**Carnegie Mellon University**
Mellon College of Science

# Thesis

Submitted in partial fulfillment of the requirements for the degree of Ph.D. in Biological Sciences

Visual Statistical Learning in Monkey Inferotemporal Cortex

Title

Presented by Suchitra Ramachandran.

Accepted by the Department of Biological Sciences

Major Professor _____ Date 12/8/14

Department Head John Woolford Date 2/16/15

Approved by the MCS College Council

Dean Frederick J. Gilman Date 2/16/15

# Visual Statistical Learning in Monkey Inferotemporal Cortex

A Thesis submitted to the Department of Biological Sciences

At Carnegie Mellon University

By

Suchitra Ramachandran

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

December, 2014

## ACKNOWLEDGEMENTS

The vision, experimental execution and writing of this dissertation would not have been possible without the contribution and support of a large number of people. I mention them below.

I owe much gratitude to Dr. Carl Olson, my dissertation advisor and mentor. He freely gave of his time and guidance, and provided me with a lot of physical and mental space to potter around. A lot of my learning in the lab came from reading Carl's grants and papers, from many discussions with him about philosophy of science, experimental design, data analysis and the principles guiding good figure making, and from his feedback to my talks. I thank him for the many things I implicitly learned from him. And I thank him most of all for the time and space to think.

Many thanks to the members of my thesis advisory committee: Dr. Alison Barth, Dr. Tai Sing Lee and Dr. Nathan Urban. They were generous with advice and suggestions, were always available for a quick discussion. In particular, I thank Dr. Lee for providing me with resources to simultaneously record from monkey visual areas V2 and V4, an experiment that generated some very exciting data outlined in Chapter VII.

I thank Dr. Sandra Kuhlman, for many general discussions about vision and for career advice.

I thank Travis Meyer, with whom I had a very productive collaboration that allowed us to acquire data relatively rapidly. During my first few months in the lab, he taught me to handle monkeys, set up the prep and record, as well as shared many tips and tricks to make the process easier. For helping with the technical details, as well as for plenty of discussions about conceptual ideas and the big picture, and refining ideas for experiments, I am very grateful to him.

The Brain Group (Olson, Colby, Lee and Cohen labs, housed in Mellon Institute) was a fantastic place to spend five years in and do science. There was always someone around to help me troubleshoot, or talk about a new idea. I acknowledge the following people: Dr. Carol Colby and Dr. Marlene Cohen, Erin Crowder, Roma Konecky and Marvin Leathers from the Olson lab

for being great colleagues and friends, Jason Samonds and Corentin Massot from the Lee lab for help with the V2/V4 recordings, Nate Hall from the Colby lab, and Doug Ruff and Gaby Costello from the Cohen lab for the enthusiasm, the sage advice, the insight and the laughs. I particularly thank them for feedback during Brain Group meetings, making sure I go out into the world endowed with extra-critical eyes and extra-thick skin.

Many thanks to Karen McCracken, who made the care of monkeys seem easier than it is.

During my time here, I benefitted from my association with two departments: the Department of Biological Sciences at Carnegie Mellon University, and the Center for the Neural Basis of Cognition (CNBC). Thanks to faculty and students in both these departments. Special thanks to Ena Miceli and Shoba Subramanian in Biology, and Barb Dorney, Rebecca Clark, Melissa Stupka and Anna Hegedus.

This dissertation would have not been possible without my support system: my parents - Appa and Amma, my extended family, my friends in Pittsburgh, particularly Phu Van, Ritika Tiwari, Madhumitha Ramesh, Salini Konikkat and Uttara Ananthakrishnan, and the folks at CMU- SPICMACAY. For keeping me sane, I thank Varun Bhaskar: resident comic, critic, coach, and partner.

Lastly I mention the subjects who participated in the studies coming up: monkeys Turk and Echo.

***Inspector Gregory (detective): "Is there any other point to which you would wish to draw my attention?"***

***Holmes: "To the curious incident of the dog in the night-time."***

***Gregory: "The dog did nothing in the night-time."***

***Holmes: "That was the curious incident."***

***– Sir Arthur Conon Doyle, 'Silver Blaze' (The Memoirs of Sherlock Holmes)***

# Contents

# List of figures

## ABSTRACT

Despite living in noisy sensory environments, humans and non-human primates have the ability to learn regularities and patterns in the environment solely on the basis of passive exposure. This ability to learn what is statistically likely and predictable in the environment is called statistical learning. Visual statistical learning of image sequences has been demonstrated at the level of single neurons in the rhesus macaque (monkey) inferotemporal cortex (IT). Upon subjecting monkeys to extensive exposure to pairs of images presented sequentially such that the display of one image always predicted the subsequent display of another image, IT neurons showed suppressed responses to images that occurred in a predicted context, but not when the same images occurred in an unpredicted context (Meyer & Olson, 2011). Upon investigating this effect, called prediction suppression, more thoroughly, we discovered that this effect depends on the conditional probability between the images presented sequentially. Further, the effect generalizes across time and space, it is domain specific, and it can be induced by training monkeys on longer sequences. These effects are long-lasting and robust: they persist at least for 20 months after initial training with no exposure to the stimuli in the interim. We have preliminary evidence for the existence of neurophysiological markers of statistical learning in areas upstream of IT in the ventral visual stream, suggesting that learning statistical regularities may be a fundamental function of sensory cortex.

# CHAPTER I

# GENERAL INTRODUCTION

## 1. Statistical Learning

### 1.1. Introduction

Despite huge advances in artificial intelligence and machine learning, the most sophisticated algorithms available at present cannot speak a natural language fluently or describe a movie by reporting the sequence of visual features in it. Machines need to be trained, mostly under supervision and feedback, with huge amounts of data for a very long period of time to achieve some modicum of success in parsing a visual scene or movie or interpreting natural language, especially in the presence of noise (Carlson et al., 2010; Fleuret et al., 2011; Lippmann, 1997). These are tasks that an average human child learns to solve quite well within a few years after birth.

What is unique about biological intelligence that makes these abilities possible? One critical feature of biological intelligence that sets it apart from artificial intelligence is the effortless ability to detect transitional probabilities in an incoming stream of data in an unsupervised manner (Griffiths, 2009). This process is broadly referred to as 'statistical learning'.

Statistical learning of sequential transitions has been behaviorally demonstrated in humans in both the auditory and visual domains. Such learning mechanisms also exist in non-human primates and rats. Neural signs of statistical learning has been observed in humans using different measurements (EEG, MEG, event-related potentials and fMRI). In the following sections, I review the evidence underlying these statements.

## 1.2 Statistical Learning in Humans

One of the earliest studies to examine transitional statistical learning in humans was in the domain of language acquisition. This study suggested that language acquisition in human infants depends on the ability of infants to learn sequential transitions of phonemic syllables. An eight month infant, upon exposure to a fluent speech stream, can find out which syllabic transitions occur within words (eg. the transition between 'pre' and 'tty', and the transition between 'ba' and 'by' in the phrase 'pretty baby'), and which syllabic transitions correspond to word-boundaries (eg. the transition between 'tty' and 'ba' in the same example) (Saffran, Aslin, & Newport, 1996). This was demonstrated by exposing infants to an artificial language consisting of sequences of nonsense syllables, however this finding has been replicated using natural grammars as well (Pelucchi, Hay, & Saffran, 2009). By manipulating the conditional probability of one syllable following another in sequence, they were able to demonstrate that this ability depended on the infants' ability to track the conditional probabilities between the syllables (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996).

Just as infants track probabilities in speech, they have also been shown to learn the transitional statistics in tonal sequences (Saffran, Johnson, Aslin, & Newport, 1999), and the such tracking depends on the absolute pitch of the tones in infancy (Saffran & Griepentrog, 2001). Adults are likewise sensitive to regularities in musical structure (Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

Transitional statistical learning has been demonstrated in the visual modality too. By means of an approach similar to that of (Saffran, Aslin, & Newport, 1996) but using sequences of visual images instead, it has been demonstrated that neonatal infants (Bulf, Johnson, &

Valenza, 2011), as well as 2-, 5- and 8-month-old infants (Kirkham, Slemmer, & Johnson, 2002) prefer to look at novel sequences of visual images over familiar sequences composed of the same elements. Likewise, human adults learn temporal sequence structure from repeatedly displayed visual shape sequences by computing joint and conditional probabilities (József Fiser & Aslin, 2002a).

In addition to transitional sequential learning of visual images, which is the focus of the studies described in this dissertation, adults (Fiser & Aslin, 2001) and infants (Fiser & Aslin, 2002b) are also able to learn regular features of static multi-element scenes as well upon exposure to them.

Statistical learning is thus ubiquitous in humans. This form of learning has been demonstrated in both human infants and adults in multiple sensory modalities, and is thought to underlie the acquisition of many abilities that are hallmarks of human intelligence, such as visual perception, spoken and written language and music (Kuhl, 2004). In addition, it has been argued that statistical learning might have a role in infants discovering causal structure about the world (Waismeyer, Meltzoff, & Gopnik, 2014).

### 1.3 Statistical Learning in Monkeys and Rats

Just as in humans, it has been observed that cotton-top tamarins (Hauser, Newport, & Aslin, 2001) and rats (Toro & Trobalón, 2005) that they can distinguish between probable and improbable syllabic transitions. These observations suggest that statistical learning may be a common, domain-general mechanism, present in many sensory modalities and species.

### 1.4 Neural Mechanisms Underlying Statistical Learning in Humans

Very little is known about the neural mechanisms underlying the learning of transitional statistics. The little information that is available comes from studies conducted in humans while measuring EEG, event-related potentials (ERP) and functional MRI responses. In the representative studies summarized below, characteristic neurophysiological signatures associated with either predictability or surprises in sequences of stimuli are discussed.

A large negative deflection is observed in human EEG response for oddball stimuli, occurring as a deviant in a sequence of regularly repeating stimuli, suggesting it signals a surprising violation of regular structure. This is called a ‘mismatch negativity’ (MMN) (Winkler, 2007). They were initially described in paradigms where a 'standard sound' was repeated, followed by a 'deviant' sound (such as AAAAB, where A and B were different sounds). The surprising oddball B elicited a greater negative ERP response. This increased responses to an unpredicted, surprising stimulus was elicited not just by deviation from repetition, but by violation of any regular sequential structure, suggesting that this may be a form of 'regularity representation' capable of creating a predictive model of the environment (Winkler, 2007).

In a study where the predictability of visual images in a sequential stream was explicitly manipulated, greater hemodynamic responses observed for predicted stimuli in a visual stream in the right anterior hippocampus (Turk-Browne, Scholl, Johnson, & Chun, 2010). Briefly, human participants were exposed to a stream of visual images (either faces or houses) and they were required to make a judgment about their category. However, unknown to them, the images were displayed with a predictable transitional structure, where the display of an image was 'paired' with the subsequent display of another particular image. Despite not explicitly realizing

the predictable structure in the design (assessed after the experiment with a questionnaire), subjects showed greater hemodynamic responses in the right anterior hippocampus for predicted stimuli compared to unpredicted stimuli. Although the sign of the effect was different in this experiment (greater responses to predicted stimuli than for surprising stimuli), the neural signal was still representing the transitional predictability of the stimuli.

A similar effect was observed in a study involving word segmentation. Humans were exposed to speech streams where concatenated nonsense syllables adhered to a certain transitional probabilities, and thus had certain regular, word-like transitions, (as in (Saffran et al., 1996)), and to other streams where the syllables were randomly presented. Increased hemodynamic activity was observed in left-lateralized temporal cortex for the predictable transitions (i.e. 'words') than the unpredictable nonsensical transitions (i.e. 'non-words') (McNealy, Mazziotta, & Dapretto, 2006). Again, the neural signal represented the transitional predictability of the stimuli, with greater responses for the predicted stimuli.

The representative phenomena discussed so far are examples of neural representations of stimulus predictability in sequences. However, depending on the task and the neural signal measured, predicted stimuli elicit decreased (EEG, ERP) or increased (fMRI) responses. Also, it is unclear what mechanisms underlie these representations, how they are learned and represented at the level of single neurons and how this knowledge is used by the brain. In order to gain an understanding of mechanism, it is imperative that we use animal models. Though monkeys and rats are not capable of behaviors as sophisticated as those of humans, such as language, it is still possible that they use mechanisms similar to those employed by humans to navigate through their native environments (Hauser et al., 2001; Toro & Trobalón, 2005). Thus, observations act

as a starting point from which we can start investigating the neural mechanisms underlying statistical learning using these organisms as model systems.

## 2. Functional Significance of Statistical Learning

*"It may often be rather hard to say how much of our perceptions as derived by the sense of sight is due directly to sensation, and how much of them, on the other hand, is due to experience and training."*(von Helmholtz, 1910/1925)

Many species including humans are able to learn statistical regularities in the sensory environment in what appears to be an implicit manner (Fiser, Berkes, Orbán, & Lengyel, 2010; Perruchet & Pacton, 2006). But why is this important? The following section makes the argument that the statistical learning is important because it allows us to predict what is going to happen next.

Briefly, the knowledge of what is 'regular, 'predictable' or 'expected' in a given context may provide information about when an event violates expectations and is thus surprising. 'Expectation' and 'surprise' in this context may not necessarily just map to cognitive expectation or cognitive surprise (such as what is subjectively experienced when a vehicle cuts in front of ours). These mechanisms may very well be below the threshold of consciousness. The statistics of the sensory environment may implicitly modify the functional neural architecture at the level of single neurons, such that upon encountering a sensory stimulus, depending on whether the stimulus fulfils or violates predictions in that context based on learned representations, neural signals representing prediction confirmation and/or prediction violation are propagated in cortex.

Why is the representation of predictions or prediction violations important?

(i) A prediction-confirming or prediction-violating signal in the brain may carry information about the ***salience*** of the event - it may be tied to the ability of the stimulus to grab attention in a particular context (or the lack of it).

(ii) A prediction-confirming or prediction-violating signal in the brain may represent a ***prediction error*** - i.e, signaling deviation from an anticipated predicted stimulus. A prediction error would be generated only for surprising stimuli, thus representing predictable stimuli with greater metabolic efficiency. Predictable events, based on expectations constructed in a 'dynamic, context-sensitive fashion' need not be represented as strongly as surprising events thus reducing the brain's free energy (Friston, 2005, 2010).

Prediction errors could also train a 'predictive model' (some of which are discussed below). The idea that learned representations influence perception in a top-down manner has been suggested by many others since and has inspired many theoretical models of cortical function. Many of these models have at their core the integration of sensory input fed forward from 'lower' areas to 'higher' areas, with internal representations fed back from 'higher' areas to 'lower' areas in a sensory pathway.

The idea that feedback loops were fundamental for perception, and the function of feedback is to match incoming sensory input to a 'template' was initially proposed by Miller and Pribram (Miller, Galanter, & Pribram, 1960) and later by David Mumford (Mumford, 1992), with anatomical evidence of massive feedback in sensory cortex to back it up. The 'templates' correspond to learned representations of environmental regularities that generate predictions. More specific instantiations of these ideas were proposed by Tai Sing Lee and David Mumford who proposed a mathematical framework to propagate Bayesian inference and

proposed that visual cortex can perform probabilistic inference based on learned priors (Lee & Mumford, 2003).

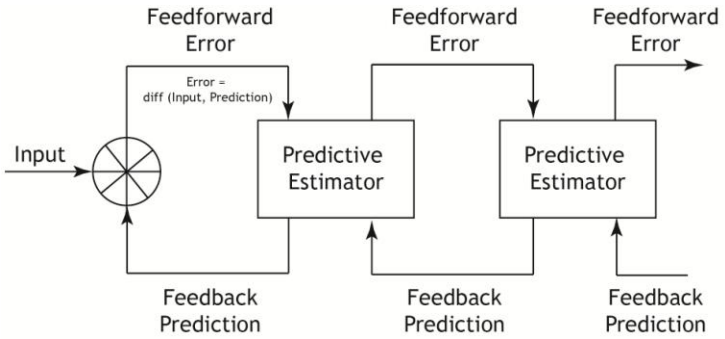


***Fig. 1.1: Hierarchical network for predictive coding*** *adapted from (Rao & Ballard, 1999). Predictions are fed back from higher areas. The different between the feed-forward input and predictions are each stage are propagated to higher areas.*

It has been proposed that the difference between the top-down prediction and the incoming sensory signal (the 'prediction error') is computed at every level and propagated forward to every subsequent layer (Lee & Mumford, 2003; Rao & Ballard, 1999). If the sensory signal matches the prediction, the prediction error propagated to areas with higher order representations of sensory stimuli (representing whole images or tone sequences) will be small. A large prediction error would signal violation of prediction to those areas (Fig.1.1, from (Rao & Ballard, 1999)).

Prediction errors have been ubiquitously observed in cortex in many contexts, including reward learning, motivational control, decision making (Hollerman & Schultz, 1998; Matsumoto & Hikosaka, 2007; Rescorla & Wagner, 1972; Roesch, Esber, Li, Daw, & Schoenbaum, 2012; Romo & Schultz, 1990; Schultz & Dickinson, 2000) as well as sensory perception (den Ouden, Kok, & de Lange, 2012; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne, Changeux, & Dehaene, 2012). Perceptual prediction errors include mismatch negativity responses in human ERPs: a large negative deflection in the EEG signal for an oddball stimulus in a sequence of repeated standard stimuli (den Ouden et al., 2012; Wacongne et al., 2012; Winkler, 2007). These prediction errors do not

signal general surprise or arousal – they are linked to a specific stimulus (den Ouden et al., 2012).

Thus, it is likely that stimulus-specific perceptual prediction errors may be computed at the level of single neurons in visual cortex. Further, such representations may be fundamental for perception, speeding up the representation of regular, predictable events and 'inferring' the visual world from the evidence provided by associative, predictable stimulus contexts. The seeds of this idea were planted by Hermann von Helmholtz in his *magnum opus* "Treatise on Physiological Optics". He was among the first to propose that perception may be a process of 'unconscious inference', with the inference based on experience (von Helmholtz, 1910/1925) (also see (Pollen, 1999)). Thus sensory systems may act as statistical inference engines, making inferences about incoming sensory information based on internally generated predictions. The core idea of the brain as a predictive machine has gained traction in recent times, and recent artificial intelligence systems are being designed based on these principles (George & Hawkins, 2005; Hawkins, George, & Niemasik, 2009; Hawkins, 2004).

To summarize, statistical learning and the generation of predictions in the visual system go hand-in-hand. In the absence of a learning mechanism that gathers statistical regularities in the environment, it would not be possible to know what events are 'irregular' and unpredicted. Neural responses based on prediction are stimulus specific: the specificity is engendered as a result of its repeated occurrence in a particular context during previous experience. Signals of prediction violation (prediction errors) may either represent arousal, stimulus salience, or may be used in the service of training a predictive model.

## 3. Macaque Inferotemporal Cortex (IT): A Plausible Site for Visual Statistical Learning

### 3. 1. Introduction to IT

Prediction coding theories posit a mechanism in which a prediction error signal is generated at each hierarchical stage of visual processing and fed back to an earlier hierarchical stage. The macaque ventral visual stream, extending from primary visual cortex to inferotemporal cortex in the primate visual system, from V1 in the occipital lobe to IT in the temporal lobe and beyond, is arranged in a sequence of hierarchical areas (Ungerleider & Mishkin, 1982). The receptive field size and complexity of visual features encoded increase at each successive level (Gross, Bender, & Rocha-Miranda, 1969). Each area sends feed-forward projections to higher areas, and feedback projections to lower areas (Felleman & Van Essen, 1991; Mumford, 1992). It is this architecture that has inspired models based on hierarchical predictive coding.

Neurons in IT, the terminus of the ventral stream, are strongly and selectively responsive to complex objects, including, faces, body parts and buildings with remarkable tolerance to object position and size (Desimone, Albright, Gross, & Bruce, 1984; Desimone & Gross, 1979; Gross et al., 1969; Gross, Rocha-Miranda, & Bender, 1972; Gross, 2008; Logothetis & Sheinberg, 1996). By virtue of IT's location in the visual processing hierarchy and its response properties, it is a possible site for the learning of statistical regularities involving images.

In accordance with this idea, extensive visual experience modifies visual responsiveness in the human temporal lobe. For example, upon extensive discrimination training between members of a novel class of objects (either an artificially created class of objects called greebles

or pre-existing classes like cars or birds), the right fusiform face area, a focal area in the human temporal lobe with predominantly face selective responses is recruited to represent the newly acquired objects of expertise (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000). Visual discrimination training was also shown to enhance fMRI response strength for trained images but not for untrained images in human extrastriate cortex (Beeck, Baker, Dicarlo, & Kanwisher, 2006). These ideas underlie the central thesis in Stanislas Dehaene's 'neuronal recycling' hypothesis, whereby that the ability of humans to read written language depends on the co-opting of existing object recognition mechanisms in the temporal lobe upon extensive experience with the visual symbols making up the written word (Dehaene, Cohen, Sigman, & Vinckier, 2005; Dehaene, Le Clec'H, Poline, Le Bihan, & Cohen, 2002; Dehaene, 2009; McCandliss, Cohen, & Dehaene, 2003).

### 3.2 Impact of Discrimination Training on Neurons in Monkey IT

Monkey IT is likewise modified by visual experience. Changes in neuronal image selectivity induced by discrimination training have been extensively studied at the level of single neurons in monkey inferotemporal cortex. Visual discrimination training has commonly been found to increase selectivity for trained images in monkey IT (Baker, Behrmann, & Olson, 2002; Kobatake et al., 1998). Likewise, requiring monkeys to make a categorical distinction at an arbitrary boundary in feature space enhances neuronal selectivity for the categorical boundary (De Baene, Ons, Wagemans, & Vogels, 2008; Sigala & Logothetis, 2002).

## 3.3 Impact of Passive Exposure on Neurons in Monkey IT

### *3.3.1 Familiarity*

These changes are not restricted to tasks requiring a behavioral response. passive exposure has been reported to enhance neuronal selectivity for trained images (Freedman, Riesenhuber, Poggio, & Miller, 2006). Passive familiarization is also sufficient for modifying the response properties of neurons such that neurons near each other are more likely than before training to respond similarly to different trained images (Erickson, Jagadeesh, & Desimone, 2000).

Passive exposure has also been shown to induce specific changes in the response strength of IT neurons. Single and multi-unit neural activity is lower in general for familiar images compared to novel images (Anderson, Mruczek, Kawasaki, & Sheinberg, 2008a; Lin Li, Miller, & Desimone, 1993; Meyer, Walker, Cho, & Olson, 2014; Mruczek & Sheinberg, 2007a). It is further known that visual experience leads to an increase in the peak responses in excitatory neurons, but results in a decrease in the peak response in inhibitory neurons (Woloszyn & Sheinberg, 2012a). Measures of local activity over some volume, such as local field potentials (LFPs) or event-related potentials (ERPs) also show an effect of familiarity: responses to familiar images are larger for familiar than for novel images (Anderson et al., 2008a; Peissig, Singer, Kawasaki, & Sheinberg, 2007).

Effects dependent on statistical learning arise from a simple form of statistical learning. Familiar images, having been seen frequently are probable, whereas novel images, having never been seen before are improbable.

### *3.3.2 Pair Coding*

The medial temporal lobe has been long implicated in associative learning. Lesions of the rhinal cortex, an area upstream of IT, compromised associative learning in monkeys (Murray, Gaffan, & Mishkin, 1993).

Passive exposure to images presented sequentially are sufficient to induce a form of association learning called pair coding in monkey IT and perirhinal cortex (PRh) neurons. Classic studies characterizing pair coding from Yasushi Miyashita's group are summarized below.

Monkeys were presented with a fractal image briefly, and after a long delay (16s), were asked to report whether a test image was the same as or different from the initially presented image ('matching task'). The fractal images could be images the monkey was extensively familiarized with in a prior training period, or completely novel images. Delay period activity in this task was observed for an over-learned familiar stimulus but not a novel stimulus held in memory during the delay (Miyashita & Chang, 1988).

Crucially, if the animal was familiarized on a set of visual stimuli but now presented in fixed sequence, and tested on each image subsequently in the matching task, cells with delay period activity for one of those stimuli were more likely to have strong delay period activity for images that were its immediate sequential neighbors during training, irrespective of the geometric similarity between the images or testing sequence order. This was the first demonstration of "experimentally controlled association between a temporally related set of stimuli" (Miyashita, 1988), an effect that was called 'pair coding'. In both these observations, it is important to note that extensive passive exposure with images was sufficient to induce the observed effects, although similar effects were also observed in a when monkeys performed a delayed match-to-sample task where the sample and match images were paired and the monkey had to retrieve the match image in a pair upon the display of the sample image (Sakai & Miyashita, 1991). The time-course of pair coding in single IT neurons seems to correlate with the time-course of learning the pair association (Messinger, Squire, Zola, & Albright, 2001). Pair coding was also observed in the perirhinal cortex (PRh) (Erickson & Desimone, 1999), an area

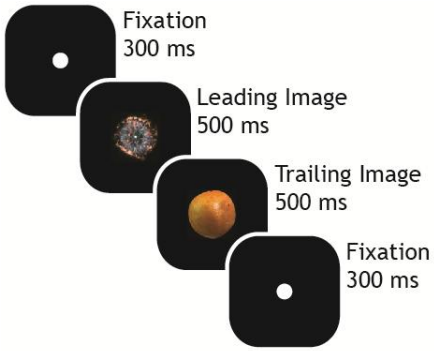


***Fig. 1.2: Trial Timing*** *Monkeys were exposed to pairs of images in fixed sequence. Each image was displayed for 500ms.*

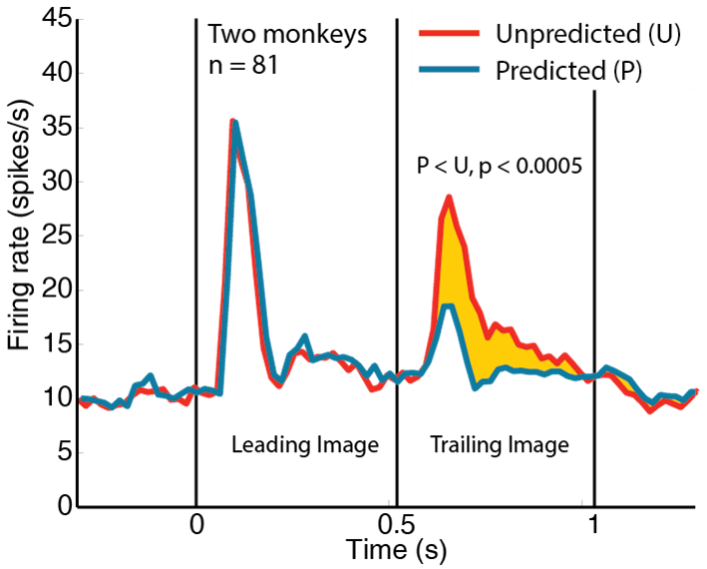


***Fig. 1.3: Prediction suppression in IT*** (Meyer & Olson, 2011) *Responses of IT neurons were suppressed when the monkey saw an image in a predicted context, but not when the same image was seen in an unpredicted context.*

downstream of IT that receives massive input from IT, and also feeds back to IT (Suzuki & Amaral, 1994).

*3.3.3 Prediction Suppression*

In studies of pair coding, neurons responsive to an image also tend to respond strongly to images presented in temporal contiguity during training (whether through passive exposure, or in the context of a delayed match-to-sample task). Such a paradigm where images are presented in sequence not only involve association (where images 'A' and 'B' are presented in temporal contiguity) but could also potentially involve prediction (where image 'A' could strongly and unilaterally predict image 'B'). We tested whether neurons in IT are sensitive to the prediction status of an image - i.e., is there a difference when an image is presented in a predicted context vs. when it is predicted in an unpredicted context. Briefly, we exposed monkeys to images paired in temporal sequence, such that the display of a 'leading' image (image $A_n$) always predicted the subsequent display of a 'trailing' image (image $B_n$) (Fig. 1.2). This exposure was carried out over a period of weeks. Post training, we measured the activity of IT neurons to both trained pairs ($A_n$ – $B_n$), where B was predicted in sequence and untrained pairs ($A_x$ - $B_n$, x =/= n), where B was unpredicted in sequence. We observed that the response to $B_n$ was much lower when it came up in a predicted context, compared to when it was seen in an unpredicted context, and the difference (unpredicted-predicted) response scaled multiplicatively with response strength to a particular image (Fig. 1.3). We call this effect 'prediction suppression'(Meyer & Olson, 2011a; Meyer, Ramachandran, & Olson, 2014). This is the first demonstration that the statistical structure of temporal sequences can be learned by neurons on the basis of long-term training. From studies in humans, it is suggestive that this form of learning is probabilistic in nature, a hypothesis that will be tested in experiments described in Chapter II.

## 4. Experimental Questions

I posed a series of questions to understand prediction suppression in further detail. The questions are summarized succinctly below. The motivation underlying each question and its specific instantiation is explained at the beginning of each chapter. The questions are:

1. Is the prediction effect due to frequency of pairing or conditional probability?
2. Is the effect dependent on suppressed responses to predicted images or enhanced responses to surprising images?
3. Is the effect sensitive to domain general (auditory-visual) prediction or domain specific (visual-visual) prediction?
4. Does the effect extend across intervening images when the monkey is trained on triplets?
5. Does the effect extend across long delays between images?
6. Is statistical learning present in areas outside IT?

Each of these questions was experimentally tested and the results are summarized in each of the six successive chapters.

## 5. A Brief Note on Methods

### 5.1. Task design

The general task design involved training monkeys on pairs of visual images in fixed sequence, (eg. Leading image A always followed by trailing image B, leading image C always followed by trailing image D etc.) (Fig.1.2). We would then test them on pairs of visual images, both conforming to the trained sequence (sequences A-B, C-D) and violating the trained sequence (sequences C-B, A-D). We included this condition as a control in all the experiments described

in subsequent chapters, and introduced variations in the design to test various hypotheses. The status of each image as leading or trailing in the sequence was always fixed (B-A, D-C etc. were not valid sequences).

### 5.2. Choice of images

Our images were sourced by making a search on Google Images. We searched for images of real-world objects with natural lighting and color of high resolution and neutral background. Images were generally not included if there was more than one object in the frame and if the object had an edge congruent with the edge of the image (i.e. if it appeared truncated). The raw images were processed in Adobe Photoshop. We foregrounded just the object against a black background and resized the image to desired dimensions (usually 88 pixels by 80 pixels).

### 5.3. Surgical procedures

All experimental procedures were approved by the Carnegie Mellon University Animal Care and Use Committee and were in compliance with the guidelines set forth in the United States Public Health Service Guide for the Care and Use of Laboratory Animals. Prior to training, each monkey underwent sterile surgery under general anesthesia so as to implant a head restraint bar. The skull was exposed, bone screws were placed around the perimeter of the skull, rapidly hardening acrylic was applied to cover the skull and a head restraint bar was embedded in the cap. After training, a 2-cm diameter of acrylic and skull overlying the left hemisphere in Monkey 1 (laboratory designation: Tu) and right hemisphere in Monkey 2 (laboratory designation: Ec) was removed and a vertically oriented cylindrical recording chamber was placed. The exact positioning of the recording chamber was calculated from MR images such that the chambers

provided direct access to ITC. Electrodes in guide tubes could be introduced through a grid placed in the chamber with grid holes placed 1-mm apart.

### 5.4. Data acquisition and analysis

Images were displayed on a PC using custom code written in Cortex software (Cortex NIMH). Eye position was monitored by means of an infrared tracking system (model RK-826, ISCAN). At the beginning of each day's session, a varnish-coated tungsten microelectrode with an initial impedance of ~1.0 MΩ at 1 kHz (FHC) was introduced into the temporal lobe through a transdural guide tube advanced to a depth such that its tip was ~1.5 cm above TE. The electrode was then advanced by use of a micromanipulator until phasic visual responses were observed. Action potentials of single neurons were isolated from the multi-neuronal trace by use of a commercially available spike-sorting system (Plexon). Action potential waveforms were recorded during the experiments and spike sorting was performed offline using commercially available software (Plexon). The raw signal was passed through a 4-pole filter with a high frequency cut-off of 170 Hz and stored continuously with 1-ms resolution for the analysis of LFP signals. Data analysis was performed on Matlab (R2010a and R2013a) with Statistics toolbox.

## REFERENCES


Anderson, B., Mruczek, R. E. B., Kawasaki, K., & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(11), 2540–52. doi:10.1093/cercor/bhn015

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. doi:10.1111/1467-9280.00063

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, *5*(11), 1210–6. doi:10.1038/nn960

Beeck, H. P. Op De, Baker, C. I., Dicarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *Journal of Neuroscience*, *26*(50), 13025–13036. doi:10.1523/JNEUROSCI.2481-06.2006

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–32. doi:10.1016/j.cognition.2011.06.010

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Jr, E. R. H., & Mitchell, T. M. (2010). Toward an Architecture for Never-Ending Language Learning. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*.

