Higher-level Representations of Natural Images

Hussain Ismail Ahamed Miflah

Submitted in partial fulfilment of the requirements of the

Degree of Doctor of Philosophy

Statement of Originality

I, Hussain Ismail Ahamed Miflah, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of collaboration and publications:

Contents in chapter 3 have previously been published in Proceedings of the Royal Society Open Science journal. Reference:

Ismail, A. M. H., Solomon, J. A., Hansard, M., & Mareschal, I. (2017). A tilt after-effect for images of buildings: Evidence of selectivity to the orientation of everyday scenes. *Proceedings of the Royal Society Open Science, 3*(1), 160551

Contents in chapter 5 are in revision for the Journal of Experimental Psychology: General. Reference:

Ismail, A. M. H. Solomon, J. A., Hansard, M., & Mareschal, I. (in revision). *A bias for city views*. Journal of Experimental Psychology: General

To my mum and dad

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my supervisor Dr. Isabelle Mareschal for her invaluable advice and guidance. I have had a wonderful time working with Isabelle. She is an irreplaceable mentor and she motivated me, encouraged my research ideas, helped me push my limits, respected my opinions, played a huge role in helping me develop a large set of research skills and cared about my academic future (sometimes more than I did). I am definitely looking forward to being her colleague in the future. Secondly, I would like to thank Prof. Josh Solomon for his contribution, insights and criticisms throughout my PhD. Josh taught me the art and elegance of doing flawless research. I also thank Dr. Miles Hansard, my co-supervisor, for his continuous feedback and support. Further, I shall always be grateful to Dr. Neil Mennie and Dr. Jess Price who kick-started my research life as an undergrad and for being great mentors. I also thank Queen Mary University of London for awarding me with the PhD studentship.

I am grateful to all my friends and peers, Keshni, Joe Florey, Miriam, Omar, Joe Cooper, Deema and many others at Queen Mary psychology department, who kept me going through these three and a half years, and for supporting me in scientific and recreational matters. Special mentions to Shahad and Aqeel, best friends since childhood, who have been my pillar of support in every step of the way. Saving the best for last, I would not have made it this far if not for my family. I am eternally grateful to my mum (Nazeema) for her unconditional love and support, and to my dad (Ismail) who is my biggest inspiration, who has guided me all my life and supported my education. I am lucky to have two amazing brothers (Murshid and Mafaz) and in-laws (Shameema and Fazila) who have always stood by my side, encouraged me and helped me in various ways.

Abstract

The traditional view of vision is that neurons in early cortical areas process information about simple features (e.g. orientation and spatial frequency) in small, spatially localised regions of visual space (the neuron's receptive field). This piecemeal information is then fed-forward into later stages of the visual system where it gets combined to form coherent and meaningful global (higher-level) representations. The overall aim of this thesis is to examine and quantify this higher level processing; how we encode global features in natural images and to understand the extent to which our perception of these global representations is determined by the local features within images. Using the *tilt after-effect* as a tool, the first chapter examined the processing of a low level, local feature and found that the orientation of a sinusoidal grating could be encoded in both a retinally and spatially non-specific manner. Chapter 2 then examined these *tilt aftereffects* to the global orientation of the image (i.e., uprightness). We found that image uprightness was also encoded in a retinally / spatially non-specific manner, but that this global property could be processed largely independently of its local orientation content. Chapter 3 investigated if our increased sensitivity to cardinal (vertical and horizontal) structures compared to inter-cardinal (45° and 135° clockwise of vertical) structures, influenced classification of unambiguous natural images. Participants required relatively less contrast to classify images when they retained near-cardinal as compared to near-inter-cardinal structures. Finally, in chapter 4, we examined category classification when images were ambiguous. Observers were biased to classify ambiguous images, created by combining structures from two distinct image categories, as carpentered (e.g., a house). This could not be explained by differences in sensitivity to local structures and is most likely the result of our long-term exposure to city views. Overall, these results show that higher-level representations are not fully dependent on the lower level features within an image. Furthermore, our knowledge

about the environment influences the extent to which we use local features to rapidly identify an image.

Thesis supervisors: Dr. Isabelle Mareschal and Dr. Miles Hansard

Table of contents

List of figures

List of tables

1. Chapter 1 - Background

1.1. Selectivity of the visual system to image features: Physiology

1.1.1. Basic architecture of the human visual system

Visual perception - the process of interpreting the visual world around us - seems effortless despite the abundance of visual information. Traditionally, perception is believed to result from information processing across several hierarchically organized stages. The process begins when light hits the retina at the back of the eye where photoreceptors (rods and cones) convert electromagnetic information into a neural signal that is transmitted to the lateral geniculate nucleus (LGN) in the thalamus and then on to the primary visual cortex (V1 / striate cortex) (Ferster & Miller, 2000), and beyond to other "higher-level" cortical regions (Felleman & Van Essen, 1991; Hochstein & Ahissar, 2002). In this account, the visual system is arranged hierarchically, where early ("low-level") areas such as V1, the first stage of visual information processing in the cortex send information in a feedforward manner to later ("higher-level") regions that encode increasingly complex information (Felleman & Van Essen, 1991; Hochstein & Ahissar, 2002; Melcher, 2005). However, information processing through the hierarchy is not simply feed-forward. Indeed, higher-level regions can modulate information processing in low-level regions through feedback connections (Hochstein & Ahissar, 2002; Xu, Dayan, Lipkin, & Qian, 2008).

1.1.2. Information units: Receptive fields

Single-cell physiology in cats has revealed that, in the early stages of the visual system, neurons only respond when light falls on their receptive field - a small region in space (with a concomitant section on the retina) where a stimulus will elicit a response from

the neuron. While the receptive fields of retinal ganglion cells and LGN cells have a circular organisation (Hubel & Wiesel, 1961; Kuffler, 1953), those in V1 have an elongated centre-surround organisation (Hubel & Wiesel, 1959) and encode information about edges. Following the discovery of receptive fields in retinal ganglion cells of cats (Kuffler, 1953), receptive fields have been reported for neurons in many visuo-cortical regions in other mammals too, mostly in non-human primates (Daniel & Whitteridge, 1961; Engel, Glover, & Wandell, 1997; Hubel & Wiesel, 1959, 1962, 1968).

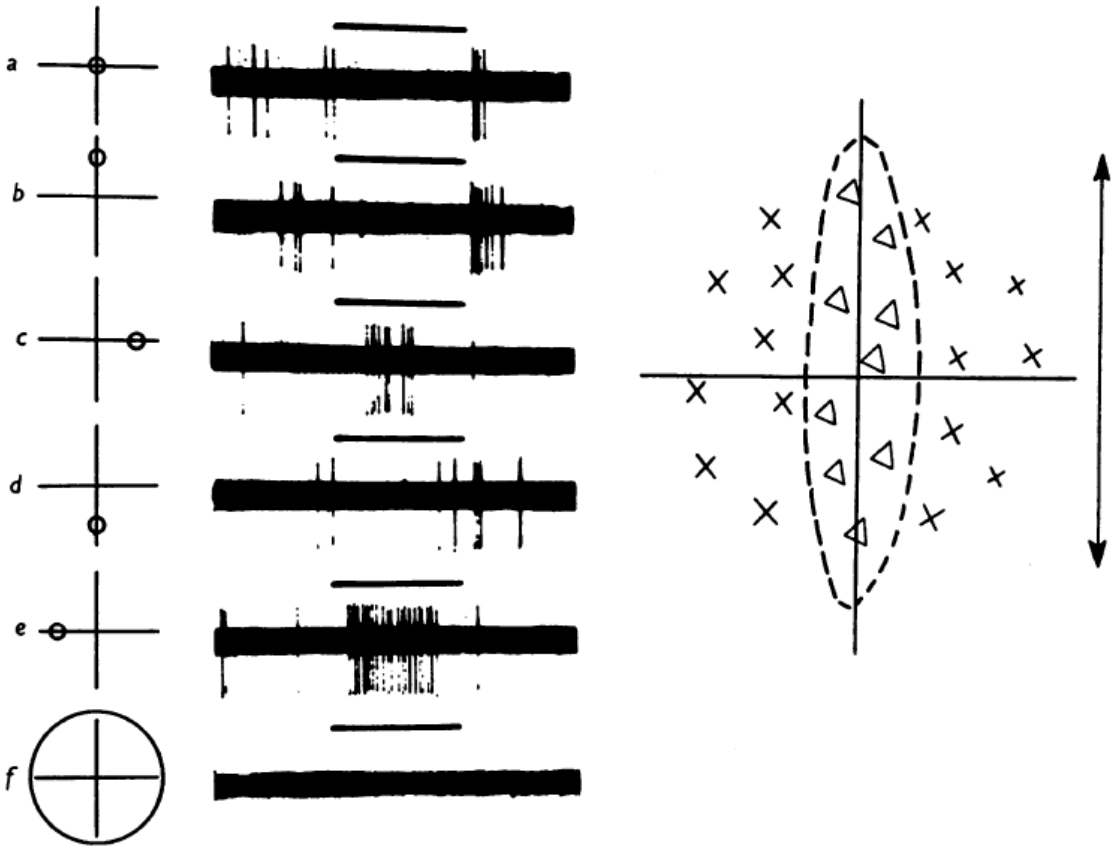


Figure 1.1. Firing patterns of an off-centre cell in the primary visual cortex of a cat. Left-hand panel: circular light spots shone on different parts of the cell's receptive field (a-e) and covering the whole receptive field (f). Middle panel: Firing patterns (vertical lines on thick horizontal bars) produced during epochs of light shone (thin horizontal line) as shown in the corresponding rows of the left-hand panel. The cell is optimally inhibited (no firing) when there is light along the vertical midline of its receptive field

and is excited (i.e., increased firing compared to the period before the epoch, the resting state) by light along the horizontal midline. No firing is observed when light covers the whole receptive field. Right-hand panel: Regions of excitation (×) and inhibition (triangles) of the cell's receptive field. As shown by the dotted ellipse, the inhibitory region takes an elongated shape along the vertical midline. *Reproduced with permission from Hubel and Wiesel (1959)*.

Early single-cell studies on non-human primates have revealed that, from V1 and beyond, the size of a neuron's receptive field varies with two factors. Firstly, the receptive fields of neurons encoding foveal stimuli (i.e., central visual field) are the smallest, but receptive fields increase in size with eccentricity (Hubel & Wiesel, 1974). The second factor is the location of the given cortical area within the hierarchical level. Neurons located in later cortical areas have larger receptive fields (Zeki, 1978). For example, MT (middle temporal) and V4 neurons have larger receptive fields than V1 neurons at any given eccentricity (Felleman & Kaas, 1984; Gattass, Sousa, & Gross, 1988). An increase in receptive field size with eccentricity and hierarchical level has also been reported in humans, using fMRI techniques that estimate the average receptive field size of neurons (Smith, Singh, Williams, & Greenlee, 2001).

1.1.3. Retinotopic organization in the visual system

In humans, the retinal image is organised in a close one-to-one mapping of the monocular visual field - neighbouring information in the visual field is represented in neighbouring parts of the retina, apart from the 'blind spot' (Le Grand, 1967 as cited in Tripathy & Levi, 1994). This type of organization of the visual field is known as 'retinotopic' and fMRI studies have shown that it is preserved in the LGN (Schneider,

Richter, & Kastner, 2004) and across many cortical regions including striate and extrastriate areas (e.g., V2/V4) (Engel et al., 1997; Sereno et al., 1995). Neurons in V1 represent neighbouring parts of the ipsilateral and contralateral visual fields through their receptive fields (Weigelt, Limbach, Singer, & Kohler, 2012). However, the presence of a blind spot in the retina produces a discontinuity in the remapping of the visual space in V1. Psychophysical results suggest that this gap is *filled-in* by intra-cortical mechanisms, using visual information coming from regions surrounding the blind spot, either from the same eye or from the opposite eye (Ramachandran & Gregory, 1991; Tripathy & Levi, 1994). In regions beyond V1, such as V2 or V4 of primates, this retinotopic organization is not as precise - it becomes coarser and more irregular (Felleman & Van Essen, 1991), but this type of organisation suggests that there are several copies of the same visual information (with different levels of precision) across the visuocortical stream.

Retinotopy, however, does not mean that visual acuity is uniform. Indeed it is sharpest for stimuli presented in the central visual field, and gradually decreases with increasing eccentricity (Weymouth, 1958). This occurs because the primate fovea contains a relatively higher density of cones and ganglion cells that results in finer sampling of the visual field (Rolls & Cowey, 1970; Weymouth, 1958). Secondly, because of "magnification", cortical regions (in humans) devoted to encoding visual information are larger for foveal regions than peripheral regions. This magnification is pervasive throughout the LGN, V1, V2 and V3 (Dougherty et al., 2003; Schneider et al., 2004; Sereno et al., 1995) and is quantified using the magnification factor (M), which denotes the size of the cortex that represents each degree of visual angle in space. In V1, M is largest (~40 mm/deg) in the fovea, decreases with slight increases in eccentricity within the foveal region and then further decreases in the periphery (Sereno et al., 1995).

1.1.4. Tuning of early cortical neurons to low-level features

In V1 and beyond, neurons encode features[1] such as edges that have a well-defined orientation. For example, a vertically tuned striate neuron of a cat would be optimally stimulated, when a stimulus that occupies the centre of its receptive field is a vertical bar, exhibiting reduced firing rates when the orientation deviates from vertical (Fig. 1.2a). Similarly, fMRI studies (in humans) and single-cell physiology (in non-human primates) have revealed that V1 neurons are also tuned to stimulus width (their spatial frequency) (De Valois, Albrecht, & Thorell, 1982; Sachs, Nachmias, & Robson, 1971; Singh, Smith, & Greenlee, 2000) and direction of stimulus motion along the receptive field (Hubel & Wiesel, 1959; Singh, Smith, & Greenlee, 2000).

[1] A clear distinction exists between a 'feature' and a 'feature value'. Features are dimensions in which a stimulus can vary (e.g., orientation and spatial frequency). On the other hand, feature values represent the variable quantities within a dimension (e.g., vertical or horizontal for orientation).

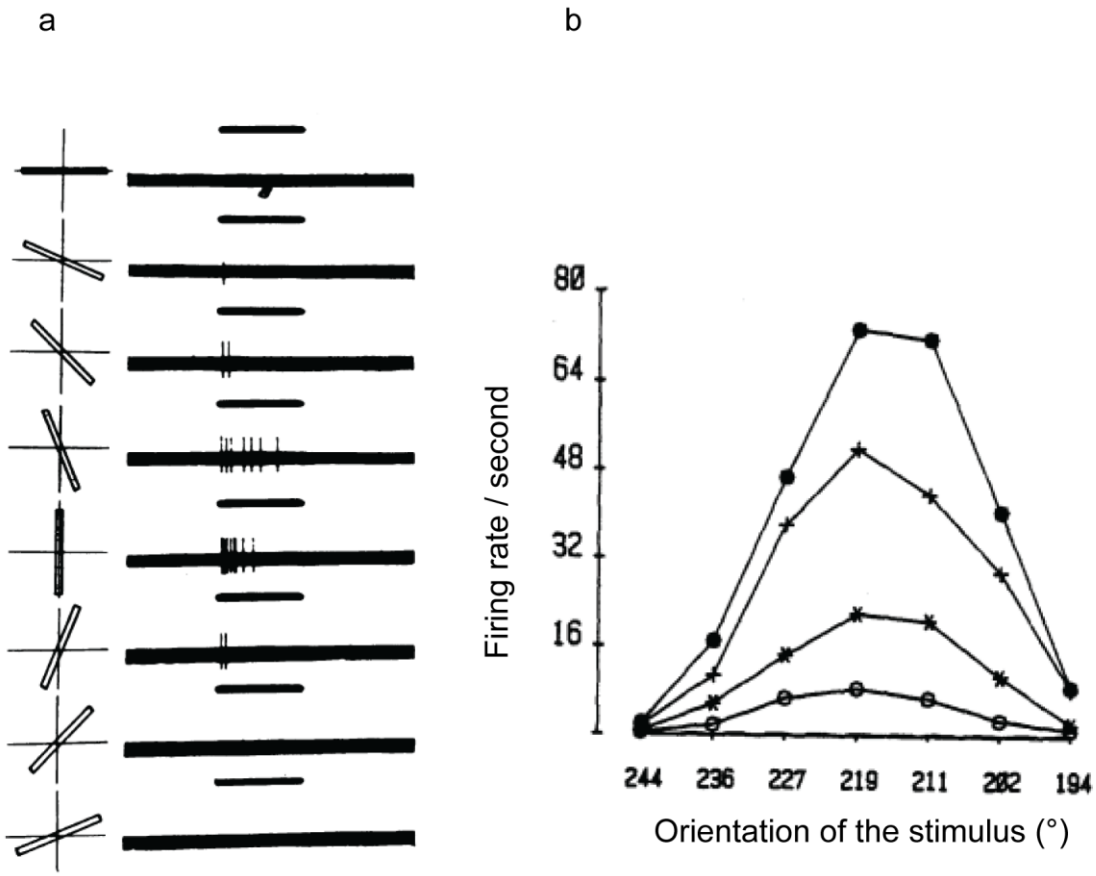


Figure 1.2. Orientation tuning of striate neurons. a) Action potentials (spikes) from a neuron in the cat's striate cortex (right-hand side panel) to a bar stimulus that is rotated clockwise from horizontal to vertical (left-hand side panel). As can be seen, spike rate is maximal for the vertical bar and it decreases as the bar is rotated. *Adapted with permission from Hubel and Wiesel (1959)*. b) Orientation tuning of a neuron in the cat's striate cortex in response to sinusoidal gratings of different orientations. The different symbols denote stimuli presented at different percentages of maximum contrast near and supra-threshold (open circles - 10%, asterisks - 20%, crosses - 40% and filled circles - 80%). This neuron fires optimally to a stimulus oriented ~220° with reduced firing rates as the orientation of the stimulus deviates from optimal. Note that despite changes in absolute firing rates, tuning remains the same across different stimulus contrasts. This demonstrates (supra-threshold) contrast invariance of orientation selectivity in striate neurons. *Adapted with permission from Sclar and Freeman (1982).*

In animals like cats, the selectivity of a neuron to a given feature (e.g., orientation) can be determined by measuring its firing rate (number of action potentials) as a function of varying the feature values (e.g., vertical / horizontal) (Gizzi, Katz, Schumer, & Movshon, 1990; Sclar & Freeman, 1982). An example is illustrated in Fig. 1.2b. The peak of such a plot denotes the neuron's preferred orientation whereas the bandwidth represents how narrowly tuned it is - e.g., a lower bandwidth represents sharp tuning or greater selectivity. A neuron with narrow tuning would only respond to a small range of orientations around its preferred value.

It is also possible to measure the selectivity of a cortical region by either calculating the number of neurons devoted to processing different feature values (single-cell physiology in cats; Li, Peterson, & Freeman, 2003), or measuring the size of neural population responses to different feature values (fMRI in humans; Furmanski & Engel, 2000; Ringach, Shapley, & Hawken, 2002). For example, there are more striate neurons that respond to vertical and horizontal orientations than to other orientations and these neurons are also more narrowly tuned (Furmanski & Engel, 2000; Li et al., 2003). These types of global measures can also be used to compare selectivity across different cortical regions (Ringach et al., 2002). In general, selectivity to low-level features (like orientation and spatial frequency) is known to gradually disappear (weaken) further up the visual system (i.e., beyond V1).

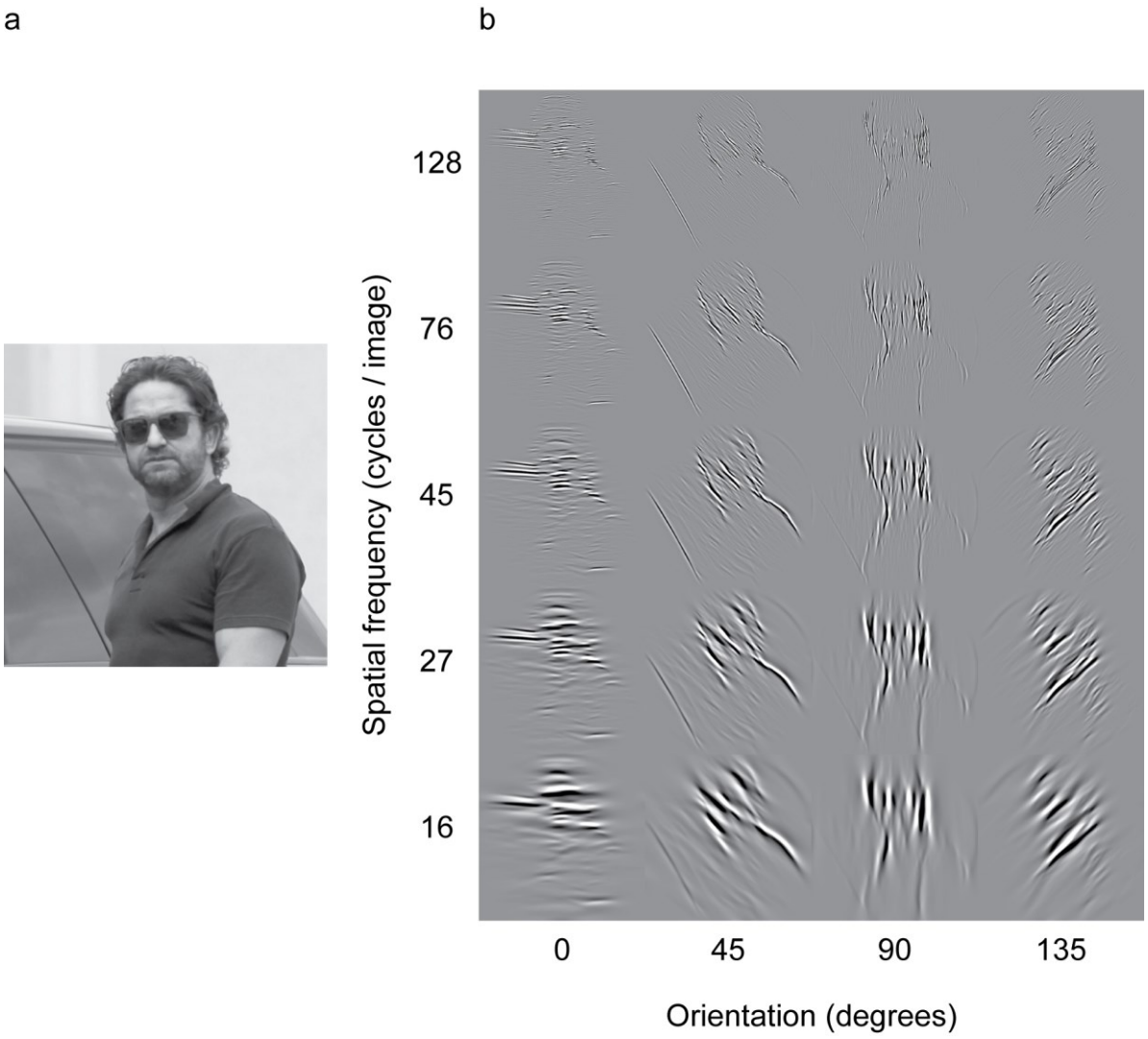
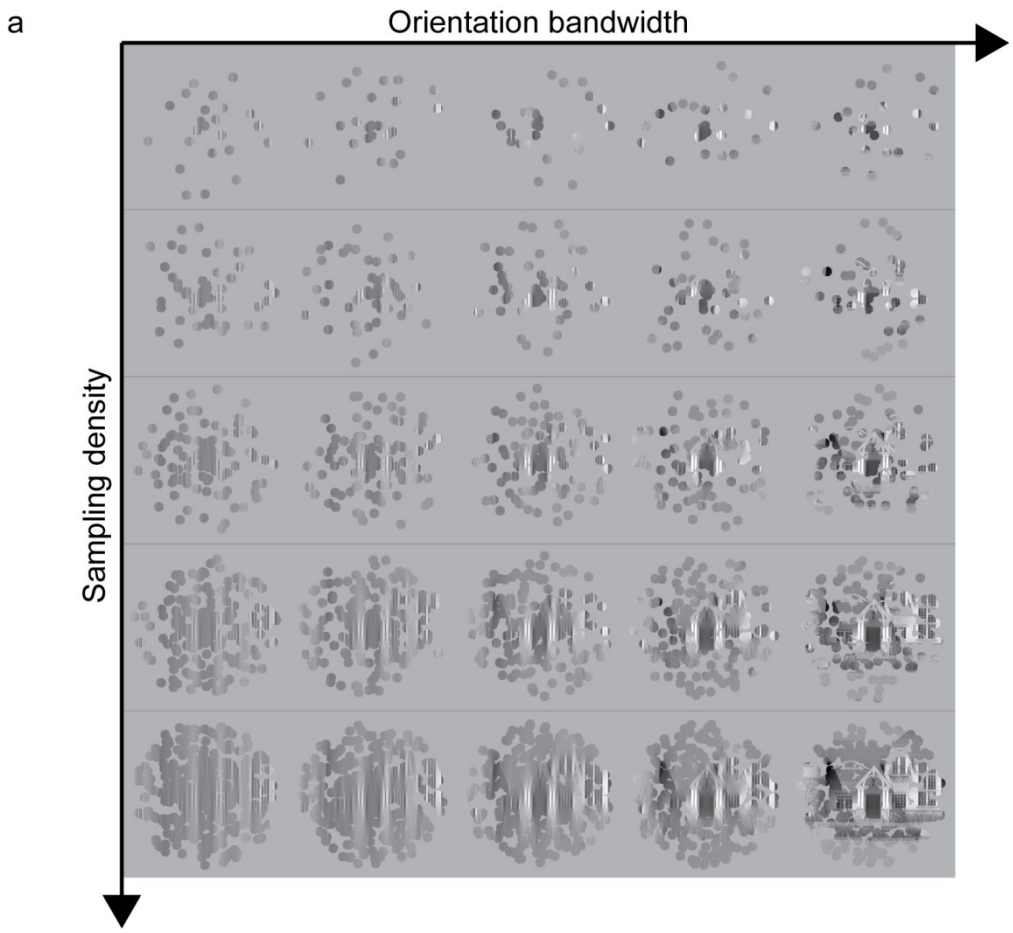


