rats, while measuring neural responses in both V1 and extrastriate area LI. The results of this study are presented in **Chapter 4.** In V1, we found clear repetition suppression for the standard, but no enhancement for the deviant. These results were very similar to what is found in monkey IT (Kaliukhovich and Vogels, 2014). However, in contrast with monkey IT and rat V1, we did find an enhanced response to the deviant in rat extrastriate area LI. In addition, we found evidence for a stronger repetition suppression for the standard. We speculated that these results might indicate a specialization in change detection of the pathway, related to the central function of visual predator detection in rats and mice (Wallace et al., 2013; Yilmaz and Meister, 2013).

7.1.3 Part III: Adaptation and expectation in macaque visual cortex

In Part III, we continued our research on the relation between neural adaptation and expectation. We moved on to monkeys to study effects of perceptual expectation in tasks and with stimuli associated with higher cognitive demands. Our motivation was to get as close as possible to the conditions of human fMRI studies in order to replicate their results and investigate what might actually be happening on a neural level.

Specifically, the results we tried to replicate in monkeys are based on a human fMRI study that showed that repetition suppression (usually of face stimuli) is stronger in blocks of high repetition probability (Summerfield et al., 2008). Despite other human fMRI replications, no such effect had been found in monkey IT neurons, using fractal images or a wide variety of natural image categories (Kaliukhovich and Vogels, 2011). In **Chapter 5**, we investigate two conditions that might be necessary in order to observe the repetition probability effect. The first is that it could be attention-dependent (Larsson and Smith, 2012), and therefore passive fixation as in Kaliukhovich et al. (2011) might not be sufficient. The second is that the effect might be restricted to specific stimulus categories such as faces (Kovács et al., 2013), perhaps based on prior experience (Grotheer and Kovács, 2014). We addressed these criteria by specifically using face stimuli combined with a stimulus-related orthogonal task, while recording neural responses in a face selective/responsive patch in macaque IT cortex. Despite these improvements, we did not find evidence for a repetition probability effect. Even in a second experiment, where repetition probability was task relevant *and* modulated task performance, it did not affect repetition suppression. Finally, in a follow-up fMRI experiment, we found opposite results in our two monkeys. We concluded that these results further call into question the generality of a role of perceptual expectation or top-down mechanisms in neural adaptation.

In contrast, a recent study did report effects of perceptual expectation on mostly face selective monkey IT neurons (Bell et al., 2016). The major problem in that study was that stimulus repetition is confounded with expectation. The authors were aware of this and used a multiple regression approach in an attempt to control for repetition suppression. In **Chapter 6**, we assessed the validity of their approach by testing their regression analysis on simulated data. Despite the fact that our simulations only implemented purely stimulus-driven effects of adaptation, we could replicate the regression results that Bell et al. (2016) interpreted as expectation effects. We concluded that their study did not provide unequivocal evidence for encoding of probabilistic information in IT.

7.2 THE RAT: A MODEL FOR VISUAL OBJECT RECOGNITION?

Almost 10 years ago, Zoccolan and colleagues argued for “an increased focus on rodents as models for studying high-level visual processing” (Zoccolan et al., 2009). In true primate-centric fashion, high-level vision is often operationally defined as object recognition, or anatomically as referring to later stages in the ventral visual stream (Cox, 2014). Over the past few years, a number of studies have been published that focused on object recognition and extrastriate visual processing in the rat, some of which constitute the first few chapters of this dissertation. In this section, I will discuss the general considerations and speculations that have followed from those studies.

In Chapters 1-3, we have adopted an object-recognition-centered approach while investigating the rat's visual system and abilities. Through this approach, we found both a similarity and differences with the primate ventral visual stream. The similarity that we found was an increasingly distinct representation of natural versus scrambled images (Chapter 2). This result can be seen as evidence for a hierarchical pathway that transforms the stimulus representation into one that is less determined by the light intensity pattern, and likely of relevance to the animal. In primates, the ultimate representation in IT cortex is usually interpreted as tailored towards the goal of object recognition (Tanaka, 1996; Vogels and Orban, 1996). However, it seems unlikely that object recognition is the main goal of the rat visual system. Quoting Marr (1982):

> The general point here is that because vision is used by different animals for such a wide variety of purposes, it is inconceivable that all seeing animals use the same representations; each can confidently be expected to use one or more representations that are nicely tailored to the owner's purposes (p. 32).

In other words, the idea is that animals that use vision for a different purpose should have a different functional specialization of their visual system. Comparing and characterizing similarities and differences of the visual system across species becomes very interesting when they can be related to similarities and differences in operational goals. If such a comparative approach is the objective, then the question of whether rats and mice are a good model for primate-based visual object recognition becomes less relevant. For example, in Chapter 4 we found that, as opposed to monkey IT, the putative rat ventral stream showed an enhanced response for visual oddball events. We speculated that this difference might be related to the emphasis of the rat visual system on predator detection.