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & memory (Cold Spring Harbor, N.Y.)*, *15*(9), 717–27. doi:10.1101/lm.1040508

Dehaene, S. (2009). *Reading in the Brain*. New York: Penguin Viking.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335–41. doi:10.1016/j.tics.2005.05.004

Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, *13*(3), 321–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11930131

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*(December), 548. doi:10.3389/fpsyg.2012.00548

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective neurons of inferior temporal in the macaque. *The Journal of Neuroscience*, *4*(8), 2051–2062.

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

Erickson, C. a, & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *19*(23), 10404–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10575038

Erickson, C. a, Jagadeesh, B., & Desimone, R. (2000). Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nature Neuroscience*, *3*(11), 1143–8. doi:10.1038/80664

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *1*(1), 1–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1822724

Fiser, J, & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11760138

Fiser, József, & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. doi:10.1037//0278-7393.28.3.458

Fiser, József, & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–6. doi:10.1073/pnas.232472899

Fiser, József, Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–30. doi:10.1016/j.tics.2010.01.003

Fleuret, F., Li, T., Dubout, C., Wampler, E. K., Yantis, S., & Geman, D. (2011). Comparing machines and humans on a visual categorization test. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(43), 17621–5. doi:10.1073/pnas.1109168108

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(11), 1631–44. doi:10.1093/cercor/bhj100

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *360*(1456), 815–36. doi:10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–38. doi:10.1038/nrn2787

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, a W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–7. doi:10.1038/72140

Gauthier, I., Tarr, M. J., Anderson, a W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–73. doi:10.1038/9224

George, D., & Hawkins, J. (2005). A hierarchical bayesian model of invariant pattern recognition in the visual cortex. *Proceedings IEEE International Joint Conference on Neural Networks, 3*, 1812–1817. doi:10.1109/IJCNN.2005.1556155

Griffiths, T. L. (2009). Connecting human and machine learning via probabilistic models of cognition. In *Technical Program. 10th Annual Conference of the International Speech Communication Association. 2009.* (pp. 9–12).

Gross, C G, Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science (New York, N.Y.), 166*(3910), 1303–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4982685

Gross, C G, Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35*(1), 96–111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4621506

Gross, Charles G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia, 46*(3), 841–52. doi:10.1016/j.neuropsychologia.2007.11.009

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11124355

Hawkins, J. (2004). *On Intelligence* (Repreint e.). Times Books.

Hawkins, J., George, D., & Niemasik, J. (2009). Sequence memory for prediction, inference and behaviour. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 364*(1521), 1203–9. doi:10.1098/rstb.2008.0322

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1*(4), 304–9. doi:10.1038/1124

Kirkham, N. Z., Slemmer, J. a, & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2), B35–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11869728

Kobatake, E., Wang, G., Tanaka, K. (1998). Effects of Shape-Discrimination Training on the Selectivity of Inferotemporal Cells in Adult Monkeys *Journal of Neurophysiology*, 324–330.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–43. doi:10.1038/nrn1533

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, image science, and vision*, *20*(7), 1434–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12868647

Li, L., Miller, E. K., & Desimone, R. (1993). The Representation of Stimulus Familiarity Temporal Cortex in Anterior Inferior. *Journal of Neurophysiology*, *69*(6), 1918–1929.

Lippmann, R. P. (1997). Speech recognition by machines and humans. *Speech Communication*, *22*(1), 1–15. doi:10.1016/S0167-6393(97)00021-6

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621. doi:10.1146/annurev.ne.19.030196.003045

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–5. doi:10.1038/nature05860

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. doi:10.1016/S1364-6613(03)00134-7

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *26*(29), 7629–39. doi:10.1523/JNEUROSCI.5501-05.2006

Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12239–44. doi:10.1073/pnas.211431098

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Meyer, T., Walker, C., Cho, R. Y., & Olson, C. R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience*, *17*(10), 1388–1394. doi:10.1038/nn.3794

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart and Winston, Inc.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, *335*(27), 817–820.

Miyashita, Y., & Chang, H. S. (1988). Neur onal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, *331*(7), 68–70.

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Context familiarity enhances target processing by inferior temporal cortex neurons. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *27*(32), 8533–45. doi:10.1523/JNEUROSCI.2106-07.2007

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251.

Murray, E. A., Gaffan, D., & Mishkin, M. (1993). Neural Substrates of Visual Stimulus-Stimulus Association in Rhesus Monkeys. *Journa of Neuroscience*, *13*(October), 4549–4561.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. a, & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, *50*(1), 302–13. doi:10.1016/j.neuroimage.2009.12.019

Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(6), 1323–34. doi:10.1093/cercor/bhl043

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–85. doi:10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–8. doi:10.1016/j.tics.2006.03.006

Pollen, D. A. (1999). Feature Article On the Neural Correlates of Visual Perception, 4–19.

Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*(4), 721–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9161021

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *The European Journal of Neuroscience*, *35*(7), 1190–200. doi:10.1111/j.1460-9568.2011.07986.x

Romo, R., & Schultz, W. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *Journal of Neurophysiology*, *63*(3), 592–606. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2329363

Saffran, J., & Griepentrog, G. J. (2001). Absolute Pitch in Infant Auditory Learning: Evidence for Developmental Reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8943209

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*(6349), 152–155. Retrieved from https://www.ft.uam.es/neurociencia/CURSOS/Biofisica-2004/CLASES/Sak+91.pdf

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500. doi:10.1146/annurev.neuro.23.1.473

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–20. doi:10.1038/415318a

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, *11*(9), 1004–6. doi:10.1038/nn.2163

Suzuki, W. a, & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *The Journal of Comparative Neurology*, *350*(4), 497–533. doi:10.1002/cne.903500402

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *31*(25), 9118–23. doi:10.1523/JNEUROSCI.1425-11.2011

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16334058

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *Journal of Neuroscience*, *30*(33), 11177–11187. doi:10.1523/JNEUROSCI.0858-10.2010

Ungerleider, L. G., & Mishkin, M. (1982). Two Cortical Visual Streams. In *Analysis of Visual Behavior* (pp. 549–586).

Von Helmholtz, H. (1925). *Helmholtz's Treatise on Physiological Optics - Vol. III (The Perceptions of Vision)*. (J. P. C. Southall, Ed.) (Electronic.). New York: Dover: The Optical Society of America. Retrieved from http//psych.upenn.edu/backuslab/helmholtz

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, *32*(11), 3665–78. doi:10.1523/JNEUROSCI.5003-11.2012

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2014). Causal learning from probabilistic events in 24-month-olds: an action measure. *Developmental Science*, 1–8. doi:10.1111/desc.12208

Winkler, I. (2007). Interpreting the Mismatch Negativity. *Journal of Psychophysiology*, *21*(1992), 147–163. doi:10.1027/0269-8803.21.3.147

Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*(1), 193–205. doi:10.1016/j.neuron.2012.01.032

# CHAPTER II

# PREDICTION SUPPRESSION IN MONKEY INFEROTEMPORAL CORTEX DEPENDS ON THE CONDITIONAL PROBABILITY BETWEEN IMAGES


## ABSTRACT

When monkeys view two images in fixed sequence repeatedly over days and weeks, neurons in area TE of inferotemporal cortex come to exhibit prediction suppression. The trailing image elicits only a weak response when presented following the leading image that preceded it during training. Induction of prediction suppression might depend either on the contiguity of the images, as determined by their co-occurrence and captured in the measure of joint probability p(A,B), or on their contingency, as determined by their correlation and as captured in the measures of conditional probability p(A|B) and p(B|A). To distinguish between these possibilities, we measured prediction suppression after imposing training regimens that held p(A,B) constant but varied p(A|B) and p(B|A). We found that reducing either p(A|B) or p(B|A) during training attenuated prediction suppression as measured during subsequent testing. This outcome supports an interpretation based on contingency as distinct from contiguity.

## INTRODUCTION

Human infants and adults are able to learn rapidly through passive experience the statistical relations governing the transition from one element to the next in a structured stream of visual (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005; Howard et al., 2008; Turk-Browne et al., 2008; Kim et al., 2009; Bulf et al., 2011) or auditory stimuli (Saffran et al., 1996; Pelucchi et al., 2009; Romberg and Saffran, 2010). The neuronal mechanisms underlying this capacity are not yet well understood (Summerfield and Egner, 2009; Meyer and Olson, 2011; Wacongne et al., 2012; Gavornik and Bear, 2014). Single-neuron recording studies in monkeys have, however, begun to cast light on the issue. Repeated viewing of two images in fixed sequence, so that the leading image becomes a strong predictor for the trailing image, induces prediction suppression among neurons of inferotemporal area TE. They respond weakly to a trailing image when it follows the leading image that preceded it during training (Meyer and Olson, 2011; Meyer et al., 2014).

In humans, transitional statistical learning depends not just on the repeated pairing between successive elements in a stimulus stream but also on their conditional probability (Aslin et al., 1998; Fiser and Aslin, 2001; Meyer and Baldwin, 2011). For instance, infants exposed to a syllable stream learn a particular sequence as legitimate if the leading syllable is always followed by the same trailing syllable. But the effect is abolished by inserting additional instances in which the leading syllable is followed by a different trailing syllable (Aslin et al. 1998). Whether similar principles apply to prediction suppression in TE is unknown. To resolve this issue, we measured prediction suppression in monkeys exposed repeatedly to displays in which leading and trailing images were paired with equal frequency but their conditional probability varied.

## MATERIALS & METHODS

### Subjects

We studied two adult rhesus macaques: monkey 1 (male; laboratory designation Tu) and monkey 2 (female; laboratory designation Ec). Procedures were in accordance with guidelines set forth by the United States Public Health Service Guide for the Care and Use of Laboratory Animals and were approved by the Carnegie Mellon University IACUC.

### Training

Each monkey received repeated exposure to pairs of images presented in fixed sequence. The succession of events in each trial was: fixation spot (300 ms), leading image at screen center (503 ms), an 18 ms delay, trailing image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery (Fig. 2.1A). A trial was aborted without reward if the monkey failed to maintain fixation within a 4° x 4° central window. On each training day, the monkey completed one or more runs. A run consisted of 60 successfully completed trials. The trials conformed to ten conditions representing all allowable pairings of leading and trailing images (filled cells in Fig. 2.1B). Each condition was imposed six times

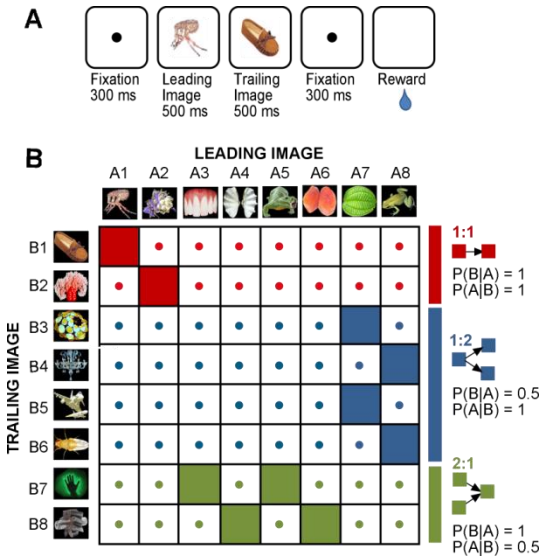


***Fig 2.1: Manipulating conditional probabilities: training paradigm A.*** *Timing of events within each trial during training and subsequent neuronal data collection sessions.* ***B.*** *Eight leading and eight trailing images were employed. During training, ten sequences were presented repeatedly with equal frequency (filled cells). During neuronal recording, each of the ten trained sequences was presented eight times and each of the 54 untrained sequences was presented once. Red, blue and green indicate the trained sequences (filled cells) and untrained sequences (dotted cells) that were compared in order to measure prediction suppression under the 1:1, 1:2 and 2:1 conditions respectively.*

during a run. The conditions were interleaved randomly subject to the constraint that within each block of ten successfully completed trials each condition had to be imposed once. The number of runs completed on a day ranged from 1 to 12 in monkey 1 and from 1 to 3 in monkey 2. Monkey 1 viewed each sequence 834 times during 139 runs extending over 27 days. Monkey 2 viewed each sequence 408 times during 68 runs extending over 40 days.

**Testing**

During neuronal data collection, the monkeys completed trials identical to training trials with regard to the timing of events (Fig. 2.1A). The status of the images as leading or trailing was the same as during training. However, any leading image might be followed by any trailing image. A run consisted of 134 trials encompassing 80 trained sequences (each of ten trained sequences occurring eight times) and 54 untrained sequences (each of 54 untrained sequences occurring once). The conditions were imposed in random order with replacement on error.

**Images**

All stimuli were digitized images of background-free objects. When presented on an LCD monitor 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. The full stimulus set for monkey 1 consisted of eight leading images and eight trailing images paired according to rules summarized in Fig. 2.1B. The same images were used in monkey 2 but with their sequential status (leading or trailing) reversed and the pattern of pairing altered so that no images paired in monkey 1 were paired in monkey 2.

**Recording**

An electrode was introduced through a vertical guide tube into left (monkey 1) or right (monkey 2) temporal lobe. Recording sites, identified by extrapolation from MRI-visible fiducial markers within the chamber, were within the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the interaural plane by 16-19 mm in monkey 1 and 13-16 mm in monkey 2.

**Database**

We recorded from 51 sites (30 and 21 in monkeys 1 and 2 respectively). Traces from these sites passed through a low-pass filter with a high frequency cut-off of 170 Hz formed the LFP database. Neurons characterized during a complete test run numbered 112 (67 from monkey 1 and 45 from monkey 2). We classified a neuron as visually responsive if, for either the leading or the trailing image, the mean firing rate in a window 50-300 ms following image onset exceeded the mean firing rate in a 100 ms baseline window centered on image onset (one-tailed t-test, $\alpha = 0.05$). The neuronal database consisted of 86 neurons meeting this criterion.

**RESULTS**

The experiment began with a training period extending over multiple weeks during which the monkeys viewed each training sequence more than 400 times (Fig. 2.1A). Ten image sequences, constructed from eight leading and eight trailing images, were presented during this period (Fig. 2.1B). Within each training session, each sequence appeared the same number of times. Thus the ten sequences were equal with regard to their absolute probability. However, across these sequences, the conditional probability between the various leading images and trailing images varied. Across the two sequences highlighted in red in Fig. 2.1B, one leading image (A) and one trailing image (B) always appeared in sequence. Accordingly, we designate this as the 1:1

condition. Under this condition, the appearance of A guaranteed that B would follow: p(B|A) = 1. Likewise, the appearance of B guaranteed that A had preceded: p(A|B) = 1. Across the four sequences highlighted in blue in Fig. 2.1B, a given leading image could precede either of two trailing images. We term this the 1:2 condition. Under this condition, p(A|B) = 1 but p(B|A) = 0.5. Across the four sequences highlighted in green in Fig. 2.1B, either of two leading images could precede a particular trailing image. We term this the 2:1 condition. Under this condition, p(B|A) = 1 but p(A|B) = 0.5. After

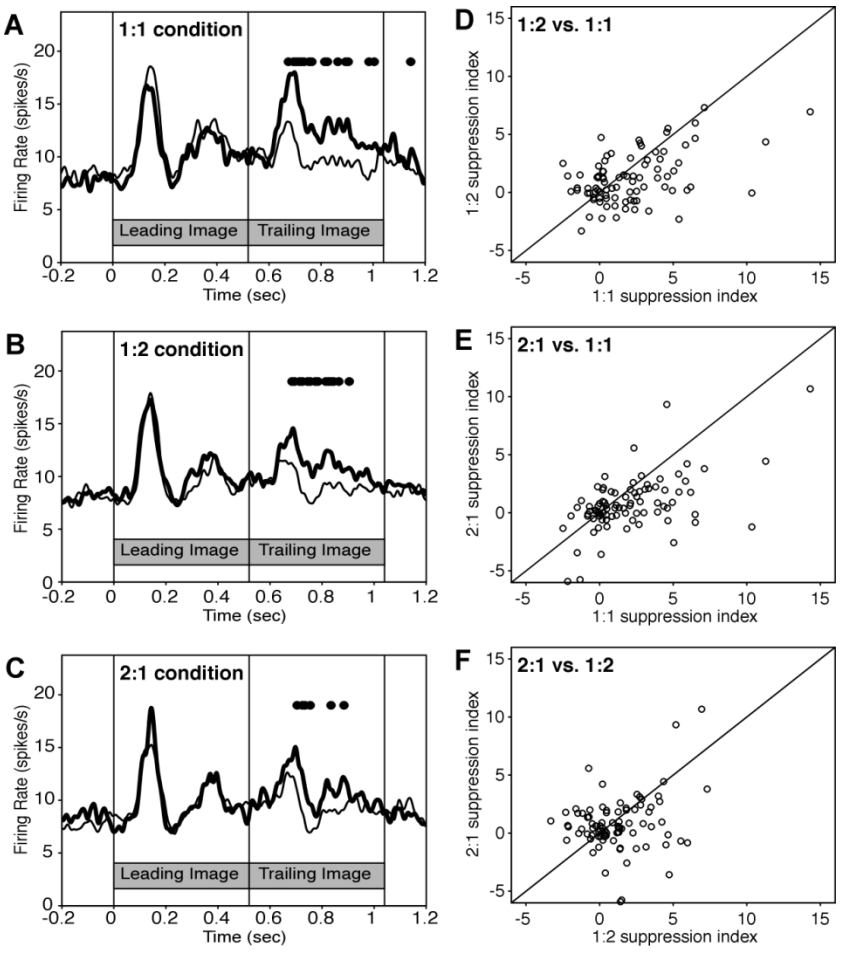


***Fig. 2.2: Neuronal population activity for conditions with different conditional probability dependencies A.*** *Mean firing rate of 86 neurons during testing with sequences containing trailing images belonging to a 1:1 training pair (red conditions in Fig. 2.1B). Thick curve represents activity under trained conditions (filled red cells in Fig. 2.1B). Thin curve represents activity under untrained conditions (dotted red cells in Fig. 2.1B). Filled symbols above the curves indicate 10 ms bins during which activity elicited by untrained sequences significantly differed from activity elicited by trained sequences (two-tailed paired t-test, $\alpha = 0.05$, $n = 86$).* ***B.*** *Population activity elicited by sequences containing trailing images belonging to a 1:2 training pair (blue conditions in Fig. 2.1B).* ***C.*** *Population activity elicited by sequences containing trailing images belonging to a 2:1 training pair (green conditions in Fig. 2.1B).* ***D.*** *The prediction suppression index measured under the 1:2 condition is plotted against the prediction suppression index plotted under the 1:1 condition.* ***E.*** *The 2:1 index plotted against the 1:1 index.* ***F.*** *The 2:1 index plotted against the 1:2 index.*

completion of training, we measured the responses of neurons in anterior TE to leading and trailing images presented in both trained and untrained sequences. During each run, the ten trained sequences (filled cells in the matrix of Fig. 2.1B) appeared eight times each and the 54 untrained sequences (dotted cells in the matrix of Fig. 2.1B) appeared once each for a total of 134 trials. We collected full data sets from 86 visually responsive TE neurons (56 in monkey 1 and 30 in monkey 2). All analyses described below were conducted on data combined across the two monkeys. Every effect described was present in both monkeys and achieved significance in at least one monkey.

To determine whether prediction suppression occurred after each training procedure, we compared population histograms representing the mean firing rate during trials in which trailing images appeared in trained sequences (and thus were predicted) or in untrained sequences (and thus were unpredicted). Prediction suppression occurred for trailing images from all three training sets (Fig. 2.2A-C). We took as a quantitative index of prediction suppression the mean firing rate 100-500 ms after image onset on trials in which the trailing images were unpredicted minus the same measure on trials in which they were predicted. This suppression index was significantly greater than zero under the 1:1 condition (mean = 3.1 spikes/sec, $p = 4.1 \text{ E-}10$), the 1:2 condition (mean = 2.0 spikes/sec, $p = 7.7 \text{ E-}8$) and the 2:1 condition (mean = 1.6 spikes/sec, $p = 4.8 \text{ E-}5$; two-tailed paired t-test, $n = 86$). The response to the unpredicted trailing image was not affected by the training status (1:1, 1:2 or 2:1) of the leading image (ANOVA, $n = 86$).

To determine whether prediction suppression was attenuated by manipulations reducing the conditional probability between the leading and trailing images, we plotted, across all neurons, the suppression index measured under each condition against the suppression index measured under each other condition. The results make clear that the suppression index under the 1:1

condition tended to be greater than under the 1:2 and 2:1 conditions (Fig. 2.2D-F). This effect was significant ($p = 8.6$ E-3 for 1:1 vs. 1:2 and 2.9 E-4 for 1:1 vs. 2:1, two-tailed paired t-test, $n = 86$). The 1:2 and 2:1 conditions did not differ significantly from each other ($p = 0.23$). We conclude that a training procedure reducing either p(A|B) or p(B|A) attenuates prediction suppression at the level of spiking activity.

To determine whether comparable effects were present at the level of the local field potential (LFP), we analyzed data collected from the 51 sites at which we had monitored neuronal activity (30 and 21 sites in monkeys 1 and 2 respectively). Prediction suppression is manifest at the level of the LFP as a reduction in the amplitude of the excursion from maximal negativity at around 200 ms to maximal positivity at

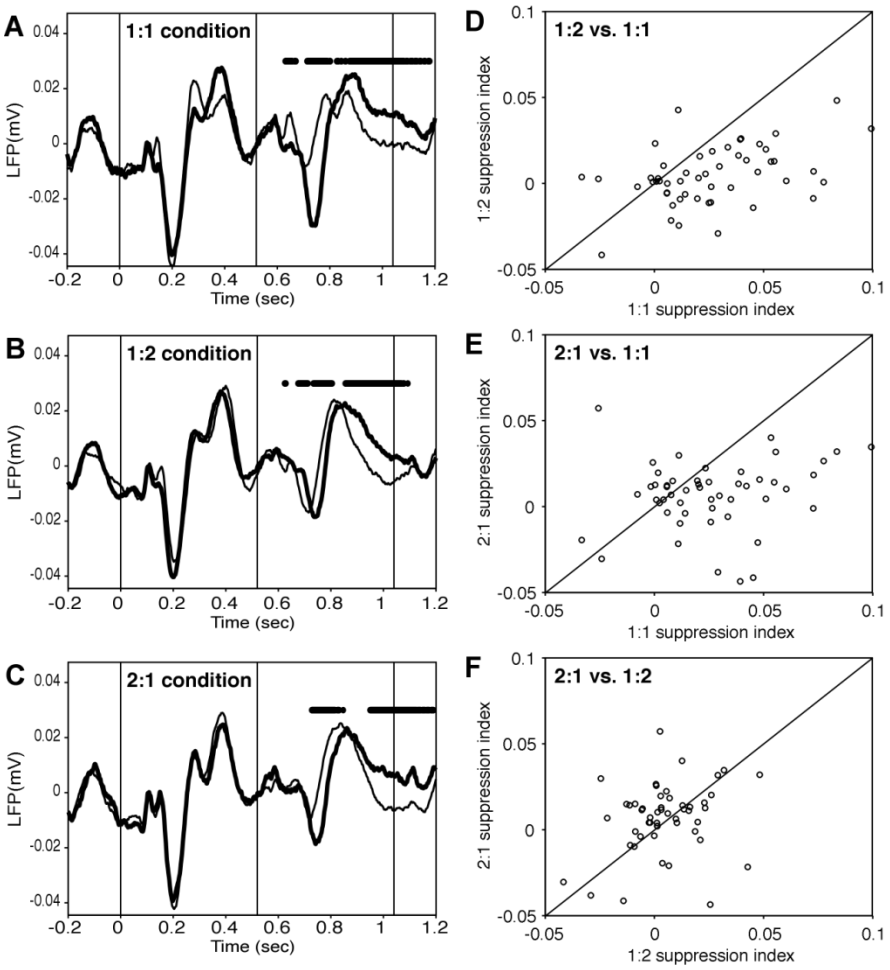


***Fig 2.3: LFP activity for conditions with different conditional probability dependencies.*** *Local field potential responses at 51 sites. Format and conventions as in Fig. 2.*

around 300 ms (Meyer and Olson, 2011; Meyer et al., 2014). This effect was highly significant in the 1:1 condition (Fig. 2.3A; mean = 25.9 μV, p = 2.8 E-8), marginal in the 1:2 condition (Fig. 2.3B; mean = 4.5 μV, p = 0.065) and significant in the 2:1 condition (Fig 2.3C; mean = 6.8 μV, p = 0.019, two-tailed paired t-test, n = 51, unpredicted minus predicted peak-trough excursion).

To determine whether the cross-condition differences were significant, we computed, for each LFP site under each condition, a suppression index equal to the unpredicted minus predicted peak-trough excursion. Then we plotted the suppression indices measured under each condition against the suppression indices measured under each other condition (Fig. 2.3D-F). Suppression under the 1:1 condition was significantly greater than under the 1:2 and 2:1 conditions (p = 3.5 E-7 and 7.2 E-5 for the 1:2 and 2:1 comparisons respectively, two-tailed paired t-test, n = 51). The 1:2 and 2:1 conditions did not differ significantly from each other (p = 0.48). We conclude that a training procedure reducing either p(A|B) or p(B|A) attenuates prediction suppression at the level of the LFP.

It was evident from the plots representing voltage as a function of time that voltage was first more negative and then more positive under the unpredicted as compared to the predicted condition (Fig. 2.3A-C). Both effects were significant under all conditions (p < 1 E-5, two-tailed paired t-test, n = 51, based on negative epoch 720-820 ms and positive epoch 860-1120 ms after trailing image onset). Prediction suppression of early negativity was greater under the 1:1 condition than under either of the other conditions, with the difference achieving significance for the 1:2 condition (p = 6.5 E-5) and approaching significance for the 2:1 condition (p = 0.070, two-tailed paired t-test, n = 51). Prediction suppression of late positivity did not differ significantly across conditions. We draw two conclusions from this analysis. First, prediction suppression is evident as a reduction both of the early negative and in the late positive LFP

response. Second, training procedures that reduce p(A|B) or p(B|A) cause an attenuation of prediction suppression largely confined to the early negative phase.

## DISCUSSION

The aim of this experiment was to determine whether prediction suppression (Meyer and Olson, 2011; Meyer et al., 2014) depends solely on the contiguity between the leading and the trailing images, as determined by their repeated pairing, or also on their mutual contingency, as determined by the ability of one to predict the other. Our results establish that contingency matters. Prediction suppression is reduced if the contingency between the images is degraded. We consider below how to interpret this finding in terms of the behavioral significance of prediction suppression and the mechanisms of synaptic plasticity that underlie it.

We suggested previously that prediction suppression in TE serves to reduce the salience of a predicted and therefore uninformative trailing image (Meyer and Olson, 2011; Meyer et al., 2014). If so, then prediction suppression can be seen as a specific phenomenon arising from the general ability of the neocortex to predict future events (Friston, 2005; Hawkins and Blakeslee, 2005; Bar, 2009). It is natural to think of prediction as giving rise to an active representation of an impending event. However, it is equally plausible to imagine it as producing a passive state with the property of filtering out the predicted event in the event of its occurrence. Damped responding to predicted events is frequently observed at the level of perceptual, cognitive and motivational systems (den Ouden et al., 2012). Filtering out could be adaptive both in preventing the capture of attention by things that require no processing (Foley et al., 2014) and in allowing the refinement of the very brain mechanisms that mediate prediction-making. The idea that surprising events (prediction errors) fine-tune the predictive apparatus lies at the heart of animal

learning theory (Kamin, 1969; Pearce and Hall, 1980; Schultz and Dickinson, 2000; Courville et al., 2006).

If prediction suppression indeed arises from the tendency of the visual system to filter out the representation of a predicted event, then it should be possible to reduce the strength of prediction suppression by reducing the predictability of the trailing image during training. That is exactly what we accomplished in the 1:2 condition. Reducing p(B|A) to 0.5 induced a corresponding reduction in prediction suppression relative to the 1:1 control. The outcome of the 2:1 condition is difficult, however, to explain in this framework. In the 2:1 condition, following a leading image, the probability of the paired trailing image, p(B|A), was 1.0 just as in the 1:1 control condition. Nevertheless, prediction suppression was reduced. The feature distinctive of the 2:1 condition was the low probability of the leading image given the trailing image: p(A|B) = 0.5. To explain this result requires considering an alternative framework.

The dependence of prediction suppression on both p(B|A) and p(A|B) can be explained parsimoniously in terms of a covariance-based synaptic learning rule (Kamin, 1969; Sejnowski, 1977; Pearce and Hall, 1980; Sejnowski et al., 1989; Schultz and Dickinson, 2000; Courville et al., 2006). Consider a network in which a neuron responsive to leading image A inhibits a neuron responsive to trailing image B (Fig. 2.4). Inhibition serves here as a proxy for the unknown mechanism underlying prediction suppression. The learning rule governing the strength of the inhibitory synapse is given by:

$$DW_{B,A}(t) = e * [Y_A(t) - \langle Y_A \rangle] * [Y_B(t) - \langle Y_B \rangle]$$

where e is the learning rate constant, $DW_{B,A}(t)$ is the change in the weight at time t, $Y_A(t)$ and $Y_B(t)$ are the firing rates of neurons A and B at time t and $\langle Y_A \rangle$ and $\langle Y_B \rangle$ are the mean firing

rates over some prior interval. Under all three training regimens, there are trials in which image A is paired with image B. Co-activation of neurons responsive to A and B induces an increase of the weight (W) of the inhibitory synapse between them (Fig. 2.4A). Under the 1:2 condition, there are also trials involving the sequence A,~B where ~B is the other trailing image paired with A. On these trials, the neuron responsive to A is active but the neuron responsive to B is not. This induces a decrease in the weight W (Fig. 2.4B). Under the 2:1 condition, there are trials involving the sequence ~A,B where ~A is the other leading image paired with B. On these trials, the neuron responsive to B is active but the neuron responsive to A is not. This induces a decrease in the weight W (Fig. 2.4C). Thus the asymptotic level of W is lower under the 1:2 and 2:1 conditions than under the 1:1 condition. Whether plasticity actually conforms to a covariance-based rule in the most commonly studied form of cortical plasticity, long-term potentiation, has been subject to debate (Stanton and Sejnowski, 1989; Kerr and Abraham, 1993; Paulsen et al., 1993). Insertion of trials in which postsynaptic activity occurs without presynaptic activity (Fig. 2.4C) has been reported to weaken LTP in accordance with the covariance principle (Pockett et al., 1990; Christofi, 1993; Bauer et

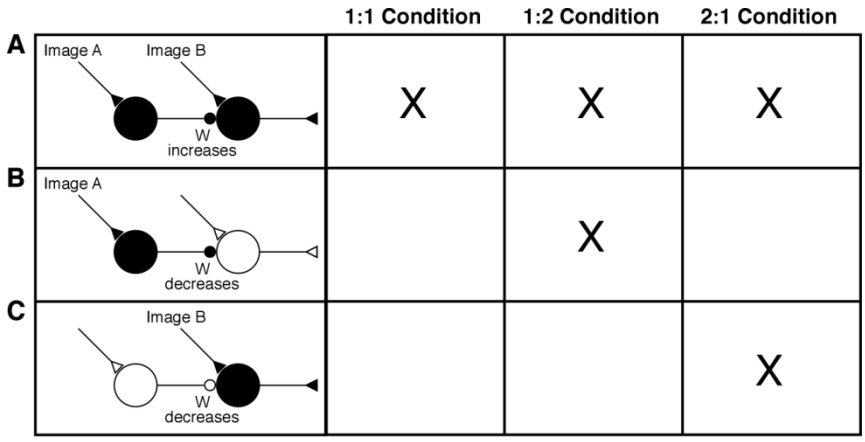


***Fig 2.4. Covariance-based model of prediction suppression***. ***A.*** *Successive activation of neurons responsive to images A and B induces an increase in the weight (W) of a synaptic link allowing A-neurons to suppress B-neurons. This occurs under all three training conditions.* ***B.*** *Activation of A-neurons in the absence of B-neuron activation induces synaptic weakening. This occurs under the 1:2 condition.* ***C.*** *Activation of B-neurons in the absence of A-neuron activation induces synaptic weakening. This occurs under the 2:1 condition.*

al., 2001) but insertion of trials in which presynaptic activity occurs without postsynaptic activity (Fig. 2.4B) has been reported not to do so (Buonomano, 1996).

Explanations of prediction suppression based on the ability of the leading image to predict the trailing image and on a covariance-based learning rule are not necessarily incompatible. It may be that TE relies, in learning to suppress responses to predicted images, on a mechanism that generally is sensitive to the predictability of the trailing image but that employs computations not perfectly suited to doing so. It is useful in considering this point to note that there are parallel quirks in the process by which, in Pavlovian conditioning, the association between the conditioned stimulus (CS) and the unconditioned stimulus (US) is acquired. The strength of the CS-US association does not depend simply on the joint probability p(CS,US) as would be expected from isolated operation of the rule depicted in Fig. 2.4A. It also depends on the conditional probability p(US|CS) as expected from operation of the rule depicted in Fig. 2.4B. This is evident in the induction of an acquisition deficit by partial reinforcement (Miguez et al., 2012). Finally, and critically, it also depends on the conditional probability p(CS|US) as expected from operation of the rule depicted in Fig. 2.4C. This is evident in the induction of an acquisition deficit by contingency degradation (Rescorla, 1968; Bermudez, 2010).