Figure 1.3. Orientation and spatial frequency filtering in primary visual cortex: a) an image of the actor Gerard Butler and b) the same image as would be encoded by sub-populations of neurons that encode edges of (or near) specific orientations (x axis; filtered to allow a half-width at half height of 5.89° around the peak orientation) and spatial frequencies (y axis; filtered to allow a half-width at half height of 0.5 octaves around the peak spatial frequency). As is evident, information within a narrow band of feature values is insufficient for a meaningful percept.


a
Orientation bandwidth
Sampling density

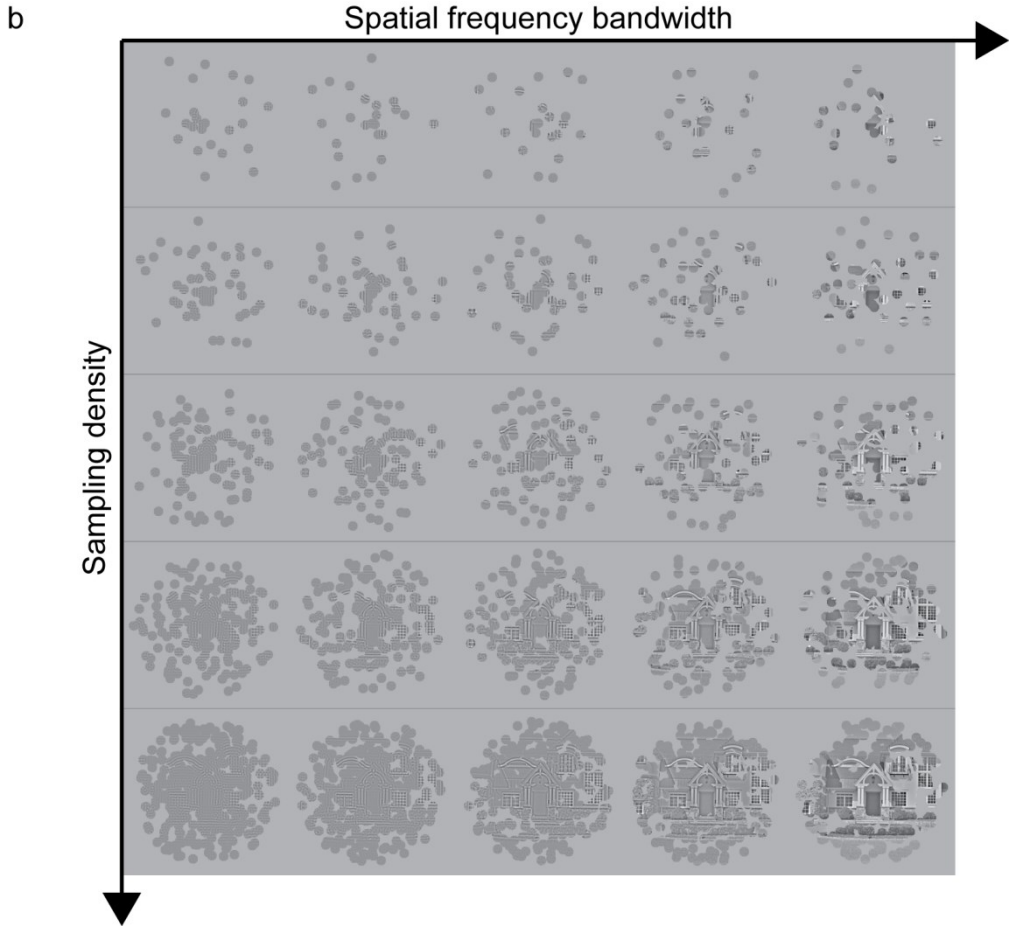

b
Spatial frequency bandwidth
Sampling density

Figure 1.4. Integrating information across different parts of the visual field and across different features and feature values: an image of a building as it would appear to a system that samples in discrete patches of information from sparse to dense sampling (y-axis of top and bottom panels; images going from top to bottom). The x-axes of the top and bottom panels illustrate how the image becomes more informative when neurons encoding edges from a broader frequency range contribute to the percept. The top panel shows this for orientation, going from a narrow band of orientations near vertical (left-most image) to a wider band (right-most image). The bottom panel illustrates it for spatial frequency, going from a narrow band of high spatial frequencies near 37.5 cycles / image (left-most image; the apparent spatial frequency differs because of image resizing) to a wider spatial frequency band (right-most image). As evident from the images, to obtain a meaningful percept, the image samples must be integrated over a large region of space and across a broader range of spatial frequencies / orientations.

1.1.5. Sample complex feature encoding in V1

Neurons in the primary visual cortex are likened to orientation and spatial frequency filters (Fig. 1.3) that respond to edges within a localised patch (the receptive field) of the image. However, edges within individual patches and/or within a narrow band of frequencies only contain limited information that usually cannot support identification of a meaningful percept (Fig. 1.4). For this to occur, information must be integrated across different parts of the image that vary in features and feature values. The process of integrating information across different parts of an image begins very early in V1 where contour perception occurs (Loffler, 2008). Intra-cortical connections in V1 support integration of local edge information that follow Gestalt rules such as

proximity, continuity and closure (Field, Hayes, & Hess, 1993; Hess, Hayes, & Field, 2003; Kovacs & Julesz, 1993; Loffler, 2008). For example, a contour is not simply formed by linking information from immediately adjacent cells but rather following geometric relationships between local edge elements, which Field et al. (1993) defined as an "association field". This includes highly deterministic properties such as the orientation and distance between local edges and the path angle determined by the arrangement of the local edges (Hess et al., 2003). The linking process is fairly tolerant to the spatial frequency difference (although the degree of tolerance reduces with increasing orientation difference between edges) and contrast difference between local edges (Dakin & Hess, 1998; Hess, Dakin, & Field, 1998). These principles by which local edges are bound into contours are well predicted by our knowledge of the statistics of how edges co-occur in natural images (Geisler, Perry, Super, & Gallogly, 2001; Hunt, Bosking, & Goodhill, 2011; Taylor, Hipp, Moser, Dickerson, & Gerhardstein, 2014). However, there are also findings that violate the predictions of the association field model. For example, although the model predicts that our ability to detect contours should monotonically deteriorate with increasing orientation difference between the orientation of the local edges and the path angle (Field et al., 1993), it has been found that this is not the case. Detection decreases up to an orientation difference of 45° but then increases beyond 45° and reaches maximum (although less than a 0° difference) with a 90° difference (orthogonal) (Ledgeway, Hess, & Geisler, 2005).

1.1.6. Selectivity to natural image features

Most of the studies discussed thus far used synthetic stimuli (e.g., Gabors, dot patterns, polygons) to infer simple or complex feature selectivity. However, the types of stimuli we experience in daily life ("natural images") are much more complex than synthetic

stimuli carefully designed to vary in just one or more feature dimensions. For example, an image of a face or a building (Figs. 1.3 and 1.4) is composed of a large number of edges or contours that vary in contrast, orientation, spatial frequency and curvature, and contain multiple elements of different shapes (e.g., eyes and nose of a face). Furthermore, many of these features also include a semantic descriptor (i.e., gender of the face or whether a built structure is a house or a building).

Brain imaging (fMRI) studies in humans have reported multiple cortical regions that are predominantly involved in processing natural or naturalistic (carefully modified natural) images (DiCarlo, Zoccolan, & Rust, 2012; Walther, Caddigan, Fei-Fei, & Beck, 2009). The term 'predominantly' stresses the fact that no perception is the product of neural activity in one single region. First, we will discuss some of the physiological studies that have identified unique brain regions that selectively respond to different types and features of natural images. Following that, we will discuss psychophysical studies reporting evidence for the selective encoding of complex and/or meaningful features of natural images.

1.1.7. A Cortical locus for natural images?

1.1.7.1. Localised representation

Physiological studies (fMRI in humans) have strongly implicated the lateral occipital cortex (LOC) in the processing of objects. The LOC is relatively more sensitive to objects (e g., tools, animals, faces) than to textures (e.g., phase-scrambled object images or patterns of randomly repeated basic geometric shapes) irrespective of the size of the object, spatial frequency content or where they are presented on the retina (Grill-Spector et al., 1998; Malach et al., 1995). On the other hand, early cortical regions (V1 to V3)

respond equally to objects and their phase-scrambled counterparts while also displaying retinotopic specificity (Grill-Spector et al., 1998). However, object encoding in the LOC appears to be largely structural, forming a bottom-up representation from low-level features constructing it, and lack a semantic description of the object. This is because the LOC is found to respond similarly to meaningful (familiar) objects and meaningless sculptures or novel stimuli that resemble familiar objects (Kanwisher, Woods, Iacoboni, & Mazziotta, 1997; Malach et al., 1995).

It is still unclear which area in the brain encodes the semantic categories of objects we classify. Some studies have revealed specific cortical regions (modules) in the brain selectively encoding specific object categories. For example, Kanwisher, McDermott, and Chun (1997) discovered an area in the fusiform gyrus (of the tempero-occipital region) that responds selectively to faces as opposed to other objects like houses and other body parts. They named it the 'fusiform face area' (FFA) and many other studies have related activity in this region to behavioural performances of specifically detecting and identifying faces (Gauthier et al., 2000; Grill-Spector, Knouf, & Kanwisher, 2004). Further evidence comes from a study showing that lesions to the FFA in prosopagnosic patients impairs the ability to discriminate faces that differ in their configuration (e.g., distance between eyes) while damage to other parts of the fusiform gyrus leaves this ability intact (Barton, Press, Keenan, & O'Connor, 2002). Another object category that is selectively encoded in the brain is buildings. Aguirre, Zarahn, and D'Esposito (1998) found a region anterior to the lingual gyrus in the ventral cortex that responds preferentially to images of buildings (like houses) compared to other objects like faces or cars.

1.1.7.2. Distributed representation

While the studies above attempt to confine selectivity of specific object categories to specific modules, others have modelled perception of any categorical object (including faces) as computations occurring across an inter-connected network of cortical regions that predominantly involves regions of the temporal lobe (Chao, Haxby, & Martin, 1999; Haxby et al., 2001; Huth, Nishimoto, Vu, & Gallant, 2012; Martin, Wiggs, Ungerleider, & Haxby, 1996). For example, it has been found that, rather than being a region selective to faces *per se*, the FFA is a region that selectively responds in the process of identifying different members of a semantic category and this occurs for both face as well as non-face objects with expertise (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Tarr & Gauthier, 2000). Therefore, the usually observed FFA selectivity to faces could be a product of participants' tendency to classify faces at an individual level but to classify other objects at a broader (semantic category) level (Grill-Spector, 2003; although see McKone, Kanwisher, & Duchaine, 2007). Further, even in studies attributing a specific area or module, there is also inter-participant variability in the precise locus of selectivity (Aguirre et al., 1998; Kanwisher, McDermott, et al., 1997) and different aspects of face processing (e.g., physical versus identity properties) may be handled by slightly different cortical regions (Rotshtein, Henson, Treves, Driver, & Dolan, 2005).

1.1.7.3. Hierarchical representation?

A distinction in the literature when it comes to how the brain encodes natural images is the difference between an object and a scene. A scene is a representation of the surroundings (most of which we can navigate in) and usually consists of a number of

objects. Whereas an image of a bedroom is considered a scene, the bed itself is treated as an object. Epstein and Kanwisher (1998) discovered the 'Parahippocampal Place Area' (PPA) that responded selectively to scenes as opposed to single objects. However, the highly deterministic property of an image that activates the PPA was the presence of a geometrical layout in the scene; an empty room or a landscape produced stronger activation than an object or even a set of objects. Moreover, we can classify both isolated objects and scenes (with multiple objects) within the first 100 to 200 milliseconds of seeing an image (Greene & Oliva, 2009; Potter, 1975; Thorpe, Fize, & Marlot, 1996). These findings raise the question of whether the most rapid representation of a scene is purely hierarchical in nature (low-level features create objects and objects in turn create a scene).

a b

Figure 1.5. The gist of a scene: a) A grayscale image of a street and b) an image showing a very coarse layout of the same street with only low spatial frequency blobs present in it (individual objects are mostly unclassifiable here). In fact, if you look back at image 'a' you will see that a part of the leftmost building's facade is made of objects usually found in a kitchen (oven and cupboards) and one of the vehicles in the middle of the street is actually a bed.

1.1.7.4. Gist representation?

If scene representation is not purely hierarchical, then how are scenes encoded? An alternative account has been proposed suggesting that rapid scene representation is holistic and can be encoded without necessarily identifying the objects present within it (Fig. 1.5). This is believed to be achieved by encoding a scene's spatial layout in one or a combination of different ways such as: 1) by using the coarse arrangement of contrast blobs of different sizes, 2) by an analysis of the global distribution of orientations and spatial frequencies within a scene or ensemble texture and/or 3) by an analysis of the basic geometric forms ('*geons*') present in it (Biederman, 1987; Brady, Shafer-Skelton, & Alvarez, 2017; Oliva & Torralba, 2001; Sanocki & Epstein, 1997; Schyns & Oliva, 1994). Therefore, it appears that the first stage of scene processing skips object processing. However, at later stages we use information like the presence of key objects (e.g., a bed in a bedroom) and the knowledge of objects typically co-occurring in scene categories (e.g., a table, chair, cabinet and computers typically co-occur in an office setting) to facilitate scene classification (Friedman, 1979; Stansbury, Naselaris, & Gallant, 2013).

In support of the view that it is the spatial layout of a scene that is most influential in determining its category, Walther, Chai, Caddigan, Beck, and Fei-Fei (2011) showed that binary (black and white) line drawings simply outlining a scene layout can be used to decode image category from brain activations in scene selective regions like PPA and retrosplenial cortex, just as well as with normal photographs. Their participants could also classify scenes significantly above chance even after the removal of 75% of contours, and removing long as opposed to short contours produced significantly worse performance, suggesting the role of global structure. Taken together, these findings

highlight the importance of global spatial layout in perceiving the "gist", a rapid semantic classification, of a scene (such as "man-made" or "natural", "indoor" or "outdoor", etc.).

1.2. Selectivity of the visual system to image features: Psychophysics

1.2.1. Adaptation after-effects

The studies discussed in section 1 involved physiological techniques such as single-cell recordings and non-invasive brain imaging (e.g., functional magnetic resonance imaging; fMRI) to examine the selectivity of neurons in different brain regions to features of differing complexity. Psychophysicists also use a number of methods to infer selectivity of neural populations to visual features by examining behavioural responses to visual stimuli, the most pervasive being "adaptation". This refers to the fact that the response of a neuron or group of neurons to a stimulus depends on previous stimulation (Kohn, 2007). Adaptation effects can occur at different timescales, from evolutionary (spanning hundreds of years) and developmental (years) to very short timescale lasting a few minutes or seconds (Simoncelli & Olshausen, 2001). While the first two may result in long-lasting changes in neural sensitivity, the last one leads to short-term (transient) changes that can be measured physiologically and/or behaviourally. This type of short-term adaptation occurs throughout the visuocortical stream (discussed in sections 2.4 - 2.6) as well in other sensory areas (e.g., primary auditory cortex neurons) (Nelken, 2004; Ulanovsky, Las, Farkas, & Nelken, 2004).

Adaptation can alter the perceived appearance of a stimulus. Neural populations are adapted by extended exposure to a specific feature value and tested with similar or slightly different feature values (e.g., adapt to an off-vertical line and test with a vertical

line). This results in a subsequently viewed test appearing different from its physical attribute, a phenomenon known as an after-effect. After-effects can take many forms affecting either the detectability of a test or its appearance (Blakemore & Campbell, 1969; Gibson & Radner, 1937) and have been reported extensively using simple stimuli such as bars, sinewave gratings or Gabor patches (a grating windowed with a Gaussian envelope). An example of an after-effect is demonstrated in Fig 1.6. Here, when a participant is adapted for a sufficiently long duration (~ 30 seconds) to a high contrast Gabor patch (the adaptor), a subsequently viewed low contrast Gabor test patch of similar size and orientation becomes difficult to detect. This is because changes in the responsiveness of contrast sensitive neurons following adaptation, skews the response of the neural population to the low-contrast stimulus making it briefly appear zero-contrast. After a short period, the after-effect disappears and the test becomes visible again.

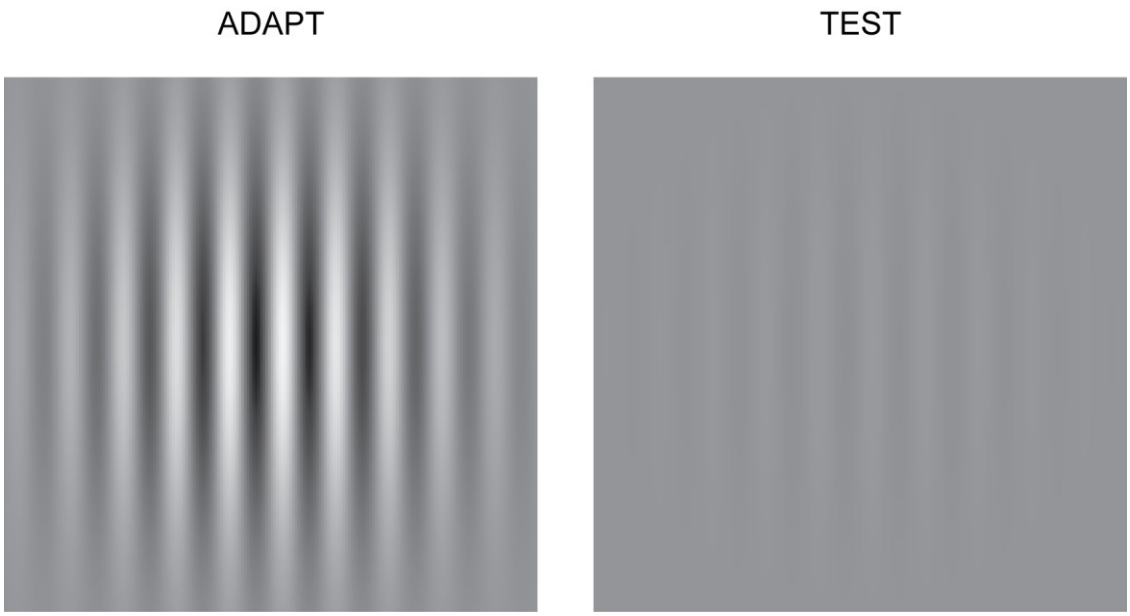


Figure 1.6. Example of a contrast adaptation after-effect. After fixating the centre of the vertically oriented Gabor patch in the left-hand side for 30 seconds, a test Gabor viewed centrally on the right-hand side should briefly appear as a uniform grey patch.

Inference of feature selectivity from a measured after-effect is based on two assumptions. Firstly, single-cell physiology in cat's striate neurons shows that the perception of any feature value is determined by the collective response of a population of neurons tuned to different feature values (Movshon & Lennie, 1979). Secondly, adaptation results from the response adjustments happening in adapted sub-populations of neurons. Based on early cortical responses to simple features like orientation, this is believed to occur in the form of one or a combination of several processes such as a) a desensitisation (reduced response) of a subpopulation of neurons tuned to the adapting (or near adapting) feature value, b) a shift in the preferred (optimally responding) feature value of adapted neurons or c) an increase in response in neurons tuned to feature values further away from the adapting value (Albrecht, Farrar, & Hamilton, 1984; Dragoi, Sharma, & Sur, 2000; Huettel & McCarthy, 2000; Kohn, 2007). Some of these adjustment mechanisms can be generalized to features of different complexities (Barlow & Hill, 1963; Engel & Furmanski, 2001; Kovács et al., 2006; Mollon, 1977). However, it is worth noting that, in some rare cases adaptation does not necessarily imply the existence of feature selective neurons. For example, Hosoya, Baccus, and Meister (2005) reported adaptation to orientation in retinal ganglion cells that were in fact not selectively encoding orientation.

In addition to revealing selectivity, adaptation studies have also revealed the mechanisms by which these features are encoded. For different features encoded by the brain, the neural populations that encode them may employ different mechanisms to represent stimuli of different feature values. For example, spatial frequency and orientation are believed to be encoded by a mechanism of central-tendency - distinct channels that are tuned to different feature values and the final response is similar to the mean of a population response of various channels (Blakemore & Campbell, 1969;

Blakemore, Nachmias, & Sutton, 1970; Clifford, Wenderoth, & Spehar, 2000; Movshon & Lennie, 1979). Whereas colour is believed to be encoded by opponent-mechanisms that encode feature values as the distance from a norm (e.g., white) (Webster & Leonard, 2008).

1.2.2. Using after-effects to infer locus of feature selectivity

After-effects can also be used to infer the cortical locus where feature selectivity occurs. For example, it is known that the minimum amount of contrast required to detect the presence of a low-contrast test grating (its detection threshold) is increased following adaptation to a high-contrast grating. The effect is maximal when the test and adapting gratings share the same spatial frequency and / or orientation (Blakemore & Campbell, 1969; Pantle & Sekuler, 1968). However, a reliable (yet weaker) after-effect is obtained even when the adapting and test gratings are presented dichoptically (Blakemore & Campbell, 1969). While the dependence of this after-effect on orientation and spatial frequency suggests the involvement of orientation and spatial frequency selective neurons, the inter-ocular transfer of the after-effect reveals the involvement of binocular neurons. Therefore, striate cortex is the most likely locus since this is the first region where binocular interaction occurs. This inference was supported by a later physiological study showing that contrast adaptation in LGN is not spatial frequency specific (Duong & Freeman, 2007).

1.2.3. Selectivity to complex features and perceptual after-effects

Although after-effects have predominantly been examined using simple stimuli (e.g., oriented bars) they can also occur for more complex features such as the orientation of

an illusory contour (a contour that is perceived in the absence of physical information; Paradiso, Shimojo, & Nakayama, 1989; Smith & Over, 1975) or the global direction of motion across the visual field (Bex & Makous, 2002; Smith, Scott-Samuel, & Singh, 2000; Snowden & Milne, 1996). Researchers have also found evidence for selectivity to global properties of explicitly defined 2-dimensional (2D) shapes such as polygons and circles, using judgements of concavity (or convexity) of hourglass-like figures (Suzuki, 2001), phase of radial frequency patterns (Anderson, Habak, Wilkinson, & Wilson, 2007), and symmetry (or aspect ratio) of squares and circles (Regan & Hamstra, 1992). Melcher (2005) also reported an after-effect for implicitly defined 2D shapes like radial or concentric patterns formed by randomly arranged dots.

A challenge to interpreting after-effects to more complex stimuli is to determine whether the after-effect results from adaptation of simple features within the stimulus or to the complex feature itself. One way to determine if an after-effect results from processing of local simple information, or is a genuine-after effect to a global, complex property, is to alter low-level properties and evaluate if that modifies the after-effect. Specifically, if an after-effect is immune to manipulations of stimulus features that are known to be encoded by early (low-level) neurons, we can infer that selectivity occurs in extrastriate regions or beyond. For example, V1 neurons are selectively tuned to spatial frequency and orientation and only respond to stimuli presented at a specific retinal location (retinotopic specificity). Accordingly, showing that after-effects persist despite changes in spatial frequency and retinal position between adaptor and test supports the idea of neural processing beyond V1 (Anderson et al., 2007; Bex & Makous, 2002; Melcher, 2005), possibly involving areas that encode complex properties of shapes (Brincat & Connor, 2004; Merigan & Pham, 1998; Pasupathy & Connor, 2001, 2002), and whose neurons are responsive to shapes irrespective of size

and retinal position (Ito, Tamura, Fujita, & Tanaka, 1995; Kobatake & Tanaka, 1994; Logothetis, Pauls, & Poggio, 1995).

There are also other ways low-level explanations can be ruled out for after-effects to complex features. For example, it is known that contrast sensitivity is largely attributed to low-level (V1) neurons, whose response profile is normalized based on stimulus contrast (Carandini & Heeger, 1994; Heeger, 1992). Some have examined the influence of changing the adaptor's contrast (relative to the test) on the after-effect to complex features. For example, Anderson et al. (2007) and Suzuki (2001) found that changing the adaptor's contrast still resulted in significant shape after-effects in a test stimulus. They interpreted this as evidence for the selectivity of ventral route (possibly inferotemporal) regions beyond V1 to 2D shapes, as these regions' responses are known to saturate at low stimulus contrasts (Rolls & Cowey, 1970).