On the other hand, our experiments of Chapter 1 and several other studies (Zoccolan, 2015) show that rats can be successfully trained to perform complex visual recognition tasks. In addition, studies of single neurons across the rat putative ventral stream have found evidence for properties typically associated with object recognition, such as tolerance for stimulus position (Vermaercke et al., 2014), but also size, rotation, and illumination (Tafazoli et al., 2017). Yet, similar properties might actually be required for successful navigation in complex environments under various lighting conditions (Cox, 2014). Indeed, it seems plausible for rats to actually use invariant representations, that they might have for navigation purposes, when they are trained in complex visual recognition tasks under experimental conditions. From this point of view, the rodent putative and primate ventral stream might be considered functional homologues because of such representational invariance, even if the purpose of rodents is not object recognition. The computational principles for gaining invariance might be preserved, even if the neurons are tuned to different features serving different purposes.

At the moment we have no idea what features are actually encoded in rodent extrastriate visual areas like LI and TO. In Chapter 3, we found that stimulus representations in LI and TO are best comparable with those in up to mid-level layers of a DNN, which were incidentally the minimum layers that could do the behavioral experiment of Chapter 1. These results are suggestive at best, but we could look at features encoded in these layers as possible candidates to investigate neural tuning in LI or TO. In a static DNN, these mid-level layers typically encode features such as shape and texture and corresponded mostly to human V4 (Güçlü and van Gerven, 2015).

In sum, we have learned that in some aspects the rat putative ventral stream is comparable to the one in primates, and in other aspects this is not the case. Ultimately, it makes sense to characterize the commonalities and differences and try to relate them to each animal's purpose of vision. In the grand scheme of this, the question of whether the rat visual system is a good model for high-level vision in primates is not really relevant. At best, it will lead to a characterization of similarities and differences anyway, at worst, it is a distraction leading commonalities to be over-emphasized by proponents and minimalized by opponents (and differences vice-versa).

7.3 In praise of simplicity: the case of adaptation

The idea of our brain as a prediction machine that constructs prior expectations of the environment has become very popular. It is a general theoretical framework for cortical responses (Friston, 2005), that emphasizes top-down mechanisms in explaining phenomena ranging from extra-classical receptive field effects (Rao and Ballard, 1999) to the mismatch negativity component in EEG (Stefanics et al., 2014), and that even found its way in DNNs (Lotter et al., 2016). Within this framework, expectation-based top down processes have been proposed to be an important contributor to repetition suppression (Summerfield et al., 2008) and the encoding of stimulus probabilities (Bell et al., 2016) in visual cortex. In this section, I will discuss the insights gained from our investigation of the relation between expectation and adaptation in these studies.

With regard to Summerfield's repetition probability effect in high-level visual cortex, there is a discrepancy between human fMRI studies and recordings of monkey neural activity. Despite multiple successful fMRI replications over the years (but see, Olkkonen et al., 2017), we couldn't even find an effect while using face stimuli and directing attention towards the stimulus repetitions (Chapter 5). When we tested the possibility that such effects are restricted to fMRI signals, we got mixed and contradictory results. Perhaps at best, those effects were too weak to warrant any useful comparison, but at worst, they suggest a difference between fMRI signals and neural activity. Neuroimaging signals can indeed contain task-related components that are not (or poorly) related to neural activity (Cardoso et al., 2012; Lima et al., 2014). Given the inconsistency and weak effects in our fMRI data, this explanation remains speculative, but it seems like a path worth further exploring. As a counterargument, Summerfield's repetition probability effect has also been reported in direct EEG measurements (Summerfield et al., 2011). However, without any source localization it is impossible to pinpoint its neuroanatomical source, which might as well not be visual cortex.

The possibility still remains that we have stumbled upon a species difference between macaques and humans, which was also considered by Kaliukhovich and Vogels (2011). Of course, this would mean that the explanation of repetition suppression/adaptation in terms of top-down expectation is restricted to certain species. Combined with previous studies suggesting that repetition probability effects are limited to certain stimulus sets or attentional states, it seems that the explanatory power of a perceptual expectation account is very restricted indeed. On a related note, in Chapter 4, we describe a surprise-related response enhancement in rat visual cortex, which was not there in monkey. Does that mean that only rodent and human higher visual cortex show effects of perceptual expectation, as opposed to monkeys? A more parsimonious interpretation would be that like in monkeys in rat visual cortex similar bottom-up and local mechanisms are at work, and that perhaps the response enhancement emerges from specific neural computations that are performed by the neural network (Solomon and Kohn, 2014).

In our final chapter, we closely investigate a paper on perceptual expectation signals in monkey IT (Bell et al., 2016). We uncovered (in addition to several methodological issues) that the results can be explained equally well by purely stimulus-driven effects of adaptation (Chapter 6). This is an excellent example of the unwarranted tendency to give results a "high-level" interpretation which emphasizes complicated cognitive concepts like expectation.

To conclude, it seems that the actual evidence does not support a general role of top-down based mechanisms of perceptual expectation in neural adaptation of sensory neurons. Such high-level explanations are often chosen in the name of a nice theory such as predictive coding, rather than by necessity. The bottom line is, no matter how attractive the theory, there is no reason to invoke it for explaining phenomena that can equally be covered by a more parsimonious account.