We note, in closing, one unresolved issue of interpretation. In manipulating the conditional probabilities p(A|B) and p(B|A) we necessarily altered the individual probabilities p(A) and p(B). In the 1:1 condition P(A) and p(B) were both 0.1 in the sense that each image appeared on one tenth of training trials. In the 1:2 condition, p(A) and p(B) were 0.2 and 0.1 respectively. In the 2:1 condition, p(A) and p(B) were 0.1 and 0.2 respectively. There is no obvious principled basis for supposing that the variations in individual probability would have given rise to the observed pattern of results. We cannot, however, absolutely rule out the possibility that the attenuation of

prediction suppression in the 1:2 and 2:1 conditions was driven by the increases in p(A) and p(B) respectively.

## REFERENCES


Aslin RN, Saffran JR, Newport EL (1998) Computation of Conditional Probability Statistics by 8-Month-Old Infants. Psychological Science 9:321-324.

Bar M (2009) Predictions: a universal principle in the operation of the human brain. Philosophical Transactions of the Royal Society B: Biological Sciences 364:1181-1182.

Bauer EP, LeDoux JE, Nader K (2001) Fear conditioning and LTP in the lateral amygdala are sensitive to the same stimulus contingencies. Nature Neuroscience 4:687-688.

Bermudez MASW (2010) Responses of Amygdala Neurons to Positive Reward-Predicting Stimuli Depend on Background Reward (Contingency) Rather Than Stimulus-Reward Pairing (Contiguity). Journal of Neurophysiology 103:1158-1170.

Bulf H, Johnson SP, Valenza E (2011) Visual statistical learning in the newborn infant. Cognition 121:127-132.

Buonomano (1996) Associative synaptic plasticity in hippocampal CA1 neurons is not sensitive to unpaired presynaptic activity. Journal of Neurophysiology 76:631-636.

Christofi (1993) The postsynaptic induction of nonassociative long-term depression of excitatory synaptic transmission in rat hippocampal slices. Journal of Neurophysiology 69:219-229.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences 10:294-300.

den Ouden HEM, Kok P, de Lange FP (2012) How Prediction Errors Shape Perception, Attention, and Motivation. Frontiers in Psychology 3.

Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition 28:458-467.

Fiser Jz, Aslin RN (2001) Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. Psychological Science 12:499-504.

Foley NC, Jangraw DC, Peck C, Gottlieb J (2014) Novelty Enhances Visual Salience Independently of Reward in the Parietal Lobe. The Journal of Neuroscience 34:7947-7957.

Friston K (2005) A Theory of Cortical Responses. Philosophical Transactions: Biological Sciences 360:815-836.

Gavornik JP, Bear MF (2014) Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat Neurosci advance online publication.

Hawkins J, Blakeslee S (2005) On Intelligence. New York: Henry Holt and Company.

Howard JH, Howard DV, Dennis NA, Kelly AJ (2008) Implicit learning of predictive relationships in three-element visual sequences by young and old adults. Journal of Experimental Psychology: Learning, Memory, and Cognition 34:1139-1157.

Kamin LJ (1969) Predictability, surprise, attention, and conditioning. In: Punishment and aversive behavior (Campbell BA, Church RM, eds), pp 279-296. New York: Appleton-Century-Crofts.

Kerr DS, Abraham WC (1993) Comparison of associative and non-associative conditioning procedures in the induction of LTD in CA1 of the hippocampus. Synapse 14:305-313.

Kim R, Seitz A, Feenstra H, Shams L (2009) Testing assumptions of statistical learning: Is it long-term and implicit? Neuroscience Letters 461:145-149.

Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition 83:B35-B42.

Meyer M, Baldwin D (2011) Statistical learning of action: The role of conditional probability. Learning & Behavior 39:383-398.

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proceedings of the National Academy of Sciences 108:19401-19406.

Meyer T, Ramachandran S, Olson CR (2014) Statistical Learning of Serial Visual Transitions by Neurons in Monkey Inferotemporal Cortex. The Journal of Neuroscience 34:9332-9337.

Miguez G, Witnauer JE, Miller RR (2012) The role of contextual associations in producing the partial reinforcement acquisition deficit. Journal of Experimental Psychology : Animal Behavior Processes 38:40-51.

Paulsen O, Li YG, Hvalby Ø, Andersen P, Bliss TVP (1993) Failure to Induce Long-term Depression by an Anti-Correlation Procedure in Area CA1 of the Rat Hippocampal Slice. European Journal of Neuroscience 5:1241-1246.

Pearce JM, Hall G (1980) A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532-552.

Pelucchi B, Hay JF, Saffran JR (2009) Statistical Learning in a Natural Language by 8-Month-Old Infants. Child Development 80:674-685.

Pockett S, Brookes NH, Bindman LJ (1990) Long-term depression at synapses in slices of rat hippocampus can be induced by bursts of postsynaptic activity. Experimental Brain Research 80:196-200.

Rescorla RA (1968) Probability of shock in the presence and absence of CS in fear conditioning. J Comp & Physiol Psychol 66:1-5.

Romberg A, Saffran J (2010) Statistical learning and language acquisition. Wiley Interdiscip Rev Cogn Sci 1:906-914.

Saffran JR, Aslin RN, Newport EL (1996) Statistical Learning by 8-Month-Old Infants. Science 274:1926-1928.

Schultz W, Dickinson A (2000) Neuronal Coding of Prediction Errors. Annual Review of Neuroscience 23:473-500.

Sejnowski TJ (1977) Storing covariance with nonlinearly interacting neurons. Journal of Mathematical Biology 4:303-321.

Sejnowski TJ, Chattarji S, Stanton P (1989) Induction of synaptic plasticity by Hebbian covariance in the hippocampus. In: The Computing Neuron (Durbin R, Miall C, Mitchison G, eds), pp 105-124. Wokingham, England: Addison-Wesley.

Stanton PK, Sejnowski TJ (1989) Associative long-term depression in the hippocampus induced by hebbian covariance. Nature 339:215-218.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends in Cognitive Sciences 13:403-409.

Turk-Browne NB, Junge JA, Scholl BJ (2005) The Automaticity of Visual Statistical Learning. Journal of Experimental Psychology: General 134:552-564.

Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA (2008) Multidimensional visual statistical learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 34:399-407.

Wacongne C, Changeux J-P, Dehaene S (2012) A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. The Journal of Neuroscience 32:3665-3678.

# CHAPTER III

# PREDICTION SUPPRESSION, NOT SURPRISE ENHANCEMENT IN MONKEY INFEROTEMPORAL CORTEX.

## INTRODUCTION

We have the extraordinary ability to learn regularities and patterns in the sensory environment, and make predictions about future events on the basis of this knowledge. This form of learning is called statistical learning, and has been demonstrated extensively in human adults, infants and non-human primates (Aslin & Newport, 2012; Fiser & Aslin, 2001; Hauser et al., 2001; Perruchet & Pacton, 2006; Saffran et al., 1996, 1999; Jenny R. Saffran, 2003a). Indeed, making contextual predictions on the basis of long-term learning of statistical regularities has been proposed to be fundamental function of neocortex (Friston, 2005; Hawkins, 2004). Active and ongoing generation of predictions may be fundamental to perception. Predictive coding models suggest that feed-forward input to the visual system is merely one side of the coin; contextual predictions generated in higher areas are fed back through the hierarchical visual pathway and perception depends on comparing the bottom-up visual input with the top-down predictions (Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999). These models suggests that 'prediction errors', i.e., the difference between the visual input (what is seen) and the top-down signal (what is predicted) are computed at every stage in the visual hierarchy.

Prediction errors are observed ubiquitously in the brain. They have been implicated in reward learning, motivational control, decision making (Hollerman & Schultz, 1998; Matsumoto & Hikosaka, 2007; Rescorla & Wagner, 1972; Roesch et al., 2012; Romo & Schultz, 1990;

Schultz & Dickinson, 2000) as well as sensory perception (den Ouden et al., 2012; Summerfield et al., 2008; Todorovic et al., 2011; Wacongne et al., 2012).

One recent example is the prediction effect in monkey inferotemporal cortex(Meyer & Olson, 2011; Meyer, Ramachandran, et al., 2014). Upon repeated exposure over weeks to pairs of visual images in fixed sequence, such that the leading image in the sequence predicts the occurrence of the trailing image, neurons in the monkey inferotemporal cortex show suppressed responses for images when they occur in a predicted context and enhanced responses when the same images occur in an unpredicted context. This could be considered as an instantiation of a perceptual prediction error.

Prediction errors in domains such as reward learning are typically signed. For example, dopamine neurons in macaque ventral tegmental area show enhanced responses the reward was better than what was expected, but suppressed responses when the reward was worse than what was expected (den Ouden et al., 2012; Klein-Flügge, Hunt, Bach, Dolan, & Behrens, 2011; Schultz & Dickinson, 2000). There is no change in the responses of these neurons in the baseline condition when the reward matches expectation. Similar phenomena are observed in the lateral habunela that signals punishment prediction errors, but with a change in sign – neurons are enhanced when the reward was worse than what was expected, suppressed when the reward was better than expected, compared to the baseline where reward matches expectation (Matsumoto & Hikosaka, 2009). Thus these prediction errors are typically deviations from a baseline commensurate with the predictability and expected value of a stimulus.

However, perceptual prediction errors such as the prediction effect should carry no information about the ‘value’ of a stimulus. A stimulus is either predicted or surprising without

any intrinsic value – there is nothing 'better' or 'worse' than the predicted outcome. What we observe are suppressed neuronal responses to predicted stimuli and enhanced responses to surprising stimuli. Thus the prediction error could have been generated in two ways. One possibility is a deviation from baseline (i.e., suppression) when a predicted event occurs and no deviation from baseline when a surprising event takes place. The other possibility is a deviation from baseline (i.e., enhancement) when an unpredicted, surprising event occurs and no deviation from baseline when a predicted event takes place. It could also be the case that what happens is a combination of both – suppression to signal prediction and enhancement to signal surprise.

The question, then, is whether the prediction effect observed in IT is due to suppressed responses to predicted stimuli or due to enhanced responses to surprising, unpredicted stimuli. This question is intractable using the design used in our earlier studies (Meyer & Olson, 2011; Meyer, Ramachandran and Olson., 2014), because there is no baseline with which we would be able to compare the predicted and unpredicted conditions. Thus it is imperative to define a baseline in order to answer this question.

In this study, we sought to answer this question by training monkeys on a prediction-neutral condition where no image in a leading position could absolutely predict the subsequent occurrence of any other image in the trailing position. The responses to these trailing images, which are absolutely prediction-neutral, would act as an appropriate baseline with which we could then compare the predicted and unpredicted responses in order to infer whether the effect is due to prediction suppression or surprise enhancement. We discovered that the observed effect is most due to prediction suppression.

## MATERIALS AND METHODS

### Subjects

Two adult rhesus macaque monkeys (*Macaca mulatta*) were used for the study: monkey 1 (adult male; laboratory designation Tu) and monkey 2 (adult male; laboratory designation Ec). All procedures used during surgery and experiments were in accordance with the guidelines set by the United States Public Health Service Guide for the Care and Use of Laboratory Animals and were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee.

### Training

Each monkey received repeated exposure to pairs of images presented in fixed sequence. The succession of events in each trial was: fixation spot (300 ms), leading image at screen center (503 ms), an 18 ms delay, trailing image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery. The monkeys had to maintain fixation within a restricted window for the duration of the trial to qualify for a water reward. A trial was aborted without reward if the monkey failed to maintain fixation within a 4° x 4° central window.

On each training day, the monkey completed one or more runs. Each run consisted of two sequential blocks: a prediction block and a prediction-neutral block. Either block could be presented first in a run. The prediction block consisted of 12 images organized into 6 unique pairs: upon the display of a particular leading image, the monkey would encounter one and only one particular trailing image (Fig. 3.1B). These six pairs were presented six times, interleaved

randomly, subject to the constraints (a) that within each set of six successfully completed trials each condition had to be imposed once and (b) the same condition could not be imposed on two successive trials. The prediction-neutral block consisted of 12 other images - six leading images and six trailing images - organized into 36 pairs. These 36 pairs were presented once each randomly interleaved. Thus a training run consisted of two blocks - prediction and prediction-neutral - with 36 trials each. The number of runs completed on a day ranged from 2-9 in monkey 1 and from 1-4 in monkey 2. Monkey 1 viewed each of the sequences in the prediction paradigm 600 times and each of the sequences in the prediction-neutral paradigm 100 times (600/6) during 100 runs extending over 34 days. Monkey 2 viewed each of the sequences in the prediction paradigm 630 times and each of the sequences in the prediction-neutral paradigm 105 times (630/6) during 105 runs extending over 59 days.

**Testing sequences**

During neuronal data collection, the monkeys completed trials identical to training trials with regard to the timing of events (Fig. 3.1A). In particular, a leading image and a trailing image appeared in immediate succession on each trial. The status of the images as leading or trailing was the same as during training. However, any leading image could be followed by any trailing image. For the prediction paradigm, this meant that 30 pairs not familiarized during training were presented in addition to the six trained pairs. To prevent erosion of training, the trained sequences were presented

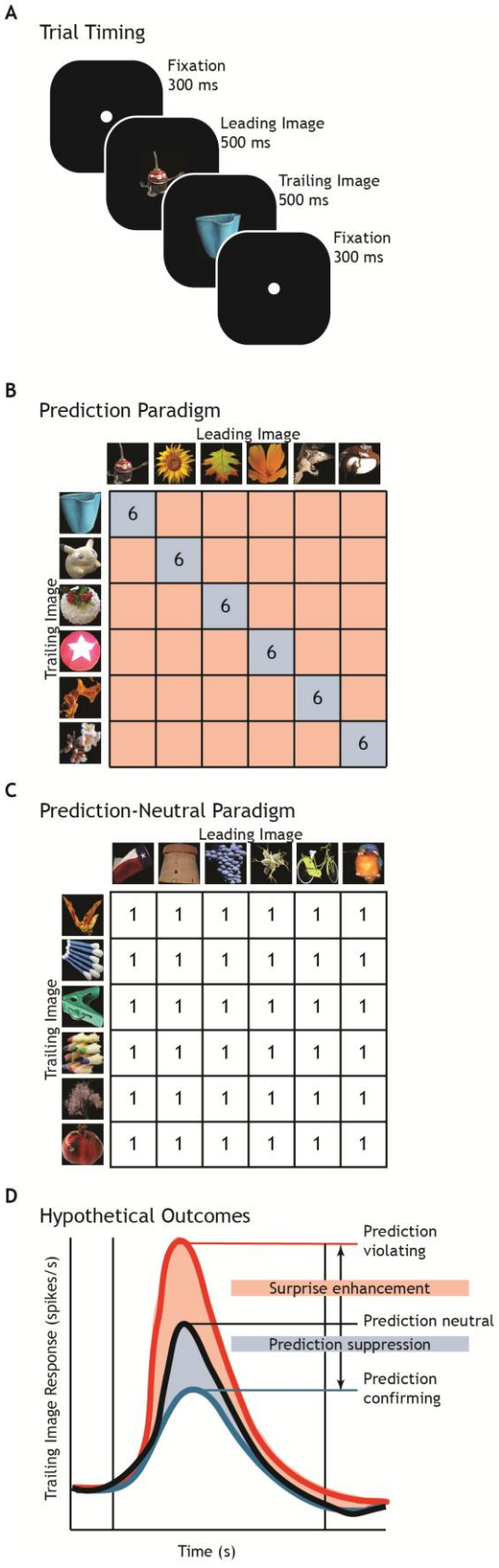


***Fig. 3.1: Suppression vs. enhancement: experimental paradigm*** *(A) Trial timing. Each trial consisted of a leading image followed by a trailing image (B) Prediction paradigm. 12 images were organized into six pairs where each leading image was followed by a, and thus predicted, unique trailing image (filled blue squares, diagonal). Other untrained pairs, where trailing images violated prediction, were shown only during testing (filled red squares, off diagonal) (C) Prediction-neutral paradigm, 12 images were organized into 36 pairs where each leading image could be followed by any other trailing image. Thus the leading images were prediction-neutral with respect to the trailing image. Numbers in black in (B) and (C) indicate ratios of training exposure. (D) Hypothetical responses to the trailing images. Responses in the prediction-neutral paradigm can be used as a baseline to distinguish between suppression and enhancement.*

more frequently than the untrained sequences - each of six trained sequences was presented five times (pink squares, Fig. 3.1(B)) while each of the 30 untrained sequence was presented only once (blue squares, Fig. 3.1(B)). For the prediction-neutral paradigm, a set of 36 trials identical to the training regimen were presented. A run thus consisted of 96 trials encompassing 30 'prediction confirming' sequences (six trained sequences five times each), 30 'prediction violating' sequences (30 untrained sequences once each) and 36 prediction-neutral sequences (36 sequences once each). The conditions were imposed in random order with replacement on error.

**Images**

All stimuli were digitized images of background-free objects. When presented at fixation at fixation on an LCD monitor 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. The stimulus set consisted of 24 unique digitized images of which 12 were used in the prediction paradigm and the other 12 were used in the prediction-neutral paradigm. Within each training paradigm, there were six leading and six trailing images. The 12 images used in the prediction paradigm in Monkey 1 were used in the prediction-neutral paradigm in Monkey 2, thus controlling for the training status of an image between monkeys.

**Recording:**

The electrode was introduced through a vertical guide tube into left (monkey 1) or right (monkey 2) ITC. Recording sites, identified by extrapolation from MRI-visible fiducial markers within the chamber, were within a region of area TE occupying the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the

interaural plane by 16-19 mm in monkey 1 and 13-16 mm in monkey 2. Prior to recording, monkey 1's chamber was tilted posteriorly 6 degrees to allow better access to IT. The recording site was located at the ventral bank of the superior temporal sulcus and inferior temporal gyrus, lateral to the rhinal sulcus.

**Database**

We recorded from 96 sites (70 and 26 in monkeys 1 and 2). Low-pass-filtered traces from these sites formed the LFP database. Neurons characterized during a complete test run numbered 227 (184 from monkey 1 and 43 from monkey 2). We classified a neuron as visually responsive if, for either the leading or the trailing image, the mean firing rate in a window 50-300ms following image onset exceeded the mean firing rate in a 100ms baseline window centered on image onset (one-tailed t-test, $\alpha = 0.05$). The neuronal database consisted of 193 neurons (155 in monkey 1 and 38 in monkey 2) meeting this criterion.

**RESULTS**

The experiment began with a training period extending over many weeks where the monkeys were exposed to training sequences in each of two training paradigms: the 'prediction' paradigm and the 'prediction-neutral' paradigm (Fig. 3.1A). The prediction paradigm consisted of six image sequences constructed from six leading and six trailing images, subject to the constraint that each leading image was followed by a unique trailing image. Each of these sequences were presented six times in a single training run (Blue squares across diagonal, Fig. 3.1B). The prediction-neutral paradigm consisted of 36 image sequences, constructed from six other leading and six other trailing images, where each leading image could be followed by each trailing image. Each of these sequences were presented once in a single trailing run (Each black-lined square, Fig.

3.1C). In a single run of paradigm, a monkey saw any of the leading images six times. However the identity of the trailing image that followed it was different on each trial: it could be one of six different trailing images. The training regimen did not predict that the occurrence any particular trailing image was more likely than another. This implied that the identity of the leading image was prediction-neutral with respect to the identity of the trailing image. Hence these constituted the 'prediction-neutral' condition. This was in contrast to the prediction paradigm, where each leading image uniquely predicted a particular trailing image. A leading image was encountered six times like in the prediction-neutral paradigm, but unlike it, every single time the leading image was followed by the same trailing image. Thus the appearance of the trailing image confirmed the prediction of the leading image. Hence, these trials constituted the 'prediction-confirming' condition.

After completion of training we measured the responses of neurons in anterior IT to images in the prediction-confirming and prediction-neutral conditions, in addition to sequences in the prediction paradigm which were never displayed to the monkey during training (red squares off diagonal, Fig. 3.1B) where the identity of the trailing images violated the predictions of the leading images on the basis of training history (prediction-violating condition). Each run consisted of each of 96 trials interleaved randomly: each of six sequences in the prediction-confirming condition repeated five times each (blue squares, Fig. 3.1B, 30 trials), each of the sequences in the prediction-violating condition presented once (red squares, Fig. 3.1B, 30 trials) and each of the sequences in the prediction-neutral condition presented once (Fig. 3.1C, 36 trials). We collected full data sets from 193 visually responsive IT neurons (155 in monkey 1 and 38 in monkey 2). We first sought to confirm that the prediction effect (comparing prediction-confirming and prediction-violating conditions) was present in this population. The goal of the

experiment was to subsequently compare the responses in these conditions to the responses in the prediction-neutral baseline condition (Fig. 3.1D). Depending the pattern of responses, we could then ask whether the effect observed was due to suppressed responses to the prediction-confirming images (prediction suppression) or enhanced responses to the prediction-violating images (surprise enhancement) with respect to the prediction-neutral responses.

To determine whether the prediction effect was present in the population, we compared population histograms representing the mean firing rate in the prediction-confirming trials

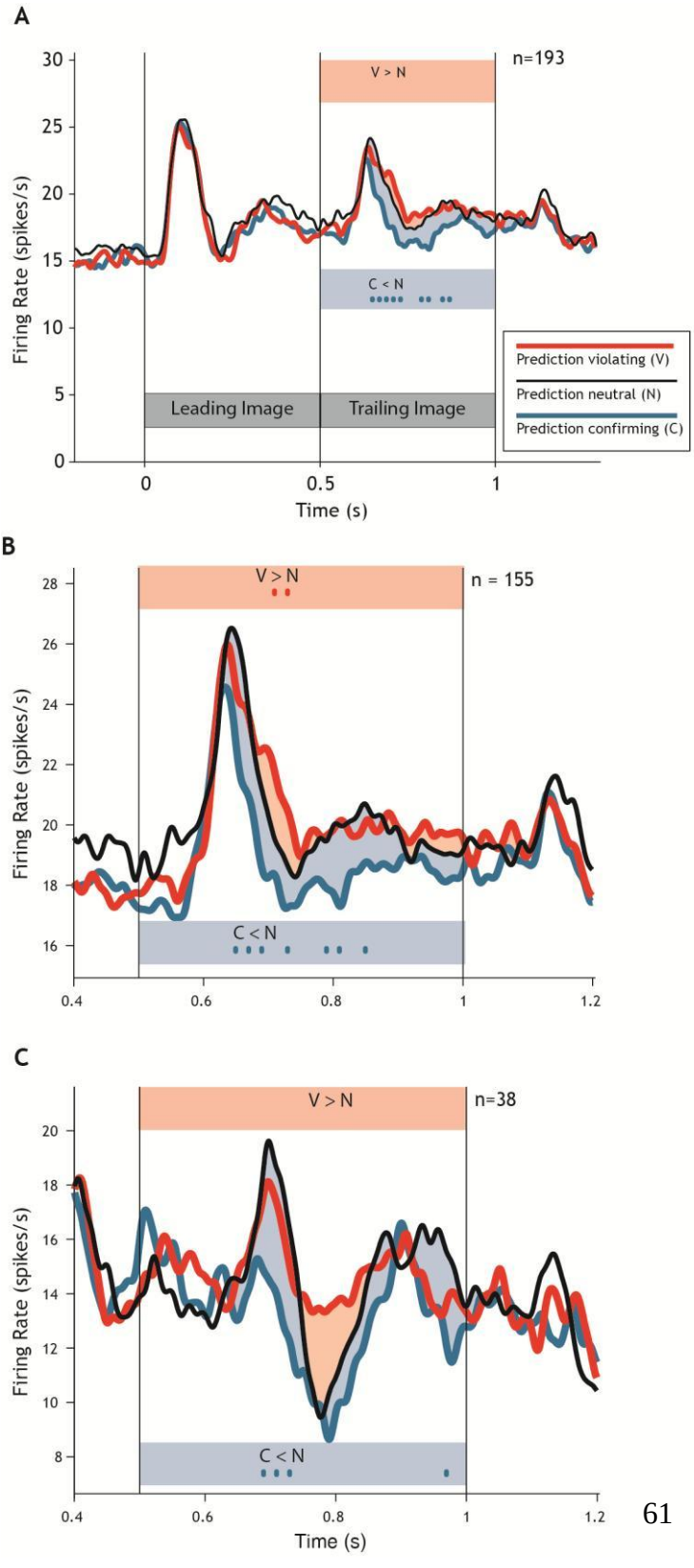


***Fig. 3.2: Neuronal population activity with prediction-neutral baseline*** *(A) Mean firing rate of 193 neurons during testing with prediction-confirming (thick blue curve, corresponding to filled blue squares in Fig. 3.1B), prediction violating (thick red curve, corresponding to filled red squares in Fig. 3.1B) and prediction-neutral (black curve, corresponding to all conditions in Fig. 3.1C) sequences. Filled symbols in blue and red indicate 10 ms bins during which activity elicited by prediction-neutral sequences significantly differed from activity elicited by prediction-confirming and prediction-violating sequences respectively (two-tailed paired t-test, $\alpha$ = 0.05, n = 193)..(B) and (C) show trailing image responses alone with same conventions for Monkey 1 and Monkey 2 individually.*

to the mean firing rate in the prediction-violating trials. The prediction effect was present in this population (Fig. 3.2A, response to trailing images, red curve over the blue curve). (This population of neurons, and identity of images used were different from the two datasets reported in Meyer and Olson, 2011 and Meyer et al, 2014), We compared the firing rates across neurons in the prediction-confirming and prediction-violating conditions, in the epoch 100-500 ms after trailing-image onset.. The firing rates in the prediction-violating condition were significantly greater than in the prediction-confirming condition ($p = 1.9$ E-11, two-tailed paired t-test, $n = 193$). This effect was present with strong significance in each monkey.

To determine whether the prediction effect was due to suppression or enhancement, we compared the population histogram representing the mean firing rate during the prediction neutral trials to the mean firing rate in the prediction confirming and violating trials. Comparing the firing rates across neurons in prediction-confirming condition with those in the prediction-neutral trials in the epoch 100-500ms after trailing image onset revealed that the firing rates in the prediction-confirming trials were significantly lower than in the prediction-neutral trials ($p = 0.0138$, two-tailed paired t-test, $n = 193$). This effect was present in each monkey, significant at the alpha = 0.05 level in Monkey 1 and approaching significance in Monkey 2. This strongly suggests that the prediction effect is largely due to suppressed responses to the predicted images.

In order to examine the contribution (if any) of surprise enhancement to the prediction effect, we next compared the firing rates across neurons in prediction-violating condition with those in the prediction-neutral trials in the epoch 100-500ms after trailing image onset. In the population, we found no differences between these conditions ($p = 0.6313$, two-tailed paired t-test, $n = 193$). This was true for each monkey. This suggests that the prediction effect is very likely not due to enhancement of responses to surprising, unpredicted images.

To determine whether comparable effects were present at the level of the local field potential (LFP), we analyzed data collected from the 96 sites at which we had monitored neuronal activity (70 and 26 sites in monkeys 1 and 2 respectively). Prediction suppression is manifest at the level of the LFP as a reduction in the amplitude of the excursion from maximal negativity at around 200 ms to maximal positivity at around 300 ms (Meyer and Olson, 2011; Meyer et al., 2014). This effect was highly significant in the population (Fig. 3.3A, p = 0.0015, two-tailed paired t-test, n = 96). Subsequently, we compared the amplitude of the N200-P300 excursion of (a) prediction-confirming and prediction neutral conditions and (b) prediction-violating and prediction neutral conditions.

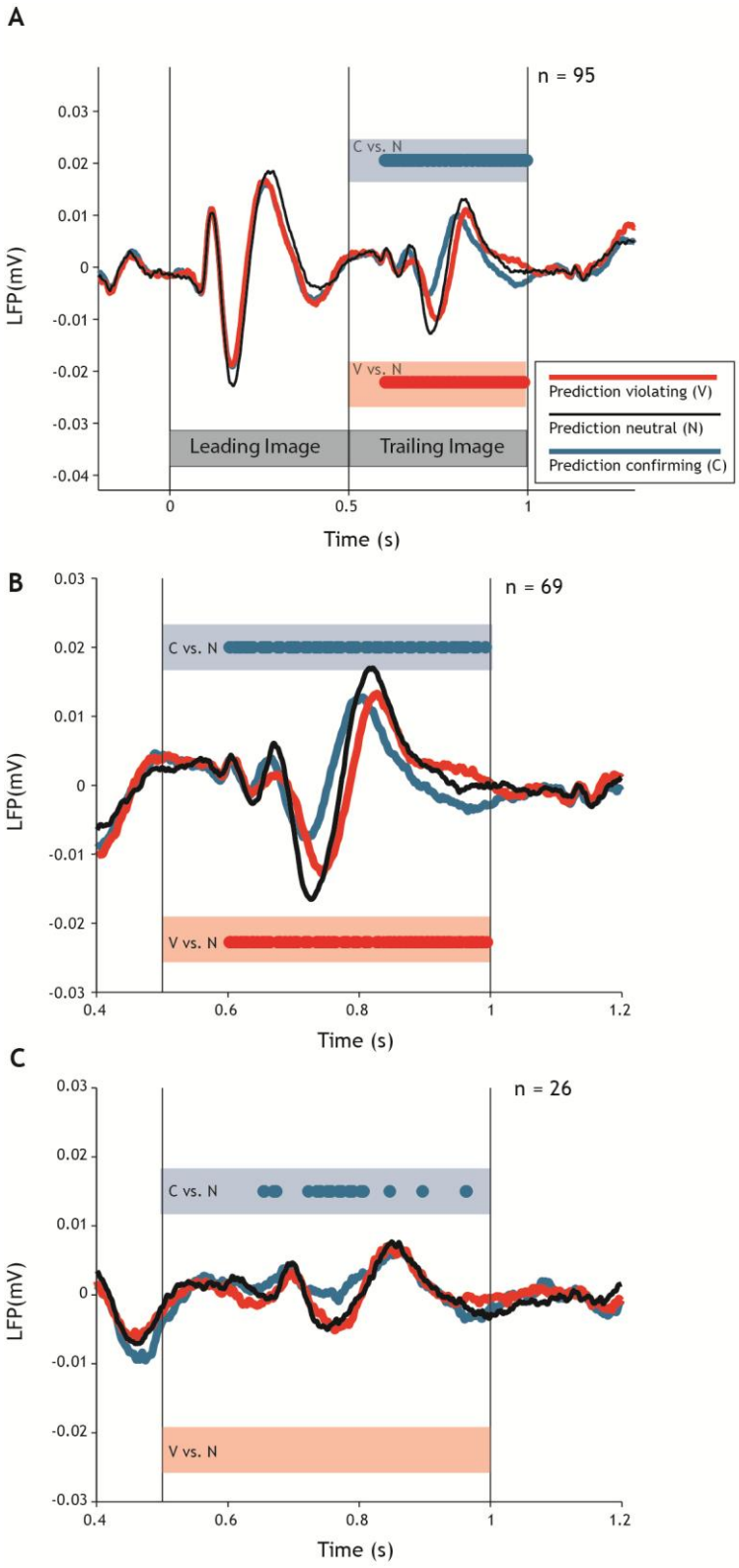


***Fig. 3.3: LFP activity with prediction-neutral baseline*** *Mean LFP activity in 95 sites. Conventions same as in Fig. 3.2*

In the first comparison, the amplitude across sites was greater in the prediction-neutral condition than in the prediction-confirming condition ($p = 9.4 E-7$, two-tailed paired t-test, $n = 96$). This is in agreement with the result from single unit analysis, suggesting a contribution of suppression to the prediction effect. In the second comparison, the amplitude across sites was greater in the prediction-neutral condition than in the prediction-violating condition ($p = 5.7 E-4$, two-tailed paired t-test, $n = 96$). Nevertheless, these observations taken together fit more consistently with suppression than enhancement.

## DISCUSSION

Neurons in inferotemporal cortex show reduced responses to predicted images and enhanced responses to unpredicted images. The aim of this experiment was to examine whether the effect was primarily due to suppression of predicted stimuli or due to enhanced responses to surprising stimuli. Our results at the population level establish that the phenomenon is mostly due to suppressed responses to predicted images.