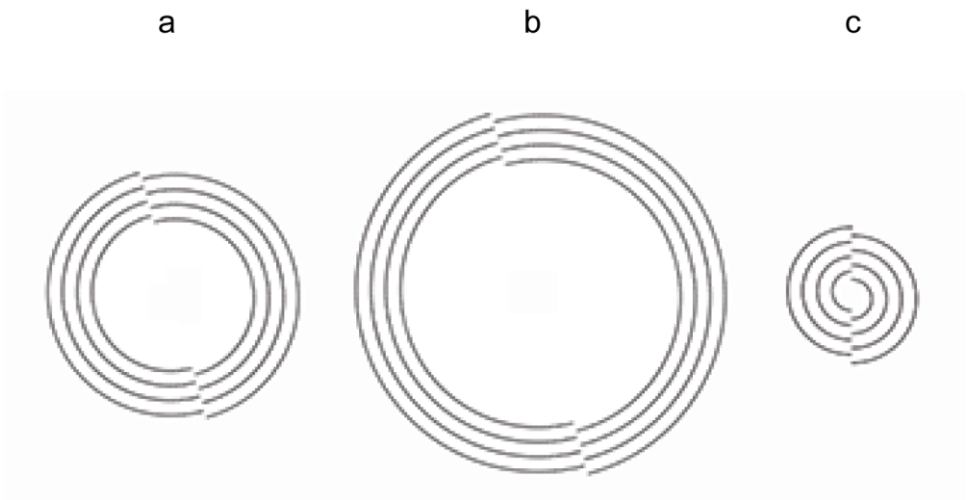


Figure 1.7. Stimuli similar to those used by Van Der Zwan and Wenderoth (1995) to examine tilt after-effects to illusory contours. The tilted adapting stimulus was either a) spatially abutting or b) spatially separated (with a gap) from c) the vertical test stimulus that is presented after one of the two adaptors.

In some cases, if after-effects do survive low-level changes between the adaptor and the test, it does not necessarily rule out the involvement of low-level regions in encoding a complex feature. For example, adapting to an illusory contour tilted slightly off vertical results in a vertical illusory contour appearing tilted slightly anticlockwise (Smith & Over, 1975). Van Der Zwan and Wenderoth (1995) extended these findings to infer the locus of selectivity to illusory contours. As shown in Fig. 1.7, they measured tilt after-effects (TAE) to illusory contours in two different conditions. It was found that, when there was an empty gap between the adaptor and the outer edge of the test pattern, although it produced a significant TAE, the magnitude of this TAE was significantly smaller than a condition where the adaptor was abutting the test. The presence of a TAE even after introducing a spatial separation suggests the involvement of neurons beyond low-level regions (V1/V2). However, it is unclear to what extent we can rule out low-level neurons in selectivity to illusory contours. This is because studies have found fMRI activations selective to illusory contours in a range of areas starting from V1/V2 to higher-levels like V7 or lateral occipital areas (e.g., LO2) (Montaser-Kouhsari, Landy, Heeger, & Larsson, 2007). This suggests that a component of the after-effect could be due to adaptation in orientation tuned low-level neurons.

1.2.4. Using adaptation to measure selectivity to natural image features

1.2.4.1. Adaptation after-effects to features of faces

Adaptation after-effects have also been used to reveal neural mechanisms selectively encoding complex, especially semantic attributes of meaningful stimuli like natural images. Among those who study natural images for feature selectivity, faces have been extensively used as stimuli because 1) they are a type of meaningful natural images we commonly encounter, 2) they possess attributes that can be manipulated along a single

dimension like gender (male to female), akin to the manipulations of attributes like spatial frequency in sinusoidal grating stimuli and 3) some physiological studies have revealed the existence of brain regions selectively responding to faces (1.1.7.1), suggesting the possibility that these regions may contain neural mechanisms encoding specific attributes of faces too. Initially, selectivity to facial features was demonstrated with adaptation to artificially distorted faces (O'Leary & McMahon, 1991; Webster & MacLin, 1999). For example, adapting to a face that is constricted results in a subsequently shown undistorted face to appear distended. Instead of creating distorted (grotesque) caricatures, later studies manipulated faces (usually by means of morphing) to produce adapting and test stimuli that resemble natural variations observed in the environment (e.g., to make a face look more male or female). Accordingly, it has been shown that adapting to a female face can alter the appearance of an ambiguous gender-neutral face to look masculine (Webster, Kaping, Mizokami, & Duhamel, 2004). These face specific aftereffects have been shown for many other feature dimensions such as age, gaze direction, identity, ethnicity and facial expression (Hsu & Young, 2004; Jenkins, Beaver, & Calder, 2006; Leopold, O'Toole, Vetter, & Blanz, 2001; O'Neil & Webster, 2011; Webster et al., 2004).

1.2.4.2. What is being adapted?

In order to demonstrate adaptation in mechanisms dedicated to processing faces, it is important to ensure that feature selectivity reported above doesn't reflect generic mechanisms encoding object shapes or simply a propagation of adaptation in low-level regions. The latter concern is addressed by presenting adaptors and tests of different sizes, in different viewpoints or in different retinal locations, so that local elements do not overlap in space. Adaptation to faces transfers robustly across such manipulations

(despite a reduction in the magnitude) and most findings largely suggest that the adaptation isn't the result of low level propagative effects (Jenkins et al., 2006; Jiang, Blanz, & O'Toole, 2006; Leopold et al., 2001; Rhodes et al., 2004); although Afraz and Cavanagh (2009) found evidence for the retinotopic dependence of the face after-effect. Watson and Clifford (2003) provided a clear demonstration to rule out low-level explanations of face after-effects. They found the axis of the face distortion after-effect to change with the orientation of the adapting face. For example, if the adapting face was distorted on a horizontal axis and tilted 45° anticlockwise of vertical, a test face tilted 45° clockwise appeared distorted along its horizontal axis (perpendicular to the distortion axis of the tilted adaptor) rather along its vertical axis (parallel to the adaptor's distortion axis) as would be predicted by adaptation at low-level mechanisms.

Next, it is important to understand if after-effects to faces are mediated by brain regions specialised in encoding faces (e.g., FFA) or by those generically encoding objects of any category. Findings on this distinction are mixed. It has been proposed that face encoding is holistic, and that only upright as opposed to inverted faces are encoded by face-specific mechanisms (McKone, Martini, & Nakayama, 2001; Yin, 1969). Activity in FFA is also significantly reduced for inverted compared to upright faces and distinct regions are found to be recruited to encode the two (Aguirre, Singh, & D'Esposito, 1999; Rossion & Gauthier, 2002). On the basis of these findings, one would expect adaptation to faces to only occur when they are upright. However, some of the findings show that the magnitudes of face distortion after-effects are similar when adaptor and test are upright or when both are inverted, suggesting that adaptation is possibly mediated by mechanisms encoding any object (Watson & Clifford, 2003). Jiang et al. (2006) added further evidence for a generic mechanism by showing that adapting to a

grotesque face that preserves features diagnostic of identity produced an identity after-effect on a normal test face.

Challenging the findings proposing a generic mechanism that encodes upright and inverted faces, some studies show that after-effects are contingent on the orientation difference between adaptors and tests. An upright adaptor and an inverted test (or vice versa) produces an after-effect that is significantly smaller than when both have identical orientations (Watson & Clifford, 2003). Rhodes et al. (2004) adapted participants to a sequence with a random mix of upright and inverted faces that were distorted differently and found the size and direction of the after-effects to be contingent on the orientation of the face. For example, when the sequence contained contracted upright faces and expanded inverted faces, normal upright test faces looked expanded and normal inverted test faces looked contracted. These findings suggest the involvement of distinct neural populations to encode upright and inverted faces. In addition to distortion, Rhodes et al. (2004) also reported orientation contingent gender after-effects, where male upright and female inverted adaptors, resulted in androgynous tests appearing slightly female when viewed upright and slightly male when inverted, respectively. Since gender is a property specific to faces and body parts alone, they interpreted these after-effects as evidence for face-specific adaptation. To summarise, a significant component of face after-effects arises from adaptation in face-specialised brain regions.

1.2.4.3. The mechanism for encoding face features

How are more complex stimulus features such as faces encoded? Some early models of face processing have suggested different channels tuned to different feature values

along a given dimension (Valentine & Endo, 1992). For example, this could be thought of as sets of neurons selectively tuned to different levels of masculinity along the gender dimension. However, later findings converge on the idea of norm-based coding, highlighting a multi-dimensional feature-space centred on a prototypical (average) face which is the norm (Bestelmeyer, Jones, DeBruine, Little, & Welling, 2010; Lee, Byatt, & Rhodes, 2000; Leopold et al., 2001; Robbins, McKone, & Edwards, 2007). Each face is therefore encoded as the distance from the norm and adaptation shifts the appearance of an existing (pre-adaptation) norm away from the adapting feature value. For example, adapting to male faces results in a pre-adaptation norm (a perfectly androgynous face) appearing feminine. Therefore, the new post-adaptation norm in face space would contain physical characteristics slightly more masculine than an androgynous face.

According to the norm-based model of face encoding, adapting to the norm itself should not affect the appearance of any non-norm face, and this is exactly what happens (Webster & MacLin, 1999). Physiological studies examining regions like the FFA also support this multidimensional face space and norm-based coding of faces (Loffler, Yourganov, Wilkinson, & Wilson, 2005; Ng, Ciaramitaro, Anstis, Boynton, & Fine, 2006). However, Storrs and Arnold (2015) found that adapting to a slightly male face made extremely male test faces look even more masculine. This finding goes against the commonly accepted norm-based coding, and is suggestive of alternative mechanisms in operation such as multi-channel coding.

1.2.4.4. Adaptation to features of scenes

Oliva and Torralba (2001) modelled the spatial layout of a scene by using statistics from the scene's power spectrum. Using these statistics, they assigned each scene a rank along a set of perceptual (and meaningful) feature dimensions such as degree of naturalness, openness and expansion. They found that, for a set of scenes, the rankings given by the model and by human participants were highly correlated on these dimensions. They suggested that these dimensions could be thought of as axes of a multidimensional space, akin to the face space described in section 1.2.4.3. By evaluating scenes along these dimensions, their models could successfully infer the basic level category of scenes (e.g., a street, a forest etc.), since scenes belonging to the same category generally have similar spatial layout. For example, a street scene is always low on naturalness and has a low degree of openness, whereas a coastal scene would rank high on naturalness and high on openness (lacking enclosed structure).

Interestingly, Greene and Oliva (2010) demonstrated that humans might also have neural mechanisms selectively encoding these feature dimensions similar to those proposed by Oliva and Torralba (2001). They found that adapting to a completely natural image with no manmade content made a subsequently viewed semi-natural scene with a bit of man-made content appear more carpentered than it did without adaptation. This could not be attributed to low-level mechanisms as the after-effect was immune to presenting adaptor and test images in different parts of the visual field. Similar after-effects were obtained by adapting to scenes at the extremes of other dimensions such as openness, depth and navigability. Moreover, they also found that adapting to extrema of these dimensions also influence basic level scene classification. For example, following adaptation to a series of images high on openness (excluding

images of forests and fields), previously ambiguous images (having a 50% chance of being classified as forest or field prior to adaptation) on a forest-field continuum were more often classified as forests. Their findings suggest the selectivity of our visual system to these dimensions of spatial layout and their importance in determining the basic level category of a scene.

Experiments conducted by Kaping, Tzvetanov, and Treue (2007) suggest that encoding these features of scene layout may not necessarily require the scene to be meaningful. They found that adapting to artificial meaningless stimuli that mimic power spectral characteristics of scenes can alter the perceived appearance of scenes. For example, adapting to stimuli resembling power spectra of highly natural scenes biased classification of semi-natural scenes as man-made. This finding is consistent with that of Greene and Oliva (2010) in the sense that adapting to images having specific spectral characteristics affects subsequent image classification. However, this raises the concern that mechanisms selectively encoding features of spatial layout based on spectral characteristics may be common to both meaningful and meaningless scenes. Greene and Oliva's (2010) findings do not provide an answer to this because they only used meaningful scenes. This could have been addressed by testing participants' classification after adapting to meaningless scenes like phase-scrambled versions that preserve the power spectra of images.

1.3. Beyond sensory representations

1.3.1. Bayesian analysis

Up to this point, we have discussed visual perception as a process of encoding sensory representations. However, the final (decoded) percept of an image is not limited to this

encoding process. Helmholtz (1925) states that "the sensations of the senses are tokens for consciousness, it being left to our intelligence to learn how to comprehend their meaning". Here, Helmholtz refers to perception as the product of interpreting visual inputs in the context of our implicit knowledge about the environment we are living in. Inspired by this view, visual perception has been modelled within a Bayesian framework as probabilistic inference. In this account our brain infers the most probable stimulus given the retinal image, by combining sensory responses with expectations that we hold about the environment (Kersten, Mamassian, & Yuille, 2004; Knill, Kersten, & Yuille, 1996).

This Bayesian framework can form the basis of a solution to many inverse problems like inferring the 3-dimensional (3D) shape of an object using the 2D retinal image and our knowledge of the 3-dimensional nature of objects in the world (Kersten et al., 2004). In the case of 3D perception, inference using our knowledge of projective geometry (i.e., perspective cues) is important because different 3D shapes can create similar retinal images while the same 3D shape can also create different retinal images (e.g., from different viewpoints), therefore resulting in ambiguity (Fig. 1.8). Interpreting perspective cues within the context of our knowledge that objects are 3D can alter the perceived appearance of retinal images when we reconstruct the object from the 2D retinal images (Pizlo & Salachgolyska, 1995; Thouless, 1931). Sometimes, the knowledge of the nature of the object alone can alter our perception of it, even in the absence of perspective cues. For example, Taylor and Mitchell (1997) showed that when viewers had to reproduce a 2D ellipse, their responses proved to exaggerate circularity when they knew that the ellipse was formed by a slanted circular disc, as opposed to when they believed it was really an ellipse. Furthermore, knowledge of environmental statistics can also affect other veridical forms of perception. For

example, knowledge of how edges co-occur in the environment (i.e., the geometrical relationship between edges) influences our ability to detect contours in complex backgrounds (Geisler et al., 2001).

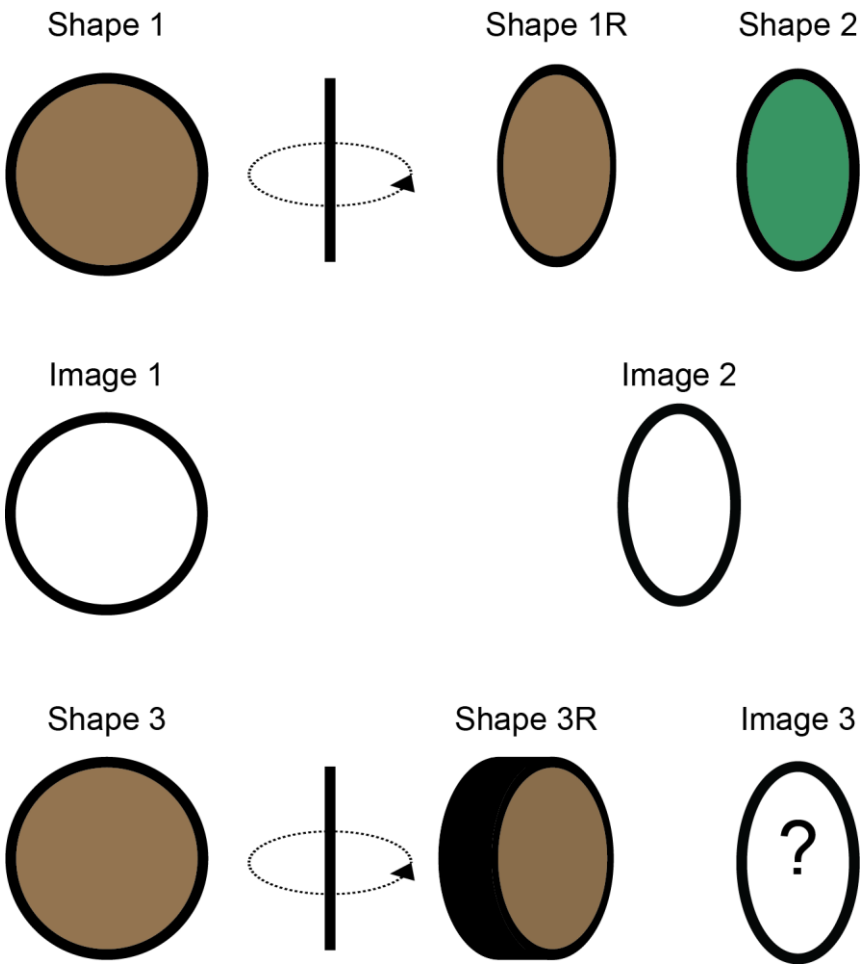


Figure 1.8. An illustration of how different shapes can produce the same retinal image and how the same shape can produce different retinal images. Imagine shape 1 is a circular disk cut out from a piece of paper (a 2D shape). Shape 1 and Shape 1R (formed by rotating shape 1 around the vertical axis) produce different retinal images, namely Image 1 and Image 2, respectively. Shape 1R and Shape 2 (an ellipse) produce the same retinal image (Image 2). Shape 1R (which lacks any cues to suggest rotation) would be perceived as an ellipse when the viewer has no knowledge of its rotation. However, when the viewer is aware of a rotation, the circularity of the ellipse is exaggerated (Taylor & Mitchell, 1997). Now imagine Shape 3 as the surface of a bass drum (a 3D

shape). When it is rotated around the vertical axis (Shape 3R), although the retinal image of the surface (Image 3) is an ellipse, contours of the back of the drum (shown in black) provides us with perspective cues suggesting rotation of a cylindrical object, and this alone or combined with our knowledge that the object is 3D results in the percept of Shape 3R's surface as more circular than Image 3 (cf. Pizlo & Salachgolyska, 1995; Thouless, 1931).

1.3.2. Perceptual biases

In addition to facilitating perception, expectations or knowledge of the environment can also result in altered or non-veridical forms of perception known as perceptual biases. Although these manifest as perceptual errors, they are believed to be the result of a visual system evolved to optimally interpret the retinal images created by the environment (Geisler & Kersten, 2002). The Bayesian account of perceptual biases proposes that the viewer's percept is the most probable estimate of a projected feature value (e.g., the speed of a moving object) and is determined by the maximum of a "posterior" distribution (Freeman, 1994; Kersten et al., 2004; Knill et al., 1996; Yuille & Bülthoff, 1996). The posterior is proportional to the product of a "prior" and the "likelihood function" at each point, where the prior is a probability distribution representing the participant's expectations about the occurrence of different feature values in the environment. For example, based on our experience, we might expect objects to be mostly stationary or moving at slow speeds (Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson, 2002). The likelihood function represents the likelihood that a particular sensory representation reflects a given feature value in the environment. The bandwidth of the likelihood function determines the precision of sensory representation; when measurements become noisier, the bandwidth increases

(Mareschal, Calder, & Clifford, 2013; Stocker & Simoncelli, 2006; Weiss et al., 2002). The likelihood function is generally assumed to be approximated by a Gaussian centred at a peak that is equivalent to the projected feature value (Mareschal et al., 2013; Stocker & Simoncelli, 2006; Weiss et al., 2002). Multiplying the prior by the likelihood function results in the posterior, which is a distribution plotting the probability of each possible environmental feature value given the retinal image and the prior.

An example of how a prior and different likelihood functions can result in different posteriors is illustrated in Fig. 1.9. According to the Bayesian theory, when the sensory measurement is noisy (Fig. 1.9b), the percept (given by the peak of the posterior-black curve) will be more influenced by the prior than when there is less noise. This occurs because a broader likelihood function multiplied by a prior, will be shifted more towards the prior than a narrow likelihood function (Mareschal et al., 2013; Stocker & Simoncelli, 2006; Weiss et al., 2002). These priors may arise from the information we have gathered about our living environment on timescales that could be evolutionary, developmental or very recent (Geisler & Diehl, 2003; Geisler & Kersten, 2002; Körding & Wolpert, 2004; Scholl, 2005).

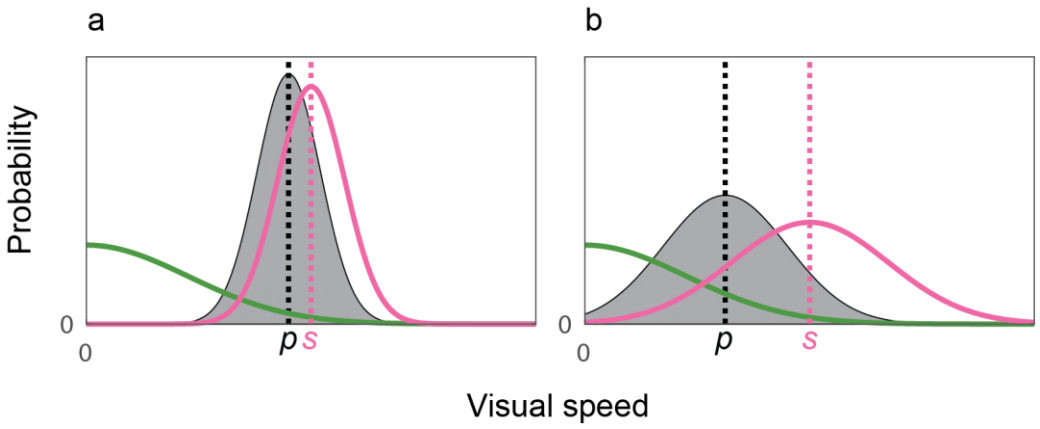


Figure 1.9. Hypothetical representations of the posterior (black curve filled in grey), the likelihood function (magenta curve) and the prior (green curve) distributions for the

perception of a stimulus moving at speed *s*, in conditions where a) the stimulus has low uncertainty and b) the stimulus has high uncertainty. In both conditions, the likelihood function is approximated by a Gaussian centred at speed *s* while the prior peaks at a low speed. When the stimulus uncertainty is high (b) the posterior shifts further away from the peak of the likelihood function towards the peak of the prior. The plots were produced based on the Bayesian model described in Stocker and Simoncelli (2006).

The influence of priors in producing perceptual biases has been demonstrated for a range of visual features of varying complexity. Two examples for low-level features are perceptual biases for orientation and motion. When people judge the average orientation of a set of local Gabor patches of different orientations close to vertical, their judgements are biased towards the cardinal (vertical and horizontal) directions, more so when uncertainty in the stimulus is increased (Girshick, Landy, & Simoncelli, 2011; Tomassini, Morgan, & Solomon, 2010). Girshick et al. (2011) showed that this results from a prior that mimics the orientation statistics of natural scenes that over-represent cardinal orientations as opposed to inter-cardinal orientations. With respect to motion, it has been shown that a prior that favours stationary or low-speeds results in participants underestimating the perceived speed of moving dot patterns (Stocker & Simoncelli, 2006; Weiss et al., 2002).

The impact of priors can also be extended to complex features. For example, we have a prior that the direction of lighting comes from above and slightly to the left (e.g., sun light) which biases the perception of differently shaded 3D shapes as either concave or convex (Mamassian & Goutcher, 2001; Stone, Kerrigan, & Porrill, 2009; Sun & Perona, 1998). Furthermore, biases have been reported for the perception of the approaching angle of a moving square pattern, which results from a prior favouring low-speed

motion (Welchman, Lam, & Bulthoff, 2008). It is also worth noting that a prior for one stimulus feature can influence the perception of another feature that is dependent on it, for example priors for illumination and speed bias the perception of 3D shape and approaching angle, respectively.

1.3.3. Biases for other natural image properties?

A number of studies have also demonstrated that we have priors for higher-level meaningful features that we encode from naturalistic stimuli. For example, Armann & Bulthoff (2012) showed that when people are uncertain about the gender of a face, they are more likely to judge it as 'male'. Watson, Otsuka and Clifford (2016) examined this within a Bayesian framework and suggested that this 'male bias' is most likely caused by a prior that assumes faces to be mainly male. A similar male bias has also been reported when people judge the gender of point-light walkers that depict biological motion (Troje, Sadr, Geyer, & Nakayama, 2006). Further, Mareschal et al. (2013) found evidence for a prior for direct gaze, that is to say we assume others are looking at us when we are uncertain about their direction of gaze. These studies have demonstrated biases for natural image properties based on their conformity of findings to Bayesian predictions. For instance, the influence of priors is strongest when stimuli display high uncertainty (see section 1.3.2). Accordingly, Watson et al. (2016) increased the uncertainty of face stimuli by phase-scrambling them and found large biases for faces with high levels of scrambling.

1.3.4. Operation and stability of priors

While most studies report the possible existence of priors for different stimulus features, some have attempted to determine where in the brain these priors are represented and how they operate in the visual system. On the one hand, priors can be represented at a level beyond where the feature is encoded and this can bias inferences via top-down feedback (Lee, 2002; Lee and Mumford, 2003). For example, Lee (2002) used single-cell recordings in monkeys to measure V1 and V2 response latencies to the perception of a Kanizsa illusion that is believed to result from prior expectations about surface occlusion. He found that V2 had a shorter latency than V1 and suggested that a prior in V2 affects neural activity in V1 to make inferences from illusory contours. On the other hand, priors at lower-levels can bias perception at higher-levels. For example, Gerardin, Kourtzi, and Mamassian (2010) showed that the direction of lighting on any object shape was well predicted by fMRI activity in humans' early retinotopic areas whereas the perceived 3D shape of an object lit from any direction was predicted by activity in later stages such as occipitotemporal and parietal regions. They concluded that, a prior that is represented at low-levels influences 3D perception at higher-levels via bottom-up connections.