7.4 CONCLUDING REMARKS

I introduced this dissertation as a presentation of the work that we did on two specific topics situated within the general context of visual perception: the study of rat vision and neural adaptation in the visual system. By now, we have first considered the implications for each chapter more or less in isolation. Subsequently, I have attempted to make an abstraction towards more general implications on each of the two research topics presented here. I am aware that by moving away from the specifics of each particular study, my considerations get more speculative. With that in mind, I will finally summarize my concluding remarks with relation to visual neuroscience in general.

The first general consideration is that, when studying a complex system such as an animal's visual system, one should not lose sight of the actual purpose of this system: what problems does it solve for the animal? This idea is of course not new, and was an important part of the "three levels of understanding" framework proposed by Marr (1982), under the name of "computational theory". In the context of animal models, this viewpoint fosters a focus on comparative studies rather than trying to prove that one is a good model of the other. But possible implications are not restricted to situations where different animals are being considered. For example, Cox (2014) has argued that object recognition is a rather limited operational definition of (primate) higher-level vision. From this point of view, it would be too restrictive to reduce the function of the primate ventral visual stream to what has been called "core object recognition" (Dicarlo et al., 2012). In this way it is not surprising that DNNs as our current best models of the ventral stream are in many ways severely limited (Kriegeskorte, 2015): they have been trained within the same confines of "core object recognition".

Of course, one should be careful to not unnecessarily impose computational theory onto an interpretation of experimental results. For example, while phenomena such as neural adaptation effects can correlate well with predictions from predictive coding theory, they often do not require high-level mechanisms such as expectation-related feedback. This brings us to a second general consideration, namely that very "high-level" interpretations are often given to neural data without an appropriate control for simpler explanations. This is in particular an issue with a theory like predictive coding that is formulated so broadly and flexibly that it can be interpreted as compatible with almost any data. Other examples that are particularly sensitive to this issue are semantic interpretations of neural selectivity. For example, there is a strong tendency to interpret IT neural representations in terms of meaningful categories such as animate versus inanimate (Kiani et al., 2007; Kriegeskorte et al., 2008a), rather than in terms of shape similarity (Baldassi et al., 2013) which is mostly independent of meaning (Op De Beeck et al., 2008). Likewise, selectivity for biologically important semantic categories such as bodies and faces seems mostly determined by shape selectivity (Srihasam et al., 2014; Popivanov et al., 2016; Kalfas et al., 2017). This second consideration is in particular relevant for interpretations of neural data in rats as evidence for the existence of a rat object-processing pathway (Tafazoli et al., 2017). This might be a tempting conclusion, but it is perhaps not very compatible with the purpose of their visual system.

REFERENCES

Aarts E, Verhage M, Veenvliet J V, Dolan C V, van der Sluis S (2014) A solution to dependency: using multilevel analysis to accommodate nested data. Nat Neurosci 17:491–496.

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299.

Aggleton J., Keen S, Warburton E., Bussey T. (1997) Extensive Cytotoxic Lesions Involving Both the Rhinal Cortices and Area TE Impair Recognition But Spare Spatial Alternation in the Rat. Brain Res Bull 43:279–287.

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

Andermann ML, Kerlin AM, Roumis DK, Glickfeld LL, Reid RC (2011) Functional specialization of mouse higher visual cortical areas. Neuron 72:1025–1039.

Aparicio PL, Issa E, DiCarlo JJ (2016) Neurophysiological organization of the middle face patch in macaque inferior temporal cortex. J Neurosci 36:12729–12745.

Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS (2017) Seeing faces is necessary for face-domain formation. Nat Neurosci.

Baker M (2013) Neuroscience: Through the eyes of a mouse. Nature 502:156–158.

Baldassi C, Alemi-Neissi A, Pagan M, DiCarlo JJ, Zecchina R, Zoccolan D (2013) 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.

Barron HC, Garvert MM, Behrens TEJ (2016) Repetition suppression: a means to index neural representations using BOLD? Philos Trans R Soc B Biol Sci 371:20150355.

Bell AH, Summerfield C, Morin EL, Malecek NJ, Ungerleider LG (2016) Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Curr Biol 26:2280–2290.

Brooks DI, Ng KH, Buss EW, Marshall AT, Freeman JH, Wasserman E a (2013) Categorization of photographic images by rats using shape-based image dimensions. J Exp Psychol Anim Behav Process 39:85–92.

Burn CC (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.

Buzsáki G, Mizuseki K (2014) The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci 15:264–278.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. J Neurophysiol 98:1733–1750.

Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10.

Calabrese E, Badea A, Coe CL, Lubach GR, Shi Y, Styner MA, Johnson GA (2015) A diffusion tensor MRI atlas of the postmortem rhesus macaque brain. Neuroimage 117:408–416.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci.

Carandini and M (1997) A Tonic Hyperpolarization Underlying Contrast Adaptation in Cat Visual Cortex. Science (80-) 276:949–952.

Cardoso MMB, Sirotin YB, Lima B, Glushenkova E, Das A (2012) The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. Nat Neurosci 15:1298–1306.

Connor CE, Brincat SL, Pasupathy A (2007) Transformation of shape information in the ventral pathway. Curr Opin Neurobiol 17:140–147.