One drawback in the design we use is that comparisons necessarily have to be made between different two different image sets; there is no other way to define a neutral baseline. IT neurons are generally highly selective (Gross, Bender, & Gerstein, 1979), and in order to minimize biases in the data due to differences in selectivity, we recorded from a large number of neurons and sites. For the same reason, it is impossible to distinguish between suppression and enhancement mechanisms at the level of single neurons as any effects observed could either be due to the prediction context (confirming, violating or neutral) or image identity. To illustrate this, we

present a scatter plot normalized firing rates for the trailing image (100-500ms after image onset) of all 193 neurons in a ternary plot (Fig. 3.4). A point at each vertex of the triangle would represent 100% responsiveness for one of the three conditions (prediction-confirming (P), prediction-violating (U) and prediction-neutral (N)) and zero responsiveness in the other two conditions. Each side of the triangle represents a continuum of responses balanced between the vertices, with the mid-point of that side representing equal responses in each condition. This way, the triangle can be divided into 6

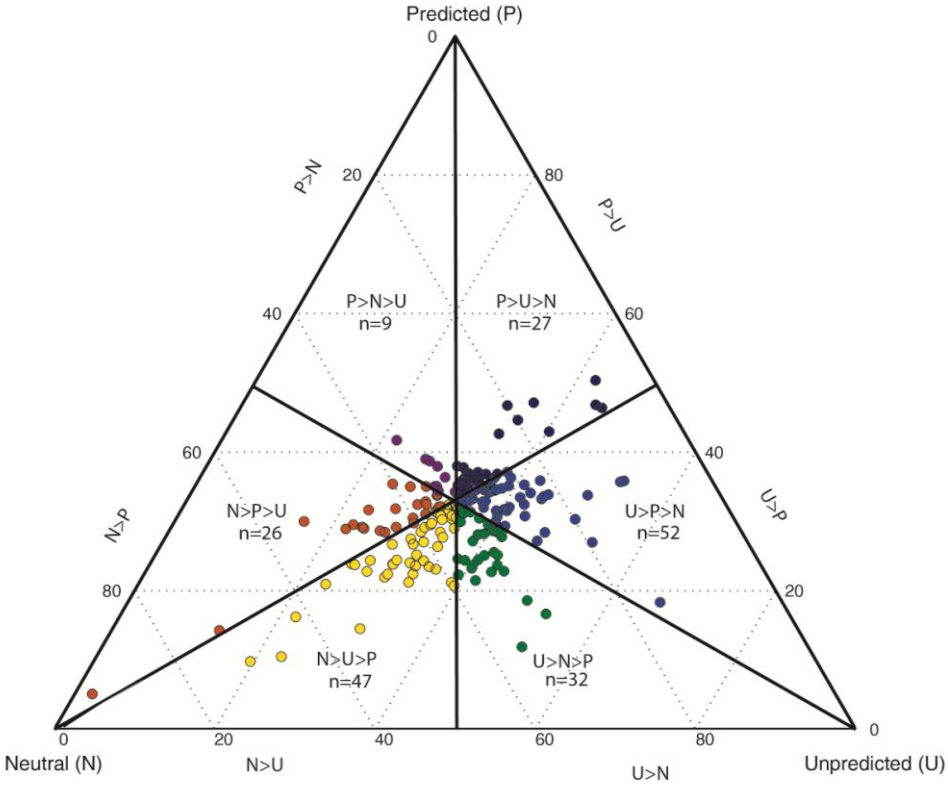


***Fig. 3.4: Ternary plot: scatter of activity of single units*** *Each point on the triangle's surface is plotted at the intersection of the normalized firing rate of a single unit for trailing images (100-500ms after onset) in each of the three conditions: predicted (P), unpredicted (U) and neutral (N). Each side of the triangle represents trends with respect to responses in the three conditions, on the basis of which the triangle is divided into six zones. For example, one zone with green filled circle contains only neurons where unpredicted responses are greater than neutral responses, which are in turn greater than predicted responses (U>N>P). Distribution of neurons in different zones provides a qualitative picture of trends in the data.*

zones depending on the ratios of responses. Most of the neurons are scattered in zones where responses in the unpredicted, prediction-violating condition is greater than in the prediction-confirming condition (U>P, yellow, green and blue filled circles), irrespective of responses to prediction-neutral (N) condition. This shows that most of the neurons indeed show prediction suppression. However, other effects observed in this plot for individual neurons, including neurons where N>P (red, yellow and green filled circles), consistent with suppression, and neurons where U>N (green, blue and indigo filled circles), consistent with enhancement, could either be a real effect, or simply a result of differences in image selectivity. Thus, further analyses at the level of individual neurons would be uninformative.

What does prediction suppression suggest in terms of mechanisms underlying the effect? Firstly, a suppressive mechanism indicates the role of local inhibition in the response. This suggests that silencing inhibition in IT would relieve suppression. Secondly, absence of enhancement suggests that statistical learning in IT is an effect modulated by prediction, and not surprise. Indeed, in a recent study in monkey IT, revealed an absence of surprise responses in an oddball paradigm suggests that mismatch responses are more in tune with stimulus-specific suppression, rather than surprise (Kaliukhovich & Vogels, 2014).Thus it may be mechanistically quite different from other physiological phenomena with sensitivity to statistical regularities observed elsewhere where graded 'surprise' signals are observed for surprising stimuli (Mars et al., 2008; Schmolesky et al., 2013; Tiitinen, May, Reinikainen, & Näätänen, 1994; Wacongne et al., 2012).

Suppressed responses can be induced on the basis of both short-term stimulus history (repetition suppression) (McMahon & Olson, 2007; Miller, Li, & Desimone, 1993;

Miller & Desimone, 1994), and on the basis of long-term training (familiarity effect) (Anderson et al., 2008a; Li, Miller, & Desimone, 1993; Meyer, Walker, et al., 2014; Mruczek & Sheinberg, 2007a; Peissig et al., 2007; Woloszyn & Sheinberg, 2012b). Suppressed responses for predicted images could be brought about by at least three different mechanisms, initially proposed to explain repetition suppression in IT (Grill-Spector, Henson, & Martin, 2006). Suppression could be due to fatigue, where the amplitude of each neuron's response decreases with repetition (Miller & Desimone, 1994), suggesting a global decrease in response across IT. It could be due to sharpened responses in a few neurons with selectivity to the predicted image while silencing everything else, so that it is represented better in the population (Desimone, 1996; Miller et al., 1993). Or it could be due to a decrease in the duration of neural processing for predicted stimuli, such that they are processed faster (Sobotka & Ringo, 1996). It remains to be tested which of these mechanisms can best explain suppression in IT.

## REFERENCES


Anderson, B., Mruczek, R. E. B., Kawasaki, K., & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(11), 2540–52. doi:10.1093/cercor/bhn015

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170–176. doi:10.1177/0963721412436806

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*(December), 548. doi:10.3389/fpsyg.2012.00548

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13494–9. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=33636&tool=pmcentrez&rendertype=abstract

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11760138

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–36. doi:10.1098/rstb.2005.1622

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Gross, C. G., Bender, D. B., & Gerstein, G. L. (1979). Activity of inferior temporal neurons in behaving monkeys. *Neuropsychologia*, *17*(2), 215–29. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/111157

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11124355

Hawkins, J. (2004). *On Intelligence* (Repreint e.). Times Books.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*(4), 304–9. doi:10.1038/1124

Kaliukhovich, D. A., & Vogels, R. (2014). Neurons in Macaque Inferior Temporal Cortex Show No Surprise Response to Deviants in Visual Oddball Sequences. *Journal of Neuroscience*, *34*(38), 12801–12815. doi:10.1523/JNEUROSCI.2154-14.2014

Klein-Flügge, M. C., Hunt, L. T., Bach, D. R., Dolan, R. J., & Behrens, T. E. J. (2011). Dissociable reward and timing signals in human midbrain and ventral striatum. *Neuron*, *72*(4), 654–64. doi:10.1016/j.neuron.2011.08.024

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12868647

Li, L. I. N., & Miller, E. K. (1993). The Representation of Stimulus Familiarity Temporal Cortex in Anterior Inferior, *69*(6).

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *The Journal of Neuroscience : The*

*Official Journal of the Society for Neuroscience*, *28*(47), 12539–45. doi:10.1523/JNEUROSCI.2925-08.2008

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–5. doi:10.1038/nature05860

Matsumoto, M., & Hikosaka, O. (2009). Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience*, *12*(1), 77–84. doi:10.1038/nn.2233

McMahon, D. B. T., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *Journal of Neurophysiology*, *97*(5), 3532–43. doi:10.1152/jn.01042.2006

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Meyer, T., Walker, C., Cho, R. Y., & Olson, C. R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience*, *17*(10), 1388–1394. doi:10.1038/nn.3794

Miller, E. K., & Desimone, R. (1994). Parallel Neuronal mechanisms for Short-Term Memory. *Science*, *263*(5146), 520–522.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *13*(4), 1460–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8463829

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Context familiarity enhances target processing by inferior temporal cortex neurons. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(32), 8533–45. doi:10.1523/JNEUROSCI.2106-07.2007

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251.

Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(6), 1323–34. doi:10.1093/cercor/bhl043

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–8. doi:10.1016/j.tics.2006.03.006

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *The European Journal of Neuroscience*, *35*(7), 1190–200. doi:10.1111/j.1460-9568.2011.07986.x

Romo, R., & Schultz, W. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *Journal of Neurophysiology*, *63*(3), 592–606. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2329363

Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. doi:10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8943209

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Schall, J. D., Leventhal, A. G., … Matthew, T. (2013). Signal Timing Across the Macaque Visual System Signal Timing Across the Macaque Visual System, 3272–3278.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500. doi:10.1146/annurev.neuro.23.1.473

Sobotka, S., & Ringo, J. L. (1996). Mnemonic Responses of Single Units Recorded from Monkey Inferotemporal Cortex , Accessed via Transcommissural Versus Direct Pathways : A Dissociation between Unit Activity and Behavior, *16*(13), 4222–4230.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–6. doi:10.1038/nn.2163

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90–2. doi:10.1038/372090a0

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(25), 9118–23. doi:10.1523/JNEUROSCI.1425-11.2011

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(11), 3665–78. doi:10.1523/JNEUROSCI.5003-11.2012

Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*(1), 193–205. doi:10.1016/j.neuron.2012.01.032

# CHAPTER IV

# STATISTICAL LEARNING IN MONKEY INFEROTEMPORAL CORTEX IS MODALITY-SPECIFIC

## INTRODUCTION

Human infants and adults rapidly and effortlessly learn statistically likely transitions in a stream of sequential stimuli solely through passive exposure, a phenomenon called 'statistical learning'(Perruchet & Pacton, 2006). This has been extensively demonstrated in the auditory domain, where human infants and adults, as well as other mammals are able to learn transitional statistics in sequences of syllables (Aslin et al., 1998; Hauser et al., 2001; Pelucchi et al., 2009; Saffran et al., 1996; Toro & Trobalón, 2005) or tones (Pearce et al., 2010; Saffran et al., 1999; J. Saffran & Griepentrog, 2001). This ability is considered to be domain general: it has also been demonstrated in the visual domain, where human infants and adults learn transitional statistics embedded in static scenes (Baker, Olson, & Behrmann, 2004; Fiser & Aslin, 2001; Fiser & Aslin, 2002b) as well as sequences of visual stimuli (Bulf et al., 2011; Fiser & Aslin, 2002a; Kirkham et al., 2002). These phenomena are considered fundamental to learning both spoken and written language (Dehaene et al., 2005; Dehaene, 2009; Kuhl, 2004).

The neuronal mechanisms underlying these learning mechanisms are just beginning to be understood. In a study aimed at uncovering the neuronal mechanisms underlying visual statistical learning in monkeys, we demonstrated that when monkeys repeatedly view pairs or triplets of digitized images in sequence over weeks, the post-exposure response of neurons in inferotemporal cortex (IT) to an image when it occurs predictably in sequence is suppressed

compared to the response to the same image when it occurs out of sequence. This effect, called prediction suppression, is dependent on probabilistic transitions in the sequence, and therefore can be considered to be a signature of visual statistical learning in monkey inferotemporal cortex (Meyer & Olson, 2011; Meyer, Ramachandran, & Olson, 2014; Ramachandran, Meyer & Olson (under review)). It is possible that a similar phenomenon may be present in higher auditory cortex that would be the basis of auditory statistical learning in monkeys.

In contrast to our laboratory situations where monkeys are passively exposed to streams of only visual stimuli or streams of only auditory stimuli, naturalistic environments typically provide an organism with multisensory, cross-modal input. The real world has regular instances where a stimulus of one sensory modality (eg., the sound of a dog barking) is congruent with, or is immediately followed by a stimulus of a different modality (eg., the visual image of a dog). The association of such cross-modal stimuli (eg., the association of visual speech cues with the auditory portion of speech) may be vital for language acquisition (McGurk & MacDonald, 1976; Mitchel, Christiansen, & Weiss, 2014; Mitchel & Weiss, 2010, 2014; Rosenblum, 2013). Inspired by predictive coding frameworks, it has been suggested that cross-modal predictions, when violated, might generate an error signal which may be used for visual learning (Jacobs & Shams, 2010).

We sought to investigate whether such cross-modal statistical regularities in auditory-visual sequences are encoded at the level of single neurons in monkey inferotemporal cortex. Specifically, just as learning statistical regularities in visual-visual sequences where an image consistently predicts another image induces prediction suppression in IT, does learning auditory-visual sequences where a sound consistently predicts a particular subsequent image also induce prediction suppression in IT?

We answer this question by training monkeys on a paradigm where they are exposed to fixed auditory-visual sequences, with the auditory stimulus and the subsequent visual stimulus each lasting for 500ms. The auditory and visual stimuli are temporally separated, i.e., incongruent, but the auditory stimulus sequentially predicts the visual stimulus. At least two previous studies report responses to just auditory stimuli in IT (Iwai, Aihara, & Hikosaka, 1987; Kaposvári, Csibri, Csete, Tompa, & Sáry, 2011), suggesting that IT could be a multimodal area, although we have not observed any such responses ourselves in a previous experiment. In any case, it is not that we are studying the responses to auditory stimuli in IT. We seek to understand the predictive effect of auditory stimuli on frequently following visual stimuli, which IT strongly and selectively encodes. We reason that even if the auditory information is processed elsewhere in the brain, 500ms would be enough time for information to be processed and directed to IT in order to exert a predictive influence on the visual image. IT has direct projections from the multi-sensory area, the superior temporal polysensory area (STP) (Saleem, Suzuki, Tanaka, & Hashikawa, 2000; Seltzer & Pandya, 1978). It has been proposed that the STP provides IT with contextual information originating in audition (Saleem et al., 2000), making it one possible route through which auditory-visual prediction suppression in IT, if it at all exists, could be mediated. If auditory-visual predictions are encoded in IT, then the phenomenon in IT is not modality specific, but could be a hallmark of abstract, domain-general prediction.

## MATERIALS AND METHODS

### Subjects

Two adult rhesus macaque monkeys (*Macaca mulatta*) were used for the study: monkey 1 (adult male; laboratory designation Tu) and monkey 2 (adult male; laboratory designation Ec).

All procedures used during surgery and experiments were in accordance with the guidelines set by the United States Public Health Service Guide for the Care and Use of Laboratory Animals and were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee.

**Training**

Each monkey received repeated exposure to pairs of stimuli presented in fixed sequence. There were six stimulus sequences in all: three visual-visual sequences where a digitized image was followed by another image and three auditory-visual sequences where a complex sound was followed by an image. The succession of events in each trial was: fixation spot (300 ms), leading stimulus at screen center (503 ms), an 18 ms delay, trailing image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery. During all three auditory-visual

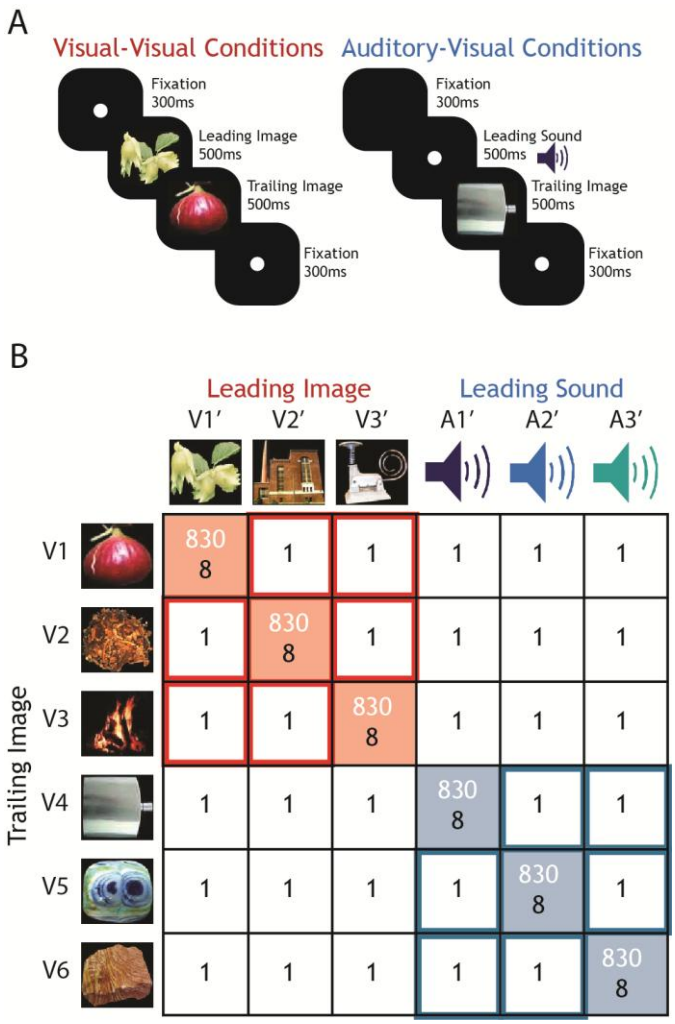


***Fig. 4.1 Testing modality specificity: experimental paradigm*** **(**A) Monkeys were trained on auditory-visual and visual-visual sequences. Each stimulus was presented for 500ms. (B) Each square represents a sequence formed by a leading image followed by a trailing image. Trained visual-visual (filled red squares) and auditory-visual (filled blue squares) sequences are shown on the diagonal. These were presented over 800 times to the monkeys. Every other sequence (off diagonal, irrespective of color) was an untrained sequence, presented only during testing. Numbers in black indicate number of times each sequence was presented during testing. Only using the untrained sequences with colored outlines - visual-visual (red outline – 6 trials) and auditory-visual (blue outline – 6 trials) were used as untrained controls in each condition in the analysis.

sequences, the onset of the complex sound was accompanied by a slight increase in the size of the fixation spot in order to couple the sound with a visual event in the trial. The monkeys had to maintain fixation within a restricted window for the duration of the trial to qualify for a water reward. A trial was aborted without reward if the monkey failed to maintain fixation within a 4° x 4° central window.

On each training day, the monkey completed one or more runs. Each run consisted of three visual-visual sequence trials and three auditory-visual sequence trials. The six unique sequences were each presented ten times, randomly interleaved, subject to the constraints (a) that within each set of six successfully completed trials each condition had to be imposed once and (b) the same condition could not be imposed on two successive trials. Thus, each run consisted of 60 trials. In each trial, upon the display of a particular leading stimulus (sound or image), the monkey would encounter one and only one particular trailing image (Fig. 4.1B). The number of runs completed on a day ranged from 1-6 in monkey 1 and from 1-4 in monkey 2. Monkey 1 viewed each of the sequences 870 times during 87 runs extending over 27 days. Monkey 2 viewed each of the sequences 830 times during 83 runs extending over 40 days.

**Testing sequences**

During neuronal data collection, the monkeys completed trials identical to training trials with regard to the timing of events (Fig. 4.1A). In particular, a leading stimulus (sound or image) and a trailing image appeared in immediate succession on each trial. The status of the stimuli as leading or trailing was the same as during training. However, any leading stimulus could be followed by any trailing image, i.e. the 30 pairs not familiarized during training were presented in addition to the six trained pairs. These untrained sequences consisted of both visual-visual and

auditory-visual sequences not experienced during training. To prevent erosion of training, the trained sequences were presented more frequently than the untrained sequences - each of six trained sequences was presented eight times (red and blue squares on the diagonal, Fig. 4.1(B)) while each of the 30 untrained sequence was presented only once (every other square off diagonal, irrespective of square color, Fig. 4.1(B)). A run thus consisted of 78 trials encompassing 24 trained visual-visual sequences (three trained sequences eight times each), 24 trained auditory-visual sequences (three trained sequences eight times each) and 30 untrained sequences (15 untrained auditory-visual and 15 untrained visual-visual sequences once each). The conditions were imposed in random order with replacement on error.

**Stimuli**

All visual stimuli (nine in all) were digitized images of background-free objects. When presented at fixation at fixation on an LCD monitor 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. All auditory stimuli (three in all) were 500ms clips of complex sounds with a sudden onset (eg. glass shattering) sourced from various sound databases on the internet and cut to size using Audacity software. These were played over speakers placed on either side of the monitor. Monkey 1 was trained on visual-visual sequences V1' – V1, V2' – V2 and V3' – V3 and auditory-visual sequences A1' - V4, A2' - V5 and A3' - V6 (where stimulus' was the leading image). This is shown in Fig. 4.1B. Monkey 2 was trained on visual-visual sequences V2' – V1, V3' – V2 and V1' – V3 and auditory-visual sequences A2' - V4, A3' - V5 and A1' - V6, thus controlling for the stimulus pairing between monkeys.

**Recording:**

The electrode was introduced through a vertical guide tube into left (monkey 1) or right (monkey 2) ITC. Recording sites, identified by extrapolation from MRI-visible fiducial markers within the chamber, were within a region of area TE occupying the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the interaural plane by 16-19 mm in monkey 1 and 13-16 mm in monkey 2. The recording site was located at the ventral bank of the superior temporal sulcus and inferior temporal gyrus, lateral to the rhinal sulcus.

**Database**

We recorded from 47 sites (22 and 25 in monkeys 1 and 2). Low-pass-filtered traces from these sites formed the LFP database. Neurons characterized during a complete test run numbered 82 (40 from monkey 1 and 42 from monkey 2). We classified a neuron as visually responsive if, for either the leading or the trailing image, the mean firing rate in a window 50-300 ms following image onset exceeded the mean firing rate in a 100 ms baseline window centered on image onset (one-tailed t-test, $\alpha = 0.05$). The neuronal database consisted of 65 neurons (30 in monkey 1 and 35 in monkey 2) meeting this criterion.

**RESULTS**

The experiment began with a training period extending over many weeks where the monkeys were exposed to training sequences: three auditory-visual sequences and six visual-visual sequences (Fig. 4.1A). Each sequence was constructed from six leading stimuli (three auditory and six visual) and six trailing stimuli (all six visual images), subject to the constraint that each leading stimulus (sound or image) was followed by a unique trailing image. Each of these

sequences were presented ten times in a single training run (Blue squares across diagonal, Fig. 4.1B).

After completion of training we measured the responses of neurons in anterior TE to both trained and untrained sequences (red squares off diagonal, Fig. 4.1B) where the identity of the trailing images violated the predictions of the leading stimuli on the basis of training history. Each run consisted of each of 78 trials interleaved randomly: each of six trained sequences repeated eight times each (red and blue squares on the diagonal, Fig. 4.1B, 48 trials) and each of the 30 untrained sequences presented once (every other square off diagonal, irrespective of square color, Fig. 4.1B , 30 trials). However, although we collected data for all the untrained sequences, only a few of them were appropriate controls for the auditory-visual and visual-visual conditions respectively. In particular, only those untrained sequences where the trailing image was preceded by an image (six sequences, squares with thick red outline in Fig. 4.1B) were considered appropriate controls in the visual-visual condition. Untrained sequences where the trailing images were preceded by a sound were not included in the comparison. Similarly, only those untrained sequences where the trailing image was preceded by a sound (six sequences, squares with thick blue outline in Fig. 4.1B) were appropriate controls in the auditory-visual condition. Untrained sequences where the trailing images were preceded by an image were not included in the comparison.

We collected full data sets from 65 visually responsive TE neurons (30 in monkey 1 and 35 in monkey 2). We first sought to confirm that the prediction effect, previously observed for visual-visual sequences was present in this population. To determine whether the prediction effect was present in the population, we compared population histograms representing the mean firing rate in the visual-visual predicted condition trials to the mean firing rate in the visual-visual unpredicted condition. The prediction effect was present in this population (Fig. 4.2A, response to trailing images, thick black curve over the thin curve). (This population of neurons, and identity of images used were different from the dataset reported in Meyer and Olson, 2011). We compared the

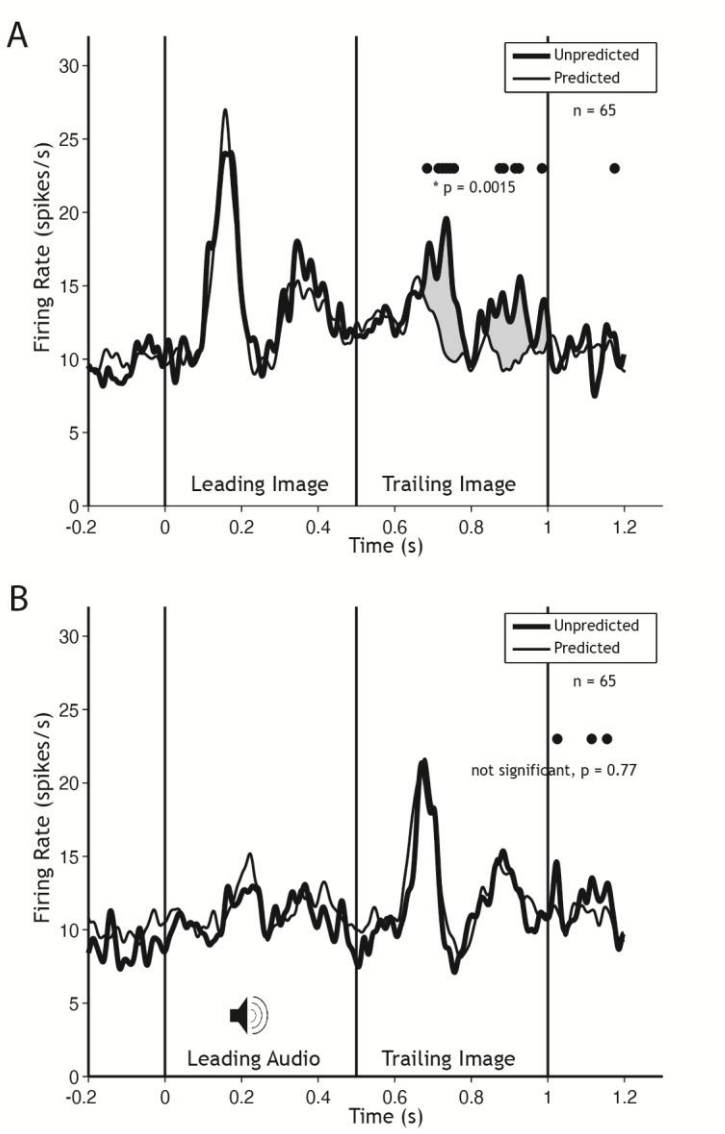


***Fig. 4.2: Neuronal population activity for (A) visual-visual and (B) auditory-visual conditions.*** *(A) Mean firing rate of 65 neurons during testing with sequences containing trailing images used in the visual-visual condition. Thick curve represents activity under trained conditions (filled red cells in Fig. 4.1B). Thin curve represents activity under untrained conditions (outlined red cells in Fig. 4.1B). Filled symbols above the curves indicate 10 ms bins during which activity elicited by untrained sequences significantly differed from activity elicited by trained sequences (two-tailed paired t-test, $\alpha = 0.05$, n = 65).(B) Same comparisons and conventions for the auditory-visual comparison. Thick curve represents activity under trained conditions (filled blue cells in Fig. 4.1B). Thin curve represents activity under untrained conditions (outlined blue cells in Fig. 4.1B).*

firing rates across neurons in the predicted and unpredicted conditions, in the epoch 100-500 ms after trailing-image onset. The firing rates in the unpredicted condition (mean firing rate = 13.73 spikes/s) were significantly greater than the predicted condition (mean firing rate = 11.13 sp/s) ($p = 7.54$ E-5, two-tailed paired t-test, $n = 65$). This effect was present with strong significance in one monkey ($p= 2.89$ E-6) and a trend towards significance in the other monkey ($p = 0.171$).

We asked the same question of the auditory-visual sequences, by comparing the population histograms representing the mean firing rate in the auditory-visual predicted condition trials (i.e, where images were predicted by the preceding sounds) to the mean firing rate in the visual-visual unpredicted condition (i.e, where images were not predicted by the preceding sounds). In the neuronal population, we saw absolutely no prediction effect: responses to the images predicted by a sound (mean firing rate = 13.11 sp/s ) were no different from the response to the same images when they were not predicted by a sound (mean firing rate = 12.39 sp/s) (two-tailed t-test, $p = 0.0688$). In fact, there was a tendency towards the opposite effect: one monkey had a tendency for slightly greater responses to the predicted condition compared to the unpredicted condition. This was not compatible with the standard prediction effect we have observed, suggesting that prediction suppression is specific only for visual stimuli predicted by other visual stimuli.

To determine whether comparable effects were present at the level of the local field potential (LFP), we analyzed data collected from the 47 sites at which we had monitored neuronal activity (22 and 25 sites in monkeys 1 and 2 respectively). The LFP signal in a population of neurons exhibiting prediction suppression typically has two distinct components: a negative excursion

between about 200-300ms where the response to the unpredicted condition is more negative than the response to the predicted condition, and a positive excursion between about 350 – 650ms where the response to the unpredicted condition is more positive than the response to the predicted condition as measured in the population (refer Fig. 4A, (Meyer & Olson, 2011a) and

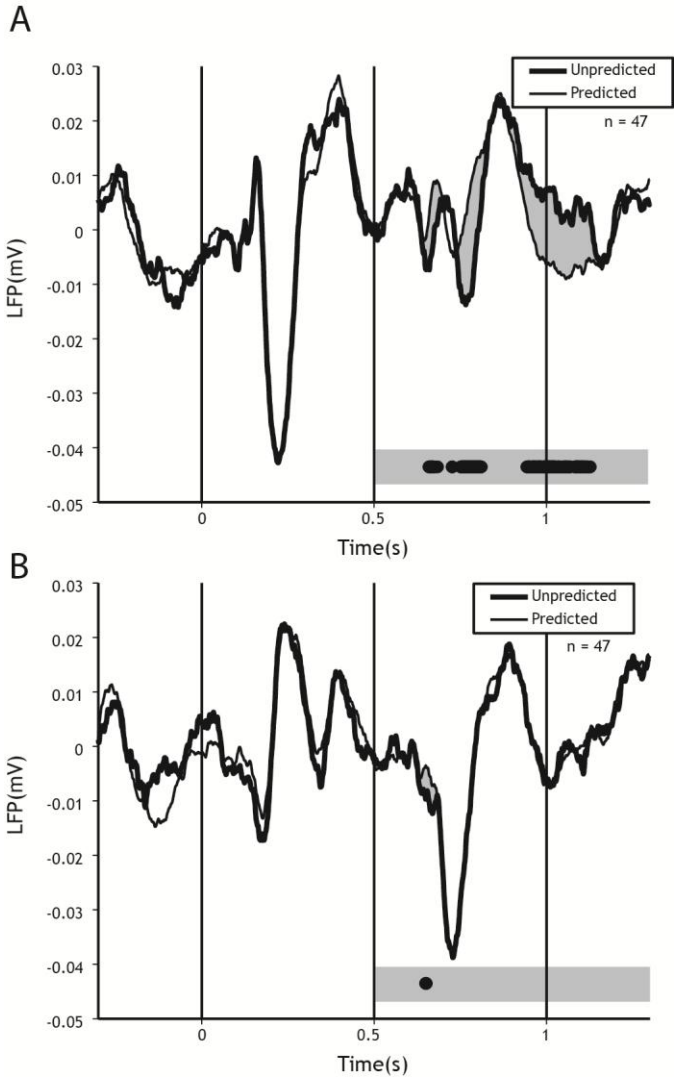


***Fig. 4.3: Local field potential activity for (A) visual-visual and (B) auditory-visual conditions.*** *Local field potentials in 47 sites. Comparisons and conventions same as Fig. 4.2.*

Fig. 4.3A and 4.3C, blue-filled negative component and red-filled positive component (Meyer, Ramachandran, et al., 2014)). We tested whether these hallmarks of prediction suppression in the LFP signal are present in the visual-visual and auditory-visual conditions in the present study. Comparing the mean voltage between the predicted and unpredicted conditions in the visual-visual trials revealed a significant difference between the conditions in both the negative component (unpredicted (mean = -0.0049mV) more negative than predicted (mean = -0.0049mV) p= 3.46 E-4, two-tailed t-test) and in the positive component (unpredicted (mean voltage = 0.0097mV) more positive than predicted (mean voltage = 0.0010mV), p= 1.41 E-5, two-tailed t-test) (Fig. 4.3A). These effects were present in each monkey. However, in the auditory-visual condition, there was no difference between the unpredicted and predicted conditions either in the initial negative component (unpredicted (mean = -0.0162mV),

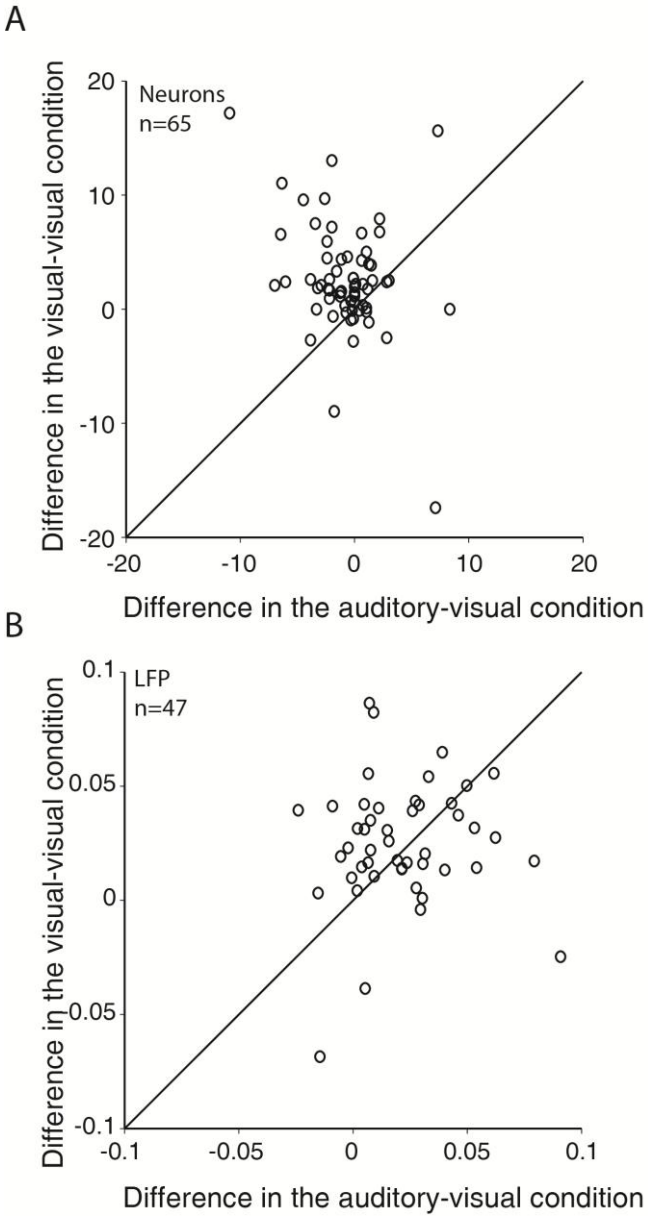


***Fig. 4.4: Scatter plots of differences between unpredicted and predicted responses for each (A) single unit (B) LFP site*** *(A) For each neuron, the difference between unpredicted and predicted responses was calculated 100-500ms after trailing image onset in the auditory-visual and visual-visual conditions. Greater scatter of points on the visual-visual side indicates that most single neurons show the prediction effect in the visual-visual condition, but not in the auditory-visual condition. (B) Similar to A, but measuring LFP responses at each site. Peak-to-peak amplitude is taken as the measure.*

predicted (mean = -0.0159mV), p = 0.898, two-tailed t-test) or in the positive component (unpredicted (mean = 0.0031mV), unpredicted (mean = 0.0030mV), p=0.8952, two-tailed t-test). Thus the hallmarks of prediction suppression consistently observed in previous studies were present only in the visual-visual condition and not in the auditory-visual condition. Prediction suppression is also manifest at the level of the LFP as a reduction in the amplitude of the excursion from maximal negativity at around 200ms to maximal positivity at around 300ms (Meyer and Olson, 2011; Meyer et al., 2014). This effect was significant in the population in the visual-visual condition (mean unpredicted response - 0.0806mV, mean predicted response – 0.0531mV, Fig. 4.3A, p = 1.18 E-7, two-tailed paired t-test, n = 47) and achieved strong significance in each monkey. We also observed that this effect was also significant in the population in the auditory-visual condition (mean unpredicted response – 0.1089mV, mean predicted response – 0.09mV, Fig. 4.3B, p = 2.25 E-7, two-tailed paired t-test, n = 47) and achieved strong significance in each monkey.