In some cases, priors are hardwired into the neural architecture. For example, to account for the over-representation of cardinal orientations in the environment, V1 contains relatively more neurons tuned to cardinal orientations and these abundant neurons are also more narrowly tuned than the neurons tuned to inter-cardinal orientations (Furmanski & Engel, 2000; Li et al., 2003). However, not all priors are permanent and these can be manipulated in the laboratory. For example, Körding and Wolpert (2004) trained participants to learn a lateral displacement in the visual feedback they received

about their finger position while they reached for a target in a virtual-reality setup. Subsequently, when no feedback was provided their reach-points were biased in the direction opposite to the learnt displacement and this showed that they updated their prior for visual feedback. Moreover, even the prior that light comes from above was found to change when participants were trained to expect lighting from a different direction (Adams, Graf, & Ernst, 2004).

1.4. Conclusion

We may have dedicated brain regions preferentially encoding natural stimuli like faces and scenes. There is also psychophysical evidence to suggest the existence of brain mechanisms that selectively encode features of natural images. However, it is still unclear to what extent these findings reflect selectivity at higher-level regions encoding meaningful attributes. Indeed, it is possible that after-effects to natural images are simply a result of adapting to some low-level physical property rather than to the meaningful attribute *per se*. Furthermore, there is also a concern that higher-level after-effects reported in the literature may be the result of non-perceptual biases such as response biases (Storrs, 2015). For instance, when participants are asked to judge the gender of a single androgynous test face following adaptation to male faces, the participant could respond "female" for two reasons. On the one hand, the perceived appearance of the androgynous test could have been genuinely altered by sensory adaptation and the participant will therefore respond "female". On the other hand the participant might have decided to respond "female" more often when s/he is unsure about a face's gender, more so given the adapting face is always a male. Alternatively, the participant could also simply decide to press a key corresponding to the "female" judgement more often without necessarily making their judgements based on the

appearance of the test. The latter two are examples of response biases that would produce the same pattern of shifts in response but for a non-perceptual reason.

Accordingly, in this thesis we aim to examine if higher-level properties of natural images can be uniquely encoded beyond their low-level image components. We address different aspects of higher-level image processing in the four empirical chapters of this thesis. Firstly, in chapters 1 & 2, we examine selectivity to a higher-level image property, specifically "uprightness", using adaptation after-effects. Notably, in order to measure after-effects, we use an experimental design that is immune to non-perceptual sources of response shifts. Secondly in chapter 3, we examine to what extent the encoding of a higher-level image property is dependent on our sensitivity to low-level features within the images. Finally, in chapter 4, we investigate how our long-term exposure to certain types of images may influence our perception of natural images.

2. Chapter 2 - Non-specific encoding of orientation: Tilt after-effects to gratings

2.1. Introduction

Most natural images contain a large number of oriented edges in localised spatial and/or retinal coordinates. Some of these natural images (e.g., houses, faces and scenes) have a global orientation ("uprightness") that we easily perceive which is simply the canonical orientation in which we are used to seeing these images (Tarr & Pinker, 1989). More details about uprightness and the functional significance of perceiving uprightness will be discussed in chapter 3. To date, it is unclear if the brain possesses a mechanism that selectively encodes uprightness, and if it does, whether it is distinct from the mechanism that encodes local edge orientations. Examining selectivity to uprightness by means of adaptation after-effects can help answer this question.

In the previous chapter (section 1.2.3) it was highlighted that an important aspect of interpreting after-effects to seemingly higher-level perceptual features is to distinguish them from after-effects to its low-level features. In the current scenario, any after-effect resulting from adaptation to uprightness must be compared with after-effects caused by adapting to local orientations, since the overall geometry of an image generally arises from the local orientations present within it. However, the literature on local orientation after-effects is inconclusive for a valid comparison. Therefore, this chapter is aimed at measuring aspects of after-effects to local orientations, prior to measuring after-effects to global scene uprightness in chapter 3.

2.1.1. Tilt after-effect and its angular function

Selectivity to local orientation has been studied using "tilt after-effects" (TAE), where adaptation to tilt alters the perceived orientation of a subsequently viewed test stimulus (Fig. 2.1a). Gibson and Radner (1937) were the first to report TAEs to the orientation of a single bar stimulus. They showed that adaptors having an angular separation of between 2.5° and 45° from either a vertical or a horizontal test produced what they termed a "direct TAE", whereby the test orientation appeared repelled away from the adaptor. For example, if the adaptor is slightly clockwise of vertical, a vertical test would appear counter-clockwise. The magnitude of the repulsive effect was maximum (~1-2°) when the adaptor was tilted between 5° and 20° from the test orientation. As the adaptor was tilted more than 45°, the test can start to appear tilted towards the adaptor, a phenomenon known as the "indirect TAE". This attractive effect was largest (~0.5°) at an angular separation of 70° between adaptor and test. Therefore the angular separation is deterministic of the magnitude and the direction of the TAE. The authors concluded that vertical and horizontal orientations are fundamental 'norms' and that adaptation shifts these subjective norms towards the adapting orientation, resulting in the biased appearance of vertical and horizontal orientations.

If Gibson and Radner's (1937) theory of normalization above is correct, then adaptation to the norms themselves should not affect the appearance of other orientations. Later findings have shown that this is not the case. For example, Mitchell and Muir (1976) showed that adapting to a vertically oriented grating made an oblique test grating oriented 45° counter-clockwise appear tilted towards the vertical axis (an attractive effect), and the magnitude of the after-effect was similar to the after-effect induced on a vertical test by an adaptor tilted slightly off horizontal. They also found attractive

effects for oblique tests that were roughly half the magnitude of the repulsive effects. These findings refute the norm based theory. Nonetheless, subsequent studies on the TAE replicated Gibson and Radner's (1937) original finding that the magnitude and direction of the TAE changes as a function of the angular separation between the adaptor and the test when both are presented in, or near, the fovea (Campbell & Maffei, 1971; Mitchell & Muir, 1976; Mitchell & Ware, 1974; Muir & Over, 1970). When stimuli are presented in the periphery, some find no evidence of an indirect TAE (e.g., Muir & Over, 1970).

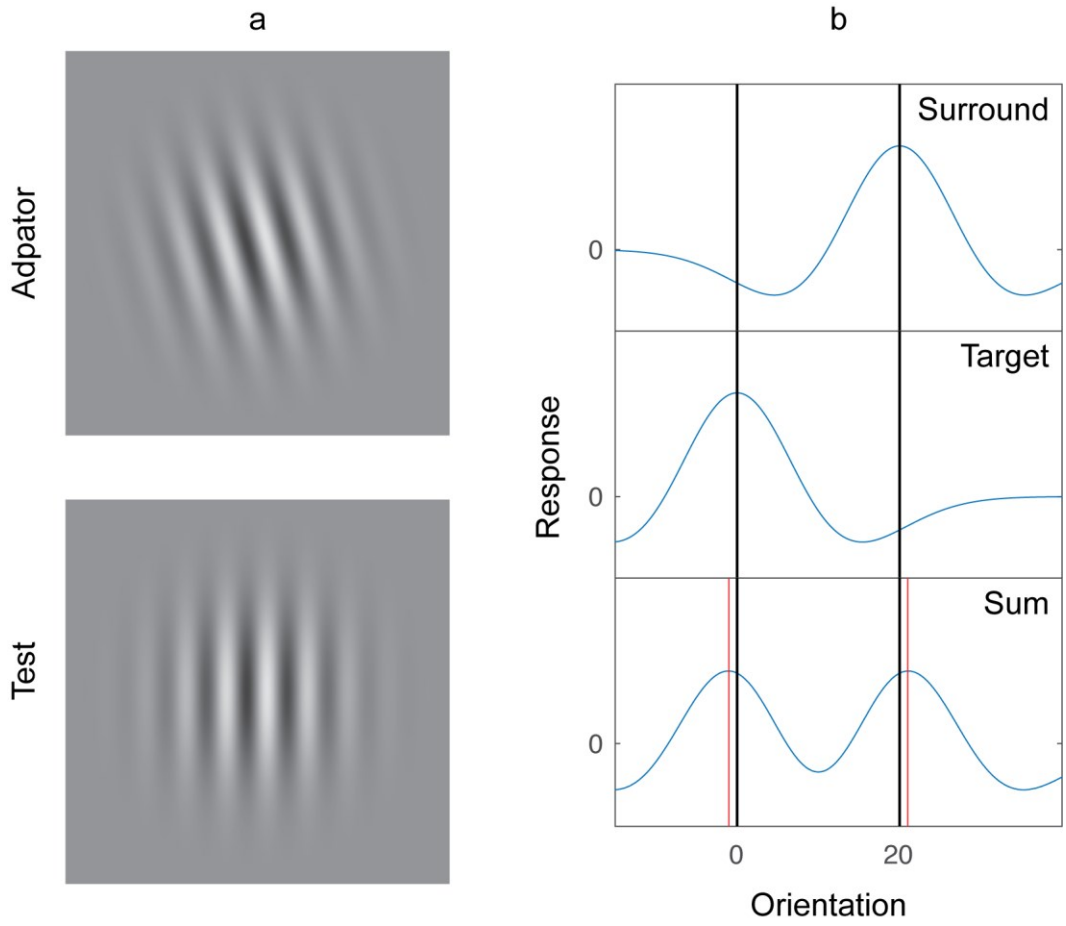


Figure 2.1. a) An illustration of the adapting and test stimuli used in a TAE experiment. If you fixate in the centre of the Adaptor for 1 minute and then immediately fixate at the centre of the Test, the vertical Test will appear tilted slightly clockwise. b) Simultaneous Tilt illusion. Hypothetical profiles of population responses to a surround stimulus tilted 20° counter-clockwise (top) and a horizontal (0°) test stimulus (middle),

and the summed population response (bottom). If the surround or the target is presented in isolation, the peaks of the response profile are aligned with the stimuli's orientations (black vertical lines in the top and middle subplots). Neurons with preferred orientations near the stimulus' orientation are excited (response > 0) and neurons whose preferred orientations are further away from the stimulus are inhibited (response < 0). When the response distributions of both surround and test are summed (i.e., when both stimuli are viewed simultaneously), it results in a compound distribution whose peaks (red lines) are displaced from the orientations of the surround and test; these peaks correspond to the apparent tilts of the stimuli.

2.1.2. Characteristics, cortical locus and the mechanism of the TAE

In addition to its dependency on the angular difference between adaptor and test, TAEs exhibit other important characteristics too. Firstly, the TAE is dependent on the retinal overlap between the adaptor and the test. Gibson and Radner (1937) found no TAE when participants fixated a tilted adapting line and were subsequently tested using a line stimulus presented outside of fixation. Secondly, when adaptors and tests are retinally overlapping, the TAE's magnitude is dependent on the spatial frequency difference between the two (Ware & Mitchell, 1974). When both have identical spatial frequencies (irrespective of the absolute spatial frequency) the TAE is maximal (~3 - 4°) and the magnitude of the TAE reduces with increasing spatial frequency difference (less than 50% of maximum TAE when they differ by 2 octaves or more). Finally, the after-effect transfers between eyes; viewing the adaptor using one eye produces TAEs of similar magnitude regardless of whether the test is viewed by the same eye or the unadapted eye (Campbell and Maffei, 1971; Gibson, 1937; Mitchell & Ware, 1974). The characteristics of the TAE described above suggest that it occurs in binocularly driven

neurons that are highly selective to the retinal position and spatial frequency of the stimulus. In light of the known physiology of the visual system, V1 appears to be the most likely locus for the TAE, since neurons' response properties predict the observed characteristics of the TAE (Blakemore & Campbell, 1969; Coltheart, 1971; Hubel & Wiesel, 1968).

Following Hubel and Wiesel's (1968) discovery of orientation selective channels in the primate striate cortex, there have been many attempts to explain the underlying mechanism of the TAE based on V1 neurons. One of the earliest accounts was the "fatigue theory" (Coltheart, 1971; Sutherland, 1961). In this account, an adaptor strongly activates V1 neurons whose orientation matches that of the adaptor and moderately activates neurons with slightly different preferred orientations. The perceived orientation is the average firing of this population of neurons and is roughly identical to the adaptor's physical orientation. With continuous activation, these neurons habituate or desensitise resulting in reduced firing rates that restore after a period of no adaptation, consistent with response properties of V1 neurons (Hubel & Wiesel, 1962). When a test stimulus of a slightly different orientation is presented, previously adapted neurons will contribute less to the test's population response, thereby skewing the population response away from the test's physical orientation in the direction opposite to that of the adaptor producing a large repulsive TAE. As the test's physical orientation gets further away from the adaptor, more unadapted neurons contribute to the population response, and therefore the average response is closer to the physical orientation producing a smaller after-effect. Coltheart (1971) explained indirect (attractive) TAEs using the fatigue theory by attributing them to "hyper-complex cells" in V1. A hyper-complex cell has two preferred orientations that are orthogonal to each other, for example vertical and horizontal (Hubel & Wiesel, 1968). Therefore, a

stimulus tilted slightly clockwise of vertical would desensitize neurons encoding orientations both near-vertical and near-horizontal. When a vertical or a horizontal test is presented the population response would be shifted anticlockwise of vertical or horizontal, respectively. This would make the vertical and horizontal test to appear tilted away (i.e., repulsed) and tilted towards (i.e., attracted) the adapting stimulus, respectively.

An alternative account of the TAE is one of "lateral inhibition", proposed by Blakemore, Carpenter, and Georgeson (1970) to explain the "tilt illusion (TI)", the spatial analogue of the TAE. In the TI, the perceived orientation of a target stimulus is biased by the presence of a surrounding stimulus of a different orientation. The TAE and the TI are believed to arise from the same underlying mechanism since both phenomena display similar characteristics, such as their angular dependence (between test and adaptor or centre and surround) as well as their selectivity to spatial position and spatial frequency (Tolhurst & Thompson, 1975; Wenderoth & Johnstone, 1988). Blakemore et al. (1970) suggested that, when an oriented stimulus is presented, neurons with orientation preferences close to the stimulus' orientation are excited whereas neurons tuned to orientations further away from the stimulus are inhibited via recurrent lateral connections. Accordingly, the resulting pattern of activity when two differently oriented lines are presented adjacent to each other produces a population response with peaks biased away from the physical orientations of the two stimuli (Fig. 2.1b).

Blakemore et al.'s (1970) findings were not the result of fatigue since introducing a 3rd stimulus next to the surround whose orientation slightly differed from that of the surround, reduced the biased appearance (TI) of the test (Blakemore, Carpenter & Georgeson, 1971; Carpenter & Blakemore, 1973). This is inconsistent with the fatigue

theory, which predicts a stronger TAE given that the presence of more stimuli should cause increased fatigue. Blakemore and Tobin (1972) provided physiological evidence for lateral inhibition in cat's striate cortex, showing that an oriented stimulus increased firing rates of neurons with a preferred orientation similar to the stimulus and decreased firing rates compared to resting state of neurons tuned to orientations further away from the stimulus. Lateral inhibition theory has been applied to explain other effects of spatial context such as Poggendorff, Müller-Lyer and Zölner Illusions (Coren, 1970; Georgeson & Blakemore, 1973; Wallace, 1969). Importantly, lateral inhibition was also extended to account for the TAE (Kurtenbach & Magnussen, 1981; Magnussen & Johnsen, 1986; Magnussen & Kurtenbach, 1980; Tolhurst & Thompson, 1975). Earlier formulations of lateral inhibitions did not account for the indirect effect. To inspect the indirect effect, Wenderoth and Johnstone (1988) applied stimulus manipulations such as spatial separation and spatial frequency difference between adaptor and test that are known to reduce the magnitude of the direct effect. They found the indirect effect to be unaffected by these manipulations. They proposed that the indirect effect is a result of adaptation in higher-level regions, like middle temporal area having larger receptive fields and broadly tuned to spatial frequency, adjusting response properties of neurons in V1 via feedback mechanisms. However, recently, Bednar and Miikkulainen (2000) modelled the TAE based on lateral inhibition in V1 and showed that it can account for the indirect effect too.

While both fatigue and lateral inhibition theories emphasise how the reduction in excitability of a set of neurons lead to a skewed population response to the test stimulus, later theories have modelled other types of changes that could possibly result in both direct and indirect TAEs. Clifford et al. (2000) showed that a model which takes into account self-calibration and decorrelation of population responses reliably fits the

observed psychophysical data on TAEs. Based on the then available literature on the physiological properties of V1, Clifford et al. (2000) proposed that these adjustments in the population response could be caused by desensitisation of adapted neurons and an increase in the bandwidth of orientation tuning in unadapted neurons. However, a single-cell study by Dragoi et al. (2000) showed that adaptation causes shifts in the preferred orientation of neurons in cat's V1. When this property of V1 neurons was incorporated into a model that accounts for desensitization of neurons following adaptation, it reliably predicted human psychophysical data on both direct and indirect TAEs (Jin, Dragoi, Sur, & Seung, 2005). In summary, the characteristics of the TAE, physiological properties of V1 neurons and computational models of the TAE, all point to V1 neurons as the locus of origin of the TAE.

2.1.3. How specific is the TAE to spatial position?

One of the characteristics of the TAE is its dependence on the retinal separation between the adaptor and test - producing the largest effect when they are completely overlapping ("retinotopic"). This led to the conclusion that response changes in V1 neurons encoding stimuli at specific retinal coordinates following adaptation results in the TAE. However, earlier studies demonstrating selectivity to retinal position did not discriminate between retinal and spatial (visual field) coordinates of the stimuli (e.g., two stimuli at different spatial coordinates can occupy the same retinal position and vice versa). Recently, some studies have found that the strict selectivity of the TAE to retinal position is not always necessary, and have shown the need to distinguish between spatial and retinal coordinate spaces. For example, Melcher (2005) conducted an experiment where a participant adapts to a grating tilted 15° off vertical at fixation, after which they make a saccade to a new location in the screen. Following this, a close to

vertical test grating is presented peripherally, either at the screen location where the adaptor was shown ("spatiotopic" condition) or at a completely new location ("non-specific" condition). In both cases the tests were at an eccentricity of 5° from fixation. Melcher found a TAE of magnitude equal to around 60% of the retinotopic TAE in the spatiotopic condition and no TAE in the non-specific condition. Melcher (2007) provided an explanation to this spatiotopic TAE by attributing it to the remapping of parts of the visual field following the participant's intention to make a saccade. Three of his major findings in which the adaptor was shown at fixation support this claim. Firstly, if the test was shown at fixation 100 ms before the onset of a saccade, the magnitude of the TAE dropped by nearly 80% compared to when the test was presented at fixation long before saccadic onset. Secondly, when the test was presented at the newly fixated position, the TAE increased from 0% to more than 60% of the retinotopic TAE as the target was shown at fixation long before to just before saccadic onset. Finally, more than 50% of the retinotopic TAE was observed when the test was shown at the adapted position, around 500ms after saccadic onset (i.e., when participant is fixating a new position). Therefore, adaptation appears to transfer to novel retinal coordinates following a remapping of spatial positions with initiating a saccade.

Subsequent studies haven't always managed to replicate Melcher's finding of a spatiotopic TAE (Knapen, Rolfs, Wexler, & Cavanagh, 2010; Mathot & Theeuwes, 2013). Knapen et al. (2010) studied it with a very similar design to Melcher (2005) and found a small spatiotopic TAE (~1°), but when corrected for non-specific TAEs, this was non-significant. A possible reconciliation has been put forward by Zimmermann, Morrone, Fink, and Burr (2013) who showed that the spatiotopic TAE depends on the duration of the interval between saccade and test stimulus onsets - a longer duration gives sufficient time to transform the stimulus location to the novel spatiotopic space.

This is consistent with Melcher (2007) who also found a larger TAE with a longer saccade-stimulus interval. On the other hand, when adaptors and tests do not overlap in both retinotopic and spatiotopic coordinates, most studies report no TAE (Gibson, 1937; Knapen et al., 2010; Melcher, 2005; Zimmermann et al., 2013). However, some studies report significant non-specific TAEs when the adaptors are separated by around 4°; nearly 30% of the retinotopic TAE (Melcher, 2007) or approximately 1-2° in size (Boi, Ogmen, & Herzog, 2011).

Given the inconsistent nature of these results, this chapter examines the positional selectivity of the TAE to local orientation, focussing on the non-specific component of the TAE. Experiments reported here only examine characteristics of the non-specific TAE that will be relevant for interpretation of TAEs to image uprightness in Chapter 3. Additionally, it should be noted that all above studies (with the exception of Boi et al., 2011) on TAEs have relied on the method of single stimuli where participants make judgements on a single test stimulus which is prone to response error (Jogan & Stocker, 2014; Morgan, Melmoth, & Solomon, 2013). Therefore, we also use a two-alternative-forced-choice (2-AFC) method developed by Morgan, Grant, Melmoth, & Solomon (2015) that eliminates many forms of non-perceptual bias.

2.2. General Methods

Participants

Four conditions were tested: (1) adaptor and tests have the same spatial frequency *(same SF)*, (2) adaptor and tests have different spatial frequencies *(different SF)*, (3) adaptors and tests have the same spatial frequency but a larger cortical separation between them compared to the first two conditions *(Same SF crossed)*, and (4) adaptor and tests have

the same spatial frequency but are, on average, orthogonally orientated *(same SF orthogonal)*. Six participants participated in each condition (including authors AM and IM). All participants except the authors were naïve to the purpose of the experiment.

Stimuli

Sinusoidal luminance modulated gratings with a Michelson contrast of 50% were used in all conditions. All gratings appeared within a hard-edged circular aperture with a diameter of 2.95°, on a uniform grey background. In the *same SF*, *same SF crossed* and *same SF orthogonal* conditions both the adaptor and test gratings had a spatial frequency of 1.6 cycles / degree (cpd). In the *different SF* condition, the adaptor (1.25 cpd) and test gratings (5 cpd) were separated by 2 octaves. In all conditions the spatial (and retinal) positions of the adaptor and tests were non-overlapping, ensuring that any after-effect is non-specific.

Experimental Setup and apparatus

Participants were seated in a dark room. Stimuli were presented on a 20" Iiama CRT monitor with a screen resolution 1600×1200 and a refresh rate of 60Hz. The viewing distance was 57cm such that each pixel subtended 1.5 arcminutes. A black aperture (diameter=24.5°) was overlaid on the monitor to eliminate the use of monitor edges as cues to vertical. Experimental design and analysis were run using Matlab and Psychtoolbox (Brainard, 1997; Pelli, 1997).

Procedure specific to *same SF, different SF and same SF orthogonal* conditions

Baseline (no adaptation):

Prior to adaptation, we measured participants' perceptual bias for gravitational vertical. Participants fixated a centrally presented black circle (diameter = 0.2°) for 1 second, followed by the presentation of two test gratings (50ms) presented on either side of fixation along the horizontal meridian. The centre-to-centre distance between fixation and each test stimulus was 3.07°. One of the test stimuli was the "pedestal" with a fixed tilt (-3° or +3°) relative to vertical and the other was the "comparison", with an offset added to the fixed tilt, randomly selected from the set {-15°,-12°,-9°,-6°,-3°,0°,3°,6°,9°,12°,15°}. The spatial position (left or right of fixation) of the pedestal and comparison was randomised on every trial. Participants judged which of the two test gratings appeared more vertical in a 2-AFC task, using keys '1' and '2' to select the test in the left or right spatial position, respectively. Each combination of pedestal and comparison tilt was tested 10 times, resulting in 220 trials per condition.

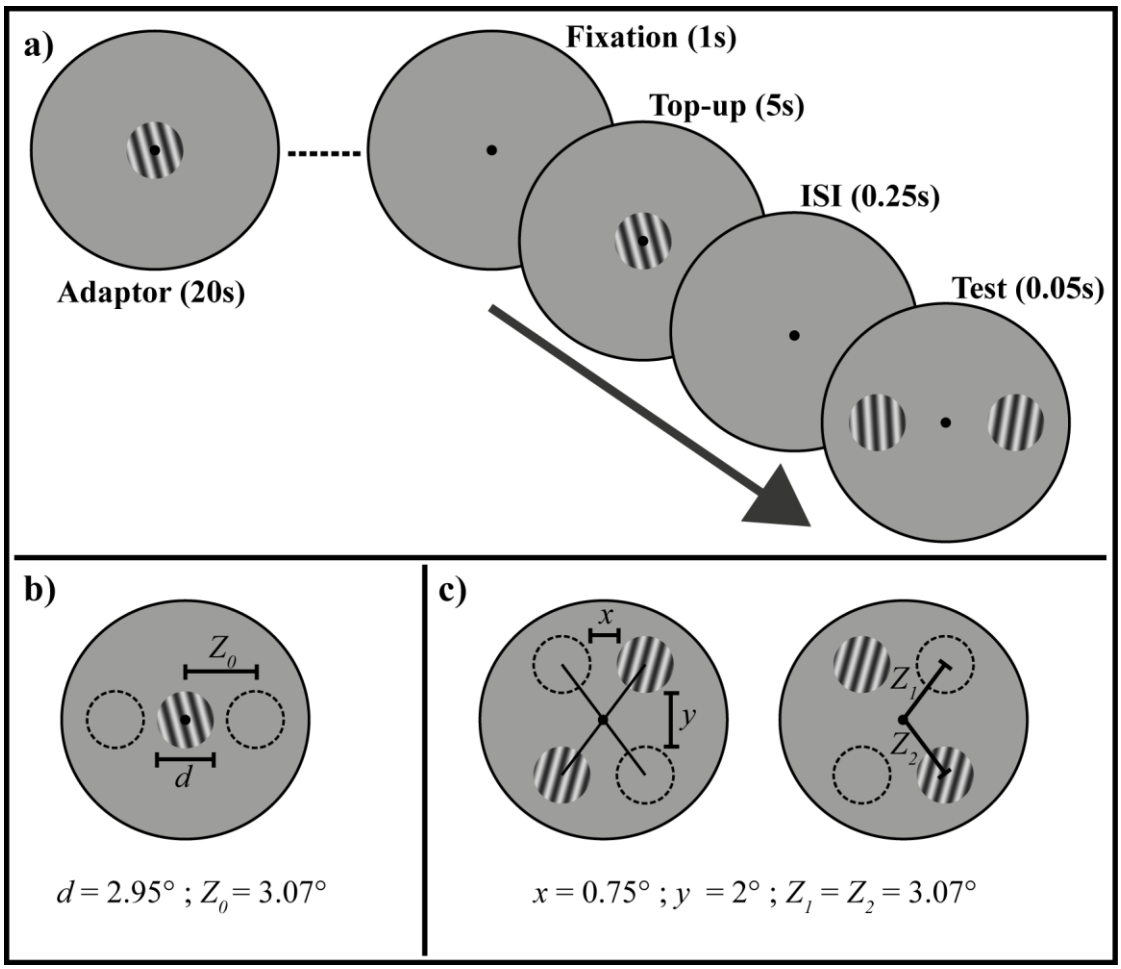


Figure 2.2. TAE to gratings methods: a) Timeline of a sample trial from the *same SF* condition. b) Stimulus configuration for all conditions except *Same SF crossed*. The grating is the adaptor and dashed circles represent the test positions. c) Stimulus configuration for *Same SF crossed* in different blocks (left and right).