Coogan T a, Burkhalter A (1993) Hierarchical organization of areas in rat visual cortex. J Neurosci 13:3749–3772.

Cooke SF, Bear MF (2015) Visual recognition memory: a view from V1. Curr Opin Neurobiol 35:57–65.

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

Czigler I, Balázs L, Winkler I (2002) Memory-based detection of task-irrelevant visual changes. Psychophysiology 39:869–873.

David S V, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18:191–212.

De Baene W, Vogels R (2010) Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. Cereb Cortex 20:2145–2165.

Desimone R (1996) Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci 93:13494–13499.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–341.

Dicarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434.

Efron B (1987) Better Bootstrap Confidence Intervals. J Am Stat Assoc 82:171–185.

Einhäuser W, König P (2010) Getting real-sensory processing of natural stimuli. Curr Opin Neurobiol 20:389–395.

Ekstrom LB, Roelfsema PR, Arsenault JT, Bonmassar G, Vanduffel W (2008) Bottom-Up Dependent Gating of Frontal Signals in Early Visual Cortex. Science (80-) 321:414–417.

Espinoza SG, Thomas HC (1983) Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. Brain Res 272:137–144.

Euler T, Wässle H (1995) Immunocytochemical identification of cone bipolar cells in the rat retina. TL - 361. J Comp Neurol 361 VN-:461–478.

Ewbank MP, von dem Hagen EAH, Powell TE, Henson RN, Calder AJ (2016) The effect of perceptual expectation on repetition suppression to faces is not modulated by variation in autistic traits. Cortex 80:51–60.

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

Farley BJ, Quirk MC, Doherty JJ, Christian EP (2010) Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. J Neurosci 30:16475–16484.

Felleman DJ, Van Essen DC (1991) Distributed hierachical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Felsen G, Dan Y (2005) A natural approach to studying vision. Nat Neurosci 8:1643–1646.

Fioravante D, Regehr WG (2011) Short-term forms of presynaptic plasticity. Curr Opin Neurobiol 21:269–274.

Fishman YI, Steinschneider M (2012) Searching for the Mismatch Negativity in Primary Auditory Cortex of the Awake Monkey: Deviance Detection or Stimulus Specific Adaptation? J Neurosci 32:15747–15758.

Fize D, Cauchoix M, Fabre-Thorpe M (2011) Humans and monkeys share visual representations. Proc Natl Acad Sci U S A 108:7635–7640.

Fraedrich EM, Glasauer S, Flanagin VL (2010) Spatiotemporal phase-scrambling increases visual cortex activity. Neuroreport 21:596–600.

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. Nat Neurosci 16:974–981.

Freiwald WA, Tsao DY (2010) Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. Science (80-) 330:845–851.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Froudarakis E, Berens P, Ecker AS, Cotton RJ, Sinz FH, Yatsenko D, Saggau P, Bethge M, Tolias AS (2014) Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. Nat Neurosci 17:851–857.

Garrett ME, Nauhaus I, Marshel JH, Callaway EM (2014) Topography and Areal Organization of Mouse Visual Cortex. J Neurosci 34:12587–12600.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol 120:453–463.

Gelman A (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Anal 1:515–534.

Gelman A, Hill J (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge, UK: Cambridge University Press.

Gilad S, Meng M, Sinha P (2009) Role of ordinal contrast relationships in face encoding. Proc Natl Acad Sci U S A 106:5353–5358.

Gilbert CD, Li W (2013) Top-down influences on visual processing. Nat Rev Neurosci 14:350–363.

Girman SV, Sauvé Y, Lund RD (1999) Receptive field properties of single neurons in rat primary visual cortex. J Neurophysiol 82:301–311.

Glickfeld LL, Reid RC, Andermann ML (2014) A mouse model of higher visual cortical function. Curr Opin Neurobiol 24:28–33.

Greene MR, Oliva A (2009) Recognition of natural scenes from global properties: seeing the forest without representing the trees. Cogn Psychol 58:137–176.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23.

Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R (1998) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp 6:316–328.

Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) 107:293–321.

Grotheer M, Kovács G (2014) Repetition probability effects depend on prior experiences. J Neurosci 34:6640–6646.

Güçlü U, van Gerven MAJ (2015) Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J Neurosci 35:10005–10014.

Hamm JP, Yuste R (2016) Somatostatin Interneurons Control a Key Component of Mismatch Negativity in Mouse Visual Cortex. Cell Rep 16:597–604.

Harvey CD, Collman F, Dombeck D a, Tank DW (2009) Intracellular dynamics of hippocampal place cells during virtual navigation. Nature 461:941–946.

Hogg R V, Craig AT (1995) The central limit theorem. In: Introduction to mathematical statistics, 5th ed., pp 233–252. Eaglewood Cliffs, NJ: Prentice-Hall.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148:574–591.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154.

Hubel DH, Wiesel TN (1968) Receptive Fields and Functional Architecture of monkey striate cortex. J Physiol 195:215–243.

Huberman AD, Niell CM (2011) What can mice tell us about how vision works? Trends Neurosci 34:464–473.

Hummel JE, Biederman I (1992) Dynamic binding in a neural network for shape recognition. Psychol Rev 99:480–517.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. Science (80-) 310:863–866.

Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. J Neurosci 32:16666–16682.

Jacobsen T, Schröger E (2001) Is there pre-attentive memory-based comparison of pitch? Psychophysiology 38:723–727.

Juavinett AL, Callaway EM (2015) Pattern and Component Motion Responses in Mouse Visual Cortical Areas. Curr Biol 25:1759–1764.

Kalfas I, Kumar S, Vogels R (2017) Shape Selectivity of Middle Superior Temporal Sulcus Body Patch Neurons. Eneuro 4:ENEURO.0113-17.2017.

Kaliukhovich DA, Vogels R (2011) Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. Cereb Cortex 21:1547–1558.

Kaliukhovich DA, Vogels R (2014) Neurons in Macaque Inferior Temporal Cortex Show No Surprise Response to Deviants in Visual Oddball Sequences. J Neurosci 34:12801–12815.

Kaliukhovich DA, Vogels R (2016) Divisive Normalization Predicts Adaptation-Induced Response Changes in Macaque Inferior Temporal Cortex. J Neurosci 36:6116–6128.

Kampa BM, Roth MM, Göbel W, Helmchen F (2011) Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex. Front Neural Circuits 5:18.

Kaposvári P, Kumar S, Vogels R (2016) Statistical Learning Signals in Macaque Inferior Temporal Cortex. Cereb Cortex:1–15.

Kayaert G, Biederman I, Op de Beeck HP, Vogels R (2005) Tuning for shape dimensions in macaque inferior temporal cortex. Eur J Neurosci 22:212–224.

Kayser C, Körding KP, König P (2004) Processing of complex stimuli and natural scenes in the visual cortex. Curr Opin Neurobiol 14:468–473.

Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97:4296–4309.

Kohn A (2007) Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. J Neurophysiol 10461:3155–3164.

Kovács G, Iffland L, Vidnyánszky Z, Greenlee MW (2012) Stimulus repetition probability effects on repetition suppression are position invariant for faces. Neuroimage 60:2128–2135.

Kovács G, Kaiser D, Kaliukhovich D a, Vidnyánszky Z, Vogels R (2013) Repetition probability does not affect fMRI repetition suppression for objects. J Neurosci 33:9805–9812.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. Nat Rev Neurosci 12:217–230.

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn Sci 17:26–49.

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

Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:1–28.

Kriegeskorte N, Mur M, Ruff D a, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst:1–9.

Krubitzer L (2009) In search of a unifying theory of complex brain evolution. Ann N Y Acad Sci 1156:44–67.

Krubitzer L, Hunt D (2007) Captured in the net of space and time: Understanding cortical field evolution. In: The Evolution of Nervous Systems: a Comprehensive Reference (Kaas J, Krubitzer L, eds), pp 49–72. London: Elsevier.

Kruschke JK (2011) Doing Bayesian Data Analysis: A Tutorial with R and BUGS, 1st ed. Burlington, MA: Elsevier.

Kubilius J, Bracci S, Op de Beeck HP (2016) Deep Neural Networks as a Computational Model for Human Shape Sensitivity. PLoS Comput Biol 12:1–26.

Lange KL, Little RJA, Taylor JMG (1989) Robust Statistical Modeling Using the t Distribution. J Am Stat Assoc 84:881–895.

Laramée M-E, Boire D (2015) Visual cortical areas of the mouse: comparison of parcellation and network structure with primates. Front Neural Circuits 8:1–16.

Larsson J, Smith AT (2012) fMRI repetition suppression: neuronal adaptation or stimulus expectation? Cereb Cortex 22:567–576.

Lazic SE (2010) The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? BMC Neurosci 11:1–17.

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

Li FF, VanRullen R, Koch C, Perona P (2002) Rapid natural scene categorization in the near absence of attention. Proc Natl Acad Sci U S A 99:9596–9601.

Lima B, Cardoso MMB, Sirotin YB, Das A (2014) Stimulus-Related Neuroimaging in Task-Engaged Subjects Is Best Predicted by Concurrent Spiking. J Neurosci 34:13878–13891.

Logothetis N, Sheinberg D (1996) Visual object recognition. Annu Rev Neurosci 19:577621.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157.

Lomber SG, Payne BR, Cornwell P, Long KD (1996) Perceptual and cognitive visual functions of parietal and temporal cortices in the cat. Cereb Cortex 6:673–695.

Lotter W, Kreiman G, Cox D (2016) Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning.

Lunn D, Spiegelhalter D, Thomas A, Best N (2009) The BUGS project: Evolution, critique and future directions. Stat Med 28:3049–3067.

Maier A, Wilke M, Aura C, Zhu C, Ye FQ, Leopold D a (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. Nat Neurosci 11:1193–1200.

Marr D (1982) The Philosophy and the Approach. In: Vision, A Computational Investigation into the Human Representation and Processing of Visual Information, pp 8–38. San Francisco: W.H. Freeman and Company.

Marshel JH, Garrett ME, Nauhaus I, Callaway EM (2011) Functional specialization of seven mouse visual cortical areas. Neuron 72:1040–1054.