We computed the difference between the unpredicted and predicted responses in single neurons (Fig. 4.4A) for the visual-visual condition and the auditory-visual condition for each neuron and plotted the scatter of the responses. Most of the neurons show a greater difference (i.e, prediction suppression) in the visual-visual condition than in the auditory-visual condition. This effect was significant (p = 1.7 E-4, two-tailed paired t-test, n = 65). Similarly, we computed the difference between the unpredicted and predicted responses in the LFP peak-to-peak amplitudes (Fig. 4.4B) for the visual-visual condition and the auditory-visual condition for each site and plotted the scatter of the responses. Some sites show a greater difference (i.e, prediction suppression) in the visual-visual condition than in the auditory-visual condition. However, this effect was not significant at the population level (p=0.5428, two-tailed paired t-test, n = 47).

## DISCUSSION

The aim of the experiment was to determine whether prediction suppression in IT (Meyer & Olson, 2011; Meyer, Ramachandran, et al., 2014) is present for a visual stimulus when it is predicted by an auditory stimulus. We trained monkeys over weeks on sequential auditory-visual pairs in fixed sequence, and tested the responses of IT neurons when monkeys were exposed to trained and untrained sequences. We did not observe any sign of prediction suppression for the auditory-visual pairs either in the single-unit responses, or in the LFP voltage responses. Prediction suppression was observed only for pairs of stimuli where both stimuli were of the visual modality. This result is compatible with the account that statistical learning and the generation of perceptual predictions is modality specific in IT, and not a hallmark of an abstract predictive process (Conway & Christiansen, 2005, 2006). This further suggests that signals carrying information about prediction of visual stimuli in IT originates somewhere earlier in the visual pathway in a bottom-up manner.

Most studies of cross-modal learning, whether in statistical learning paradigms or otherwise, have used paradigms where the auditory and visual stimulus were presented simultaneously. Upon exposure to congruent auditory-visual pairings, humans can learn the associations implicitly (Seitz, Kim, van Wassenhove, & Shams, 2007). Exposure to such congruent auditory-visual pairs has been shown to improve perception: information presented in two congruent modalities (auditory-visual) as opposed to just one (either auditory or visual alone) was remembered better (Lehmann & Murray, 2005; Moran et al., 2013), and the congruence of an visual stimulus with a unique sound improves visual learning (Kim, Seitz, & Shams, 2008) and the learning of abstract rules in infants (Frank, Slemmer, Marcus, & Johnson, 2009). Cross-modal associations have also been shown to improve performance in

statistical learning tasks in humans (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010; Glicksohn & Cohen, 2013; Thiessen, 2010). In particular, the presence of auditory stimuli has been shown to facilitate visual statistical learning in IT (Robinson & Sloutsky, 2007). It is still an open question whether repeated congruent association of an image with a sound facilitates responses in IT in comparison to incongruent auditory-visual pairs.

Very few human behavioral studies have studied cross-modal statistical learning using paradigms where the auditory and visual stimuli were presented in temporal sequence, as in our study. In one study, humans were exposed for 7-8 minutes to sequences of different stimuli, some of them auditory (pure tones) and some of them visual (shapes). The order of presentation was probabilistically determined: an auditory stimulus could be followed 50% of the time by one other particular auditory stimulus, and the other 50% of the time by one particular visual stimulus. Thus the sequence consisted of certain valid auditory-auditory transitions, certain valid visual-visual transitions and certain valid auditory-visual transitions. When asked to report 'grammatical correctness' upon presentation with valid and invalid sequences of all three combinations, subjects performed above chance in reporting the validity of transitions of the same domain, but performed only at chance in reporting the validity of auditory-visual transitions (Walk & Conway, 2008). Thus, there is precedence to the idea that auditory-visual statistics cannot be parsed easily in sequences. The subjects were exposed to the training sequences only for 7-8 minutes and the authors argue that performance might have been better given more training time. In our experiments, we exposed monkeys to auditory-visual sequences over weeks. While we do not have a behavioral report about how familiar monkeys found the trained sequences to be, at any rate, training for such a long period does not seem to have induced prediction suppression in IT. It is possible that the learning might have taken place

elsewhere, for example in a multi-sensory area like STP or in higher association areas, and would need further experiments for confirmation.

**References**


Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. doi:10.1111/1467-9280.00063

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460–6. doi:10.1111/j.0956-7976.2004.00702.x

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–32. doi:10.1016/j.cognition.2011.06.010

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–12. doi:10.1111/j.1467-9280.2006.01801.x

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology (2006)*, *63*(2), 260–74. doi:10.1080/17470210902888809

Dehaene, S. (2009). *Reading in the Brain*. New York: Penguin Viking.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335–41. doi:10.1016/j.tics.2005.05.004

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11760138

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. doi:10.1037//0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–6. doi:10.1073/pnas.232472899

Frank, M. C., Slemmer, J. a, Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science*, *12*(4), 504–9. doi:10.1111/j.1467-7687.2008.00794.x

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*(6), 1161–9. doi:10.3758/s13423-013-0458-4

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11124355

Iwai, E., Aihara, T., & Hikosaka, K. (1987). Inferotemporal neurons of the monkey responsive to auditory signal. *Brain Research*, *410*(1), 121–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3580890

Jacobs, R. a, & Shams, L. (2010). Visual learning in multisensory environments. *Topics in Cognitive Science*, *2*(2), 217–25. doi:10.1111/j.1756-8765.2009.01056.x

Kaposvári, P., Csibri, P., Csete, G., Tompa, T., & Sáry, G. Y. (2011). Auditory Modulation of the Inferior Temporal Cortex Neurons in Rhesus Monkey. *Physiol. Res.*, *60*, 93–99.

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PloS One*, *3*(1), e1532. doi:10.1371/journal.pone.0001532

Kirkham, N. Z., Slemmer, J. a, & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11869728

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–43. doi:10.1038/nrn1533

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Brain Research. Cognitive Brain Research*, *24*(2), 326–34. doi:10.1016/j.cogbrainres.2005.02.005

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. doi:10.1038/264746a0

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Mitchel, A. D., Christiansen, M. H., & Weiss, D. J. (2014). Multimodal integration in statistical learning: evidence from the McGurk illusion. *Frontiers in Psychology*, *5*(May), 407. doi:10.3389/fpsyg.2014.00407

Mitchel, A. D., & Weiss, D. J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, *25*(4), 456–482. doi:10.1080/01690960903209888

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. *Language and Cognitive Processes*, *29*(7), 771–780. doi:10.1080/01690965.2013.791703

Moran, Z. D., Bachman, P., Pham, P., Hah Cho, S., Cannon, T. D., & Shams, L. (2013). Multisensory Encoding Improves Auditory Recognition. *Multisensory Research*, *26*(6), 581–592. doi:10.1163/22134808-00002436

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. a, & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, *50*(1), 302–13. doi:10.1016/j.neuroimage.2009.12.019

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–85. doi:10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–8. doi:10.1016/j.tics.2006.03.006

Robinson, C. W., & Sloutsky, V. M. (2007). Visual Statistical Learning : Getting Some Help from the Auditory Modality. In McNamara D. S. & Trafton J. G. (Ed.), *Proceedings of the 29th Annual Cognitive Science Society.* (pp. 611–616).

Rosenblum, L. D. (2013). Speech Perception as a Multimodal Phenomenon. *Curr Dir Psychol Sci*, *17*(6), 405–409. doi:10.1111/j.1467-8721.2008.00615.x.Speech

Saffran, J., & Griepentrog, G. J. (2001). Absolute Pitch in Infant Auditory Learning: Evidence for Developmental Reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8943209

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Saleem, K. S., Suzuki, W., Tanaka, K., & Hashikawa, T. (2000). Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(13), 5083–101. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10864966

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10), 1445–1453. doi:10.1068/p5843

Seltzer, B., & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research*, *149*(1), 1–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/418850

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*(6), 1093–106. doi:10.1111/j.1551-6709.2010.01118.x

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16334058

Walk, A. M., & Conway, C. M. (2008). Multisensory Statistical Learning : Can Associations between Perceptual Categories Be Acquired ?, 3337–3342.

# CHAPTER V

# STATISTICAL LEARNING OF SERIAL VISUAL TRANSITIONS BY NEURONS IN MONKEY INFEROTEMPORAL CORTEX


## ABSTRACT

If monkeys repeatedly, over the course of weeks, view displays in which two images appear in fixed sequence, then neurons of inferotemporal cortex come to exhibit prediction suppression. The response to the trailing image is markedly weaker if it follows the leading image with which it was paired during training than if it follows some other leading image. Prediction suppression is a plausible neural mechanism for statistical learning of visual transitions such as has been demonstrated in behavioral studies of human infants and adults. However, in the human studies, subjects are exposed to continuous sequences in which the same image can be both predicted and predicting and statistical dependency can exist between nonadjacent items. The aim of the present study was to investigate whether prediction suppression in ITC develops under such circumstances. To resolve this issue, we exposed monkeys repeatedly to triplets of images presented in fixed order. Microelectrode recording in inferotemporal cortex revealed that the second image, although itself subject to prediction suppression, was able to suppress the response to the third image. Furthermore, there were hints of a nonadjacent effect whereby the first image modulated the response to the third image. We conclude that prediction suppression can be induced by training not only with pairs of images but also with longer sequences.

## INTRODUCTION

Human infants and adults are able to learn rapidly through passive experience the statistical relations governing the transitions from one element to the next in a structured stream of visual stimuli (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005; Howard et al., 2008; Turk-Browne et al., 2008; Kim et al., 2009; Bulf et al., 2011) or auditory stimuli (Saffran et al., 1996; Pelucchi et al., 2009; Romberg and Saffran, 2010) (Gomez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009). The neuronal mechanisms underlying this capacity are not yet well understood (Summerfield and Egner, 2009; Meyer and Olson, 2011; Wacongne et al., 2012; Gavornik and Bear, 2014).

It is reasonable to suppose that inferotemporal cortex (ITC) contributes to the learning of transitional statistics in the visual domain. As the terminus of the ventral stream of visual areas (Ungerleider and Mishkin, 1982), ITC plays a critical role in object vision (Buckley et al., 1997; De Renzi, 2000) which depends on the ability of its neurons to respond selectively to complex images (Kobatake and Tanaka, 1994). Neurons in ITC exhibit statistical learning. Repeated viewing of a single image leads to familiarity suppression: the experienced image elicits comparatively weak responses (Freedman et al., 2006; Mruczek and Sheinberg, 2007; Meyer and Olson, 2014). Repeated viewing of two images close together in time leads to pair coding: neurons responsive to one image tend to respond to the other (Miyashita, 1988; Erickson and Desimone, 1999; Li and DiCarlo, 2008). Finally, and critically, repeated viewing of two images in fixed sequence, so that the leading image becomes a strong predictor for the trailing image, leads to prediction suppression: the trailing image, when presented in the trained context, elicits only a weak response (Meyer and Olson, 2011).

Prediction suppression is a plausible mechanism for sensitivity to transitional statistics at the behavioral level. However, there is a difference between the circumstances under which prediction suppression has been demonstrated – presentation of two images in sequence – and circumstances under which statistical learning is studied in humans – presentation of long strings of images. Long sequences possess two distinctive properties. First, each image can play a dual role, not only confirming or violating a prediction conveyed by a preceding image but also conveying a prediction about a subsequent image. Second, each image can condition the probability not only of the immediately succeeding image but also of later images. The aim of the present study was to determine whether prediction suppression is induced by training with sequences possessing these properties (Meyer, Ramachandran, et al., 2014).

## MATERIALS & METHODS

### Subjects

We studied two adult rhesus macaque monkeys (monkey 1, male, laboratory designation Tu, and monkey 2, female, laboratory designation Ec). All experimental procedures were approved by the Carnegie Mellon University IACUC and were in compliance with the guidelines set forth in the USPHS Guide for the Care and Use of Laboratory Animals.

### Images

All stimuli were digitized images of background-free objects. When presented at fixation 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. Eighteen images were used for training each monkey. The image sets used for monkeys 1 and 2 contained no items in common.

**Training**

Each monkey received repeated exposure to six triplets of images. The images in each triplet were always presented in the same sequence. The succession of events in each trial was: fixation spot (300 ms), first image at screen center (503 ms), an 18 ms delay, second image at screen center (503 ms), an 18 ms delay, third image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery. A trial was aborted without reward if the monkey broke central fixation at any time. On each training day, the monkey completed one or more runs. During a run, each triplet was presented ten times for a total of 60 trials. The sequence of trials within a run was random with the exception that during each block of six successfully completed trials each triplet must be

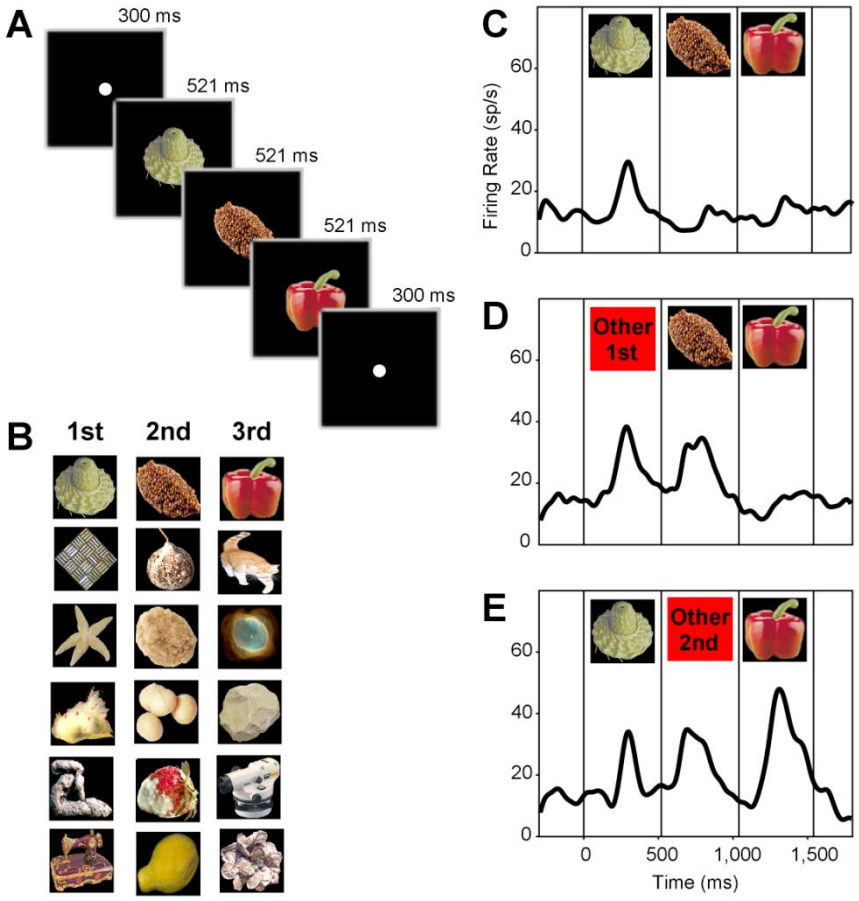


***Fig 5.1: Serial transitions: training paradigm*** A. Timing of events in each trial conducted during training and testing. B. Six triplets used in training monkey 1. The triplets used in training monkey 2 were formed from different images. C-E. Mean firing rate of a typical neuron in response to (C) trained sequences, (D) untrained sequences in which the first stimulus had been replaced with the first item from another training triplet, and (E) untrained sequences in which the second stimulus had been replaced with the second item from another triplet.

presented once. Monkey 1 viewed each triplet 1,090 times over the course of 32 days. Monkey 2 viewed each triplet 830 times over the course of 40 days.

**Testing**

During neuronal data collection, the monkeys performed a task identical to the one used during training with the sole exception that images were presented not only in trained sequences but also in sequences created by substitution of an item occupying the first or second position in another trained triplet. This was the smallest set of sequences required to ensure that the prediction status of an image was fully counterbalanced against other factors likely to influence neuronal firing. In addition to the six trained triplets (presented five times each), the monkeys viewed thirty untrained sequences in which the first image was borrowed from another triplet (each presented once) and thirty untrained sequences in which the second image was borrowed from another triplet (each presented once). Thus a run consisted of 90 trials divided evenly among trained sequences, sequences with a misfit first image and sequences with a misfit second image. The order of the trials was random. This design included only sequences required to ensure that the prediction status of an image was fully counterbalanced against other factors likely to influence neuronal firing. These factors included the identity of the preceding image, the identity of the current image, the ordinal position of the current image and general effects carried over from earlier in the trial, such as adaptation or an off-response. Other potentially informative sequences, for example presentation of the images as singletons, were omitted so as to minimize the danger that exposure to untrained sequences would attenuate the training effect. There was a trend toward attenuation of the effect over the course of the recording sessions, but the trend did not achieve significance and the effect remained robust even during late sessions.

**Recording**

In each monkey, a surgically installed cranial implant held a post for head restraint and a vertically oriented chamber through which the electrode could be introduced via a guide tube into ITC along tracks forming a square grid with 1 mm spacing. Recording was carried out in the left hemisphere of monkey 1 and the right hemisphere of monkey 2. The location of recording sites relative to morphological landmarks was determined by extrapolation from MRI-visible fiducial markers at known locations within the chamber. The recording sites occupied the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the interaural plane by 16-19 mm in monkey 1 and 13-16 mm in monkey 2.

**Database**

We recorded from 52 sites (27 in monkey 1 and 25 in monkey 2). Low-pass-filtered traces from these sites formed the LFP database. Neurons characterized during a complete test run numbered 112 (67 from monkey 1 and 45 from monkey 2). We classified a neuron as visually responsive if off-line analysis revealed a significant difference (paired t-test, $\alpha = 0.05$) between the mean firing rate during a pre-image period (300 ms to 50 ms before onset of the first image) and the mean firing rate during the image period (50 ms after onset of the first image to 50 ms after offset of the last image). The neuronal database consisted of 75 neurons (39 from monkey 1 and 36 from monkey 2) classified as visually responsive by this criterion.

**Statistical Analysis**

To demarcate periods during the trial when the population firing rate was different for untrained sequences and trained sequences, we compared the instantaneous difference signal (untrained firing rate minus trained firing rate) to an instantaneous statistical threshold based on a Monte Carlo analysis. We applied this procedure independently to untrained sequences

containing a misfit first image and those containing a misfit second image. For each neuron, we considered all 30 trained-sequence trials and all 30 untrained-sequence trials. Working with data at 1 ms resolution, we converted the discrete spike events in each trial to a spike-density function by convolution with a 10 ms Gaussian kernel. Then, over one thousand iterations, we labeled 30 randomly selected trials from each neuron as "pseudo-trained", labeled the 30 remaining trials from each neuron as "pseudo-untrained" and computed, for each 1 ms bin, the mean across all neurons of the signed difference between "pseudo-trained" and "pseudo-untrained" trials. Upon completion of the iterative procedure, we computed the standard deviation of the 1000 values in each 1 ms bin. We defined +2.58 and –2.58 standard deviations as the upper and lower confidence limits at that point in time. For the observed signal to cross either of these limits implied a likelihood of $p < 0.01$ that it would have occurred through random shuffling. We applied an identical procedure, including smoothing with a 10 ms Gaussian kernel, to the LFP data.

**RESULTS**

We exposed monkeys during a training period extending over several weeks to triplets of images presented in fixed back-to-back sequence for half a second each (Fig. 5.1A). The monkeys were rewarded at the end of each trial if they had maintained central fixation throughout the display. Each monkey viewed each of the six triplets more than 800 times during training (Fig. 5.1B). In ensuing microelectrode recording sessions, we measured neuronal responses elicited in anterior ITC not only by the trained triplets but also by untrained triplets created through substitution in a trained sequence of one element from another sequence. For each trained triplet, we created five untrained variants by replacing the first image with the first image from another trained sequence and five untrained variants by replacing the trained second

image with the second image from another trained sequence. During a recording session, each of the six trained triplets was presented five times and each of the sixty untrained triplets was presented once for a total of 90 trials. Using this procedure, we collected data from 75 visually responsive neurons (39 in monkey 1 and 36 in monkey 2).

The responses of a typical neuron are displayed in Fig. 5.1C-E. Presented with a trained sequence, this neuron responded strongly to the first image but

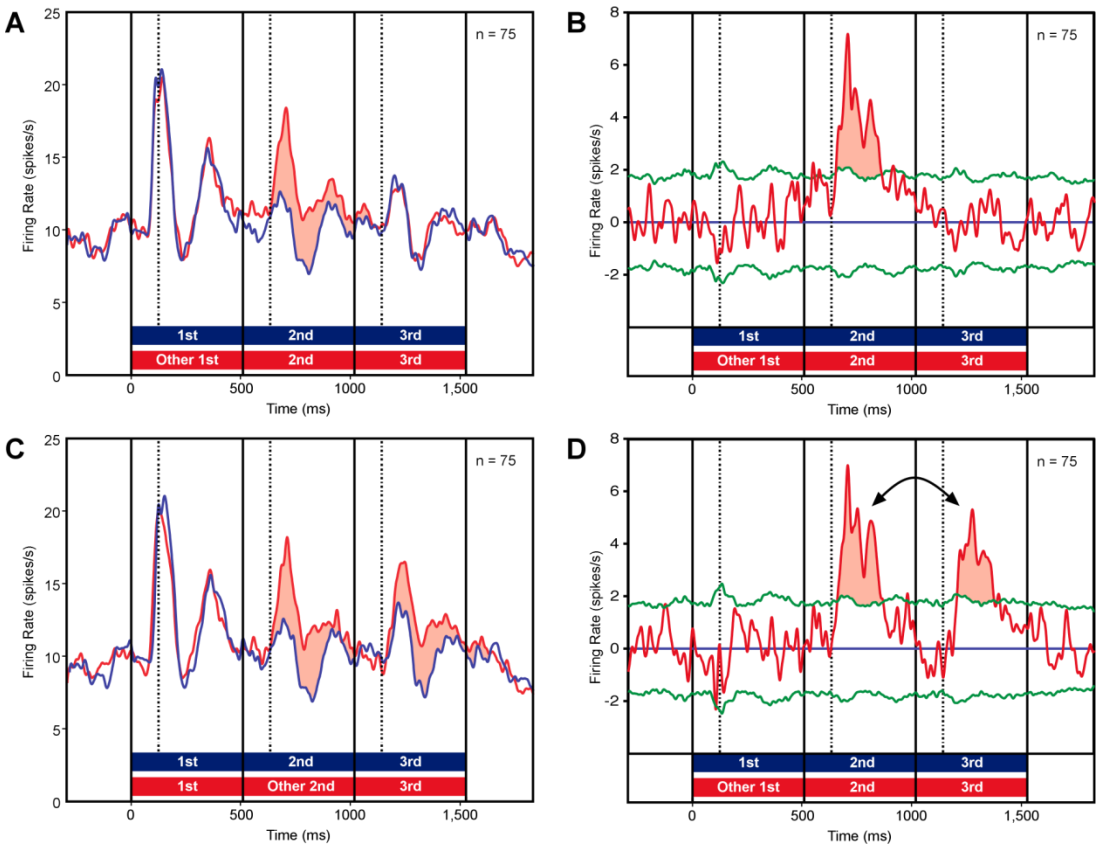


***Fig 5.2: Neuronal population activity for trained and untrained triplets*** A. Population firing rate elicited by trained sequences (blue curve) and by untrained sequences with a misfit first image (red curve). Red fill indicates the period during which the response to the untrained sequence was greater than the response to the trained sequence. B. The difference between the two population firing rates. Green curves represent confidence limits ($p < 0.01$) based on a Monte Carlo shuffling test. Red fill indicates the period during which the response to the untrained sequence was significantly greater than the response to the trained sequence. C-D. These plots compare the population firing rate elicited by trained sequences (blue curve) to the population firing rate elicited by untrained sequences with a misfit second image (red curve). Conventions as in A-B. Arrows are discussed in text. The dashed lines, each marking a point in time 125 ms after stimulus onset, are included to facilitate visual estimation of effect latency. The biphasic response pattern is typical of neurons in ITC (Rollenhagen and Olson, 2005).

weakly to the subsequent two images (Fig. 5.1C). This outcome can be explained as a consequence of prediction suppression arising from adjacent dependencies: the second image confirmed a prediction conveyed by the first image and the third image confirmed a prediction conveyed by the second image. Presented with an untrained sequence containing a misfit first image, the neuron responded weakly to the third image (Fig. 5.1D). This outcome can also be explained in terms of adjacent dependencies: the second image violated a prediction based on the first image whereas the third image confirmed a prediction based on the second image. Presented with an untrained sequence containing a misfit second image, the neuron responded strongly to both the second and the third image (Fig. 5.1E). This outcome likewise allows an explanation based on adjacent dependencies: the second image violated a prediction based on the first image and the third image violated a prediction based on the second image.

To determine whether the activity of the neuronal population as a whole conformed to this pattern, we carried out a paired t-test (n = 75) on the firing rate 50-500 ms following presentation of each image. Introduction of a misfit first image enhanced the response to the second image ($p = 1.5$ e-6) but not the third image ($p = 0.93$). Introduction of a misfit second image enhanced the responses to both the second image ($p = 3.2$ e-6) and the third image ($p = 6.4$ e-5). These results were present and significant ($\alpha = 0.05$) in each monkey considered individually. To examine the time-course of the effect, we constructed curves representing mean population firing rate as a function of time during the trial under all three conditions. When the first image was a misfit, the response to the second image was enhanced relative to trained-sequence baseline (red fill in Fig. 5.2 A). When the second image was a misfit, the responses to the second and third images were visibly enhanced (red fill in Fig. 5.2 C). To analyze the timing of the effect, we computed the instantaneous difference in firing rate between untrained and trained sequences (red curves in

Fig. 5.2B, D) and determined when it crossed a confidence limit ($p = 0.01$) established by a Monte Carlo procedure (green curves in Fig. 5.2 B, D). On average, across the three instances in which an image was unpredicted by the item immediately preceding it, enhancement became significant 131 ms after image onset. Further incidental observations conform to prior report based on two-item sequences: responses became weaker and occurred at longer latency as the sequence progressed, the response to an image was unaffected by the strength of the response to the preceding image, and the response to an unpredicted image was scaled up multiplicatively from the response elicited by the same image when predicted (Meyer and Olson, 2011). We do not yet know whether the mechanism of the prediction effect is "surprise enhancement" or "prediction suppression" because neither the original experiment nor this one contained a prediction-neutral control.

The local field potential (LFP) responses recorded at all 52 sites (27 in monkey 1 and 25 in monkey 2) depended in a similar fashion on the prediction status of each image. The response to each unpredicted image deviated from the response to the same image in a trained triplet by coursing first more negatively (blue fill in Fig. 5.3) and then more positively (red fill in Fig. 5.3). Each effect achieved statistical significance ($\alpha = 0.01$) as indicated by its exceeding Monte-Carlo-based confidence limits (green curves in Fig. 3B, D). The fact that the untrained sequences contained three images violating a prediction conveyed by the immediately preceding item allowed us to judge which features of the prediction effect were consistently present. Of particular note is the negative deflection that achieved brief significance shortly after image onset (all three instances are marked by asterisks in Fig. 5.3 B, D). Although this event was of low amplitude, it was absolutely consistent. The average time of attainment of significance across three conditions was 121 ms. This was 10 ms earlier than the onset of the spiking effect.

On the assumption that the earliest phase of the LFP is generated by bottom-up synaptic input to ITC, this observation raises the possibility that neurons afferent to ITC are sensitive to the prediction status of an image.