Adaptation:

Following the baseline presentation, we measured the TAE in participants who adapted to a grating tilted in the counter-clockwise (CCW) or clockwise (CW) direction by 15° of vertical. CCW and CW adaptations were tested in separate blocks, with block order pseudorandomized across participants. As illustrated in Fig. 2.2a, participants fixated a central circle (diameter = 0.2°) centered on an oriented grating (adaptor) for 20s. After the grating was removed, a top-up adaptor appeared for 5s, followed by a grey screen

(250ms) and the two test stimuli (50ms). The participants' task was the same as in the baseline. Following response, the next trial began with a fixation followed by a top-up adaptor. To avoid Troxler fading (disappearance of stimuli with stabilized gaze), adaptors were counter-phase flickering at 2.5Hz in all conditions.

In the *same SF orthogonal* condition, the tests were tilted relative to horizontal (90°) during baseline as well as adaptation blocks and we measured the perceptual bias for gravitational horizontal. The participants' task was to choose the test stimulus that appeared more horizontal. The adaptor was tilted by 15° relative to vertical which resulted in a mean angular separation of 75° between the adaptor and tests.

Procedure specific to the *same SF crossed* condition

I divided the baseline as well as each adaptation condition into two blocks. For the baseline, the two test gratings appeared in opposite visual fields with respect to both vertical (left and right of fixation) and horizontal (above and below fixation) meridians for one block. For the second block the test positions switched to the opposite quadrants. The block order was pseudorandomized. During adaptation, two diagonally located adaptors counter-phase flickering in-phase at 2.5Hz were presented in (for example) the upper left and lower right quadrants and the two tests were presented in the opposite (upper right and lower left) quadrants (Fig. 2.2c). Stimulus presentation timings were identical to the other three conditions.

The centre-to-centre (CTC) radial distance between fixation and adaptor or test stimuli was maintained at 3.07°. The horizontal CTC distance between the adaptor and test in the left and right visual fields was 3.70°, and the vertical CTC distance between the adaptor and test in the upper and lower visual fields was 4.95°. The procedure was the

same as above and participants indicated whether the test grating above or below the horizontal midline appeared more vertical using the arrow keys 'up' and 'down'.

Fixation monitoring

To ensure fixation stability, three participants in the *same SF* (AM, IM, SB), *different SF* (AM, IM, JF) and *same SF crossed* (AM, JS, JP) conditions had their gaze position monitored during both baseline and adaptation blocks. Binocular gaze position was recorded using an EyeTribe table mount infra-red eye tracker sampling at 30Hz with an accuracy of 0.5 degrees. For those participants who were gaze tracked, all experimental blocks began with a 9-point calibration. During experimental trials, gaze position was monitored with an online rejection criterion, from the beginning of the fixation stimulus until the test gratings disappeared. Starting from the 20th gaze sample (excluding blinks), we used a sliding temporal window that computed the mean gaze position of the 5 preceding samples at every point. If the mean horizontal gaze position exceeded ±1.5° from the fixation point, the trial was aborted and restarted. In all cases, participants held fixation and rejections occurred infrequently.

Psychophysical model

Data were analyzed within the context of signal-detection theory, as described by Morgan et al. (2015). Within this model, the appearances of pedestal (*S*) and comparison (*C*) are normally distributed, i.e., $S \sim N(p + \mu, \sigma^2/2)$ and $C \sim N(p + \mu + t, \sigma^2/2)$, where σ^2 is the variance of the performance-limiting noise, p is the pedestal tilt, t is the offset added to the comparison, and μ is the perceptual bias specific to each test block. If there were no perceptual bias, the distributions for pedestal and comparison would have means of p and $p + t$ respectively. The participant chooses the pedestal as closer to vertical (or horizontal in the *Same SF orthogonal* condition) when

it appears less tilted than the comparison. Accordingly, the probability of this choice $P("S") = P(|S| < |C|) = P(S^2/C^2 < 1)$, has a doubly non-central *F* distribution. This distribution's denominator's noncentrality parameter is $2(p + \mu + t)^2/\sigma^2$, its numerator's noncentrality parameter is $2(p + \mu)^2/\sigma^2$, and both denominator and numerator have 1 degree of freedom.

2.3. Results

For each condition and each test block (baseline, CCW and CW), we plotted the proportion of times the pedestal was chosen to appear more vertical (or horizontal in the *Same SF orthogonal*) as a function of the offset tilt added to the pedestal (i.e., the comparison's tilt relative to the pedestal). From these plots, we obtained maximum likelihood estimates of bias μ and the variance of performance limiting noise σ^2 by fitting the above mentioned psyhophysical model to the data (Fig. 2.3). For conditions *Same SF*, *Different SF* and *Same SF crossed*, negative biases with CCW adaptors and positive biases with CW adaptors (relative to the baseline's bias) are indicative of repulsive (i.e., direct) TAEs. On the other hand, for the *Same SF orthogonal* condition, negative biases with CCW adaptors and positive biases with CW adaptors (relative to the baseline's bias) are indicative of attractive (i.e., indirect) TAEs. To quantify the reliability in individual estimates of μ, we performed non-parametric bootstrapping (Efron & Tibshirani, 1994). First, 1000 maximum likelihood estimates of μ were derived by randomly sampling the (proportion choosing pedestal) data with replacement. The bootstrapped estimates were bias-corrected for asymmetry around the maximum likelihood estimate of μ on the observed data. Finally, upper and lower bounds of the distributions 95% confidence intervals were calculated.

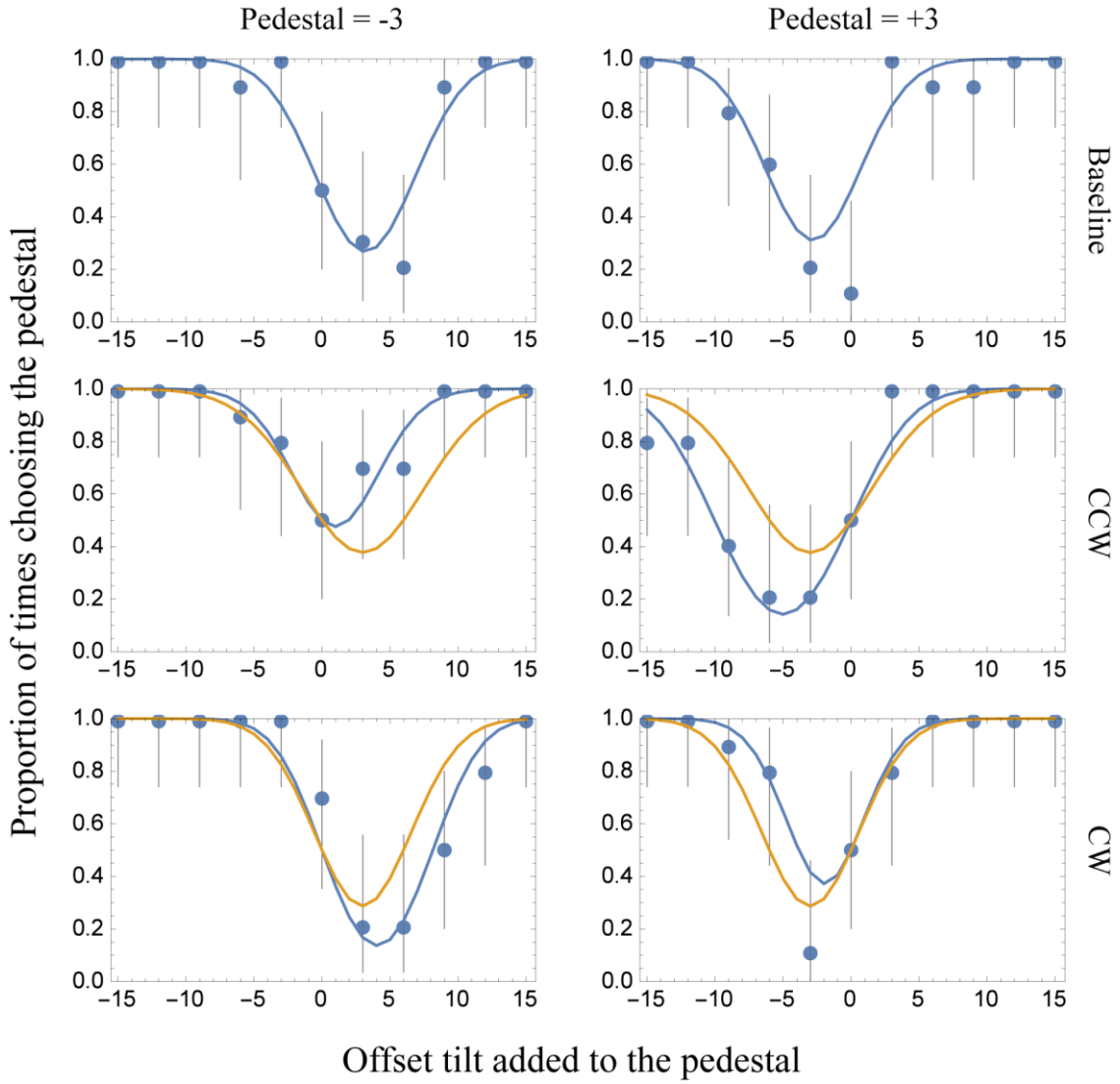


Figure 2.3. Psychometric fits for IM from the *Same SF* condition. The ordinate refers to the probability of choosing the pedestal while the abscissa is the offset tilt added to the pedestal. Top, middle and bottom panels show fits for the baseline, CCW and CW test blocks, respectively, collapsed between the two pedestals (left and right panel). Blue curves represent separate fits for each test block. The nested model fit for CCW and CW conditions are denoted by orange curves. In all subplots, error bars denote approximate Bernoulli confidence intervals (95%) around each data point.

We also fit each participant's data from CCW-adaptor and CW-adaptor blocks simultaneously, forcing the bias parameter μ to be the same in both cases, but allowing σ to vary (Fig. 2.3). The ratio L, between the likelihood of this nested model fit

and the joint likelihood of the aforementioned separate fits to the same data is necessarily no greater than 1. To evaluate the 'null' hypothesis of no significant TAE in individual participants, we compare the criteria $\alpha = 0.05$ and $\alpha = 0.001$ to the value $1 - F(-2 \ln L)$, where F is the cumulative χ^2 distribution, with 1 degree of freedom. This is known as the generalized likelihood-ratio test (Mood, Graybill and Boes, 1974, pp. 440–441).

To evaluate the null hypotheses at the group level, we performed two-tailed one-sample *t*-tests using estimates of repulsion, which can be quantified either in degrees of tilt or in terms of the 'just-noticeable difference' (JND). A single value for repulsion, in degrees of tilt, can be obtained by subtracting one maximum-likelihood estimate of μ (the one obtained with CW adaptors) from the complementary estimate (obtained with CCW adaptors), and dividing the difference by 2. The 'conspicuousness' of repulsion can be quantified by further dividing this quotient by the JND (Eq. 2.1). For the latter, we use the root mean-square of the maximum-likelihood estimates of σ. Results of the group-level *t*-tests appear in Table 2.1.

$$conspicuousness = \frac{(\mu_{CCW} - \mu_{CW})}{2 \times \sqrt{\frac{(\sigma_{CCW})^2 + (\sigma_{CW})^2}{2}}} \tag{2.1}$$

Table 2.1. *Group level statistics of two-tailed one-sample t-tests conducted on mean repulsion and mean conspicuousness across participants*.

Condition	Mean (°)	*t*-statistic	*p*-value	Cohen's *d*
		Repulsion		
Same SF	1.07	7.85	0.001	3.20
Different SF	1.19	4.94	0.004	2.02
Same SF crossed	0.88	9.15	<0.001	3.74
Same SF orthogonal	0.00	0.01	0.993	0.00
		Conspicuousness		
Same SF	0.34	5.64	0.002	2.30
Different SF	0.33	4.22	0.008	1.72
Same SF crossed	0.18	5.01	0.004	2.04
Same SF orthogonal	-0.01	-0.14	0.891	-0.06

Individual biases for each condition and test block are plotted in Fig. 2.4. In the *same SF*, *Different SF* and *Same SF orthogonal* conditions, TAEs were repulsive in general. As revealed by likelihood-ratio tests, TAEs were significantly repulsive for 6/6 participants in the *Same SF*, 5/6 in the *Different SF* and 4/6 in the *Same SF crossed* conditions. Group level analyses revealed that both mean repulsion and mean conspicuousness across participants were significantly different from zero (no TAE) at the level of $p < 0.01$, for all three conditions. In the *Same SF orthogonal* condition, only one participant experienced a repulsive TAE that was significant based on a likelihood-ratio test. Both mean repulsion and conspicuousness did not differ from zero for this condition.

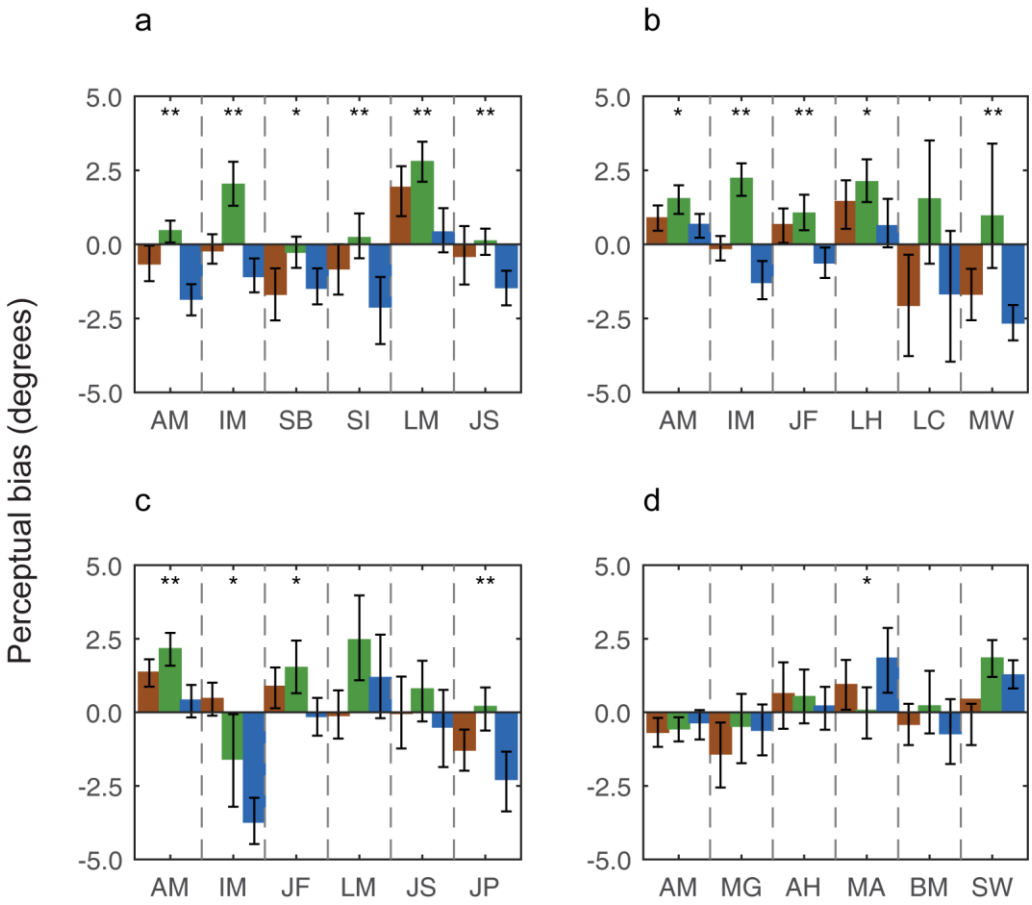


Figure 2.4. Maximum likelihood estimates of perceptual bias for baseline (brown), CW (green) and CCW (blue) test blocks from a) *Same SF*, b) *Different SF*, c) *Same SF crossed* and d) *Same SF orthogonal* conditions. Error bars denote bootstrapped 95% confidence intervals. Single asterisks (*) denote TAEs significant at the $\alpha = 0.05$ level for repulsion. Double asterisks (**) denote TAEs significant at the $\alpha = 0.001$ level for repulsion.

2.4. Discussion

Traditionally, TAEs to gratings are believed to originate in the early stages of the visual system, more specifically in V1 neurons encoding information from confined retinal positions. Contrary to the strict retinal/spatial selectivity of TAEs mostly reported in the literature, here we find significant TAEs induced by adapting gratings that do not overlap with the tests in both retinal and spatial coordinates. These non-specific TAEs

produced by adaptors with a mean angular separation of 15° from the tests, were repulsive in nature. Mean TAEs were roughly 1° in magnitude and reliably survived further manipulations of spatial frequency and cortical distance between adaptor(s) and tests. The pattern of results was highly consistent across participants, with most participants experiencing significant TAEs in the *Same SF*, *Different SF* and *Same SF crossed* conditions. However, when the mean angular separation between adaptor and tests was 75° (*Same SF orthogonal*), no reliable TAEs were observed. These non-specific TAEs could not have arisen simply due to some response bias, as we adopted a design that is immune to non-perceptual sources of bias (Morgan et al., 2015).

Can these non-specific TAEs be attributed to spillage from retinotopic adaptation? Firstly, it is possible that participants moved their eyes during the adaptation period, which could have resulted in retinal overlap between the adaptor(s) and tests. However, in all but the *Same SF orthogonal* condition, half the participants' gaze positions were monitored throughout each trial. Yet, all gaze monitored participants (except JS in *Same SF crossed*) experienced significant TAEs. Secondly, one could argue that the fixation monitoring criteria used here is less stringent given that the adaptor in the *Same SF* and *Different SF conditions* were abutting the tests, although not overlapping. If that is the case, and the non-specific TAEs here were mediated by retinotopic adaptation, it would be expected that adding a 2-octave spatial frequency difference between the adaptor and tests would reduce the size of the TAE by less than 50% compared to a case where both have identical spatial frequencies (Ware & Mitchell, 1974). However, we found that this is not the case. Mean TAEs were similar in both *Same SF* and *Different SF* conditions. This supports the fact that some form of weak retinotopic adaptation in V1 cannot explain non-specific TAEs reported here.

Alternatively, non-specific TAEs could result from adaptation in orientation selective units beyond V1 with larger receptive fields, such as those found in V2, V3 or V4 (Boynton & Finney, 2003; Smith et al., 2001). For example, V4 neurons are orientation and spatial frequency selective, with broader spatial frequency tuning (Desimone & Schein, 1987), and therefore similar TAEs for the *Same SF* and *Different SF* does not really rule out the possibility that adaptation might be happening in these regions. However, if this is in fact the case, then further increasing the cortical separation between adaptors and tests should dissipate any small-magnitude TAEs. In fact, studies that measured the spatial extent of the retinotopic TAE have reported TAEs of ~1° when adaptors and tests were spatially separated by a CTC distance of 4°, but no TAE with larger separations of 7° (Melcher, 2007).

To assess if adaptation is occurring in orientation selective units with relatively larger receptive fields, in the *Same SF crossed* condition, the cortical separation between adaptor and test was increased by presenting them in opposite sides of the vertical and horizontal meridians. Although the CTC spatial separation was either 3.70° or 4.95°, the cortical separation was much larger. This is because of discontinuities in how the visual field is mapped onto V1, V2, V3 and V4 across hemifields (Sereno et al., 1995). Stimuli adjacent in space would be represented by neighbouring neurons if they are presented within the same hemifield (left or right of vertical meridian or above or below horizontal meridian), but if they are presented in different hemifields, despite being close in space, they would be encoded by neurons that are cortically much farther apart. In fact, Liu, Jiang, Sun, and He (2009) successfully utilised this dissociation between spatial and cortical distance to examine the mechanisms of crowding. Further, the crossed stimulus configuration was chosen in order to maintain identical test stimulus eccentricity (from fixation) across all 4 experimental conditions. Nonetheless, despite a

large cortical separation, most participants experienced significant TAEs, where mean repulsion and conspicuousness was significant, thus showing that non-specific TAEs could not have been mediated by orientation selective neurons with larger receptive fields than those found in V1.

A further possibility is that adaptation in one cortical level could have modulated activity of neurons in other cortical levels. For example, Xu et al. (2008) showed that adapting to a concave or a convex curve, a low-level feature, biased the perception of facial expressions of emotion, a higher-level feature. This shows that adaptation transfers between cortical levels, in this case in the form of feedforward propagation. With regard to orientation processing, Roach, Webb, and McGraw (2008) showed that adapting to concentric or radial patterns displaying global form caused TAEs, where the perceived orientation of subsequently presented gratings at unadapted locations (that matched empty regions of adapting patterns) was biased. These non-specific TAEs were not tuned to spatial frequency. It is believed that such TAEs are likely caused by higher-level form processing mechanisms attempting to fill-in empty regions within global patterns by means of extrapolation, which in turn results in feedback-modulation of activity in orientation selective V1 neurons from which form processing regions pool information (Roach & Webb, 2013; Roach et al., 2008). However, when the adapting patterns did not contain a global structure (e.g., a large sinusoidal grating) no non-specific TAE was observed. Therefore, it is unlikely that top-down feedback mechanisms from form processing regions account for non-specific TAEs reported here.

On the other hand, Liu and Hou (2011) reported non-specific TAEs of the magnitude 1-1.5°. They presented adaptors in one of the 4 quadrants of the screen at an eccentricity of 7° from fixation. Adaptors always had two oppositely oriented gratings (15° CCW

and 15° CW of vertical) superimposed and participants were asked to pay attention to one of them. The attended grating caused repulsive TAEs on vertical tests presented in unadapted quadrants at similar eccentricity. The size of these non-specific TAEs were unaffected by presenting adaptor and test in either the same or in different hemifields. Hence, for attention modulated TAEs, the cortical distance appears to be negligible. The authors attributed the TAEs to attentional modulation of neural activity within V1, a process that they believe is global (i.e., not restricted to the adapted cortical location). Moreover, attention is also found to result in motion after-effects that are spatially non-specific, showing that attentional modulation can be generalised across many low-level features (Liu and Mance, 2011). With the current set of data, it is difficult to firmly attribute TAEs reported here to attentional modulation, given that attention was not manipulated. However, it still remains a plausible candidate given that no other stimuli competed for attention during adaptation.

When participants adapt to low-level features like orientation, in addition to adapted neurons, it is possible that unadapted neurons (in distant cortical locations) with preferred orientations closer to the adaptor also change their response state. In support of this, a recent physiological study on cats' striate neurons reported that adapting to oriented stimuli changes the orientation preferences of cortically distant unstimulated channels up to 15° away from the adapted site (Bachatene, Bharmauria, Cattan, Rouat, & Molotchnikoff, 2015). This is similar to the changes in orientation preference reported in orientation-adapted V1 channels (Dragoi et al., 2000; Jin et al., 2005). However, it is unclear how such distant modulations occur within V1. A possible mechanism may involve long-range horizontal connections within V1 that connect orientation columns of like orientation preference (Gilbert & Wiesel, 1979, 1989; Ts'o, Gilbert, & Wiesel, 1986; Weliky, Kandler, Fitzpatrick, & Katz, 1995). These extend

over several degrees of visual angle between cells with non-overlapping receptive fields (Livingstone & Hubel, 1984). However, this connectivity must be carefully examined by means of physiological measures in the context of TAEs to arrive at a valid conclusion about its role in non-specific TAEs.

In summary, our results reveal some important characteristics of TAEs to local orientation with respect to positional selectivity. Although the origin of these non-specific TAEs remains unclear, these findings provide useful guidance in interpreting after-effects to image uprightness in the next chapter.

3. Chapter 3 - Selective encoding of uprightness: Tilt after-effects to images of buildings

3.1. Introduction

The aim of this chapter is to examine the selectivity of our visual system to the global orientation ("uprightness") of natural images which is suggested to correspond to the canonical representation of some categories of images that we hold in our long-term memory (Tarr & Pinker, 1989). Other images such as a ball do not have a clear upright posture. Although the representation of an image, and consequently the uprightness of it, is determined by the geometry formed by the local orientations present within it, uprightness of an image and the orientation of its local edges can be perceptually dissociated. For example, an image of a face or a bottle will have an upright percept in its canonical posture but will appear inverted when it is rotated 180° in the fronto-parallel plane. However, the percept of a local edge in one of those images that is initially vertical will remain vertical even after rotating 180°.