Mayrhauser L, Bergmann J, Crone J, Kronbichler M (2014) Neural repetition suppression: evidence for perceptual expectation in object-selective regions. Front Hum Neurosci 8:1–8.

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci 108:19401–19406.

Mill R (2014) Stimulus-Specific Adaptation, Models. In: Encyclopedia of Computational Neuroscience, pp 1–7. New York, NY: Springer New York.

Minini L, Jeffery KJ (2006) Do rats use shape to solve "shape discriminations"? Learn Mem 13:287–297.

Mishkin M, Ungerleider LG (1982) Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res 6:57–77.

Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414–417.

Moeller S, Crapse T, Chang L, Tsao DY (2017) The effect of face patch microstimulation on perception of faces and objects. Nat Neurosci 20:743–752.

Moeller S, Freiwald WA, Tsao DY (2008) Patches with Links: A Unified System for Processing Faces in the Macaque Temporal Lobe. Science (80-) 320:1355–1359.

Musall S, von der Behrens W, Mayrhofer JM, Weber B, Helmchen F, Haiss F (2014) Tactile frequency discrimination is enhanced by circumventing neocortical adaptation. Nat Neurosci 17:1567–1573.

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118:2544–2590.

Niell CM (2011) Exploring the next frontier of mouse vision. Neuron 72:889–892.

Nieto-Diego J, Malmierca MS (2016) Topographic Distribution of Stimulus-Specific Adaptation across Auditory Cortical Fields in the Anesthetized Rat Zatorre R, ed. PLOS Biol 14:e1002397.

Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J Neurosci 31:14551–14564.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646.

O'Connor DH, Huber D, Svoboda K (2009) Reverse engineering the mouse brain. Nature 461:923–929.

Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567–581.

Olkkonen M, Aguirre GK, Epstein RA (2017) Expectation modulates repetition priming under high stimulus variability. J Vis 17:10.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011.

Op De Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ (2008) A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. Cereb Cortex 18:1676–1694.

Orban GA (2008) Higher order visual processing in macaque extrastriate cortex. Physiol Rev 88:59–89.

Payne BR (1993) Evidence for visual cortical area homologs in cat and macaque monkey. Cereb Cortex 3:1–25.

Peelen M V, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature 460:94–97.

Plummer M (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003) (Hornik K, Leisch F, Zeileis A, eds). Vienna: Technische Universität Wien.

Popivanov ID, Schyns PG, Vogels R (2016) Stimulus features coded by single neurons of a macaque body category selective patch. Proc Natl Acad Sci 113:E2450–E2459.

Prusky GT, Harker KT, Douglas RM, Whishaw IQ (2002) Variation in visual acuity within pigmented, and between pigmented and albino rat strains. Behav Brain Res 136:339–348.

Prusky GT, West PW, Douglas RM (2000) Behavioral assessment of visual acuity in mice and rats. Vision Res 40:2201–2209.

R Core Team (2015) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rainer G, Augath M, Trinath T, Logothetis NK (2002) The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. Neuroimage 16:607–616.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3 Suppl:1199–1204.

Rosselli FB, Alemi A, Ansuini A, Zoccolan D (2015) Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. Front Neural Circuits 9:1–22.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 16:225–237.

Rust NC, Dicarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J Neurosci 30:12978–12995.

Rust NC, Movshon JA (2005) In praise of artifice. Nat Neurosci 8:1647–1650.

Sanchez-Vives M V, Nowak LG, McCormick DA (2000a) Membrane mechanisms underlying contrast adaptation in cat area 17 in vivo. J Neurosci 20:4267–4285.

Sanchez-Vives M V, Nowak LG, McCormick D a (2000b) Cellular mechanisms of long-lasting adaptation in visual cortical neurons in vitro. J Neurosci 20:4286–4299.

Sánchez RF, Montero VM, Espinoza SG, Díaz E, Canitrot M, Pinto-Hamuy T (1997) Visuospatial Discrimination Deficit in Rats after Ibotenate Lesions in Anteromedial Visual Cortex. Physiol Behav 62:989–994.

Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. Neuron 49:307–318.

Schultz J, Pilz KS (2009) Natural facial motion enhances cortical responses to faces. Exp brain Res 194:465–475.

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A 104:6424–6429.

Simonyan K, Zisserman A (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Inf Softw Technol 51:769–784.

Simpson EL, Gaffan EA (1999) Scene and object vision in rats. Q J Exp Psychol B 52:1–29.

Solomon SG, Kohn A (2014) Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons. Curr Biol 24:R1012–R1022.

Srihasam K, Vincent JL, Livingstone MS (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat Neurosci 17:1776–1783.

Stan Development Team (2016a) Stan Modeling Language User's Guide and Reference Manual, Version 2.9.0.

Stan Development Team (2016b) RStan: The R Interface to Stan, Version 2.9.0.

Stefanics G, Jan Kremláček J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 8:1–19.

Stojanoski B, Cusack R (2014) Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. J Vis 14:1–16.

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci 15:745–756.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006.

Summerfield C, Wyart V, Johnen VM, de Gardelle V (2011) Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. Front Hum Neurosci 5:1–13.