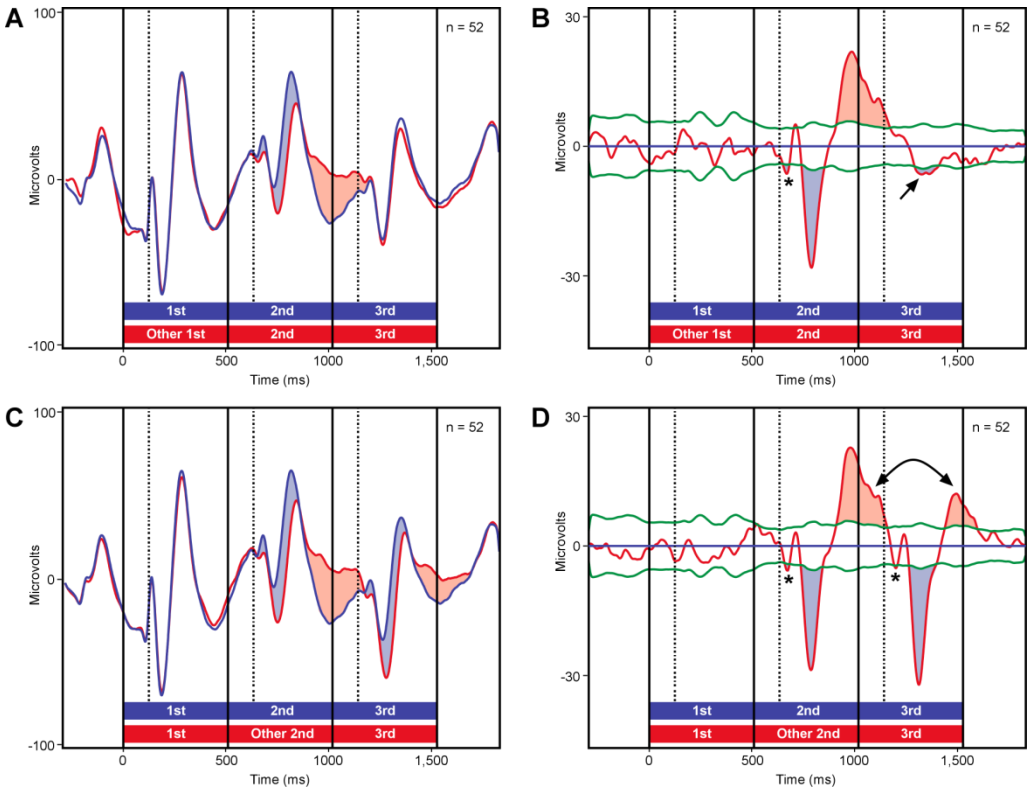


***Fig 5.3: LFP activity for trained and untrained triplets A.*** Mean LFP response elicited by trained sequences (blue curve) and by untrained sequences with a misfit first image (red curve). Blue (red) fill indicates periods during which the response to the untrained sequence was more negative (positive) than the response to the trained sequence. B. The difference between the two LFP responses. Blue (red) fill indicates periods during which the response to the untrained sequence was significantly more negative (positive) than the response to the trained sequence. C-D. These plots compare the mean LFP response elicited by trained sequences (blue curve) to the mean LFP response elicited by untrained sequences with a misfit second image (red curve). Asterisks and arrows indicate events discussed in the main text. Other conventions as in Fig. 2.

The effects described up to this point can be explained entirely in terms of adjacent dependencies. The response to each image was strong if it violated a prediction conveyed by the immediately preceding image and weak otherwise. Nonadjacent dependencies could, however, have exerted a superadded effect on neural activity because the first image in each triplet strongly predicted the final image. There are trends in the data suggesting that the response to the third image was indeed affected by its violating or confirming a prediction conveyed by the first image. In sequences containing a

misfit first image, the LFP evoked by the third image exhibited a small but significant negative deflection (arrow in Fig. 5.3B). The effect occurred in both monkeys. It is consistent with an interpretation based on the third image's violating a prediction conveyed by the first image. In sequences containing a misfit second image, the second image violated a prediction conveyed by the first image and the third image violated a prediction conveyed by the second image. Nevertheless, the prediction effect was weaker for the third than for the second image both at the level of spiking activity (double arrow in Fig. 5.2D) and at the level of the positive deflection of the LFP (double arrow in Fig. 5.3D). The difference fell short of statistical significance in the case of spiking activity ($p = 0.70$; paired t-test; unpredicted minus predicted firing rate 150-500 ms after stimulus onset; $n = 75$) but did achieve significance in the case of the LFP ($p = 0.014$; paired t-test on unpredicted minus predicted voltage 450-650 ms after stimulus onset; $n = 52$). The effect occurred and achieved significance ($\alpha = 0.05$) in each monkey considered individually. It is consistent with an interpretation based on the third image's confirming a prediction conveyed by the first image.

## DISCUSSION

The stream of experience is far from random. Events that have just occurred carry with them dependable predictions about events that will occur next. Some predictions are based on physical principles so fundamental that they may possess a hard-wired representation in the brain (Alink et al., 2010). Other predictions must be learned. Music and language are cases in point. The brain of an acculturated listener listening to a melody or a sentence is sensitive to how it will probably unfold. This is manifest in the fact that events violating reasonable expectation elicit strong neural responses (Dien et al., 2003; James et al., 2008; Vuust et al., 2009; Pearce et al., 2010; Kim et al., 2011). The ability of the brain to detect and respond to improbable events is thought

to depend not only on mastery of complex rules that govern the event stream but also on the encoding of simple statistical relations (Aslin and Newport, 2012). Melodies and sentences exhibit tonotactic and phonotactic regularities: the value of an upcoming note or phoneme is probabilistically related to the values of at least the two preceding elements (Pearce and Wiggins, 2004; Gonzalez-Gomez and Nazzi, 2013). Human infants and adults are able to learn rapidly, during passive listening, the statistical relations between immediately adjacent items in a structured auditory stream (Saffran et al., 1996; Pelucchi et al., 2009; Romberg and Saffran, 2010). They are also able, under favorable circumstances, to learn nonadjacent dependencies between events separated by an intervening item (Gomez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009). The capacity for the learning of transitional statistics, although most studied in the auditory domain, extends to visual sequences as well (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005; Howard et al., 2008; Turk-Browne et al., 2008; Kim et al., 2009; Bulf et al., 2011).

Prediction suppression, as observed in ITC, is a plausible mechanism for behavioral sensitivity to visual transitional statistics, as demonstrated in the human studies. However, for it to serve this role would require that it develop in response to the presentation of image sequences longer than the two-image displays previously employed in studies of ITC (Meyer and Olson, 2011). In other words, it must occur under conditions in which the same image can be both predicted and predicting and in which the potential for nonadjacent predictions exists. The key observation of this study is that prediction suppression does indeed occur when these conditions are met.

The fact that prediction suppression occurred under circumstances in which the second image was both predicted and predicting casts light on the nature of neuronal mechanisms mediating

suppression. In the simplest possible model of the phenomenon, ITC neurons responsive to a given leading image induce a state of suppression among neurons responsive to the predicted trailing image. In this framework, the rate of firing of neurons representing the leading image could reasonably be expected to determine the degree of suppression of the response to the trailing image. However, it did not. Regardless of whether the second image confirmed a prediction and elicited a weak response (low blue second-image response in Fig. 5.2A) or violated a prediction and elicited a strong response (high red second-image response in Fig. 5.2A), its occurrence suppressed the response to the predicted third image (overlapping low third-image responses in Fig. 5.2A). The induction of the suppressive state must be a highly nonlinear process or must depend on neurons other than those whose activity is captured in the neuronal population response.

The occurrence of subtle effects apparently dependent on whether the third image violated or confirmed a prediction conveyed by the first image is consistent with findings indicating that human observers learn nonadjacent dependencies (Gomez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009). However, it is surprising in light of the fact that learning of nonadjacent statistics by human observers depends on use of a training display that emphasizes relations between nonadjacent items, either by making them physically similar (Creel et al., 2004; Onnis et al., 2005; Gebhart et al., 2009) or by randomizing the identity of the intervening element (Gomez, 2002; Newport and Aslin, 2004). The use of prolonged training in our study may have allowed nonadjacent dependencies to exert an effect even without these manipulations. This interpretation fits with the observation that humans are sensitive to nonadjacent musical and linguistic dependencies after prolonged natural exposure (Pearce and Wiggins, 2004; Gonzalez-Gomez and Nazzi, 2013).

**REFERENCES**


Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus Predictability Reduces Responses in Primary Visual Cortex. J Neurosci 30:2960-2966.

Aslin RN, Newport EL (2012) Statistical Learning: From Acquiring Specific Items to Forming General Rules. Current Directions in Psychological Science 21:170-176.

Buckley MJ, Gaffan D, Murray EA (1997) Functional Double Dissociation Between Two Inferior Temporal Cortical Areas: Perirhinal Cortex Versus Middle Temporal Gyrus. Journal of Neurophysiology 77:587-598.

Bulf H, Johnson SP, Valenza E (2011) Visual statistical learning in the newborn infant. Cognition 121:127-132.

Creel SC, Newport EL, Aslin RN (2004) Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition 30:1119-1130.

De Renzi E (2000) Disorders of Visual Recognition. Seminars in Neurology 20:479-486.

Dien J, Frishkoff GA, Cerbone A, Tucker DM (2003) Parametric analysis of event-related potentials in semantic comprehension: evidence for parallel brain mechanisms. Cognitive Brain Research 15:137-153.

Erickson CA, Desimone R (1999) Responses of Macaque Perirhinal Neurons during and after Visual Stimulus Association Learning. The Journal of Neuroscience 19:10404-10416.

Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition 28:458-467.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. Cereb Cortex 16:1631-1644.

Gavornik JP, Bear MF (2014) Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat Neurosci advance online publication.

Gebhart A, Newport E, Aslin R (2009) Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. Psychonomic Bulletin & Review 16:486-490.

Gomez RL (2002) Variability and Detection of Invariant Structure. Psychological Science 13:431-436.

Gonzalez-Gomez N, Nazzi T (2013) Effects of Prior Phonotactic Knowledge on Infant Word Segmentation: The Case of Nonadjacent Dependencies. Journal of Speech, Language, and Hearing Research 56:840-849.

Howard JH, Howard DV, Dennis NA, Kelly AJ (2008) Implicit learning of predictive relationships in three-element visual sequences by young and old adults. Journal of Experimental Psychology: Learning, Memory, and Cognition 34:1139-1157.

James CE, Britz J, Vuilleumier P, Hauert C-A, Michel CM (2008) Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. NeuroImage 42:1597-1608.

Kim R, Seitz A, Feenstra H, Shams L (2009) Testing assumptions of statistical learning: Is it long-term and implicit? Neuroscience Letters 461:145-149.

Kim S-G, Kim JS, Chung CK (2011) The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study. PLoS ONE 6:e17337.

Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition 83:B35-B42.

Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. Journal of Neurophysiology 71:856-867.

Li N, DiCarlo JJ (2008) Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. Science 321:1502-1507.

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proceedings of the National Academy of Sciences 108:19401-19406.

Meyer T, Olson C (2014) Image Familiarization Sharpens Response Dynamics of Neurons in Inferotemporal Cortex. Under review.

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

Mruczek REB, Sheinberg DL (2007) Context Familiarity Enhances Target Processing by Inferior Temporal Cortex Neurons. The Journal of Neuroscience 27:8533-8545.

Newport EL, Aslin RN (2004) Learning at a distance I. Statistical learning of non-adjacent dependencies. Cognitive Psychology 48:127-162.

Onnis L, Monaghan P, Richmond K, Chater N (2005) Phonology impacts segmentation in online speech processing. Journal of Memory and Language 53:225-237.

Pearce MT, Wiggins GA (2004) Improved methods for statistical modeling of monophonic music. Journal of New Music Research 33:367-385.

Pearce MT, Ruiz MaH, Kapasi S, Wiggins GA, Bhattacharya J (2010) Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. NeuroImage 50:302-313.

Pelucchi B, Hay JF, Saffran JR (2009) Statistical Learning in a Natural Language by 8-Month-Old Infants. Child Development 80:674-685.

Rollenhagen JE, Olson CR (2005) Low-Frequency Oscillations Arising From Competitive Interactions Between Visual Stimuli in Macaque Inferotemporal Cortex. Journal of Neurophysiology 94:3368-3387.

Romberg A, Saffran J (2010) Statistical learning and language acquisition. Wiley Interdiscip Rev Cogn Sci 1:906-914.

Saffran JR, Aslin RN, Newport EL (1996) Statistical Learning by 8-Month-Old Infants. Science 274:1926-1928.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends in Cognitive Sciences 13:403-409.

Turk-Browne NB, Junge JA, Scholl BJ (2005) The Automaticity of Visual Statistical Learning. Journal of Experimental Psychology: General 134:552-564.

Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA (2008) Multidimensional visual statistical learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 34:399-407.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of Visual Behavior (Ingle DJ, Goodale MA, Mansfield RJW, eds), pp 549-586. Cambridge, MA: MIT Press

Vuust P, Ostergaard L, Pallesen KJ, Bailey C, Roepstorff A (2009) Predictive coding of music--Brain responses to rhythmic incongruity. Cortex 45:80-92.

Wacongne C, Changeux J-P, Dehaene S (2012) A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. The Journal of Neuroscience 32:3665-3678.

# THE ROLE OF TIMING IN STATISTICAL LEARNING IN MONKEY INFEROTEMEPORAL CORTEX

## INTRODUCTION

When monkeys are passively exposed to pairs of visual images in fixed sequence over many weeks, neurons in inferotemporal cortex are induced to express a phenomenon called prediction suppression: the responses to an image that is predicted on the basis of long-term exposure are suppressed compared to the responses to unpredicted images (Meyer & Olson, 2011; Meyer, Ramachandran, et al., 2014). In studies investigating prediction suppression so far, we typically expose monkeys to two images (a 'leading' image followed by a 'trailing' image) in rapid sequence in a single trial. Each image is presented for 500ms. The presentation of the leading image in the sequence is immediately followed by the presentation of the trailing image; the only delay between the images is the 16ms refresh rate of the monitor.

Prediction suppression is essentially a phenomenon predicted trailing images are suppressed by predicting leading images. The predicting influence of the leading image in the sequence clearly lasts in the system for at least 16ms for leading images strongly suppress trailing images that follow it immediately. It is not clear whether the predicting effect of the leading effect extends across longer delays. We sought to test this by training monkeys on pairs of images where there was a 300ms between the images. In order to test the limits of this effect, we also included conditions where there was a 600ms delay between the images. Each image was still displayed for 500ms, so a 600ms delay was longer than the duration of image display.

After extensive exposure to pairs of different images displayed with different delays between them, we would then test them by comparing the responses to trailing images when they were preceded by a predicting leading image ('predicted' condition), and the responses to the same trailing images when they were preceded by a non-predicting leading image ('unpredicted' condition. Testing would be done for image pairs not only with the delays they were trained at, but also at delays they were not trained at, to ask whether the effect generalizes across delays.

Thus this study would have the potential to answer many different questions:

1. Does prediction suppression extend across long delays of 300ms and 600ms?
2. Does prediction suppression for images occur at a delay different from the delay they were trained at?
3. Does the omission of an image at a 'predicted' time (eg. When the trailing image that is supposed to be displayed with no delay is displayed 600ms later) cause the occurrence of a surprise signal in IT neurons?

## MATERIALS AND METHODS

### Subjects

Two adult rhesus macaque monkeys (*Macaca mulatta*) were used for the study: monkey 1 (adult male; laboratory designation Tu) and monkey 2 (adult male; laboratory designation Ec). All procedures used during surgery and experiments were in accordance with the guidelines set by the United States Public Health Service Guide for the Care and Use of Laboratory Animals and were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee.

**Training**

Each monkey received repeated exposure to pairs of images presented in fixed sequence. The general succession of events in each trial was: fixation spot (300 ms), leading image at screen center (503 ms), a variable delay, trailing image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery (Fig. 6.1B). We trained our monkeys on six pairs of images. Two of those pairs were presented with an 18ms delay between them (short delay - A1B1 and A2B2 in Fig. 6.1A - red). The next two pairs were presented with a 300ms delay between them (medium delay - A3B3 and A4B4 in Fig. 6.1A - blue). The last two pairs were presented with a 600ms delay between them (long delay - A5B5 and A6B6 in Fig. 6.1A - green). A trial was aborted without reward if the monkey failed to maintain fixation within a 4° x 4° central window. On each training day, the monkey completed one or more runs. A run consisted of 60 successfully completed trials. Each trained pair (Fig. 6.1A, gray filled squares along the diagonal) was presented to the monkey ten times during a training run. The conditions were interleaved randomly subject to the constraints (a) that within each block of six successfully completed trials each condition had to be imposed once and (b) the same condition could not be imposed on two successive trials. The number of runs completed on a day ranged from 1 to 6 in monkey 1 and from 1 to 5 in monkey 2. Monkey 1 viewed each of the ten sequences 840 times during 84 runs extending over 26 days. Monkey 2 viewed each sequence 860 times during 86 runs extending over 41 days.

**Testing**

During neuronal data collection, the monkeys were presented with both trained and untrained sequences, both with trained and untrained delays. In other words, trained sequences (filled gray

squares along the diagonal), as well as untrained sequences (filled squares of different colors, off diagonal in Fig. 6.1A) were presented to the monkeys with three gap conditions: no delay, 300ms delay and 600ms delay. Each trained sequence was presented 5 times during a data collection block with a particular timing delay (colored numbers in squares on the diagonal, Fig. 6.1A). Each untrained sequence was presented once (white numbers in squares, Fig. 6.1 A). Thus, each block consisted of 60 trials. The conditions were imposed in random order with replacement on error. There were three consecutive blocks with the same image pairs (i.e., all pairs shown in Fig. 6.1 A); the difference between blocks was that within each block, image pairs were was displayed with a different delay (18ms, 300ms or 600ms) between them. The order of the blocks was randomized.

Thus within each block, there were two image pairs that were tested with the trained delay and 4 image pairs that were tested with untrained delays. The fundamental comparison within each block was between untrained and trained pairs tested with the same delay they were trained at (i.e., testing the prediction effect at different trained delays). In order to test whether prediction suppression manifests with delays different from the trained delay between the stimuli, we would then compare responses to trained (gray squares in Fig. 6.1A) and untrained (red, blue and green squares in Fig. 6.1A) pairs tested with a delay different from the delay they were trained at. We refer to the three blocks in the following way henceforth: the 'tested short' (block with 18ms delay), 'tested medium' (block with 300ms delay), and 'tested long' (block with 600ms delay).

**Images**

All stimuli were digitized images of background-free objects. When presented on an LCD monitor 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. The full stimulus set for monkey 1 consisted

of eight leading images and eight trailing images paired according to rules summarized in Fig. 1B. The same images were used in monkey 2 but with their sequential status (leading or trailing) reversed and the pattern of pairing altered so that no images paired in monkey 1 were paired in monkey 2.

**Recording**

An electrode was introduced through a vertical guide tube into left (monkey 1) or right (monkey 2) temporal lobe. Recording sites, identified by extrapolation from MRI-visible fiducial markers within the chamber, were within the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the interaural plane by 16-19 mm in monkey 1 and 13-16 mm in monkey 2.

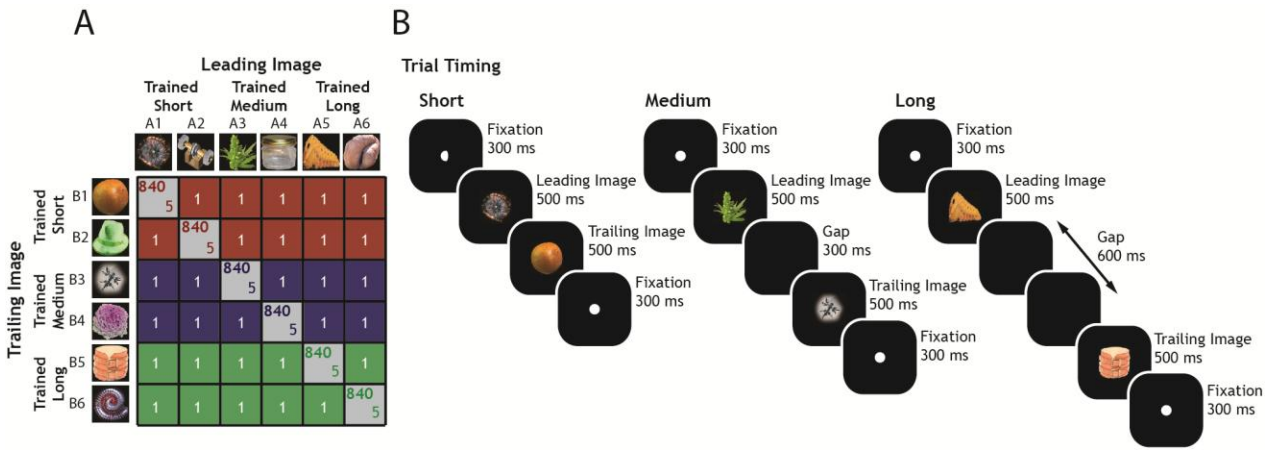


***Fig. 6.1 Training and Testing Protocol*** Six image pairs were paired in sequence (An followed by Bn) and were present to monkeys over weeks with different delays (short - no delay, medium - 300ms delay and long - 600ms delay) between the images. *(A)* Trained sequences in gray along diagonal were presented over 800 times each. All possible sequences were presented during data acquisition (60 trials) with all possible delays between images (60 trials x 3 delays = 180 trials). Squares shaded in red, blue and green were unpredicted controls for images trained with short, medium and long times respectively. *(B)*Trial timing for different delays.

**Database**

Often, neuronal isolation was lost after testing on one or two of the blocks. So the data for each of the three blocks (tested short, tested medium and tested long) was obtained from different numbers of neurons, though there is considerable overlap in the populations. We recorded responses in the 'tested short' condition from 41 sites (18 and 23 in monkeys 1 and 2 respectively). We recorded responses in the 'tested medium' condition from 40 sites (18 and 22 in monkeys 1 and 2 respectively). We recorded responses in the 'tested long' condition from 39 sites (16 and 23 in monkeys 1 and 2 respectively). Traces from these sites passed through a low-pass filter with a high frequency cut-off of 170 Hz formed the LFP database. Neurons characterized during a complete test run numbered 73 (36 from monkey 1 and 37 from monkey 2), 72 (36 from monkey 1 and 36 from monkey 2) and 69 (31 from monkey 1 and 38 from monkey 2) in the tested short, tested medium and tested long conditions respectively. We

classified a neuron as visually responsive if, for either the leading or the trailing image, the mean firing rate in a window 50-300 ms following image onset exceeded the mean firing rate in a 100 ms baseline window centered on image onset (one-tailed t-test, $\alpha = 0.05$). The neuronal database consisted of 65 neurons (34 and 31 in monkeys 1 and 2 respectively) meeting this criterion in the

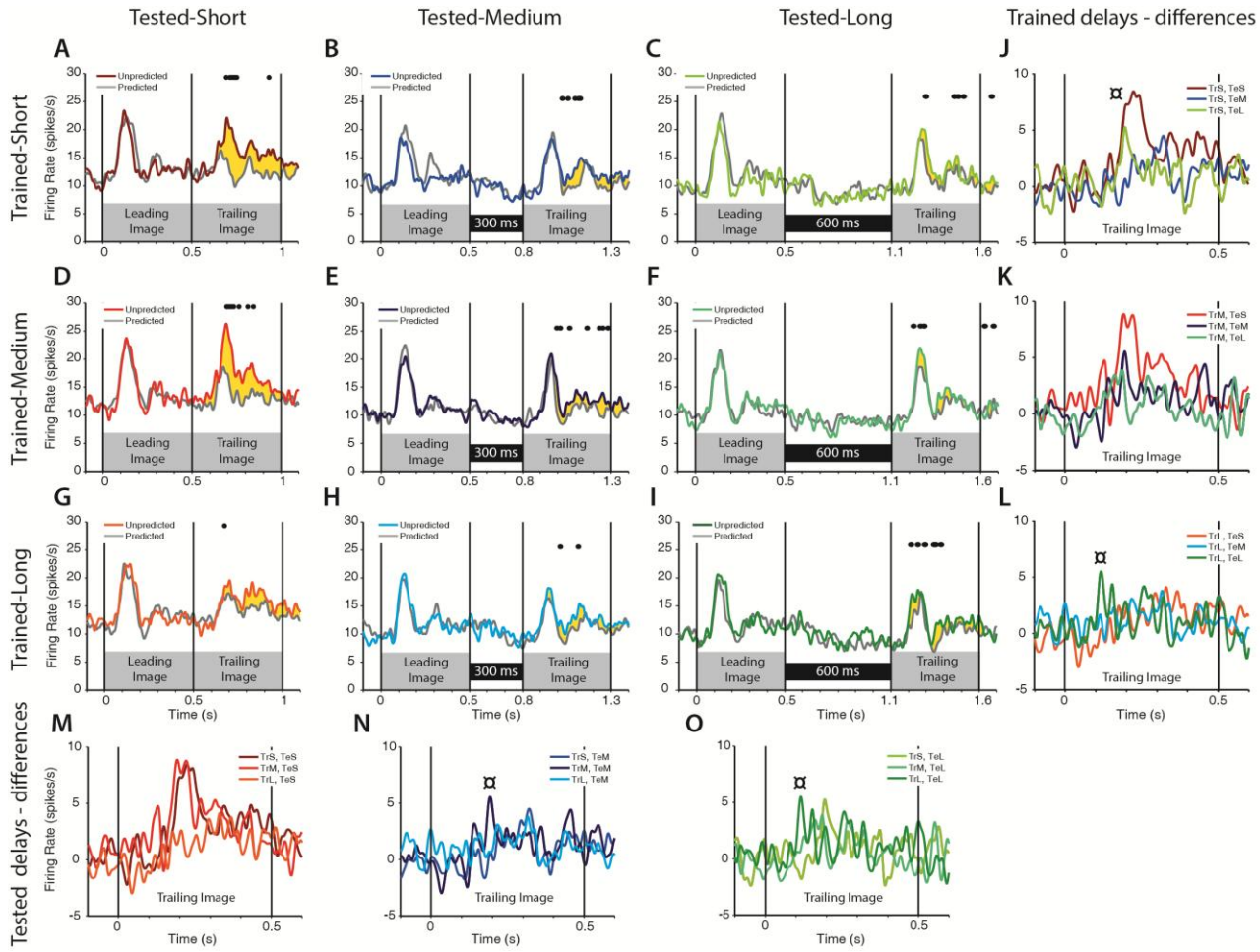


***Fig. 6.2 Population PSTHs with different delays*** Population PSTHs are plotted for sequences trained with three different delays and tested with different delays. Sequences tested at the trained delay are presented in panels A (short), E (medium) and I (long), with unpredicted condition represented in the darkest shades of red, blue and green respectively. Prediction suppression is manifest for sequences tested with both trained and untrained delays. Significant 10ms windows for the duration of the trailing image with a significant difference between the predicted and unpredicted conditions are marked with black circles. The difference curves (unpredicted minus predicted) for the trailing image are plotted, organized by tested delays (panels M,N,O) and trained delays (J, K, L) with colors of the curves the same as the color of the unpredicted curves in each panel.

trained short condition, 62 neurons sites (32 and 30 in monkeys 1 and 2 respectively) meeting this criterion in the trained medium condition and 55 neurons (26 and 29 in monkeys 1 and 2 respectively) meeting this criterion in the trained long condition.

**RESULTS**

The experiment began with a training period extending over multiple weeks during which the monkeys viewed each training sequence with a specific delay between the images more than 800 times (Fig. 6.1A). Six image sequences, constructed from six leading and six trailing images, were presented during this period. During the training phase, two of these sequences were presented with no delay between them (the only delay was the natural refresh rate of the monitor – 18ms) Two other sequences were presented with a 300ms delay between them. The last two sequences were presented with a 600ms delay between them (Fig. 6.1B). We call each of these delays 'short', 'medium' and 'long' delays respectively. These sequences were interleaved and repeatedly presented to the monkeys over days. After completion of training, we measured the responses of neurons in anterior TE to leading and trailing images presented in both trained and untrained sequences (gray and colored squares in Fig. 6.1A), with both trained and untrained delays between the images. This was accomplished by recording neuronal responses while monkeys were presented with all the sequences in Fig. 6.1A in three blocks with three inter-image delays: short, medium and long. In each block, each trained sequence was presented 5 times and each untrained sequence was presented once. Thus the three separate blocks consisted of 180 trials in all. Due to limitations in the ability to keep a neuron well-isolated for a long period of time, data or the tested short, tested medium and tested

| | Tested Short (n=65) | Tested Medium (n=62) | Tested Long (n=55) |
|---|---|---|---|
| Trained Short | 6.33 E-6 | 2.245 E-4 | 0.0262 |
| Trained Medium | 7.94 E-6 | 0.0037 | 0.0452 |
| Trained Long | 0.0085 | 0.0106 | 2.31 E-4 |

***Table 6.1*** P-values for different combinations of trained and tested delays when comparing neuronal responses to the predicted and unpredicted conditions (two-tailed t-test,, 100-400ms after trailing image onset, α=0.05). Significant p-values in magenta.

and achieved significance in at least one monkey.

long trials were obtained from different numbers of neurons. We collected full data sets from 65 neurons in the trained short condition (34 in Monkey 1, 31 in Monkey 2), 62 neurons in the trained medium condition (32 in Monkey 1, 30 in Monkey 2) and 55 neurons in the trained long condition (26 in Monkey 1, 29 in Monkey 2). All analyses described below were conducted on data combined across the two monkeys. Every effect described was present in both monkeys

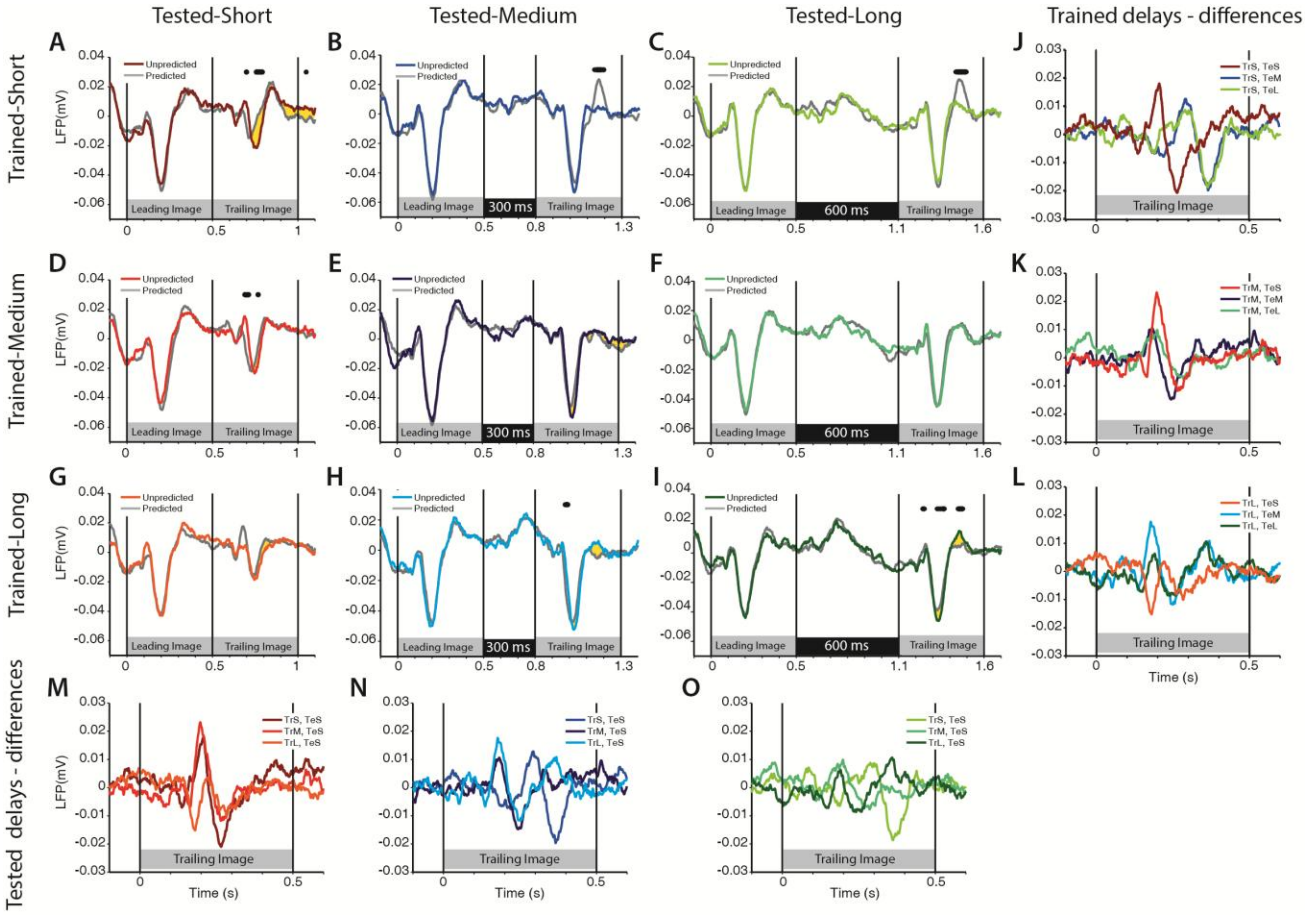


***Fig. 6.3 Population LFPs with different delays*** Conventions same as in Fig. 6.2

To determine whether prediction suppression occurred at each of short, medium and long delays, we compared population histograms representing the mean firing rate during trials in which trailing images appeared in trained sequences (and thus were predicted) or in untrained sequences (and thus were unpredicted). These comparisons were initially on data where the delay between a pair of images was the same as the delay they were trained on initially.