3.1.1. Uprightness versus subjective vertical

Dyde, Jenkin, and Harris (2006) distinguished between uprightness and subjective vertical and measured uprightness using a letter naming task. The idea was that recognition of a letter depends on its orientation (e.g., when the letter 'd' is rotated 180° it becomes letter 'p'). An index of uprightness was obtained by measuring the orientation at which participants were equally likely to judge the letter as a d or a p. Subjective vertical (SV) was measured by asking participants if a line stimulus was clockwise or counter-clockwise of vertical. An index of SV was obtained by measuring the orientation at which participants were equally likely to judge a line to be clockwise

or counter-clockwise. The authors found that uprightness and SV are perceptually distinct by showing that when participants laid right-side down (a body posture parallel to the ground surface), their SV judgements were biased towards gravity (perpendicular to the ground) and uprightness judgements were biased towards the axis of the body (parallel to the ground). They found gravity and body axis to differently affect uprightness and SV. Therefore, they suggested that uprightness is a unique perceptual property.

3.1.2. Functional significance of uprightness

Image uprightness is important for several reasons. Firstly, perceived global orientation of an image provides visual information about the direction of gravity, which in turn informs self-orientation relative to gravity. This is particularly relevant when gravity information provided by other sensory sources is discordant (Howard & Childerson, 1994). Secondly, judgements of subjective visual vertical are affected by the uprightness of background images, which serve as a global frame of reference for perceptual judgements (Asch & Witkin, 1948; Haji-Khamneh & Harris, 2010). Finally, it has been reported that scene orientation affects how people deploy overt attention within a scene, where scene-centric directional asymmetries of eye movements always remain aligned with the orientation of the scene (Foulsham & Kingstone, 2010; Foulsham, Kingstone, & Underwood, 2008).

Furthermore, image uprightness is crucial for the classification of many image types. Faces for example, are recognized more accurately when they are upright as opposed to when they deviate from upright (Hochberg & Galper, 1967; Yin, 1969). This also applies to other types of natural images. Jolicoeur (1985) showed that when line

drawings of images (e.g., animals, furniture, vehicles) deviated from canonical upright postures, the time required to name images increased monotonically with increasing deviations within 120° from upright. Recently, Loschky, Ringer, Ellis, and Hansen (2015) studied the effect of uprightness on scene classification and found that classification accuracy reduced as scene orientations deviated from upright up until 135°. One of the theories of image classification posits that we hold templates of images in our long-term memory, and we attempt to classify images in different orientations by means of mentally rotating these images via the shortest angular route to match canonical orientations (Jolicoeur, 1985; Jolicoeur & Milliken, 1989; Tarr & Pinker, 1989). However, to be able to optimally perform mental rotation one should be aware of the image's current state of uprightness. This suggests the need for a mechanism that encodes image uprightness and to date it is unclear if we possess one.

In order to examine if we have a mechanism for scene uprightness, we will use adaptation to measure TAEs to "uprightness" of natural images. Recently, Dekel and Sagi (2015) demonstrated that, adapting to synthetic $1/f^{2.5}$ noise patterns and patches of unaltered natural images (e.g., animals, plants) with Fourier power distributions biased at 25° off vertical induced TAEs on vertical Gabor tests. In a similar vein, Goddard, Clifford, and Solomon (2008) demonstrated tilt illusions that mimic the classic angular function of TAEs using orientation filtered natural images as surround and test stimuli. Although these studies demonstrate that the characteristics of the classic TAE obtained with synthetic stimuli extend to natural / naturalistic stimuli, they can be accounted for exclusively by local orientations within the stimuli, and do not speak to uprightness *per se*.

Here we examine if we have a mechanism selective for uprightness, by studying TAEs to uprightness of natural images. Using after-effects, psychophysicists have inferred the existence of neural selectivity to a range of complex natural image attributes like the gender of a face (see section 1.2.4). However, it is possible that some of these after-effects might be the result of adaptation in "low-level" visual mechanisms, tuned to stimulus values that have nothing to do with the complex property *per se*. For example, in the case of after-effects to face gender, if adapting to a thick, masculine eyebrow suppresses a few neurons that prefer (low spatial frequency) shapes like that, then a subsequently viewed, androgynous eyebrow (with a slightly higher spatial frequency) will appear much thinner, making a face it is on appear more feminine. Thus, inferring neural mechanisms from perceptual after-effects is not always as straightforward as one might hope.

Inferring neural selectivity from psychophysics is complicated, not only because after-effects can reflect adaptation by low-level mechanisms, but also because many conventional measurements of appearance are susceptible to contamination from non-perceptual sources of bias (e.g., expectation effects and response biases; (Storrs, 2015)). In this study, we minimize the influence of low-level adaptation by presenting adaptor and tests in different regions of the visual field and / or different regions in frequency space. Adaptor and tests were separated in frequency space by filtering adaptors and tests to have different spatial frequency content or different orientation content. We minimize the influence of non-perceptual sources of bias by adopting the recently developed, two-alternative, forced-choice (2AFC) comparison-of-comparisons paradigm, with roving pedestals (Morgan et al., 2013; Yarrow, Martin, Di Costa, Solomon, & Arnold, 2016).

In Experiment 1 we confirm that the TAE for natural scenes can be obtained with different (and differently sized) adapting and test images, which are presented in a partially overlapping spatial configuration but share minimal spatial frequency components due to spatial frequency filtering. In Experiment 2, we examine whether the TAE for uprightness arises because of interactions between mechanisms selective for natural scenes, or whether it is simply a by-product of suppression between lower-level mechanisms. To disentangle these possibilities, we use orientation-filtered and phase-scrambled stimuli. Vertically filtered images are designed to have a negligible effect on the responsivity of low-level mechanisms tuned to near-horizontal orientations. Phase-scrambled stimuli are designed to have a similarly negligible effect on the responsivity of mechanisms selective for natural scenes.

3.2. Methods

Participants

A total of 23 participants (18 – 46 years of age), each having a unique two-character set of initials (see figures 3.2 and 3.3), from Queen Mary University of London, with normal or corrected-to-normal visual acuity took part in the experiments. Procedures were approved by the Queen Mary University of London research ethics committee and written informed consent was obtained from all participants. The number of participants for each experimental condition was determined based on previous studies investigating higher-level visual after-effects, which involved from 5 to 10 participants per condition (Melcher, 2007; Roach et al., 2008; Xu et al., 2008).

Experimental set-up and apparatus

Participants were seated in a dark room, and were instructed to keep their head upright and maintain the same distance from the screen throughout the experiment. Stimuli were presented on a 20" Iiyama CRT monitor with a 1600 × 1200 screen resolution and a refresh rate of 60 Hz. The viewing distance was approximately 57 cm, such that each pixel subtended 1.5 arcminutes. A black mask with a circular aperture (diameter = 24.5°) was overlaid on the monitor to eliminate the use of monitor edges as cues to vertical or horizontal. Stimulus presentation and data collection used Matlab (Mathworks) and Psychtoolbox (Brainard, 1997; Pelli, 1997).

Stimuli

Images of 5 different houses (Fig. 3.1B), in their frontal views, appearing to be at eye level from a standing position, were obtained from an archive of the Caltech Computational Vision Group (available online at http://www.vision.caltech.edu/archive.html). We used images of houses because: 1) scene orientation of man-made scenes is judged with better discrimination precision than non-man-made scenes (Haji-Khamneh & Harris, 2010) and 2) houses have a clear frontal facade and cover limited depth, resulting in minimal linear perspectives. The images were initially cropped to a square aspect ratio and then resized to 300 × 300 pixels using bicubic interpolation. Cropped images were converted to grayscale by independently weighting and summing the red, green and blue channels of the image according to the CIE procedure (0.299 × R + 0.587 × G + 0.114 × B; Hughes et al., 2013). These images were presented as adaptors within a hard-edged circular aperture (diameter = 7.5°; Fig. 3.1A). The test images were resized to 75% of the adaptor's size and presented within a hard-edged window of diameter 5.7°.

Images of houses were tilted and, in some cases, filtered. Filtering was a 7-step procedure. In step 1 the mean graylevel of a tilted image was subtracted, creating a difference image with no DC component. In step 2 this difference image was multiplied with a 2-dimensional, separable cosine window of the same size. In step 3 the windowed image was Fourier transformed (applying the cosine window before Fourier transformation helps to reduce wrap-around artefacts). In step 4 the transformed image was multiplied by one of the filters described below. In step 5 the product was inverse-Fourier transformed. In step 6 the image was scaled such that adaptors would have a root mean square (RMS) contrast of 0.10 and tests would have an RMS contrast of 0.18. Finally, in step 7, a graylevel of 0.50 was added to each image. This matched the graylevel of the screen background.

Procedure

Trials were blocked by condition (there were three conditions in Experiment 1 and two conditions in Experiment 2) and adaptor orientation: either –15° or +15°. By convention, we consider tilts clockwise (CW) from vertical to be negative and tilts counter-clockwise (CCW) from vertical to be positive. Each condition in Experiment 1 and 2 was also associated with a "baseline block" in which no adaptor was shown.

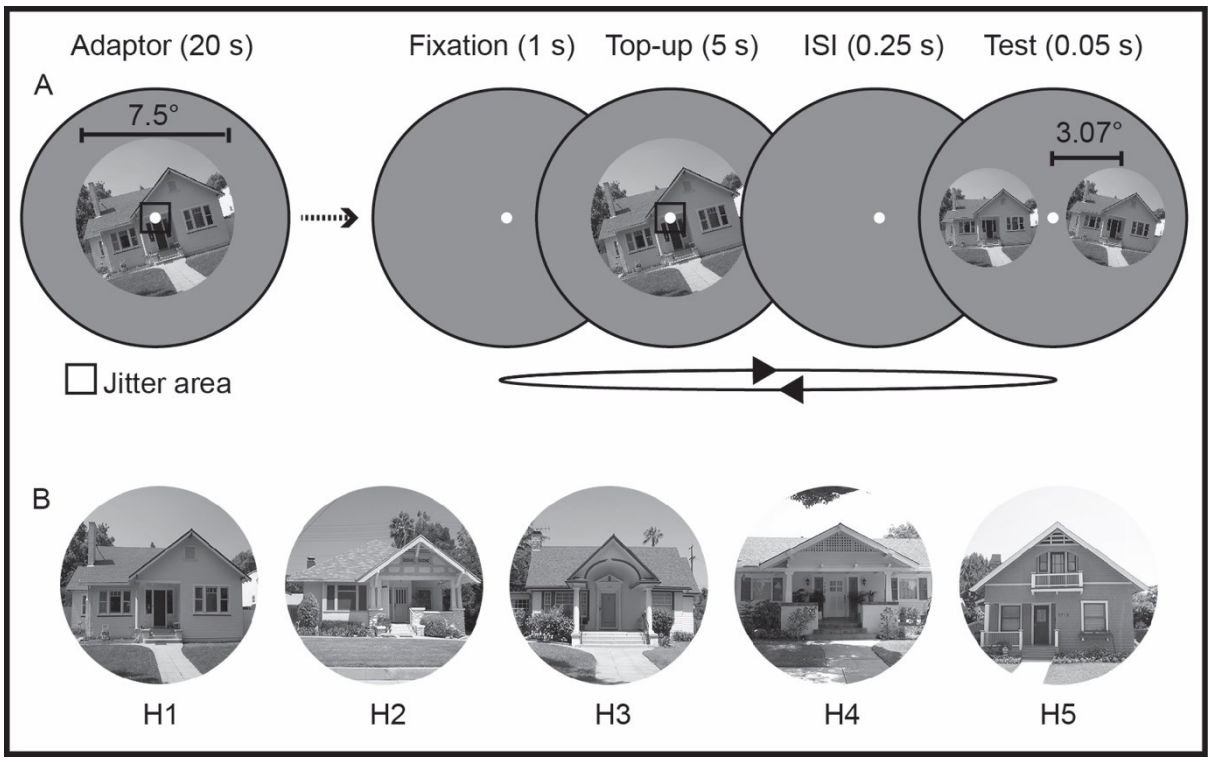


Figure 3.1. TAEs to uprightness methods: (A) Stimulus configuration and timeline of a sample trial from Experiment 1. (B) Five different house scenes used across the different conditions in the study.

The general procedure is outlined in Fig. 3.1A. Participants were instructed to fixate a centrally presented white circle (diameter = 0.2°) for the duration of each block. All blocks (except baseline blocks) began with an initial adaptation phase of 20 s. Following this, each test trial started with a "top-up" adaptation phase of 5 s. During adaptation phases, the adaptor was jittered every 0.5 s by recentering it on a random pixel within a predefined jitter area of 0.25° × 0.25° surrounding fixation. Top-up adaptors were followed, after 0.25 s, by two test houses, presented immediately to the left and right of fixation, for 0.05 s. One of the test houses was the "pedestal," with one of two fixed tilts: –3° or +3°. The other test was the "comparison," with an offset added to the fixed tilt, randomly selected from the set {–15°, –12°, –9°, –6°, –3°, 0°, +3°, +6°, +9°, +12°, +15°}. Each combination of pedestal and comparison tilt was tested 10 times, resulting in 220 trials per block. The spatial positions (left and right of fixation)

of the pedestal and comparison were randomized on every trial. Participants chose which of the two test houses appeared more upright, using keys "1" (for left) and "2" (for right). Participants were told that an upright house is how they would imagine it to appear, if they stood in front of it with their head held straight.

As is evident from Fig. 3.1A, there was a small amount of spatial overlap between the adaptor and tests. However, the overlapping parts of the images were not the same (e.g., the right half of the adaptor overlapped with the left half of one test) and images were of different sizes to reduce retinotopic adaptation (Webster & MacLeod, 2011).

Methods specific to Experiment 1

In the *same house* condition image H1 was used for both adaptor and test stimuli. In the *different house* condition image H2 was the adaptor and image H3 was used for the tests (Fig. 3.1B). In the *different SF house* condition the adaptor and test stimuli were images of the same house, but filtered to separate them for their spatial frequency (SF) content (Fig. 3.2B). In this condition, three different house images were used (H2, H4 & H5; Fig. 3.1B). Two participants were tested with H2, two with H4 and two with H5.

Log-normal filters were used for the *different SF house* condition. The filter used for adaptors had a peak SF of 10 cycles / degree. The filter used for the tests had a peak SF of 1.25 cycles / degree. Both filters had a half-bandwidth at half-height of 1.5 octaves.

Methods specific to Experiment 2

All 10 participants participated in both the *orthogonal house* condition and the *phase-scrambled house* condition. In both conditions adaptors were first tilted (either CW or CCW) and then filtered to retain Fourier amplitudes close to vertical orientations (Fig.

3.3). Tests were upright images of the same house, initially filtered horizontally and then tilted by different amounts in each trial, as in Experiment 1. Five participants were tested using H1; the other five were tested using H2. For each participant, the adapting and test stimuli were differently filtered versions of the same house image. In the orientation domain, each filter was a Gaussian function of angle, centred on 0° (for the vertically filtered adaptors) or 90° (for the horizontally filtered tests); with a half-bandwidth at half-height of 23.5° and was clipped at ± 40° from the peak, resulting in zero gain at orientations beyond the clip. In the *phase-scrambled* condition, tilted adaptors were phase-scrambled prior to orientation filtering, by adding the Fourier phase spectrum of a unique white noise pattern (having the same dimensions as the image and with a uniform distribution of pixel intensities ranging from 0 to 1) to the phase spectrum of the image. The amplitude spectra and RMS contrast of adaptors in the *phase-scrambled house* condition matched the amplitude spectra and RMS contrast of adaptors in the *orthogonal house* condition. Identical (unscrambled), horizontally filtered, tilted tests were used in both conditions.

Psychophysical model

Data were analysed within the context of signal-detection theory, as described by Morgan et al. (2015). Within this model, the appearances of pedestal (*S*) and comparison (*C*) are normally distributed, i.e., $S \sim N(p + \mu, \sigma^2/2)$ and $C \sim N(p + \mu + t, \sigma^2/2)$, where σ^2 is the variance of the performance-limiting noise, p is the pedestal tilt, t is the offset added to the comparison, and μ is the perceptual bias specific to each test block. If there were no perceptual bias, then the distributions for pedestal and comparison would have means of p and $p + t$ respectively. The participant chooses the pedestal as closer to upright when it appears less tilted than the comparison. Accordingly, the probability of this choice $P("S") = P(|S| < |C|) = P(S^2/C^2 < 1)$,

has a doubly non-central *F* distribution. This distribution's denominator's noncentrality parameter is $2(p + \mu + t)^2 / \sigma^2$, its numerator's noncentrality parameter is $2(p + \mu)^2 / \sigma^2$, and both denominator and numerator have 1 degree of freedom.

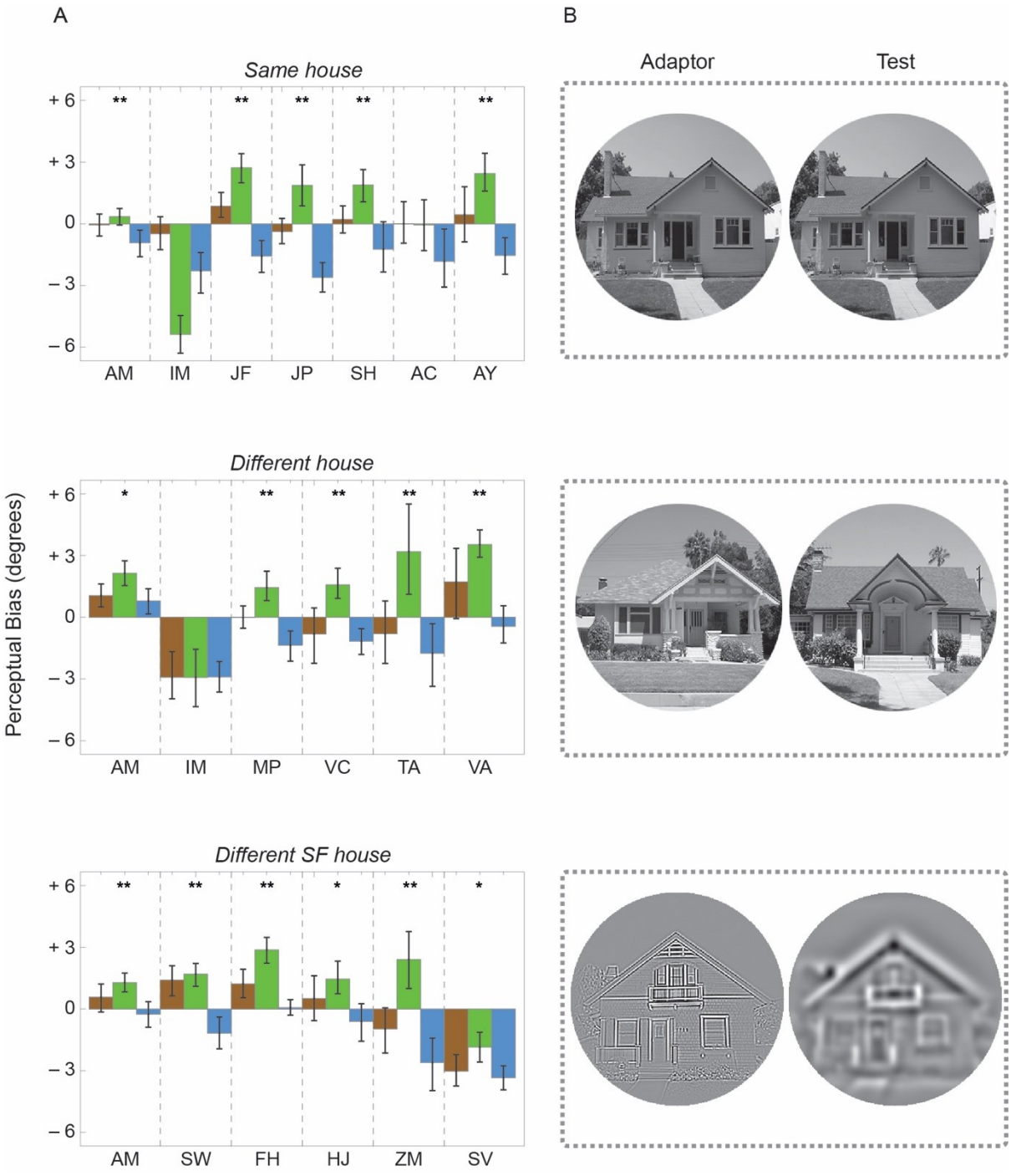


Figure 3.2. (A) Maximum likelihood estimates of perceptual bias for baseline (brown), CW (green) and CCW (blue) blocks from the same house, different house and different SF house conditions in Experiment 1. Error bars are bootstrapped 95% confidence

intervals. Single asterisks (*) denote after-effects significant at the $\alpha = 0.05$ level for repulsion. Double asterisks (**) denote after-effects also significant at the $\alpha = 0.001$ level for repulsion. (B) Examples of adaptors and test stimuli used in each of the conditions tested (where necessary, contrast has been amplified for visibility).

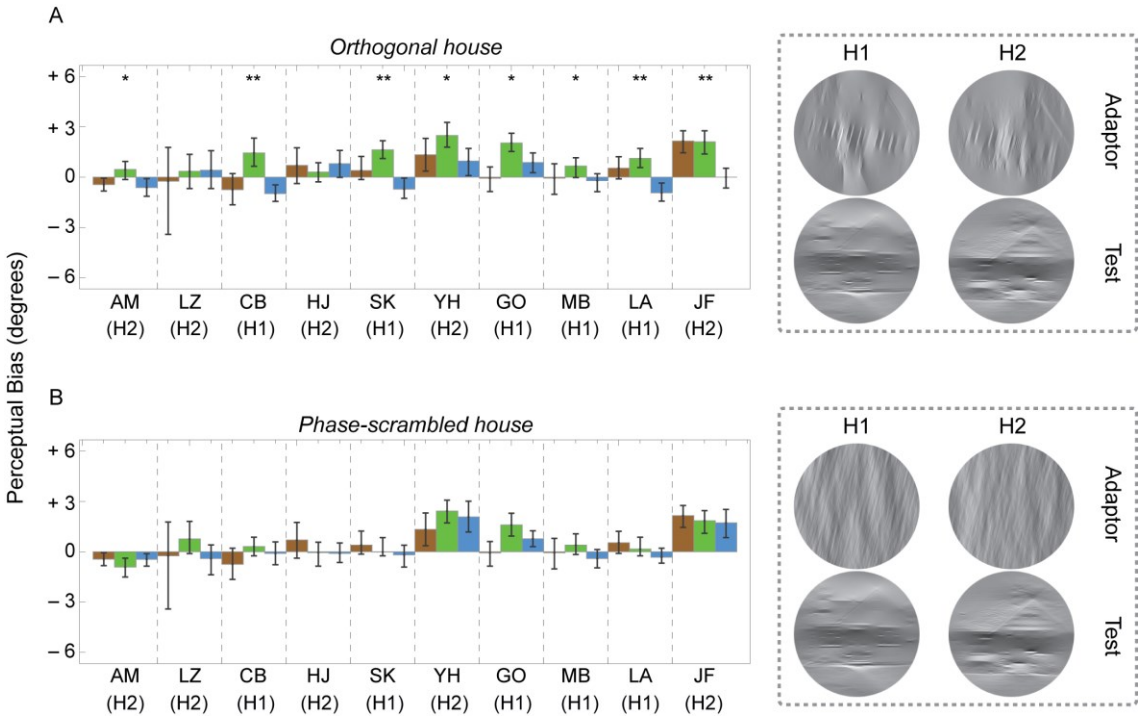


Figure 3.3. Maximum likelihood estimates of perceptual bias for baseline (brown), CW (green) and CCW (blue) blocks from (A) the *orthogonal house* and (B) the *phase-scrambled house* conditions in Experiment 2. Error bars are bootstrapped 95% confidence intervals. Single asterisks (*) denote after-effects significant at the $\alpha = 0.05$ level for repulsion. Double asterisks (**) denote after-effects also significant at the $\alpha = 0.001$ level for repulsion. Examples of CW-tilted adaptors with untilted test stimuli used in each condition are illustrated to the right. The image number used for each participant is given below their initials.

3.3. Results

From each block of trials (baseline, CCW and CW), we obtained maximum-likelihood estimates of bias μ and the variance of performance-limiting noise σ^2. Negative biases with CCW adaptors and positive biases with CW adaptors are indicative of the repulsive TAE. Non-parametric bootstrapping (with bias-correction (Efron and Tibshirani, 1994)) was used to quantify the reliability of parameter estimates. The error bars shown in figures 3.2 and 3.3 contain the resultant 95% confidence intervals.

We also fit each participant's data from CCW-adaptor and CW-adaptor blocks simultaneously, forcing the bias parameter μ to be the same in both cases, but allowing σ to vary. The ratio L, between the likelihood of this nested model fit and the joint likelihood of the aforementioned separate fits to the same data is necessarily no greater than 1. To evaluate the "null" hypothesis of no significant TAE in individual participants, we compare the criteria $\alpha = 0.05$ and $\alpha = 0.001$ to the value $1 - F(-2 \ln L)$, where F is the cumulative chi-square distribution, with 1 degree of freedom. This is known as the generalized likelihood-ratio test (see Mood, Graybill & Boes, 1974, p.440–441).