Tafazoli S, Di Filippo A, Zoccolan D (2012) Transformation-tolerant object recognition in rats revealed by visual priming. J Neurosci 32:21–34.

Tafazoli S, Safaai H, De Franceschi G, Rosselli FB, Vanzella W, Riggi M, Buffolo F, Panzeri S, Zoccolan D (2017) Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. Elife 6:1–39.

Talebi V, Baker CL (2012) Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. J Neurosci 32:1560–1576.

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

Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R (2015) The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. J Neurophysiol 113:1644–1655.

Tees RC (1999) The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. Behav Brain Res 106:55–73.

Thomas HC, Espinoza SG (1987) Relationships between interhemispheric cortical connections and visual areas in hooded rats. Brain Res 417:214–224.

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

Todorovic A, de Lange FP (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J Neurosci 32:13389–13395.

Towe AL, Harding GW (1970) Extracellular microelectrode sampling bias. Exp Neurol 29:366–381.

Tran D, Bourdev L, Fergus R, Torresani L, Paluri M (2014) Learning Spatiotemporal Features with 3D Convolutional Networks. Int J Comput Vis 101:6–21.

Tsao DY, Freiwald W a, Knutsen T a, Mandeville JB, Tootell RBH (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6:989–995.

Tsao DY, Freiwald W a, Tootell RBH, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674.

Tsao DY, Moeller S, Freiwald W a (2008) Comparing face patch systems in macaques and humans. Proc Natl Acad Sci 105:19514–19519.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398.

Utzerath C, St. John-Saaltink E, Buitelaar J, de Lange FP (2017) Repetition suppression to objects is modulated by stimulus-specific expectations. Sci Rep 7:8781.

Valdés-Hernández PA (2011) An in vivo MRI template set for morphometry, tissue segmentation, and fMRI localization in rats. Front Neuroinform 5:1–19.

Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RBH, Orban GA (2001) Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron 32:565–577.

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

Vermaercke B, Op de Beeck HP (2012) A multivariate approach reveals the behavioral templates underlying visual discrimination in rats. Curr Biol 22:50–55.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276.

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

Vinken K, Vermaercke B, Op de Beeck HP (2014) Visual categorization of natural movies by rats. J Neurosci 34:10645–10658.

Vinken K, Vogels R (2017) Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex. Curr Biol in press:R1–R3.

Vinken K, Vogels R, 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.

Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proc 2001 IEEE Comput Soc Conf Comput Vis Pattern Recognition CVPR 2001 1:I-511-I-518.

Viola P, Way OM, Jones MJ (2004) Robust Real-Time Face Detection. Int J Comput Vis 57:137–154.

Vogels R (1999a) Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur J Neurosci 11:1223–1238.

Vogels R (1999b) Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci 11:1239–1255.

Vogels R (1999c) Effect of image scrambling on inferior temporal cortical responses. Neuroreport 10:1811–1816.

Vogels R (2016) Sources of adaptation of inferior temporal cortical responses. Cortex 80:185–195.

Vogels R, Orban G a (1996) Coding of stimulus invariances by inferior temporal neurons. Prog Brain Res 112:195–211.

Wallace DJ, Greenberg DS, Sawinski J, Rulla S, Notaro G, Kerr JND (2013) Rats maintain an overhead binocular field at the expense of constant fusion. Nature 498:65–69.

Walther DB, Caddigan E, Fei-Fei L, Beck DM (2009) Natural scene categories revealed in distributed patterns of activity in the human brain. J Neurosci 29:10573–10581.

Wang L, Narayan R, Graña G, Shamir M, Sen K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? J Neurosci 27:582–589.

Wang Q, Burkhalter A (2007) Area map of mouse visual cortex. J Comp Neurol 502:339–357.

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.

Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. Curr Opin Neurobiol 17:423–429.

Warton DI, Hui FKC (2011) The arcsine is asinine: the analysis of proportions in ecology. Ecology 92:3–10.

Wörtwein G, Mogensen J, Williams G, Carlos JH, Divac I (1994) Cortical area in the rat that mediates visual pattern discrimination. Acta Neurobiol Exp (Wars) 54:365–376.

Wurtz RH (2009) Recounting the impact of Hubel and Wiesel. J Physiol 587:2817–2823.

Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci 111:8619–8624.

Yilmaz M, Meister M (2013) Rapid Innate Defensive Responses of Mice to Looming Visual Stimuli. Curr Biol 23:2011–2015.

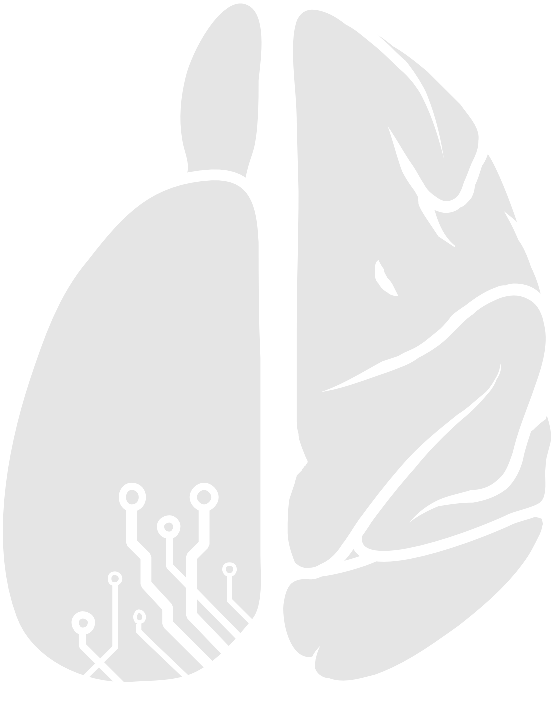
Yovel G, Freiwald WA (2013) Face recognition systems in monkey and human: are they the same thing? F1000Prime Rep 5:10.