Prediction suppression occurred for trailing images from all three delay blocks (Fig. 6.2) tested at all both the delays they were trained at (Fig. 6.2 A, E, I) as well as at untrained delays (Fig. 6.2 B,C, D, F, G,H) (paired t-test between unpredicted and predicted conditions, 100-400ms after onset of the trailing image, $\alpha$=0.05). We took as a quantitative index of prediction suppression the mean firing rate 100-500 ms after image onset on trials in which the trailing images were unpredicted minus the same measure on trials in which they were predicted for each

of the nine plots in Fig. 6.2 A-I. When the difference curves were plotted by tested time (short, medium and long) for different trained timings (Fig. 6.2 M,N,O), it was observed that there was a trend for the earliest prediction effect to emerge at the trained timing (darkest blue and darkest green curves appearing early in Fig. 6.2 N and O). Similarly when the difference curves were plotted according to the trained time for different tested timings, the largest effect size in the trained small and trained long conditions was when the images were tested at the same timing (Fig. 6.2 J, L). Numbers of significant neurons in each combination of trained and tested delay condition is summarized in Fig. 6.5.

| | Tested Short (n=41) | Tested Medium (n=40) | Tested Long (n=39) |
|---|---|---|---|
| Trained Short | 0.0018 | 0.1518 | 0.7196 |
| Trained Medium | 0.0027 | 0.0085 | 0.9151 |
| Trained Long | 0.0223 | 8.73 E-4 | 1.42 E-5 |

***Table 6.2*** P-values for different combinations of trained and tested delays when comparing LFP responses to the predicted and unpredicted conditions (two-tailed t-test, 100-400ms after trailing image onset, α=0.05). Significant p-values in magenta.

Prediction suppression was also observed at the level of the LFP for images in all three delay blocks at the delays they were trained at (Fig. 6.3 A, E, I) (paired t-test between amplitude of deflection (N200-P300) for unpredicted and predicted conditions, 100-400ms after onset of the trailing image, α=0.01). When tested with short delay, pairs trained with other delays also exhibited significant prediction suppression (α=0.05). When tested with medium delay, pairs trained with a long delay exhibited significant prediction suppression (α=0.05). When tested with long delay, only, pairs trained with a long delay exhibited significant prediction suppression. Difference plots grouped by tested delay (Fig. 6.3 M,N,O) and by trained delay (Fig. 6.3 J, K,L) are also plotted.

In order to assess whether there is an advantage for images tested at the times they were trained at, we computed the normalized area under the curve for the unpredicted condition (colored curves, Fig. 6.2 A-I) and from it, subtracted the area under the curve for the predicted condition (gray curves, Fig. 6.2 A-I). The resultant values provide a measure of the size of the prediction effect (yellow shaded region in each panel). We scaled these values by their diameter and plotted them as circles (Fig. 6.4A). There is a trend where the biggest size of the effect was observed for images tested at the trained timings, (diagonals, Fig. 6.4A), however, it is not significant.

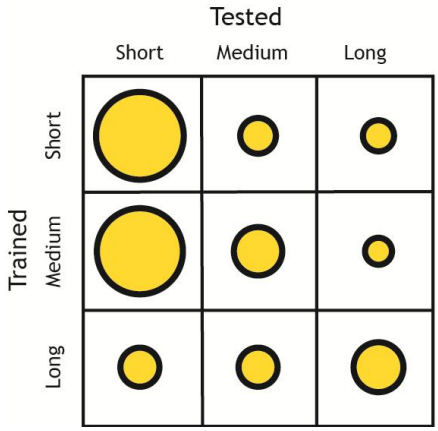


***Fig. 6.4 Effect size across conditions*** The size of the prediction effect (differences between the normalized areas under the unpredicted and predicted curves ) for the different conditions are plotted. No significant differences were observed in the effect size between conditions, but there was a trend where the biggest size of the effect within each tested delay was for mages trained at that delay (diagonal).

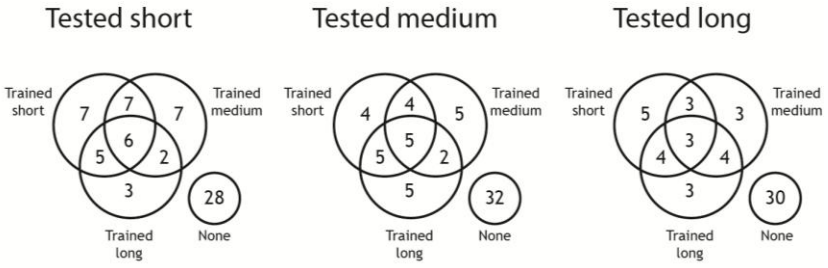


***Fig. 6.5 Numbers of significant neurons*** Numbers of neurons with a significant effect of prediction status for each of the nine conditions scored in a two-factor ANOVA with image identity and prediction identity as factors.

Neither in the single units, not in the LFPs, did we notice a ‘surprise’ response when an image was omitted at a time when it was predictably displayed (Fig. 6.2 B, C, F).

**Discussion**

We started out seeking to test whether the prediction effect still persists when there are long delays between the leading and trailing images. In other words, what is the longest duration over which the predictive effect of the leading image is exerted over the trailing image? We tested with different image pairs trained with a short delay (0ms), medium delay (300ms) and a long delay (600ms), and observed robust prediction suppression up to 600ms.

We then tested whether prediction suppression is observed for images displayed with delays that they were not trained with. At the level of single neurons, prediction suppression extends across timing, and is expressed for images when the delay between them was both shorter and longer than the trained delay. At the level of LFPs, prediction suppression seems to manifest only with delays other than trained delays when they are shorter than the trained delay. These findings reveal that prediction suppression is highly flexible and robust mechanism that is reliable across fairly variable and long delays. This suggests that visual statistical learning is capable of being generalized to novel temporal contexts regardless of training conditions.

There was no indication of a surprise response when an image was not presented at its predictable timing, i.e., when we compared the responses of images presented at trained vs. untrained timings.

There are very few studies that explicitly test the temporal limits of visual statistical learning. The effects of temporal constraints such as presentation duration (Conway & Christiansen, 2009) and ordinal position of image in time (Turk-Browne & Scholl, 2009) have been studied. Slower presentation rate of a sequence of stimuli (i.e. with longer delays between each of the stimuli) has been suggested to improve performance in visual

statistical learning, but worsen performance in auditory statistical learning (Emberson, Conway, & Christiansen, 2011). However, none of these studies really ask the question of whether statistical learning at longer delays generalizes across delays. Our study suggests that visual statistical learning in monkey IT is far more robust and flexible than previously thought.

## REFERENCES


Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561–580. doi:10.1080/09541440802097951

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology (2006)*, *64*(5), 1021–40. doi:10.1080/17470218.2010.538972

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: transfer across space and time. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 195–202. doi:10.1037/0096-1523.35.1.195

# CHAPTER VII

# FAMILIARITY EFFECT IN MONKEY AREAS V2 & V4

## INTRODUCTION

Response properties of neurons in monkey inferotemporal cortex (IT) undergo changes when monkeys are provided long-term experience with visual stimuli (Baker et al., 2002; Cox & DiCarlo, 2008; Freedman et al., 2006; Kobatake et al., 1998; Sigala & Logothetis, 2002). Even when the monkey is not explicitly required to use the long-term experience to make a behavioral response, these changes occur and persist in IT. Two particular interesting examples involve extensive passive exposure to regular events in the environment, upon which IT neurons show responses modulated by the regularities. First, if monkeys are exposed repeatedly to a set of images such that they are rendered familiar over days, responses in IT to these familiar images is lower and sharply truncated after an initial peak response, compared to responses to novel images. This has been extensively documented in IT in a variety of studies and is broadly called the familiarity effect (Anderson & Sheinberg, 2008; Meyer, Walker, et al., 2014; Miller et al., 1993; Mruczek & Sheinberg, 2007b; Peissig et al., 2007; Woloszyn & Sheinberg, 2012b). Second, if monkeys are exposed repeatedly to a sequential pair of images where the leading image in the sequence is always followed by the trailing image such that the display of the trailing image is predicted by the leading image, response of IT neurons to such sequentially predicted images is suppressed compared to the responses to the same images occurring in an unpredicted context (Meyer & Olson, 2011; Meyer, Ramachandran, et al., 2014). This phenomenon whereby sequential regularities are learned and represented in IT is called

prediction suppression. IT is thus exquisitely sensitive to the statistics of the environment and represents frequently seen images and predictable images in special ways. These phenomena are broadly called 'statistical learning'.

Statistical learning phenomena in IT have been demonstrated to be modality-specific; i.e., the frequency and predictability of just visual stimuli are represented in IT (Chapter IV). IT is at the terminus of the ventral visual stream in monkeys and inherits its properties from earlier visual areas (Gross et al., 1969, 1972; Ungerleider & Mishkin, 1982). The substrate for statistical learning should have originated in neural projections originating in earlier visual areas like V1, V2 and V4 that feed into IT sequentially and directly. Thus a very pertinent question at this juncture is whether statistical learning also occurs in earlier areas of the visual cortex.

Theoretical frameworks of predictive coding suggest that perception is essentially a process of predictive inference occurring at each stage of processing in the visual hierarchy, where predictions are generated based on learned regularities and compared with incoming visual input (Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999). Thus the hypothesis that statistical learning takes place in earlier visual areas fits in very nicely under this theoretical framework. Preliminary evidence indicates that neurons in V1 are in fact modulated by stimulus predictability in many model organisms. Visual stimuli evoke smaller BOLD responses in human V1 when they can be predicted in a context (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010). Rodent V1 is also sensitive to regularities: neurons in mouse V1 show increased responses for familiar stimuli and predictable sequences of stimuli (Cooke & Bear, 2012; Frenkel et al., 2006; Gavornik & Bear, 2014).

In this study, we sought to investigate whether single neurons in monkey V2 and V4 are sensitive to statistical regularities imposed over long-term training. In particular, we used the familiarity response observed in IT (a marker of the frequency of exposure to a particular stimulus) as a measure. We exposed a monkey to a set of images over days and subsequently, recorded from single units in monkey visual areas V2 and V4. We discovered that the responses in these areas to familiarized images were sharply truncated after an initial burst of spikes, just like we observe in IT. Thus, we have preliminary data suggesting that statistical learning is manifested at the level of single neurons in areas outside IT.

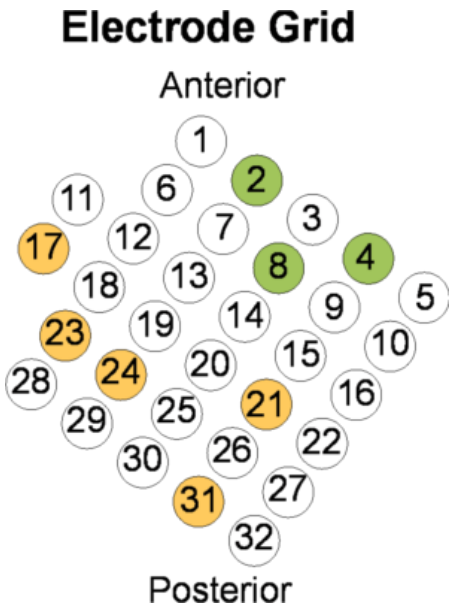


***Fig. 7.1: 32-channel semi-chronic recording array*** The array consisted of 32 electrodes each of which could be independently manipulated. Colored circles indicate channels from which we got consistent recordings during the course of the experiment.

## METHODS

### Subjects

The data in this study is from one adult rhesus macaque: monkey 3 (male; laboratory designation Ga). Procedures were in accordance with guidelines set forth by the United States Public Health Service Guide for the Care and Use of Laboratory Animals and were approved by the Carnegie Mellon University IACUC.

### Recording array

The monkey used in this study had been previously surgically implanted with a 32-channel semi-chronic recording array with multichannel micromanipulators (Gray, Goodell, & Lear, 2007) over right visual cortex and was participating in another experiment when this study began

(Fig. 7.1). The array placed over cortex activated by stimuli in the lower left visual field (roughly centered at -3°,-3°). The advantage of this recording system was that different electrodes could penetrate cortex to different lengths. Thus, it was possible to simultaneously and semi-chronically record from multiple visual areas. When we started the study, 8 channels in the array provided consistent single-unit responses. After mapping the receptive fields of these units, we classified 3 channels as V4 units and 5 as V2 units (Fig. 7.2). Responses from these channels had fairly consistent selectivity profiles over the 12 days we performed this study.

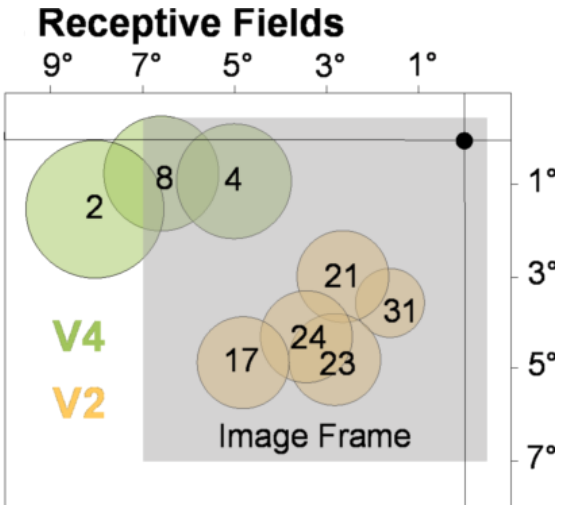


***Fig. 7.2: Receptive fields of some recorded neurons*** Receptive fields (RFs) of single neurons from eight channels were mapped prior to the experiment. We isolated three3 V4 units (green) and five V2 units (orange). The images were places covering most of the RFs and the fovea (gray square)

**Training**

The monkey was trained to fixate prior to the start of the study. For the duration of the study, we exposed out monkey to 25 images ('familiar images', Fig. 7.3A). All stimuli were digitized images of background-free objects. When presented on an LCD monitor 57 cm from the monkey's eyes, each image subtended 7.4° of visual angle along whichever axis, vertical or horizontal, was longer. Images were displayed such that they covered the receptive fields of the neurons in


***Fig. 7.3: Images and trial timing*** The stimulus set used for familiarization consisted of 25 digitized images. Each image was presented for 400ms in the lower left quadrant while the monkey maintained fixation on fovea for the duration of the trial.

channels where we obtained consistent responses prior to starting the experiment (lower left visual field, centered on (-3.2°, -3.2°)) as well as the fovea. The monkey was rewarded for maintaining fixation on a small dot at the fovea (that was present even when the image was displayed) for the duration of the trial. Thus each trial consisted of 200ms fixation spot display, followed by 400ms of image display at a location that drove the neurons under study while the fixation spot persisted, followed by 200ms of fixation spot display. We trained the monkey on familiar images for 12 days. On each training day, the monkey was exposed to one or two 'runs' of familiar images. Each run consisted of 15 blocks. Within each block, each of 25 familiar images was presented once in a pseudoramdom order. The entire run thus consisted of 375 trials.

**Testing**

On a subset of the 12 days (i.e., days 3, 8, 9, 10, 11 and 12), we presented the monkey with novel images (that she had never seen before in any context) in addition to familiar images. Each testing run consisted of 25 familiar images and 25 novel images, displayed 15 times each. Familiar and novel images were interleaved randomly in 15 consecutive blocks of 50 trials each.The entire run thus consisted of 750 trials.

**Recording**

On each day, prior to the start of training, we connected the array to the recording system (Cerebus neural Signal processing System, Blackrock Microsystems) and tested each of the 32 channels for spiking activity on the electrodes at a fixed depth. We recorded from channels where single units were isolated on the electrodes on each day. Each electrode could be independently advanced, although for the duration of the experiment, electrodes where we observed consistent responses prior to starting the experiment were not advanced. On each day, we isolated between 2 and 11 neurons from 2-8 channels. We classified a neuron as visually responsive if, for either the leading or the trailing image, the mean firing rate in a window 0-200 ms following image onset exceeded the mean firing rate in computed 100 ms before image onset (one-tailed t-test, $\alpha = 0.05$).

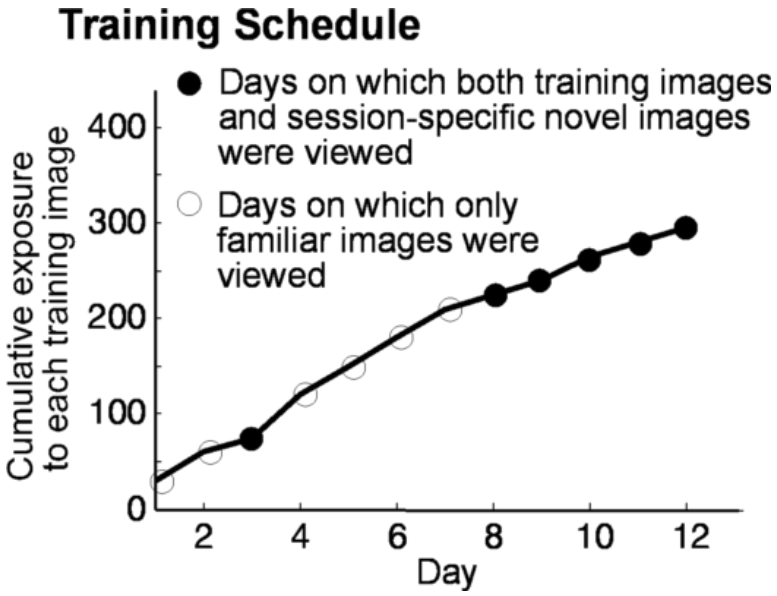


***Fig. 7.4: Training schedule*** Monkeys were exposed to the images in Fig. 7.3 over days. Cumulative exposure to familiar images is plotted over days. On a subset of days (filled black circles) session-specific novel images were also displayed.

**RESULTS**

The experimental paradigm consisted of passively exposing monkeys to 25 images ('familiar images') day after day for 12 consecutive days (Fig. 7.3). Simultaneously, we recorded the activity of V2 and V4 neurons on each day. On a subset of days (day 3, 8, 9, 10, 11, 12), we

also displayed a set of session-specific ‘novel images’ that the monkey had never seen before (not shown). The time course of training is summarized in Fig. 7.4.

On each of the 12 days, the familiar images were presented during the course of 1-2 runs, either alone, or interleaved with session-specific novel images. On each run, the monkey was exposed to each of the familiar images 15 times. By Day 3 of the experiment, they monkey had been exposed to each familiar image 90 times. By Day 8, the cumulative number of exposures to familiar images had increased to 225. Over all 12 days, the monkey was exposed to each familiar image 295 times. Novel images on the other hand were session specific – a new set was used every day. Each novel image was presented 15 times during data collection on each day.

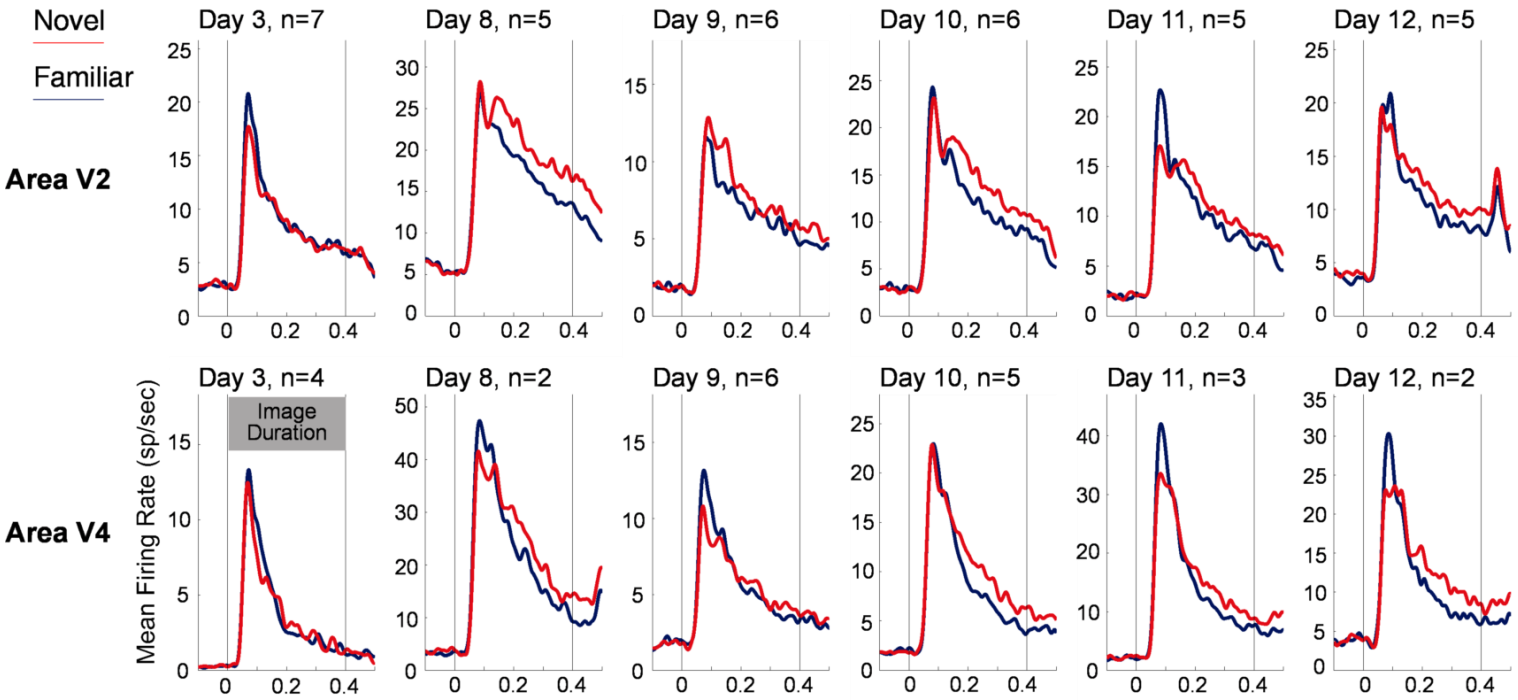


***Fig. 7.5: Neuronal responses for familiar and novel images over days in monkey areas V2 and V4.*** Responses of single neurons in areas V2 (top row) and V4 (bottom row) recorded on training days (columns – Day 3, 8-12) for novel (red) and familiar (blue) images. 'n' demotes the number of neurons recorded on each day in that area. No differences between responses to familiar and novel images were observed on Day 3. Consistent trends towards differences (early facilitation and late truncation of responses to familiar images) were observed from days 8-12 on each day and significant in the population (Fig. 7.6). Stats summarized in Table 7.1.

We obtained consistent single unit recordings from 8 channels, whose receptive fields are plotted in Fig. 7.2. Sometimes, we were able to spike-sort more than one neuron from the same channel. Sometimes, we got single units on other channels too, and based on their receptive fields, we classified them as V2 or V4 neurons. Thus there was natural variability in the exact number of neurons we were able to isolate each day. We were able to record between 5-7 V2 neurons (mean: 5.6 neurons/day) and 2-6 V4 neurons (mean: 3.6 neurons/day) on each of the 6 recording days when novel images were displayed.

On Day 3 of the experiment (Fig. 7.5, Day 3), at 90 exposures to familiar images, we compared the responses to the familiar images to the response to novel images. No differences were observed between the responses to familiar and novel images on Day 3, either in V2 ($p=0.6118$,

two-tailed t-test, 0-400ms of image exposure period, n=7) or in V4 (p=0.6424, two-tailed t-test, 0-400ms of image exposure period, n=4) for the entire duration of the response.

After training without the display of novel images on days 4-7, on Day 8 of the experiment (Fig. 7.5, Day 8) at 225 familiar exposures, we compared the responses to the by now highly familiar images to the response to novel images. We observed an initial facilitated peak response for the familiar images (more pronounced in V4 than in V2, approaching significance with p=0.0817, two-tailed t-test, 0-100ms of image exposure period, n=2). Later in the trial, we observed that the response to the familiar image was sharply truncated. This effect was prominent both in V2 (p=0.0551, two-tailed t-test, 0-400ms of image exposure period, n=5) (Fig. 7.5, Day 8, V2) as well as

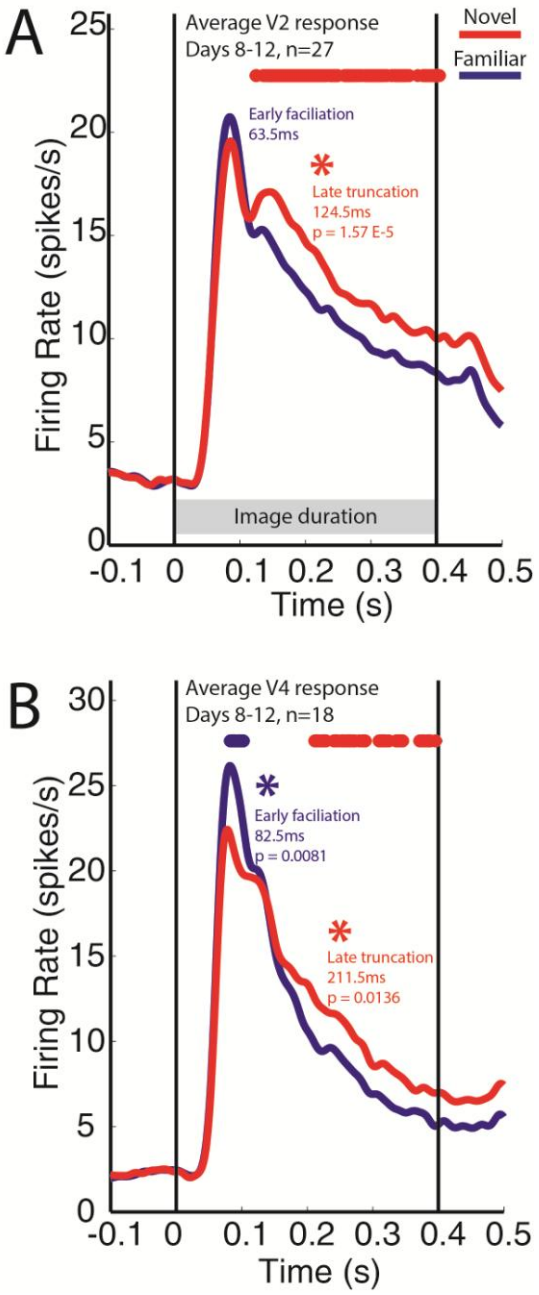


***Fig. 7.6: Averaged population neuronal responses for familiar and novel images over days in monkey areas V2 and V4.*** Averaged responses of single neurons in (A) area V2 and (B) area V4 averaged over neurons recorded on days 3-8 for novel (red) and familiar (blue) images. Early facilitation and late truncation of responses to the long-term familiarized images were observed, indicated by blue circles and red circles respectively. The circles represent 10ms-windows where the responses to the familiar images were significantly greater than (early facilitation, blue circles) or lesser than (late truncation, red circles) the responses to the novel images in a two-tailed t-test (□=0.01). Latency of the facilitation and truncation responses were calculated on the basis of the same analysis (i.e., the first significant 10ms-window). Stats and latency analysis (Table 7.2) summarized in Table 7.1 and 7.2 respectively.

V4 (p=0.3284, two-tailed t-test, 0-400ms of image exposure period, n=2) (Fig. 7.5, Day 8, V4). Thus there were trends towards two key findings: early facilitation and late truncation of responses for familiar images. These trends were consistent in both areas on subsequent days as well (Table 7.1).

While the neurons we obtained on each day are of excellent quantity, we were able to record only from a few neurons each day and thus lack statistical power. In order to compensate for the lack of statistical power, we combined the responses of all V2 neurons recorded on Days 8-12 as similar phenomena are consistently observed on all days (n=27). Similarly, we combined the responses of all V4 neurons recorded on Days 8-12 (n=18) (Fig. 7.6). For these populations, for each neuron, we computed the mean firing rate for the familiar condition and the mean firing rate for the novel condition over the duration of image display (0-400ms). We then compared the mean firing rates of the familiar and novel conditions in the V2 population and V4 population separately.

In both populations, we observed both early facilitation and late truncation of responses for familiar images (Fig. 7.6). When we compared just the early periods of the response (0-100ms), we observed strongly significant early facilitation of familiar responses in the population of V4 neurons (n=18, p=0.0081, two-tailed t-test, 0-100ms of image exposure period).This effect was present, but did not reach significance in the population of V2 neurons (n=27, p=0.1496, two-tailed t-test, 0-100ms of image exposure period). In the late period of the response (100-400ms), we observed truncation of responses for familiar images both in V4 (n=18, p=0.0136, two-tailed t-test, 100-400ms of image exposure period) and V2 (n=27, p=1.57 E-5, two-tailed t-test, 100-400ms of image exposure period) populations. We calculated the latency of onset of both the effects in both populations by performing a two-tailed t-test on 10ms

windows stepped every ms between familiar and novel image responses. Results are summarized in Table 7.2.

| | V2 | | | V4 | | |
|---|---|---|---|---|---|---|
| Days | Number of neurons | **Early facilitation response** p-values (0-100ms of image display) | **Late truncation response** p-values (100-400ms of image display) | Number of neurons | **Early facilitation response** p-values (0-100ms of image display) | **Late truncation response** p-values (100-400ms of image display) |
| 3 | 7 | 0.2944 | 0.8440 | 4 | 0.5201 | 0.8640 |
| 8 | 5 | 0.7349 | **0.0428** | 2 | 0.0817 | 0.2267 |
| 9 | 6 | 0.8056 | 0.1018 | 6 | 0.1745 | 0.7845 |
| 10 | 6 | 0.4236 | 0.0513 | 5 | 0.3573 | 0.2055 |
| 11 | 5 | 0.1933 | **0.0049** | 3 | 0.2149 | 0.2297 |
| 12 | 5 | 0.8382 | 0.0806 | 2 | 0.5819 | 0.5186 |
| Days 8-12 | 27 | 0.1496 | **1.57 E-5** | 18 | **0.0081** | **0.0136** |

***Table 7.1: Statistical analyses on V2 and V4 by day*** *P-values in two-tailed t-tests conducted between familiar and novel conditions on the (a) 0-100ms early response and (b) 100-400ms late response across neurons by area. Significant p-values (α=0.05) are in bold type. P-values were also calculated for the same conditions over the population of V2 and V4 neurons (all neurons from day 8-12).*

| | **Onset of visual response** | **Onset of early facilitation response** | **Onset of late truncation response** |
|---|---|---|---|
| **V4** | 46.5ms | 82.5ms | 211.5ms |
| **V2** | 50.5ms | 63.5ms* | 124.5ms |

***Table 7.2: Latency analysis*** *Latency of visual response onset in the V4 population (n=18) and V2 population (n=27) was calculated based on a one-tailed t-test between a 100ms-window response before novel image onset and 10ms-window responses stepped every ms after novel image onset (α=0.01). Latency of early facilitation response onset and late truncation response onset were calculated based on a two-tailed t-test between responses to familiar and novel images in 10ms windows stepped every ms (α=0.01).*

**Significant early facilitation was observed from 63.5ms -70.5ms in V2 neurons when α=0.05.*

## DISCUSSION

Neurons in monkey inferotemporal cortex are strongly modulated by regularities in the sensory environment on the basis on long-term experience. Here, we present evidence that similar phenomena can be observed in areas upstream of IT. When monkeys are provided extensive exposure (over days) to images such that they are rendered familiar, neurons in areas V2 and V4 come to represent familiar images differently from images the monkey sees for the first time. We observed two important components in the effect: an early facilitation in the response to familiar images, followed by a subsequent truncation of the same responses. While the truncation effect is consistently observed after a few days of training both in V2 and V4 and is significant in the population, the facilitatory early enhancement was significant in the V4 population with strong trends towards the effect in the V2 population. The observed effects are compatible with familiarity effects described in IT. This is the first demonstration that such effects can be elicited in areas outside V1, and that these monkey areas are sensitive to frequency statistics.

The response to familiar images seems to be made up of two components: early facilitation and late truncation. In the next section we speculate on the functional roles these physiological components could play based on the available evidence. The early facilitation is essentially a burst of stimulus-selective spikes. It is possible that familiarization improves the representation of the stimuli either by facilitation – representing the familiar image faster by reducing the neural processing time (Sobotka & Ringo, 1996), or by sharpening – where the representation of the image in the population becomes sparser (Desimone, 1996; Miller et al., 1993) , or by a combination of both these mechanisms. The effect could reflect network dynamics: upon encountering a familiar image, it is represented strongly by a sparse population

of neurons that are selective for the image, that inhibit themselves and other non-selective neurons in the population resulting in truncated responses a few tens of milliseconds later. This could save metabolic energy, paving the way for faster processing and/or reduce salience for familiar, potentially boring stimuli. Such selective facilitation of responses for familiar images has been observed and is thought to put neurons 'in a state of readiness' to respond to surprising stimuli (Meyer, Walker et al., 2014). Further, there is evidence that separate populations of neurons may contribute to each of these effects: excitatory neurons in IT show facilitated responses to familiar images, whereas inhibitory neurons in IT show suppressed responses to familiar images in the early period (Woloszyn & Sheinberg, 2012b). It remains to be tested where excitatory and inhibitory neurons show such functional asymmetry in V2 and V4 as well.

We have provided the first neurophysiological evidence in monkeys that statistical learning is not the special property of IT; it is to be found even in very early visual cortex. The functional role of these computations and their influence on statistical learning in IT need to be studied further. In particular, it would be very interesting to test whether prediction suppression is also present in earlier visual areas.