To evaluate null hypotheses at the group level, we performed two-tailed one-sample *t*-tests using estimates of repulsion, which can be quantified either in degrees of tilt or in terms of the "just-noticeable difference" (JND). A single value for repulsion, in degrees of tilt, can be obtained by subtracting one maximum-likelihood estimate of μ (the one obtained with CCW adaptors) from the complimentary estimate (obtained with CW adaptors), and dividing the difference by 2. The "conspicuousness" of repulsion can be quantified by further dividing this quotient by the JND. For the latter, we use the root-

mean-square of the maximum-likelihood estimates of σ. Results of the group-level *t*-tests appear in tables 3.1 and 3.2.

Table 3.1. *Group level statistics for repulsion in Experiment 1 and 2*

			Repulsion (R)						
	Condition	N	mean R (°)	t - statistic ($R > 0$)	p - value	Cohen's d	paired t - statistic	p - value	Cohen's d
Experiment 1	*Same house*	7	1.13	2.25	0.066*	0.85			
	Different house	6	1.31	3.62	0.015	1.48			
	Different SF house	6	1.31	4.90	0.004	2.00			
Experiment 2	*Orthogonal house*	10	0.65	4.11	0.003	1.30	2.42	0.039	1.16
	Phase-scrambled house	10	0.20	2.68	0.025	0.85			

Notes: N denotes the number of observers in each condition. The asterisk (*) denotes that the p value was approaching significance. Removing observer IM from analysis makes the $p = 0.002$. All t-tests are two-tailed.

Table 3.2. *Group level statistics for conspicuousness in Experiment 1 and 2*

			Conspicuousness (*CI*)						
	Condition	*N*	mean *CI* (JND)	*t* - statistic (*CI* > 0)	*p* - value	Cohen's *d*	paired *t* - statistic	*p* - value	Cohen's *d*
Experiment 1	*Same house*	7	0.26	2.42	0.052*	0.91			
	Different house	6	0.27	4.24	0.008	1.73			
	Different SF house	6	0.33	5.84	0.002	2.38			
Experiment 2	*Orthogonal house*	10	0.21	4.36	0.002	1.38	2.88	0.018	1.30
	Phase-scrambled house	10	0.06	2.45	0.037	0.77			

Notes: *N* denotes the number of observers in each condition. The asterisk (*) denotes that the *p* value was approaching significance. Removing observer IM from analysis makes the $p = 0.003$. All *t*-tests are two-tailed.

Experiment 1

Estimates of bias (μ) from Experiment 1 are plotted in Fig. 3.2A. For the majority of participants, adaptation to a house tilted 15° (CCW of upright) produced a negative bias (relative to the baseline's bias) in subsequently viewed test houses, and adaptation to a house tilted –15° produced a positive bias. Generalized likelihood ratio tests suggest after-effects significant at the $\alpha = 0.05$ level for repulsion in the data from 5 of the 7 participants in the *same house* condition, 5 of the 6 participants in the *different house* condition, and all 6 of the 6 participants in the *different SF house* condition. Group-level statistics appear in tables 3.1 and 3.2.

Experiment 2

Estimates of bias from Experiment 2 are plotted in Fig. 3.3. Generalized likelihood ratio tests suggest after-effects significant at the $\alpha = 0.05$ level for repulsion in the data from 8 of the 10 participants in the *orthogonal house* condition and none of the (same) 10

participants in the *phase-scrambled house* condition. Group-level statistics appear in tables 3.1 and 3.2. At the group level, both conditions produced mean repulsion and conspicuousness significantly larger than zero. However, a comparison using a paired-samples *t*-test between the means of the two conditions revealed that the *orthogonal house* condition produced a significantly larger repulsion compared to the *phase-scrambled house* condition (tables 3.1 & 3.2).

3.4. Discussion

Results reported in Experiment 1 demonstrate that the TAE for natural images (houses) can be obtained with partially overlapping, yet different (and differently sized) adapting and test images, widely separated in spatial frequency content. Similar results have been obtained with sinusoidal gratings (in chapter 2 and also by others such as Melcher, 2007; Liu and Hou, 2011) and with circular / radial patterns (Roach et al., 2008). When after-effects survive manipulations of image, size and spatial frequency, their origin cannot be attributed to low-level visual mechanisms (Webster & MacLeod, 2011). Experiment 1's results extend Dekel & Sagi's (2015) findings of TAEs with natural images as adaptors and sinusoidal gratings as tests, by showing that adaptation to global orientation can occur between adaptors and tests that are natural images. However, it is unclear from Experiment 1 whether the TAE for natural images arises because of interactions between high-level mechanisms selective for natural images, or whether it is simply a by-product of suppression between mid-level mechanisms, selective for spatial orientation in general.

To distinguish between these alternatives, in Experiment 2, we applied perpendicular filters to the stimuli, widely separating the orientation contents of adaptor and tests. The

finding of a repulsive TAE in this condition qualitatively differs from the assimilative "indirect effect" found when retinally overlapping lines or gratings are separated between 60° and 87.5° (Gibson & Radner, 1937). We attribute this repulsion to the images' recognisability as slightly tilted scenes, rather than their Fourier image components. In support of this viewpoint, we found no after-effect at the individual participant level when the Fourier phases of the adaptors were scrambled. However, the group level analyses did reveal a relatively small but significant TAE (tables 3.1 & 3.2), with phase-scrambled adaptors. This must be attributed to Fourier image components. A possible reason for this is that since man-made images are usually dominated by cardinal orientations, a sense of global tilt is still apparent in the images even after randomizing Fourier phase information (see figure 3.3B, where randomized images might appear tilted CW).

The most interesting finding is that vertically filtered houses induce repulsive TAEs on horizontally filtered houses. These TAEs were not only evident in most participants, but they were also much larger than the TAEs from phase-scrambled adaptors at the group level. Comparing this with non-specific adaptation to meaningless gratings reported in chapter 1, when adaptor and test gratings were orthogonal, it did not produce any TAE. Although the orientation-filtered houses are not as easily recognizable as their unfiltered counterparts, they possess clear higher-order structure, which is lacking in the phase-scrambled versions used for adaptation. Textures with similar higher-order (meaningless) structure are also more effective than phase-scrambled scenes as backward masks of 'scene gist' (Loschky, Hansen, Sethi, & Pydimarri, 2010). This suggests that textures with higher-order structure are fundamentally different from phase-randomized stimuli with similar orientation statistics. Nonetheless, the after-effect of adapting to tilted buildings is different from the after-effect elicited by the

perception of a global form contained in meaningless textures. Whereas Experiment 2 showed that the former can survive large differences between the orientation contents of adaptor and test, the latter cannot (Roach et al., 2008).

This chapter's results are unique in the literature on the appearance of uprightness, because they show that the global orientation of a scene can be encoded separately from its local feature content. It is assumed that information about image orientation is embedded in the early global percept of image layout, a property which is rapidly extracted when looking at a meaningful image like a scene (Foulsham & Kingstone, 2010; Greene & Oliva, 2010). Based on this assumption, at present, we can only speculate regarding where selectivity for the orientation of natural images arises in the brain. One possible candidate is the Parahippocampal Place Area, which is thought to encode scene layout rather than object content and shows greater sensitivity to images of buildings like houses as opposed to other objects like faces (Epstein & Kanwisher, 1998). In support of this, such scene selective regions are known to respond equally to scenes containing only close-to-vertical or close-to-horizontal orientations (Watson, Hymers, Hartley, & Andrews, 2016), akin to the stimuli we used here. Different local feature content can therefore lead to the encoding of similar global spatial layout in scenes, which presumably is what led to a repulsive TAE from vertically filtered adaptors on horizontally filtered tests.

As noted in chapter 2, the TAE is routinely invoked as a manifestation of the mutual inhibition between visual mechanisms selective for orientation. Consequently, the natural conclusion to draw from the results is that there must be mechanisms selective for the orientations of images with meaningful, higher-order structure. Of course, we cannot say whether those mechanisms are mutually inhibitory, or whether the TAE for

natural scenes should be attributed to their modulation of lower-level mechanisms. Indeed, other authors have invoked pre-saccadic remapping in space (Melcher, 2007), top-down modulation of low-level feature detectors through feedback from form processing regions (Roach et al., 2008) and selective attention (Liu and Hou, 2011) in attempts to explain how the TAE can survive the spatial separation of adaptor and tests.

One further possibility is normalization. Extensive real-world experience with close-to-upright scenes (canonical orientation) may have resulted in the establishment of uprightness as a norm against which other orientations are compared. Exposure to tilted scenes may simply shift the subjective norm of uprightness towards the tilted direction, which then results in an objectively upright scene seen as tilted away. Indeed, Asch and Witkin (1948) report that tilted scenes eventually appear upright over extended viewing, implying normalizing towards uprightness.

4. Chapter 4 – Orientation sensitivity during image classification

4.1. Introduction

Perception of a natural image necessarily involves encoding of (some of) its low-level structures that vary in contrast, orientation and spatial frequency (SF). Consequently, it is imperative to understand to what extent our perception or identification of faces, objects and scenes is dependent on the low-level information contained within these images. This can be examined from two perspectives. Firstly, we can ask to what extent invariable higher-level representations can be achieved despite variability in the types and amounts of low-level information present in an image, such as when an image is partially degraded or filtered. Secondly, we can ask to what extent higher-level representations are influenced by non-uniformities in how the brain encodes low-level features and by non-uniformities in the distribution of low-level features present in the environment. These two sources of non-uniformity are linked; it is believed that the visual system has evolved to efficiently capture information from our environment, and this results in non-uniformities in how the visual system encodes this environmental information (Simoncelli & Olshausen, 2001).

4.1.1. Image classification as a tool to study higher-level representations

Studying categorical perception of natural images, either psychophysically or physiologically, is one way to answer the aforementioned questions. Consider the set of images shown in Fig. 4.1a. Images with obvious large variability in their low-level structures can be classified as belonging to the same category (e.g., “a forest”), yet small variations in structures between images can lead to distinct categorical percepts (e.g., “a pool” and “a bay”; Fig. 4.1b). Despite such differences underlying perception,

humans are remarkable in accurately classifying images, typically within the first 200 ms of seeing the image (Kirchner & Thorpe, 2006; Potter & Levy, 1969; Seeck et al., 1997; Thorpe et al., 1996). It has also been shown that we can rapidly classify images even in the near-absence of attention (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002). Given that such a complicated task can be achieved within a short period, this allows researchers to uncover the extent to which image classification relies on encoding low-level features, by systematically manipulating the low-level information in an image and measuring rapid classification performance.

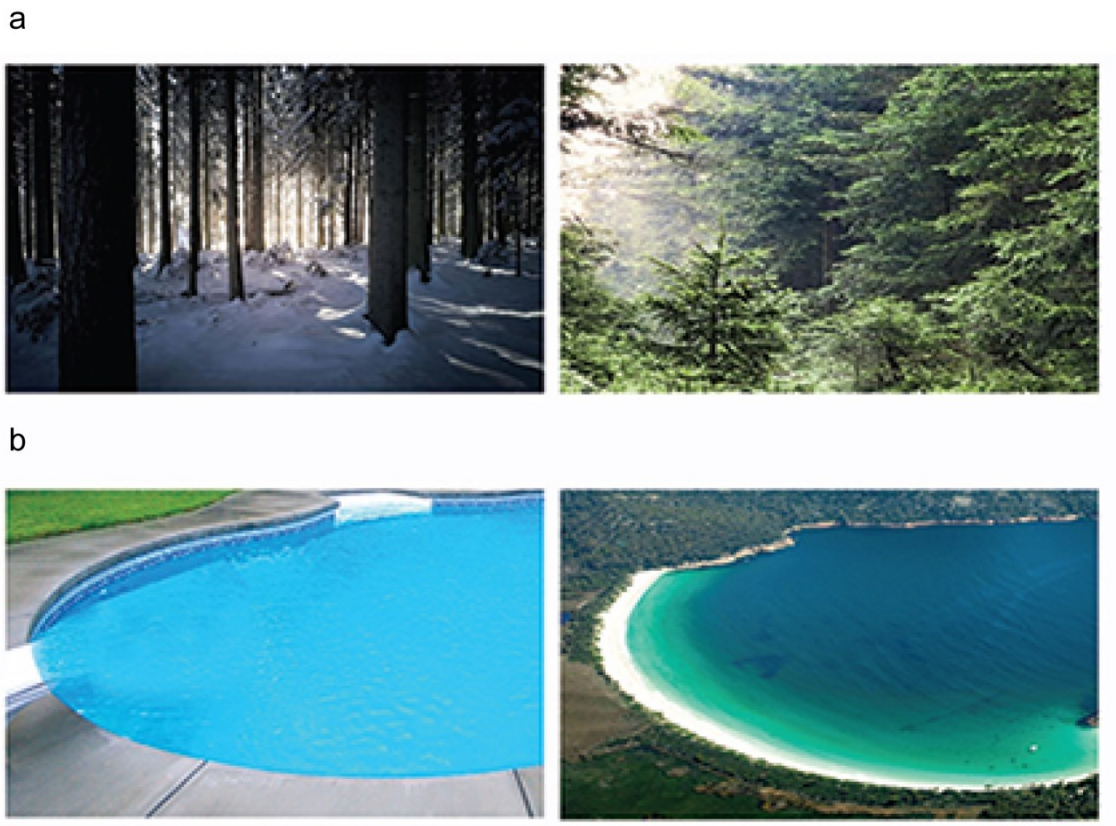


Figure 4.1. Variabilities between images and semantic categorisation: a) Large variabilities in local structures between two images can lead to the same categorical percept. Both images would be classified as a "forest". b) Small variabilities in local structures can lead to different categorical percepts of the two images as "a pool" (left) and "a bay" (right).

4.1.2. Robustness of image classification to manipulated low-level information

Image classification generally survives large manipulations in an image's low-level information. Broadly speaking, the "inverse problem" discussed in section 1.3.1 is a clear example of this. We can assign the same categorical label to images that produce different 2D retinal images as a result of changes in scale (e.g., image size changes with increasing depth), lighting conditions, and changes in observer's viewpoint (DiCarlo et al., 2012). Interestingly, we can also recognize newly learnt objects from different viewpoints, showing that viewpoint invariance is not simply achieved by matching different retinal images to image templates we might hold in our memory (Biederman & Bar, 1999). Therefore, irrespective of our experience, many different low-level representations can trigger a categorical percept.

More specific examples of image classification also provide evidence to support the robustness of categorical perception despite alterations to low-level information. For instance, it is believed that classifying an image as a face as opposed to a non-face depends more on the configuration or the arrangement of its low-level features, than the precise nature of these features (Tsao & Livingstone, 2008). Two findings support this. Firstly, even though all low-level information is preserved in an image, when the top and bottom halves of faces are misaligned, our ability to discriminate faces is impaired (Young, Hellawell, & Hay, 2013). Secondly, an image can be classified as a face even with very limited low-level information as long as the (face) specific configuration between the low level features is maintained. For example, a circle, two dots, a vertical line and a curved line do not convey any meaning on their own. However, they can be arranged in a specific manner to elicit not just the percept of a face, but also its

emotional state (smiling or sad). Churches, Nicholls, Thiessen, Kohler, and Keage (2014) showed using electrophysiological recordings that emoticons conveying an emotional expression (e.g., ":-)"), are processed in a similar manner to real faces in the brain, recruiting the same cortical sites that encode facial configuration. Further, Xu et al. (2008) showed that adaptation after-effects of facial expression (e.g., happy or sad) can be obtained from adapting to smiley faces as well as to photographs of real faces. Therefore, a simple edge-representation of an emoticon can sometimes be comparable to a real photograph of a face that is rich in many surface properties like shading and texture.

The robustness of categorical percepts to limited low-level information also applies to other natural images like objects and scenes. For example, Biederman and Ju (1988) showed that people classified images of commonly encountered objects (e.g., chair, telephone) with similar speed and accuracy when they were either line drawings or actual photographs. With respect to scenes, Walther et al. (2011) showed that scene category can be successfully decoded from brain activations in scene selective regions in response to both line drawings and normal photographs. In summary, the findings discussed in this section reveal that higher-level representations underlying accurate image categorisation can occur despite large manipulations in the low-level information typically present in a real-world image.

4.1.3. Preferential encoding of low-level features in artificial stimuli

At the earliest stages of our visual system, neurons are selective to the SF and orientation of edges (see section 1.1.4). However, these neurons do not encode all

feature values uniformly which might cause us to be more sensitive to some feature values more than others. In humans, non-uniformities in sensitivity have been examined in psychophysics by measuring contrast sensitivity to artificial stimuli like sinusoidal gratings that vary in SF and/or orientation. Contrast sensitivity for gratings is usually measured by obtaining the reciprocal of the minimum Michelson contrast required for the viewer to detect the presence of a grating, where the Michelson contrast of a grating is defined as $C_M = L_{max} - L_{min}/2\bar{L}$. Here C_M denotes the Michelson contrast, L_{max} denotes the maximum luminance of the stimulus, L_{min} denotes the minimum luminance and $\bar{L}$ denotes the mean luminance.

Campbell and Robson (1968) found that Michelson contrast sensitivity was highest for gratings with a SF of 4 cycles / degree (cpd). Sensitivity decreased monotonically as the SF increased or decreased from the optimal frequency of 4 cpd. This dependence of contrast sensitivity on SF, commonly known as the "contrast sensitivity function" (CSF), maintains the same pattern irrespective of the orientation of the grating; vertical, horizontal and 45° or 135° clockwise of vertical (Campbell, Kulikowski, & Levinson, 1966). Although the shape of the CSF remains similar, absolute values of sensitivity also depend on the orientation of the grating. Campbell et al. (1966) showed that, for a grating with a SF of 25 cpd, contrast sensitivity is best at cardinal (vertical and horizontal) orientations and decreases with increasing deviation from cardinal orientations, with the lowest sensitivity reported for inter-cardinal (45° and 135° clockwise of vertical) orientations. However, increased sensitivity to cardinals is negligible when the SF is near or less than the optimal SF and is only present at SFs higher than the optimal SF; the higher the SF the larger the cardinal advantage (Berkley, Kitterle, & Watkins, 1975; Campbell et al., 1966; Freeman & Thibos, 1975). In support of this, Li et al. (2003) found that, in the cat's striate cortex, there are more high SF

neurons tuned to cardinal orientations, and that they also have narrower orientation tuning widths, than low SF tuned neurons that generally display no orientation anisotropies.

Increased contrast sensitivity to cardinal orientations above the optimal SF is one example of the "oblique effect" proposed by Appelle (1972) where detection is in many cases found to be superior for cardinally compared to inter-cardinally oriented stimuli. Another example of the oblique effect is that our ability to discriminate the orientations of two stimuli is better near cardinal orientations, requiring a relatively smaller orientation difference between the two stimuli, compared to when stimuli are oriented near the inter-cardinal axes (Caelli, Brettel, Rentschler, & Hilz, 1983; Girshick et al., 2011).

4.1.4. Preferential encoding of low-level features in naturalistic stimuli

Much of our understanding of the early visual system is derived from experiments using stimuli such as gratings, where results are sometimes generalised to more complex, ecologically relevant stimuli. However, this approach has been challenged by recent inconsistencies with findings obtained using more naturalistic stimuli (Felsen & Dan, 2005; Olshausen & Field, 2005; Rust & Movshon, 2005). For example, some of the previously discussed studies, measuring contrast sensitivity to gratings of different orientations and SFs, assumed that perceived contrast correlates with the physical (Michelson) contrast of the stimulus. Indeed in early cortical regions (V1, V2 and V3) there is a strong association between a grating's Michelson contrast, its perceived contrast and fMRI BOLD amplitudes (Boynton, Demb, Glover, & Heeger, 1999; Campbell & Kulikowski, 1972). However, Michelson contrast is not a good predictor of perceived contrast for complex stimuli with a broader band of orientations and/or SFs

(Bex & Makous, 2002; Meese, Baker, & Summers, 2017). Bex and Makous (2002) showed that contrast detection thresholds to natural images were best predicted by root-mean-squared (RMS) contrast of the stimulus, rather than its Michelson contrast. Further, Olman, Ugurbil, Schrater, and Kersten (2004) showed that BOLD responses in V1 are proportional to the RMS contrast of natural images when they contained their unaltered amplitude spectrum, compared to when they had an amplitude spectrum that is uniform across all SFs. Natural images typically contain a characteristic $1/f^{\alpha}$ amplitude spectrum, displaying greater power at low SFs and decreasing power with increasing SF (Van der Schaaf & van Hateren, 1996), where α, the slope of a power against SF plot can range between 0.6 and 1.6 (Hansen, Haun, & Essock, 2008). Olman et al.'s (2004) findings suggest that this characteristic strongly determines the BOLD response patterns in V1.

Some studies have also challenged findings regarding the early visual system's anisotropic sensitivity to low-level features like spatial frequency and orientation, obtained using simple stimuli. Bex, Solomon, and Dakin (2009) showed that contrast sensitivity to structures of different spatial frequencies embedded in natural images could not be predicted by the CSF obtained with gratings presented in backgrounds of uniform luminance. Because of the broadband nature of natural images, interactions between different spatial frequency channels resulted in sensitivity that was disproportionately suppressed at the lower than the higher spatial frequencies. They also proposed that the CSF obtained with natural images can be partly attributed to the $1/f^{\alpha}$ characteristic of natural scenes and partly to the density of edges surrounding the target structure to be detected. Moreover, Tadmor and Tolhurst (1994) conducted an experiment where participants discriminated between two natural images that varied in the slope of the amplitude spectrum but were otherwise identical in total power. They

found that, discrimination threshold was highest for slope values typical of natural scenes. They interpreted this as evidence for a visual system that is optimised to encode naturally occurring amplitude spectra, by having stronger tolerance to variations near the optimal slope.

Regarding orientation, the oblique effect reported with narrowband stimuli does not hold for stimuli resembling naturally encountered (broadband) images. Essock, DeFord, Hansen, and Sinai (2003) showed that when participants had to detect increments of spectral amplitude at specific orientations within broadband noise patterns, sensitivity was highest for inter-cardinal orientations, with relatively reduced sensitivity to vertical orientations. They found that sensitivity was lowest for horizontal orientations, a finding they called the "horizontal effect". Even with the use of natural images with an amplitude spectrum roughly isotropic for orientation, the horizontal effect occurs (Hansen & Essock, 2004). Natural scenes, both carpentered (man-made) and uncarpentered (natural), typically display an anisotropic power spectrum with greater power near cardinal than inter-cardinal orientations (Switkes, Mayer, & Sloan, 1978). It has also been reported that there is higher spectral power near horizontal orientations than near vertical orientations in most natural scenes (Baddeley & Hancock, 1991; Hansen & Essock, 2004). Hansen et al. (2008) proposed that the horizontal effect is a result of early cortical regions optimized to encode orientations that are generally *lacking* in natural scenes, the inter-cardinals. They showed that a model of divisive contrast normalization in V1 could account for both an oblique effect with narrowband stimuli and a horizontal effect with broadband stimuli. This is essentially based on a physiological over-representation of cardinal orientations in V1, with more, and more narrowly tuned neurons encoding cardinals (Furmanski & Engel, 2000; Li et al. 2003). An over-representation, coupled with a normalizing mechanism produces BOLD

responses in V1 that mimic the psychophysical horizontal effect (Mannion, McDonald, & Clifford, 2010).

Collectively, the studies discussed in this section reveal that responsivity of the early visual system is different to natural and artificial stimuli. More importantly, this difference highlights that the visual system is optimized to encode low-level information typically present or sometimes lacking in natural scenes.

4.1.5. Do some orientations facilitate image classification?

During natural image perception, there is preferential encoding for specific edge orientations. For example, it has been demonstrated that performance in either the identification of familiar faces or discriminating between unfamiliar faces is significantly better when faces retain close to horizontal structures as opposed to structures close to any other orientations including vertical or inter-cardinal (Dakin & Watt, 2009; Goffaux & Dakin, 2010). This horizontal advantage cannot be attributed to increased sensitivity to cardinal orientations for two reasons: 1) identification performance is worse for faces retaining vertical as compared to inter-cardinal structures (Dakin & Watt, 2009) and 2) only faces retaining horizontal information preserve the face inversion effect, where face discrimination performance is poorer for inverted compared to upright faces (Goffaux & Dakin, 2010). Notably, the face inversion effect is generally used to indicate the involvement of special face-specific mechanisms in classification, because inverting faces disproportionately impairs their classification as compared to other types of objects (Valentine, 1988). However, this horizontal advantage for faces is not the result of increased sensitivity to horizontal orientations *per se*. Structures near the horizontal band are more informative during face

perception because of a characteristic arrangement of horizontal structures uniquely present in faces compared to other natural images. Dakin and Watt (2009) defined this arrangement of horizontal structures as "bar-codes" of faces and showed that these bar codes are tolerant to many everyday transformations of faces we experience such as changes in pose, viewpoint and illumination.

Anisotropic sensitivity to orientations has also been reported for natural images other than faces. Recently, Nasr and Tootell (2012) found stronger BOLD responses in the Parahippocampal Place Area (PPA) (a brain region that selectively encodes natural scenes, as opposed to faces or other artificial stimuli (Epstein & Kanwisher, 1998)), in response to cardinal orientations in natural scenes as opposed to inter-cardinal orientations. They interpreted this as evidence that areas processing natural images exploit our knowledge of the orientation statistics in the environment, specifically the dominance of cardinal orientations. However, their difference in BOLD response was not unique to natural image stimuli. Even meaningless stimuli made of geometrical structures such as lines elicited this effect. Therefore, it is unclear if higher-level mechanisms prioritise cardinal information during natural image perception *per se,* by increasing sensitivity to cardinal orientations.