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

Zoccolan D, Cox DD, Benucci A (2015) Editorial: What can simple brains teach us about how vision works.

Zoccolan D, Oertelt N, DiCarlo JJ, Cox DD (2009) A rodent model for the study of invariant visual object recognition. Proc Natl Acad Sci U S A 106:8748–8753.

APPENDIX



8.1 Scientific acknowledgement and personal contribution

8.1.1 Part I: Visual object recognition in rats

Chapter 1: A behavioral investigation of rat visual abilities

Vinken K., Vermaercke B., Op de Beeck H. (2014). Visual categorization of natural movies by rats. Journal of Neuroscience, 34 (32), 10645-10658.

We thank Steven Zwaenepoel and Kim Gijbels for assistance in training and testing the rats. This work was supported by the Fund for Scientific Research (FWO) Flanders (PhD Fellowship of K.V. and Grants G.0562.10 and G.0819.11) and the University of Leuven Research Council (Grant GOA/12/008). K.V., B.V., and H.O.d.B. designed research; K.V. performed research; B.V. contributed unpublished reagents/analytic tools; K.V. analyzed data; K.V. and H.O.d.B. wrote the paper.

Chapter 2: Natural stimulus representations in rat visual cortex

Vinken K., Van den Bergh G., Vermaercke B., Op de Beeck H. (2016). Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex. Cerebral Cortex, 26 (7), 3310-3322.

This work was supported by the Fund for Scientific Research Flanders (PhD Fellowship of K.V. and grant numbers G.0819.11, and G0A3913N) and the KU Leuven Research Council (grant number GOA/12/008). Funding to pay the Open Access publication charges for this article was provided by Fonds voor Wetenschappelijk Onderzoek (FWO) Flanders. Conceptualization and Writing: K.V., R.V., and H.O.d.B.; Resources and Software: B.V.; Investigation: K.V.; Formal Analysis: K.V.; Visualization: K.V.; Supervision: H.O.d.B. and G.V.d.B.

Chapter 3: A bridge between behavior and neurons

Vinken K., Op de Beeck H. (2017). Deep Neural Networks and Visual Processing in the Rat. Annual Conference on Cognitive Computational Neuroscience. New York City, NY, USA(conference proceedings)

Funding: FWO Flanders (Fellowship K.V., G0A3913N) and KU Leuven Research Council (C14/16/031). Conceptualization and Writing: K.V., H.O.d.B.; Formal Analysis and Visualization: K.V.

8.1.2 PART II: ADAPTATION AND EXPECTATION IN RAT VISUAL CORTEX

Chapter 4: Change detection in rat visual cortex

Vinken K., Vogels R., Op de Beeck H. (2017). Recent Visual Experience Shapes Visual Processing in Rats Through Stimulus Specific Adaptation and Response Enhancement. Current Biology, 27 (6), 914-919.

We would like to thank Gert Van den Bergh for performing the surgeries. This work was supported by Fonds voor Wetenschappelijk Onderzoek (FWO) Flanders (PhD Fellowship of K.V. and grant numbers G0A3913N, G058212N, G.00007.12-Odysseus, and G088916N) and the KU Leuven Research Council (grant number C14/16/031). Conceptualization and Writing: K.V., R.V., and H.O.d.B.; Investigation: K.V.; Formal Analysis: K.V.; Visualization: K.V.; Supervision: H.O.d.B.

8.1.3 PART III: ADAPTATION AND EXPECTATION IN MACAQUE VISUAL CORTEX

Chapter 5: The perceptual expectation account of neural adaptation

Not published.

We are grateful to C. Ulens, I. Puttemans, P. Kayenbergh, G. Meulemans, S. Verstraeten, W. Depuydt, and M. De Paep for technical support. This work was supported by Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO). K.V. is a research fellow of FWO. Conceptualization and Writing: K.V., H.O.d.B., R.V.; Investigation: K.V.; Formal Analysis and Visualization: K.V. ; Supervision: R.V.

Chapter 6: Adaptation confounded as expectation

Vinken K., Vogels R. (2017). Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex. Current Biology, in press.

K.V. is a research fellow of Fonds voor Wetenschappelijk Onderzoek Vlaanderen. This work was supported by Fonds voor Wetenschappelijk Onderzoek Vlaanderen. Conceptualization and Writing: K.V., R.V.; Data Simulation, Formal Analysis, and Visualization: K.V.

8.2 CONFLICT OF INTEREST STATEMENT

None declared.