## REFERENCES


Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(8), 2960–6. doi:10.1523/JNEUROSCI.3730-10.2010

Anderson, B., & Sheinberg, D. L. (2008). Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia*, *46*(4), 947–57. doi:10.1016/j.neuropsychologia.2007.11.025

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, *5*(11), 1210–6. doi:10.1038/nn960

Cooke, S. F., & Bear, M. F. (2012). Stimulus-selective response plasticity in the visual cortex: an assay for the assessment of pathophysiology and treatment of cognitive impairment associated with psychiatric disorders. *Biological Psychiatry*, *71*(6), 487–95. doi:10.1016/j.biopsych.2011.09.006

Cox, D. D., & DiCarlo, J. J. (2008). Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(40), 10045–55. doi:10.1523/JNEUROSCI.2142-08.2008

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13494–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=33636&tool=pmcentrez&rendertype=abstract

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(11), 1631–44. doi:10.1093/cercor/bhj100

Frenkel, M. Y., Sawtell, N. B., Diogo, A. C. M., Yoon, B., Neve, R. L., & Bear, M. F. (2006). Instructive effect of visual experience in mouse visual cortex. *Neuron*, *51*, 339–349. doi:10.1016/j.neuron.2006.06.026

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, *17*(5), 732–7. doi:10.1038/nn.3683

Gray, C. M., Goodell, B., & Lear, A. (2007). Multichannel micromanipulator and chamber system for recording multineuronal activity in alert, non-human primates. *Journal of Neurophysiology*, *98*(1), 527–36. doi:10.1152/jn.00259.2007

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science (New York, N.Y.)*, *166*(3910), 1303–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4982685

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4621506

Kobatake, E., Wang, G., Tanaka, K. (1998). Effects of Shape-Discrimination Training on the Selectivity of Inferotemporal Cells in Adult Monkeys. *Journal of Neurophysiology*, 324–330.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12868647

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Meyer, T., Walker, C., Cho, R. Y., & Olson, C. R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience*, *17*(10), 1388–1394. doi:10.1038/nn.3794

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *13*(4), 1460–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8463829

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(11), 2825–36. doi:10.1523/JNEUROSCI.4102-06.2007

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251.

Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(6), 1323–34. doi:10.1093/cercor/bhl043

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–20. doi:10.1038/415318a

Sobotka, S., & Ringo, J. L. (1996). Mnemonic Responses of Single Units Recorded from Monkey Inferotemporal Cortex , Accessed via Transcommissural Versus Direct Pathways : A Dissociation between Unit Activity and Behavior, *16*(13), 4222–4230.

Ungerleider, L. G., & Mishkin, M. (1982). Two Cortical Visual Streams. In *Analysis of Visual Behavior* (pp. 549–586).

Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*(1), 193–205. doi:10.1016/j.neuron.2012.01.032

# DISCUSSION

## 1. Summary of results

Humans and non-human primates have the ability to learn statistically regular events in the environment through passive exposure, a phenomenon called statistical learning. This has been demonstrated in multiple sensory domains, in both static scenes and in sequential stimuli (Aslin et al., 1998; Fiser & Aslin, 2001; Fiser & Aslin, 2002a; Hauser et al., 2001; Saffran et al., 1996b, 1999; Saffran, 2003b; Toro & Trobalón, 2005). This project sought to uncover the neuronal mechanisms underlying statistical learning, in particular the learning of visual sequences. Our lab recently discovered a phenomenon called statistical learning in monkey inferotemporal cortex. Upon repeated exposure over weeks to pairs of visual images in fixed sequence, such that the leading image in the sequence predicts the occurrence of the trailing image, neurons in the monkey inferotemporal cortex show suppressed responses for images when they occur in a predicted context but not when the same images occur in an unpredicted context (Meyer & Olson, 2011a). Subsequently, we have discovered the following features about the prediction effect.

1. The prediction effect is modulated by the conditional probabilities between the images (Ramachandran, Meyer and Olson, (under review at Journal of Neuroscience)).
2. The prediction effect is due to suppression of responses to predicted images, suggesting that inhibition of responses to the predicted image may be instrumental for the effect.
3. The prediction effect is modality-specific, i.e., neurons in IT show the prediction effect only for visual-visual pairs and not for cross-modal pairs of stimuli.

4. The prediction effect can be induced by training monkeys not only on pairs of images, but also on longer sequences (Meyer, Ramachandran, et al., 2014).
5. The prediction effect persists even with long delays between the leading and trailing image, and is observed even if the delay between the images was different from the trained delay.

In addition, we performed an experiment to ask whether statistical learning is present in areas outside of IT. We tested this using the familiarity effect as a measure of statistical learning (Anderson, Mruczek, Kawasaki, & Sheinberg, 2008b; Lin Li et al., 1993; Meyer, Walker, et al., 2014; Mruczek & Sheinberg, 2007b; Peissig et al., 2007; Woloszyn & Sheinberg, 2012b), in areas V2 and V4 of the monkey visual cortex. We observed that the familiarity effect was present in both areas V2 and V4, suggesting that statistical learning is not confined to IT. In fact, this study suggests that regularities in the sensory environment may be learned by neurons in all the hierarchical areas along the ventral visual stream. Thus this discovery may pave way for directly testing hypotheses suggested by predictive coding models (Lee & Mumford, 2003; Rao & Ballard, 1999).

**2. Is statistical learning truly implicit?**

The idea that regularities in a stimulus set can be learned implicitly is pervasive (Lewicki, Hill, & Czyzewska, 1992; Perruchet & Pacton, 2006). Artificial grammar learning paradigms (Conway & Christiansen, 2005; Conway & Pisoni, 2008; Reber, 1967) are early examples where human subjects, upon exposure to many 'example' sequences conforming to a particular artificial grammar that governs the sequential transitions, implicitly learn the rules. These

paradigms and statistical learning paradigms provide no explicit supervision or feedback, and are hence thought to be implicit. But does statistical learning truly proceed in an implicit manner?

A recent study (Kim, Seitz, Feenstra, & Shams, 2009) shows that implicit statistical learning did take place, even in the absence of explicit learning. After 5 minutes of exposure to sets of predictable triplets, subjects were presented with a target stimulus from the set of stimuli used for training. They were then randomly presented with all the triplets they were trained on in a rapid serial visual presentation paradigm, and asked to press a key when the target stimulus came on. This was a measure of implicit learning. In a consecutive experiment, they were presented with one stimulus from a triplet set and asked to pick out which two of the other stimuli were associated with that stimulus, as a measure of explicit recall. Even when subjects could not recall which stimuli had occurred together, they were much faster at responding to the second and third stimuli in a sequence in the RSVP test (which occurred predictably) compared to the first stimulus (which was not predictable in its occurrence). This is argued as evidence that statistical learning is acquired implicitly, even in the absence of explicit learning.

A number of other studies show that implicit learning takes place only when relevant, predictive information is selectively attended to (Baker et al., 2004; Jiang & Chun, 2001; Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne, Jungé, & Scholl, 2005). Some of these results do not preclude the occurrence of implicit learning, but add that selective attention gates what is implicitly learned.

In our paradigm, we only require the monkeys to maintain fixation. We also train our monkeys over weeks, while most statistical learning paradigms in humans involve exposure over only a few minutes. So we do not know if they pay attention to the images, or if they explicitly

learn the associations. It is also possible that attention is required for initial learning but not for the manifestation of the neuronal effects.

### 3. Duration of exposure training: hours or days?

A survey of statistical learning paradigms in humans reveals that they are generally conducted with very short visual experience, often on the scale of minutes to the scale of hours: 2 min (Saffran et al., 1996c), 3 min (Aslin et al., 1998; Saffran & Griepentrog, 2001), 5 min ( Kim et al., 2009), 7 min (Fiser & Aslin, 2001), 21 minutes (Newport & Aslin, 2004) etc. However there is evidence that statistical learning consolidates during sleep, making a case for more powerful learning with long-term exposure to the regularities (Durrant, Taylor, Cairney, & Lewis, 2011).

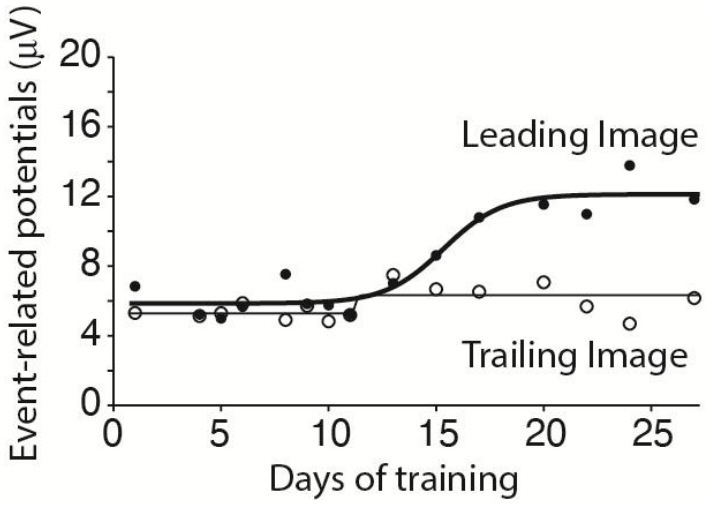


***Fig. 8.1: Time-course of learning (unpublished results)*** *Event-related potentials (ERPs) measured on each training day over multiple sessions were averaged and plotted separately for the set of leading images and the set of trailing images. After day 10 (200 exposures to each sequence), differences between the increase in response to the leading and trailing image are observed, suggesting onset of suppression.*

We opted for long-term learning (over weeks) for our experiments. Since we did not measure behavior, we do not know whether it took the monkey that long to learn the associations. However, during the course of training, we did measure event-related potentials in IT by lowering a blunt electrode to the skull and retracting a few microns. The data we acquired

for the trained sequences (Fig. 8.1) shows baseline responses for both leading and trailing images for the first 10 days (corresponding to about 200 exposures). However around day 10, the response to the leading image increases sharply in a sinusoidal manner, and the increase in response is maintained till the end of training. Such enhanced ERP responses to familiar images has been reported before (Peissig et al., 2007). Around the same time, the responses to the trailing images also increase, but not as much as the responses to the leading images. Thus around day 10, ERPs start showing consistently suppressed responses to predicted trailing images, suggesting the onset of the neuronal effect around that time. Similarly, when we measured the difference between familiar and novel images over days in monkey V2 and V4, the effect was not apparent on day 3 (at about 75 exposures) but was clearly present starting from day 8 onwards (after about 150 exposures). Thus while statistical learning as measured by some behavioral read-outs seems to manifest in humans within a few minutes of training, it takes multiple days to manifest at the neuronal level in monkeys.

4. **How long does statistical learning persist in the system?** Statistical learning seems to proceed efffortlessly and implicitly. Furthur, it persists in the system for a long time, making statistical learning a highly robust mechanism. Behaviorally, implicit statistical learning of artificial grammar has been shown to persist for at least 2 years (Allen & Reber, 1980). Physiologically, we observe that prediction suppression in IT at the level of LFPs was also present 20 months years after initial training, with absolutely no exposure to the tested pairs in the interim. Fig.8.2 shows the prediction effect at the level of LFPs for a set of six stimulus pairs. Training was conducted for three months in early 2010, and data acquisition proceeded till July 2010. Data from the initial recording phase (mid 2010) is shown in Fig. 8.2A. Responses to unpredicted stimuli (red) are greater than responses to predicted images (blue). The monkey was not exposed to these stimulus pairs for almost 2 years. In March 2012 while performing another experiment, we recorded local field potentials for these pairs at

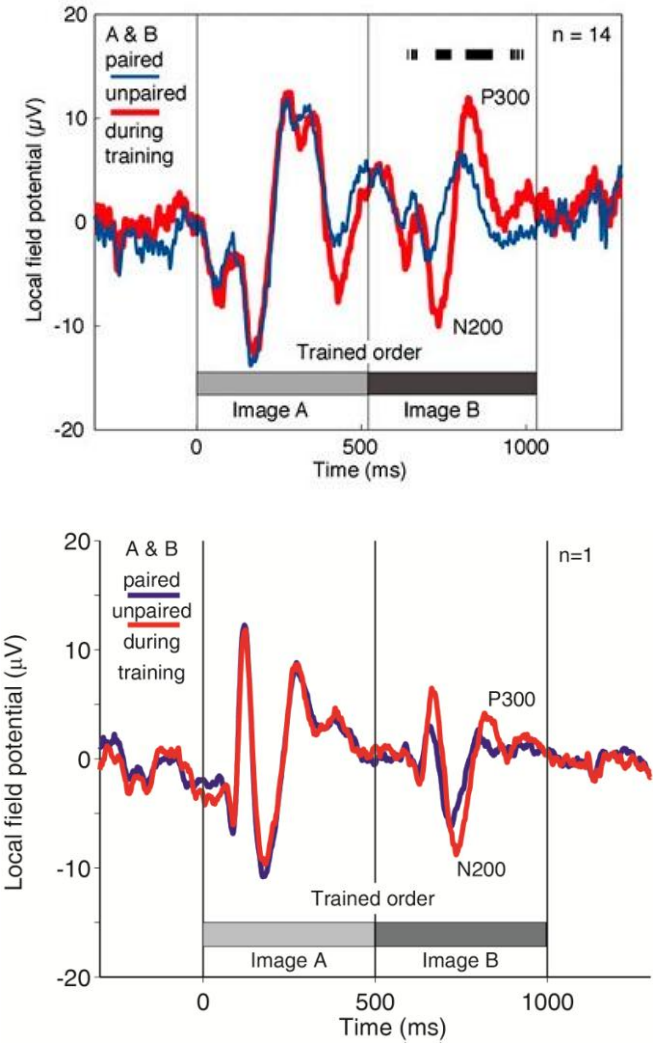


***Fig. 8.2: Prediction suppression persists in the system for at least 20 months (unpublished results)*** *(A) Prediction suppression at the level of LFPs (n=14 sites) measured in 2010. Responses to unpredicted trailing images (red) was greater than responses to trailing images (blue) with characteristic negative and positive deflections (N200 and P300 respectively).(B) Prediction suppression persists in IT with the same characteristics 20 months later with no exposure to the images in the interim.*

one site in one monkey. We observed the persistence of prediction suppression (unpredicted (red) greater than predicted (blue)) in IT even after 20 months (Fig. 8.2B).

The stimulus-specific response potentiation effect in rodents (SRP), a phenomenon similar to the familiarity effect in macaques, has been shown to persist over weeks without any sign of degradation (Cooke & Bear, 2012; Frenkel et al., 2006).

## 5. Speculations on mechanisms underlying the prediction effect in IT

The following section makes some remarks on mechanisms that may underlie the prediction effect.

### a) Relation to repetition suppression

Prediction suppression is very similar to another phenomenon widely documented in human and monkey temporal cortex, called repetition suppression, that seems to indicate sensitivity to short-term stimulus history. When a visual stimulus is shown repeatedly, the electrophysiological and BOLD responses in human temporal cortex and single unit activity in monkey inferotemporal cortex neurons are reduced for repeated displays of the stimulus compared to the response to the initial display of the stimulus(Grill-Spector et al., 2006; Liu, Murray, & Jagadeesh, 2009; McMahon & Olson, 2007; Miller, Li, & Desimone, 1991; Miller & Desimone, 1994). This effect is stimulus specific (Sawamura, Orban, & Vogels, 2006), and, contrary to prediction suppression, proceeds in the absence of any long-term training. However, the core effect is still stimulus-specific response suppression. So it is likely that repetition suppression and prediction suppression may share common core mechanisms.

Initially, repetition suppression was thought to be no different from adaptation, and was thought to be caused by the biophysical properties of neuronal discharge. For example, the repeated stimulus may occur during a tonic hyperpolarization phase that follows the initial presentation (Carandini & Ferster, 1997; Summerfield, Wyart, Johnen, & de Gardelle, 2011). However, if that is the case, then any stimulus (eg. Image B) that elicits the same response from a neuron as the initial image (eg. Image A), when presented second (i.e., the trial A-B), should elicit the same amount of suppression irrespective of image identity (as in the trial A-A). When this hypothesis was tested (Sawamura et al., 2006), it was revealed that repetition suppression is stimulus specific: suppression is far greater when the same image is presented a second time sequentially (trial A-A) than when a different image eliciting the same initial response from the neuron is presented second (trial A-B). Prediction suppression, too, is not simply adaptation. Both predicted and unpredicted responses are less than the response to the image leading the sequence, which may be due to adaptation (Fig. 1.3). However, adaptation-related response reduction apart, the responses to the same images are far lesser when predicted than when unpredicted.

Further there is evidence that repetition suppression is by itself modulated by predictability of the repetition (Summerfield & Egner, 2009; Summerfield et al., 2008, 2011). Repetition suppression is relieved in contexts where the repetition of a stimulus is rare, and therefore surprising. This suggests that repetition suppression may be a specific case of (local, stimulus-specific) prediction or an associated neurobiological phenomenon where ‘expected’ events are represented less strongly than surprising ones. Prediction

suppression may be a global, stimulus specific, and a more robust form of encoding predictions.

This hypothesis can be explicitly tested by training monkeys long-term on trials where an image is always followed by the same image (eg., the sequence A-A) (repetition block) and trials where an image is always followed by a particular, different image (eg., the sequence C-D) (prediction block). We could then ask whether training in a 'prediction' context relieves repetition suppression when an image in the prediction block is followed by itself (C-C), instead of its paired associate trailing image (C-D).Further, we could ask how repetition suppression (D-D) fares against prediction suppression (C-D). It should be noted that a similar study in monkey IT was carried out recently (Kaliukhovich & Vogels, 2011) and found that prediction does not modulate repetition suppression. However, they did not train monkeys long-term like we do; the results were based on predictive contexts specified just over a recording session. It would be interesting to test this hypothesis in the context of long-term training and prediction suppression. This experiment would clarify whether repetition suppression and prediction suppression are served by similar mechanisms.

Current ideas about the mechanisms underlying repetition suppression may be informative in thinking about mechanisms underlying prediction suppression. In particular, three mechanisms have been proposed (Grill-Spector et al., 2006): (i) the **fatigue model**, where the amplitude of each neuron's response decreases with repetition (Miller & Desimone, 1994) (ii) the **sharpening model**, a model of sparse representations where fewer neurons, mostly ones responding selectively to the image, respond and other neurons are silenced (Desimone, 1996; Miller et al., 1993)and

(iii) the **facilitation model**, where the duration of neural processing for repeated stimuli is shortened (Sobotka & Ringo, 1996). Each of these models can be explained on the basis of various cellular-level and synaptic mechanisms (Grill-Spector et al., 2006). Similar models can be invoked to explain prediction suppression.

**b) Suppression: the role of inhibition**

Since these phenomenon hinges on suppression, stimulus-selective inhibition of responses has been suggested to play a role in shaping the neural response (Norman & O'Reilly, 2003). This can be tested by using GABA antagonists during testing and observing whether prediction suppression persists.

**c) Cellular and molecular mechanisms revealed in rodent studies**

A very interesting observation in recent years is that statistical learning has been observed at the neurophysiological level in rodents. When mice were exposed over days to a grating stimulus of a particular orientation, it was observed that visual evoked potentials in V1 consistently increased over training, specifically to the trained orientation. This effect was called stimulus-specific response potentiation (SRP) ((Cooke & Bear, 2012; Frenkel et al., 2006). This is a familiarity effect that is stimulus-specific, similar to our observations in the macaque. More recently, effects of long-term training on sequences of gratings have been demonstrated in rodent V1 (Gavornik & Bear, 2014). Extensive exposure to oriented gratings in fixed sequence induced increased evoked potential responses only for the learned sequences. The SRP effect was shown to be eye-specific (when stimuli were presented only to one eye, the effect did not transfer to the other eye, suggesting that plasticity takes place within layer 4, at thalamo-cortical synapses. Further, these phenomena have been shown to be dependent on NMDA

dependent LTP (Beste & Dinse, 2013; Cooke & Bear, 2012, 2014; Frenkel et al., 2006; Gavornik & Bear, 2014).

LTP has been demonstrated to be long-lasting, over a period of many months (Abraham, Logan, Greenwood, & Dragunow, 2002). This is consistent with our observations that statistical learning in IT persists over months. Based on our observations in the experiment testing the role of conditional probability on prediction suppression, it is likely that synaptic strength scales with the fidelity of association between stimuli, a mechanism that can be effected by LTP and LTD. Thus it is a reasonable hypothesis that LTP in IT plays a role in prediction suppression. The role of LTP in maintaining statistical learning in the form of prediction suppression in monkeys can potentially be tested by introducing NMDA blockers in IT before testing prediction suppression.

## 6. Proposed future experiments

### a) Prediction suppression and cognitive expectation: the need for behavioral correlates

One important question our studies do not address is whether prediction suppression has behavioral relevance. Humans can acquire these associations within a few sessions of training, but it takes at least 10 days or 200 exposures to a pair for the neuronal phenomena to develop in monkeys. These observations raise a number of questions.

i. Do monkeys cognitively 'expect' the display of a paired trailing image? Is prediction suppression correlated with the behavior?

ii. How long to monkeys take to learn these paired associates only on the basis of passive training? It would be necessary to define an appropriate behavioral read-out. As observed

in the human studies, it may be the case that explicit recall of a paired stimulus may take longer to manifest while the animals may be faster to recognize a stimulus in sequence within a few exposures (Kim et al., 2009). Is prediction suppression in IT necessary for (a) faster recognition of the paired stimulus (b) recall of the paired stimulus?

iii. If prediction suppression is not necessary for cognitive expectation, recall or recognition memory (and indeed, if these phenomena proceed even before prediction suppression manifests in IT), then what function does prediction suppression serve in IT? It is possible that the effect may be a computation that is a property of IT neurons, taking place below the threshold of consciousness. It may be a marker of long-term plasticity, recording the statistical regularities in the environment accrued over a long time period. A future step could be to devise experiments to disrupt prediction suppression after training, and ask how that affects cortical function and behavior.

iv. Can monkeys pick up structure embedded in an artificial grammar paradigm? In our training paradigm, we presented monkeys with fixed sequences or triplets within individual trials and rewarded at the end of the trial for maintaining fixation. We have not tested whether, if presented with a sequence of stimuli following more complex transitional rules (as in the artificial grammar paradigm), monkeys can pick up on those rules. Whether such transitional rule learning is represented at the level of IT neurons is also not known and is open for testing.

**b) Is attention required for prediction suppression?**

It has been suggested in a number of studies that attention may gate implicit, statistical learning (Jiang & Chun, 2001; Toro et al., 2005; Turk-Browne et al., 2005).

We could test whether attention is necessary for prediction suppression in IT in a future

experiment. One way to test this would be by flashing a salient distractor stimulus in the contralateral visual field when the predicted stimulus is displayed and testing whether prediction suppression is still intact to the same degree.

**c) Investigating the role of inhibition in prediction suppression**

The prediction effect seems to be due to suppressed responses to predicted stimuli. This observation suggests the role of inhibition in bringing about the effect. We can test this by introducing GABA antagonists like bicuculline into IT prior to testing prediction suppression.

**d) Is prediction suppression induced in areas upstream of IT in the ventral visual stream?**

We have evidence that the familiarity effect is present in areas in the ventral visual stream upstream of IT such as V2 and V4. Whether prediction suppression is also induced in these areas is an outstanding question. If so, by measuring the latency of the onset of suppression in each of the areas in the visual system, we can ask whether predictions are propagated back through the ventral visual stream. Further, these experiments could be starting point from where we can probe the role of feedback in the visual system.

**e) Is statistical learning induced in other sensory systems?**

Statistical learning has been extensively demonstrated behaviorally in the auditory system (Hauser et al., 2001; Pelucchi et al., 2009; J R Saffran et al., 1999; Toro & Trobalón, 2005; Winkler, 2007) and even in the somatosensory system (Conway & Christiansen, 2005). It is possible that the auditory and somatosensory cortex as well as other sensory cortices in monkeys are also capable of statistical learning, and may manifest effects such

as the ones we observe. Prediction coding has been suggested to be a fundamental computation in all of cortex (Friston, 2005; Hawkins, 2004). Investigating statistical learning at the level of single neurons in other sensory systems may be the most direct way to test this idea. Further, it may reveal that common computations and mechanisms (Bastos et al., 2012) underlie sensory processing irrespective of modality and may allow us to study those commonalities and differences.

**f) Experiments bridging statistical learning in monkeys and rodents**

Effects similar to statistical learning have been observed in rodents. At the behavioral level, rats are able to parse statistical regularities in syllabic transitions, just like human infants (Toro & Trobalón, 2005). But more pertinently, neural signatures of statistical learning comparable with our observations in monkey IT have been discovered. Upon exposure over days to a grating of a particular orientation, evoked potentials in mouse V1 show increased responses just to that particular stimulus. This phenomenon, called stimulus-specific response potentiation, is induced by extensive familiarity with a particular visual stimulus (Cooke & Bear, 2012, 2014). Similarly, extensive exposure to oriented gratings in fixed sequence induced increased evoked potential responses only for the learned sequences (Gavornik & Bear, 2014). Thus mouse V1 shows sensitivity to regularities learned over a long period, even if the phenomena observed are different from what we observe in the monkey. These observations, taken together, have important implications. Firstly, the natures of these computations in different organisms may have commonalities and differences, and a detailed study of these features might be informative in understanding cortical function. Secondly, these phenomena might be dependent on feedback from higher visual areas. In the mouse, areas outside V1 (i.e.,

'higher' visual areas) have been mapped out only very recently (Andermann, Kerlin, Roumis, Glickfeld, & Reid, 2011; Marshel, Garrett, Nauhaus, & Callaway, 2011; Niell, 2011). The functional roles of these areas are yet to be understood. In this regard, statistical learning phenomena may be used as assays to study feedback from higher areas in the mouse visual cortex. Thirdly, the circuit-level and molecular basis for these computations can be studied thoroughly in a rodent model, unlike in the monkey. Lastly, these studies suggest that the computation of statistical regularities may very well be a common feature of mammalian cortices cutting across phylogeny (Gavornik & Bear, 2014).

**Concluding remarks**

Humans, monkeys and rodents are sensitive to regularities and patterns in sensory stimuli. In particular, they are able to learn the relationships governing regular transitions in sequences of stimuli. We discovered the neural mechanisms underlying a particular instantiation of visual statistical learning in monkey inferotemporal cortex. Neurons in IT show reduced responses to predicted images, a phenomenon we call prediction suppression. In this work, I have presented evidence showing that prediction suppression (i) depends on the conditional probability between the images presented sequentially, (ii) the effect is indeed due to suppressed responses to predicted images, (iii) the effect is modality-specific in IT, (iv) the effect can be induced by training monkeys on longer sequences and (v) the effect persists across long delays between the sequential stimuli. These effects are long-lasting and robust: they persist at least for 20 months after initial training with no exposure to the stimuli in the interim. Further, we present evidence for statistical learning outside IT, in macaque V2 and V4 neurons. These results suggest that

statistical learning is a robust phenomenon in the macaque visual system, and may be a fundamental computation in all sensory systems.

## REFERENCES


Abraham, W. C., Logan, B., Greenwood, J. M., & Dragunow, M. (2002). Induction and Experience-Dependent Consolidation of Stable Long-Term Potentiation Lasting Months in the Hippocampus, *Journal of Neuroscience*, *22*(21), 9626–9634.

Allen, R., & Reber, A. S. (1980). Very long term memory for tacit knowledge. *Cognition*, *8*(2), 175–185. doi:10.1016/0010-0277(80)90011-6

Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L., & Reid, R. C. (2011). Functional specialization of mouse higher visual cortical areas. *Neuron, 72*, 1025–1039. doi:10.1016/j.neuron.2011.11.013

Anderson, B., Mruczek, R. E. B., Kawasaki, K., & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(11), 2540–52. doi:10.1093/cercor/bhn015

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. doi:10.1111/1467-9280.00063

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460–6. doi:10.1111/j.0956-7976.2004.00702.x

Bastos, A. M., Usrey, W. M., Adams, R. a, Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. doi:10.1016/j.neuron.2012.10.038

Beste, C., & Dinse, H. R. (2013). Learning without training. *Current Biology : CB*, *23*(11), R489–99. doi:10.1016/j.cub.2013.04.044

Carandini, M., & Ferster, D. (1997). A Tonic Hyperpolarization Underlying Contrast Adaptation in Cat Visual Cortex. *Science*, *276*(5314), 949–952. doi:10.1126/science.276.5314.949

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, *1145*, 113–31. doi:10.1196/annals.1416.009

Cooke, S. F., & Bear, M. F. (2012). Stimulus-selective response plasticity in the visual cortex: an assay for the assessment of pathophysiology and treatment of cognitive impairment associated with psychiatric disorders. *Biological Psychiatry*, *71*(6), 487–95. doi:10.1016/j.biopsych.2011.09.006

Cooke, S. F., & Bear, M. F. (2014). How the mechanisms of long-term synaptic potentiation and depression serve experience-dependent plasticity in primary visual cortex. *Philosophical Transactions of the Royal Society of London B*, *360*(20130284). doi:http://dx.doi.org/10.1098/rstb.2013.0284

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13494–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=33636&tool=pmcentrez&rendertype=abstract

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. a. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–31. doi:10.1016/j.neuropsychologia.2011.02.015

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11760138

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. doi:10.1037//0278-7393.28.3.458

Frenkel, M. Y., Sawtell, N. B., Diogo, A. C. M., Yoon, B., Neve, R. L., & Bear, M. F. (2006). Instructive effect of visual experience in mouse visual cortex. *Neuron*, *51*, 339–349. doi:10.1016/j.neuron.2006.06.026

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–36. doi:10.1098/rstb.2005.1622

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, *17*(5), 732–7. doi:10.1038/nn.3683

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11124355

Hawkins, J. (2004). *On Intelligence* (Repreint e.). Times Books.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *54*(4), 1105–24. doi:10.1080/713756001

Kaliukhovich, D. a, & Vogels, R. (2011). Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(7), 1547–58. doi:10.1093/cercor/bhq207

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–9. doi:10.1016/j.neulet.2009.06.030

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12868647

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *The American Psychologist*, *47*(6), 796–801. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1616179

Li, L., Miller, E. K., & Desimone, R. (1993). The Representation of Stimulus Familiarity Temporal Cortex in Anterior Inferior. *Journal of Neurophysiology*, *69*(6), 1918–1929.

Liu, Y., Murray, S. O., & Jagadeesh, B. (2009). Time course and stimulus dependence of repetition-induced response suppression in inferotemporal cortex. *Journal of Neurophysiology*, *101*(1), 418–36. doi:10.1152/jn.90960.2008

Marshel, J. H., Garrett, M. E., Nauhaus, I., & Callaway, E. M. (2011). Functional specialization of seven mouse visual cortical areas. *Neuron*, *72*(6), 1040–54. doi:10.1016/j.neuron.2011.12.004

McMahon, D. B. T., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *Journal of Neurophysiology*, *97*(5), 3532–43. doi:10.1152/jn.01042.2006

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1112895108

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(28), 9332–7. doi:10.1523/JNEUROSCI.1215-14.2014

Meyer, T., Walker, C., Cho, R. Y., & Olson, C. R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience*, *17*(10), 1388–1394. doi:10.1038/nn.3794

Miller, E. K., & Desimone, R. (1994). Parallel Neuronal mechanisms for Short-Term Memory. *Science*, *263*(5146), 520–522.

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. *Science*, *254*(5036), 1377–1379.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *13*(4), 1460–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8463829

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(11), 2825–36. doi:10.1523/JNEUROSCI.4102-06.2007

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. doi:10.1016/S0010-0285(03)00128-2

Niell, C. M. (2011). Exploring the next frontier of mouse vision. *Neuron*, *72*(6), 889–92. doi:10.1016/j.neuron.2011.12.011

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. doi:http://dx.doi.org/10.1037/0033-295X.110.4.611

Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(6), 1323–34. doi:10.1093/cercor/bhl043

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–85. doi:10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–8. doi:10.1016/j.tics.2006.03.006

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*(6), 855–863.

Saffran, J., & Griepentrog, G. J. (2001). Absolute Pitch in Infant Auditory Learning: Evidence for Developmental Reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. doi:10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8943209

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8943209

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Sawamura, H., Orban, G. a, & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron*, *49*(2), 307–18. doi:10.1016/j.neuron.2005.11.028

Sobotka, S., & Ringo, J. L. (1996). Mnemonic Responses of Single Units Recorded from Monkey Inferotemporal Cortex , Accessed via Transcommissural Versus Direct Pathways : A Dissociation between Unit Activity and Behavior, *16*(13), 4222–4230.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–9. doi:10.1016/j.tics.2009.06.003

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–6. doi:10.1038/nn.2163

Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Frontiers in Human Neuroscience*, *5*(July), 67. doi:10.3389/fnhum.2011.00067

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–34. doi:10.1016/j.cognition.2005.01.006

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16334058

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–64. doi:10.1037/0096-3445.134.4.552

Winkler, I. (2007). Interpreting the Mismatch Negativity. *Journal of Psychophysiology*, *21*(1992), 147–163. doi:10.1027/0269-8803.21.3.147

Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*(1), 193–205. doi:10.1016/j.neuron.2012.01.032