To our knowledge, no previous study has directly examined if humans prioritise cardinal orientations during the classification of natural scenes, by increasing the visual system's sensitivity to cardinals. Given that the natural images we experience are typically dominated by cardinal structures (Coppola, Purves, McCoy, & Purves, 1998; Switkes et al., 1978), it might be advantageous for image classifying mechanisms to prioritise this information. Accordingly, we measured RMS contrast thresholds required for participants to correctly classify scenes as either outdoor or indoor. Ideally this

would involve measuring contrast thresholds for scenes retaining near-cardinal orientations only and compare it to thresholds obtained with scenes retaining near-inter-cardinal orientations only. However, given the anisotropic distribution of orientations typically present in scenes, if an increased threshold for inter-cardinally filtered images is obtained, that could be due to the lack of structure near inter-cardinals. Therefore, we filtered images cardinally and measured thresholds while the scene is upright and tilted 45° clockwise. The latter would provide thresholds for scenes with inter-cardinals only, but any difference in thresholds would also be contaminated by tilting the image. To disentangle differences in thresholds purely due to the presence of different orientations from differences that arise due to tilting, we also quantified the effect of tilting alone by measuring thresholds for unfiltered images that are upright and tilted. A difference in threshold over and above differences arising due to filtering and/or tilting can be attributed to a difference in sensitivity.

A further obstacle is that according to "Bloch's law", contrast detection thresholds for simple stimuli are inversely proportional to the stimulus duration when stimuli are presented for short durations (Gorea, 2015). At longer stimulus durations (greater than 100ms), temporal summation ceases and contrast thresholds only depend on luminance (Barlow, 1958; Kelly & Savoie, 1978; Roufs, 1974). This suggest that measuring contrast thresholds for stimulus durations less than 100 ms could encompass differences due to differences in our ability to detect specific orientations and differences in temporal summation of specific orientations. For example, in the pathway between the retina and V1, macaque physiology shows that V1 receives connections from the magnocellular pathway that carries low SF information around 20 ms earlier than from the parvocellular pathway that carries high SF information (Nowak, Munk, Girard, & Bullier, 1995). Breitmeyer (1975) showed that when viewers had to detect the presence

of vertically oriented gratings having a range of SFs from 0.5 - 11 cpd, response latencies increased with increasing SF. This pattern survived even after matching gratings of different SFs for subjective contrast. Although Li et al. (2003) did not find a difference in response latency of V1 simple cells to gratings of different orientations, it would not completely discount the possibility that response latencies would not differ when measured with oriented structures from natural images, because V1 behaves differently to artificial narrowband and natural (or naturalistic) broadband stimuli (Bex et al., 2009; Hansen et al., 2008). For these reasons, we measured thresholds for stimuli presented at durations above 100 ms and further increased their processing time by adding an interval between the offset of a stimulus and the onset of a backward mask that ceases stimulus processing. Moreover, we also presented stimuli at two different durations to examine if any differences found in thresholds simply represent participants requiring longer durations to classify filtered and/or tilted images.

4.2. General Methods

Participants

We performed an image classification experiment at two durations: a short duration (Exp 1) and a longer duration (Exp 2). We recruited thirteen participants for each experiment and one participant (MS) did both. In Experiment 3 (Exp 3), we measured detection thresholds in four participants. Observer IM and AV participated in both Exp 2 and Exp 3. Experimental protocols were approved by the Ethics committee of Queen Mary University of London and all participants provided informed consent.

Apparatus

Participants were seated in a dimly lit room. Stimuli were presented using *Psychtoolbox* (Brainard, 1997; Pelli, 1997) and custom-written scripts for *Matlab* (*Mathworks*) on a 16" Dell CRT monitor with a screen resolution of 1024 × 768 and a refresh rate of 60 Hz. Display luminance was linearized to a pseudo 12-bit accuracy with an ISR video attenuator (Pelli & Zhang, 1991). Mean display luminance was set to 50 cd/m^2. A chinrest placed 57 cm in front of the screen was used to maintain observers' head upright and a constant viewing distance at which each pixel subtended 1.8 arcmin.

4.2.1. Methods specific to Experiments 1 and 2 - classification threshold

Stimuli

525 indoor and 525 outdoor scenes were collected from the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). As shown in Fig. 4.2a, the indoor scenes included images belonging to a range of indoor environments including offices, bedrooms, airports, auditoriums, living rooms and restaurants. The outdoor scenes comprised of castles, places of worship, fuel stations, houses, and commercial buildings (Fig. 4.2a). The images were altered from their originals in the following manner. Firstly, all images were converted to grayscale by computing the weighted sum of red, green and blue channels of an image ($0.299R + 0.587G + 0.114B$; Hughes et al., 2013). Secondly, all images were cropped to a square of side length 300 pixels. Thirdly, a 2-dimensional, circularly symmetric, raised cosine window was applied to each image (Eq. 4.1).

$$W_{x,y} = \left(0.5 + 0.5cos\left(\frac{r_{x,y}\pi}{R}\right)\right)^p \quad (4.1)$$

where W is the window, r is the distance of each pixel from the centre of a 2-dimensional array whose column and row numbers are denoted by x and y, respectively, R is the radius of the window (150 pixels) and p is the power to which the cosine function is raised (0.5).

As suggested by Van der Schaaf and van Hateren (1996), the window was applied after subtracting the weighted mean intensity from the image and normalizing it as in Eq. 4.2.

$$C_{x,y} = \left(\frac{I_{x,y} - \mu}{\mu}\right) W_{x,y} \tag{4.2}$$

Where $C_{x,y}$ is the windowed image, $\mu = \sum_{x,y}(I_{x,y} - W_{x,y})/\sum_{x,y} W_{x,y}$, $I_{x,y}$ is the image to be windowed and $W_{x,y}$ is the cosine window. x and y denote the column and row number of pixels, respectively.

In some cases (see procedure), the images were filtered with a cardinal orientation filter created using two wrapped Gaussian functions in the frequency domain. Each Gaussian function had a half-width at half-height of 14.7° and was clipped to have zero gain beyond ±30° from its peak. One of the Gaussians peaked at 0° (horizontal) and the other peaked at 90° (vertical). To filter an image, its amplitude spectrum was obtained by Fourier transformation and was then multiplied by the cardinal orientation filter. The product of the two was combined with the image's original phase spectrum to obtain the filtered image by inverse Fourier transforming.

In all cases the image presented as a test stimulus was windowed (either unfiltered or orientation filtered), and was assigned a specific RMS contrast (see procedure). Since images were in the range -1 to 1, with a mean of 0, the assigned RMS contrast was equal to the standard deviation of pixel intensities. For every test image, a unique backward mask was created by phase-scrambling a different image from the same category. This was achieved by adding the Fourier phase spectrum of a unique white noise pattern (having the same dimensions as the image and with a uniform distribution of pixel intensities ranging from 0 to 1) to the phase spectrum of an image. The RMS contrast of the backwards mask was always yoked to that of the test image. Finally, both the test and mask were scaled to have a pixel intensity range of 0 - 255, with a mean pixel intensity of 127.5.

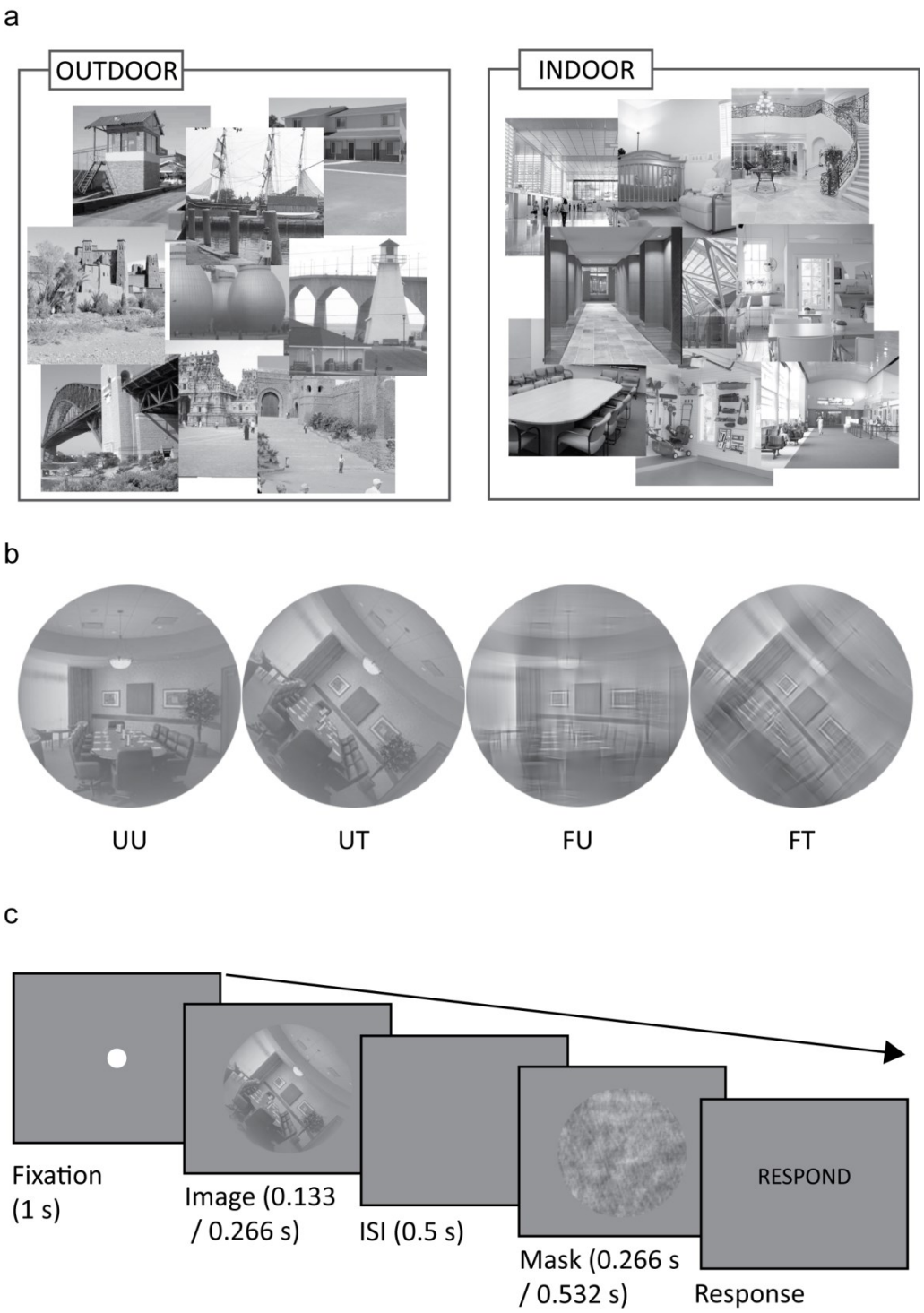


Figure 4.2. Orientation sensitivity during image classification methods: a) A representative selection of images obtained for each image category. b) Example stimuli used in the 4 different conditions ('UU' - unfiltered upright, 'UT' - unfiltered tilted, 'FU' - filtered upright and 'FT' - filtered tilted). c) Timeline of a trial in Experiments 1 and 2.

Procedure

Four conditions were tested in both experiments. In the first two conditions, test images were unfiltered and were presented either in an upright orientation (UU) or tilted 45° clockwise in the fronto-parallel plane (UT). In the other two conditions, orientation filtered images were presented as tests, in either an upright orientation (FU) or tilted 45° clockwise (FT). In each trial, a test stimulus was created by taking an image from one of the two image categories, indoor and outdoor, and was given an RMS contrast that was pseudo-randomly picked from a set of 11 possible values {0.005, 0.014, 0.023, 0.033, 0.042, 0.051, 0.060, 0.098, 0.0135, 0.0173, 0.210}.

In each condition, each combination of image category and RMS contrast was repeated 10 times. This resulted in a total of 880 trials per experiment (4 conditions × 2 categories × 11 RMS contrasts × 10 repeats). As shown in Fig. 4.2c, an experimental trial began with a white central fixation circle (diameter = 0.3°) presented on a uniform grey background for 1 s. Subsequently, a unique test stimulus, selected randomly from one of the two categories and at one of the contrast levels, was presented followed by a phase-scrambled mask. Both test and mask were presented within a hard-edged circular window of diameter 9.4° (300 pixels). In the short experiment, the test was shown for 0.133 s and the mask was shown for 0.266 s. In the long experiment, the test was shown for 0.266 s and the mask for 0.532 s. In both experiments, there was an inter-stimulus interval (ISI) of 0.5 s that displayed a uniform grey screen. After the offset of the mask, participants were prompted with a 'RESPOND' text in the screen and they indicated the category of the test stimulus they saw by pressing one of two keys, '1' for outdoor and '2' for indoor.

4.2.2. Methods specific to Experiment 3 - detection threshold

Stimuli

All test stimuli and masks were created in the same way as in Experiment 1 and 2. However, test stimuli were always orientation filtered.

Procedure

I varied the RMS contrast and the tilt of a test stimulus to measure the minimum RMS contrast required for a participant to detect the presence of an upright and 45° test. In each trial a unique image was randomly selected from one of the two categories, such that in half the trials the test depicted an indoor scene while in the other half the test depicted an outdoor scene. As in Experiments 1 and 2, for each test, a unique image from the same category was selected to create the backward mask. Once images for the test and mask were selected, both were assigned an RMS contrast of one of 11 possible values $\{2.00, 2.57, 3.31, 4.27, 5.49, 7.07, 9.10, 11.72, 15.09, 19.42, 25\} \times 10^{-3}$ and was presented either upright or tilted 45° clockwise. Each combination of image category, RMS contrast and image orientation was repeated 5 times. Therefore, an experimental block contained a total of 220 trials (2 categories × 11 contrasts × 2 orientations).

Figure 4.3 shows the timeline for a trial. Each trial consisted of two stimulus intervals and the interval in which the test stimulus is presented was randomized across trials. In half the trials, the first interval contained the test and began with an auditory beep played for 0.2 s followed by a black fixation circle (diameter = 0.3°) presented in the centre of the screen on a uniform grey background for 1 s. Subsequently, the test image was shown for 0.133 s in the screen centre followed by a uniform grey screen presented

for 0.5 s. The first interval ended with a mask presented for 0.266 s. The second interval began with an auditory beep played for 0.2 s followed by a white fixation circle (diameter 0.3°) presented for 1 s. Following fixation, a grey screen was shown for 0.633 s and the second interval ended with a 0.266 s presentation of the same mask used in the first interval. In the other half of the trials, the test was shown in the second interval, following a similar procedure and timeline as described above. At the end of the second interval, the participant viewed a 'RESPOND' text on the screen and judged which interval contained the test stimulus by pressing '1' for the first and '2' for the second interval.

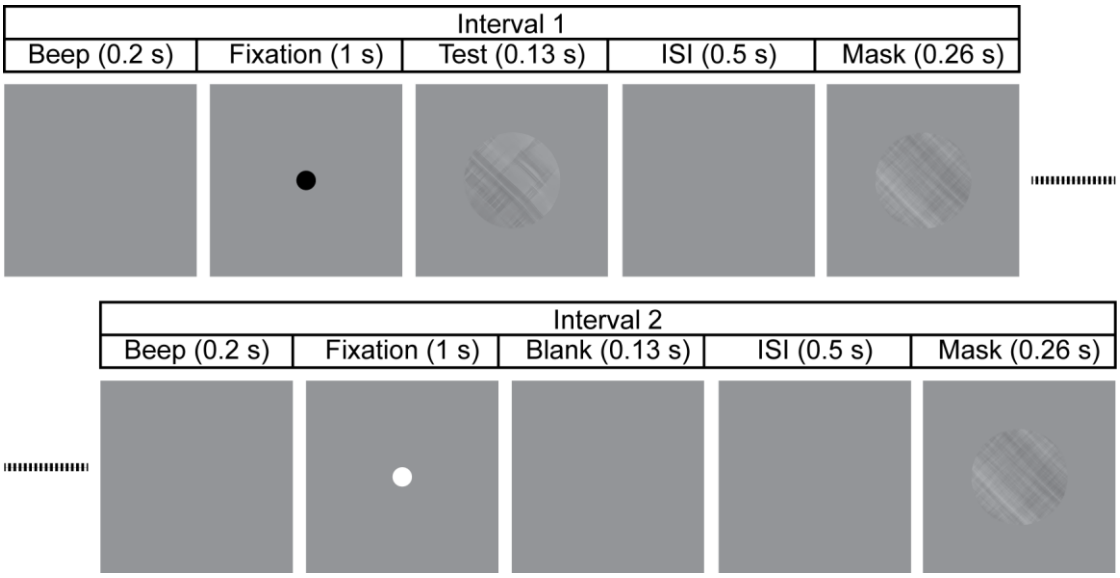


Figure. 4.3. Timeline of a trial in Experiment 3 measuring detection thresholds, in which the test image was presented in the first interval.

4.3. Results

4.3.1. Experiment 1 and 2 - classification threshold

The proportion of correct classification of the test stimulus was plotted against the RMS contrast for each experiment, each condition and each participant separately. By fitting a Weibull function to this data using the *Psignifit 4* toolbox (Schütt, Harmeling, Macke, & Wichmann, 2016), we obtained maximum likelihood estimates of the threshold (63% of the unscaled sigmoid) which corresponds to the point of inflection (Fig. 4.4a). The threshold was used as an estimate for the minimum RMS contrast required to reliably classify a scene. For each condition and duration, mean estimates of thresholds across participants are given in Fig. 4.4b. A full-factorial mixed-subjects analysis of variance (ANOVA) was performed on the estimated thresholds, with image orientation (upright and 45° clockwise) and filtering (unfiltered and filtered) as within-subjects factors and stimulus presentation duration (short and long) as a between-subjects factor.

There was no main effect of duration, $F(1,22) = 1.71$, $p = 0.205$, revealing that presentation duration did not affect the estimated thresholds. There was a main effect of image orientation $F(1,22) = 32.73$, $p < 0.001$. This shows that, in general, thresholds for tilted images were higher than thresholds for upright images. There was also a main effect of filtering, showing that filtered images had higher thresholds than unfiltered images, $F(1,22) = 29.19$, $p < 0.001$. Inspecting the two-way interactions, we found no interaction between orientation and duration $F(1,22) = 0.92$, $p = 0.348$, or between filtering and duration, $F(1,22) = 1.36$, $p = 0.256$. However, there was a significant interaction between orientation and filtering, $F(1,22) = 28.56$, $p < 0.001$. This interaction was further analysed by paired samples *t*-tests. For both short and long durations, when the images were unfiltered, there was no significant difference in the

thresholds between upright and tilted images; short: $t(11) = -1.48$, $p = 0.167$ and long: $t(11) = -0.19$, $p = 0.854$. However, when the images were filtered, there was a significant difference in threshold between upright and tilted images, for both durations; short: $t(11) = -4.55$, $p = 0.001$ and long: $t(11) = -3.45$, $p = 0.006$. These p-values are significant after Bonferroni corrections too ($p < 0.05 / 4$). The three-way interaction between duration, orientation and filtering was not significant, $F(1,22) = 0.00$, $p = 0.722$.

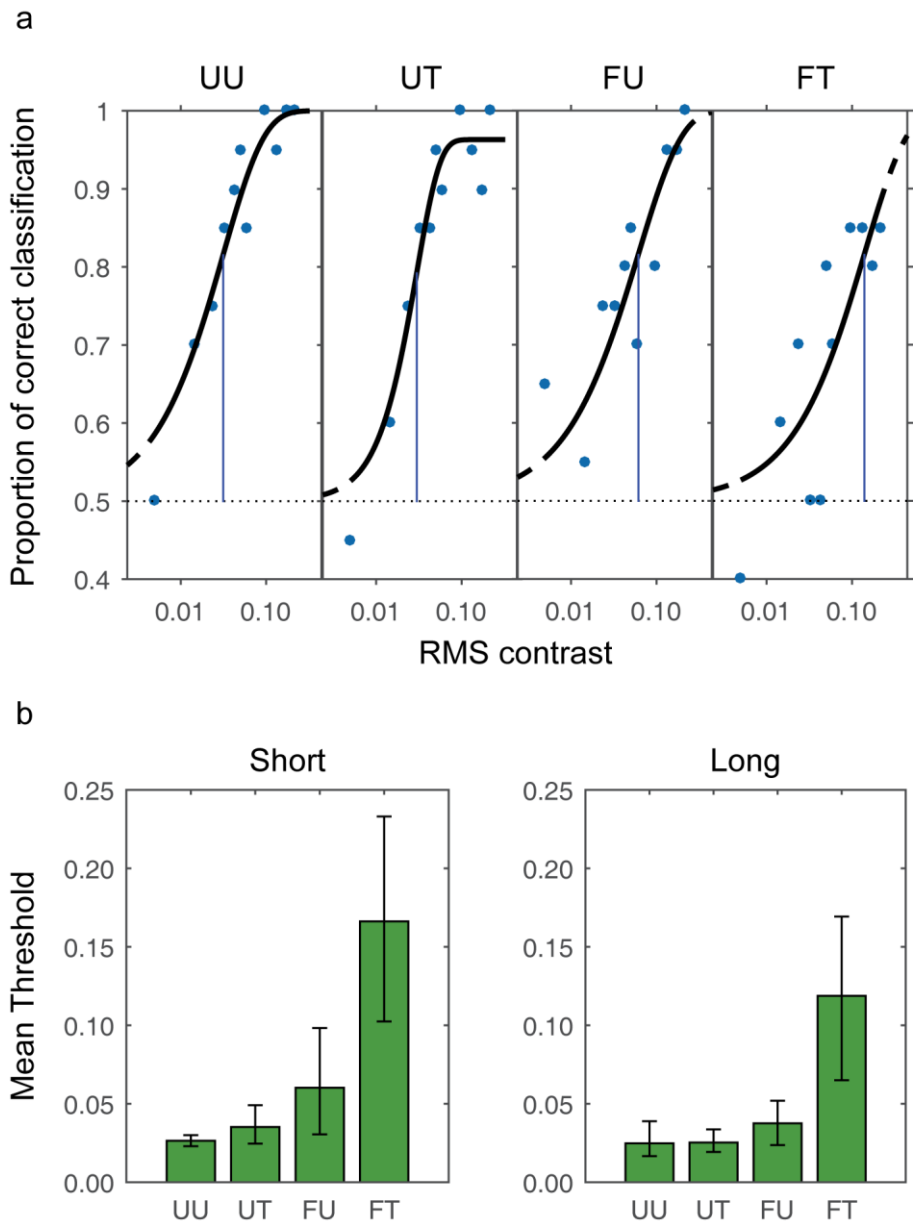


Figure 4.4. Orientation sensitivity during image classification results: a) Sample psychometric functions from participant IM for each condition from the long experiment. Data points plot the proportion of correct classification as a function of RMS contrast and black curves are best-fitting Weibull functions. Blue vertical lines denote the maximum likelihood estimates of threshold. b) Bar plots showing mean threshold across participants for each condition in the short (left) and long (right) experiments. Error bars denote 95% confidence intervals obtained from parametric bootstrapping.

4.3.2. Experiment 3 - detection threshold

After pooling trials from both image categories, the proportion of choosing the correct interval with the test was plotted as a function of the RMS contrast of the test, for each image orientation and each participant, separately. Each data point would contain responses from 10 trials. By fitting a Weibull function to these data using the *Psignifit 4* toolbox (Schütt et al., 2016), we obtained maximum likelihood estimates of 63% threshold (of unscaled sigmoid) for each image orientation. This threshold denoted the minimum RMS contrast required to reliably detect the presence of a scene retaining specific orientations only. Table 4.1 provides individual and mean estimates of threshold across observers for upright and 45° clockwise tilted images.

Table 4.1. *Individual and mean detection thresholds for upright and 45° clockwise tilted images*.

Participant	Upright		Tilted 45° clockwise	
	Individual	Mean	Individual	Mean
AM	0.004	0.009	0.005	0.008
IM	0.009		0.007	
DA	0.009		0.007	
AV	0.012		0.012	

4.4. Discussion

Experiments 1 and 2 revealed three major results. Firstly, contrast thresholds increased when scenes were tilted 45° away from an upright position, irrespective of whether they retained all or a limited band of orientations, and irrespective of their presentation duration. This shows that scene classification is disrupted when scenes do not appear in their canonical orientations. This supports the findings of Loschky, Ringer, Ellis, and Hansen (2015) who showed that the accuracy of scene classification decreases monotonically with increasing tilt in the fronto-parallel plane up to 135° away from upright. Therefore, when we classify natural scenes within a single fixation, the global orientation of the scene appears to play a crucial role. This would suggest that a mechanism that is selectively encoding the global orientation or the uprightness of a scene may also contribute to mechanisms encoding scene category.

Secondly, we find that filtering images to retain near-cardinal structures alone increased contrast thresholds, irrespective of the presentation duration and the tilt of images. This might appear unsurprising given that filtered images retained limited structural information compared to unfiltered images. Although thresholds increased, these findings also show that orientation-filtered scenes did not eliminate participant's ability to classify scenes. Scenes were reliably classified 63% above chance with limited structural information present, irrespective of whether they were presented upright or tilted. Notably, the cardinal orientation filter used here removed structures within a 60° band. However, this could be because carpentered scenes like the ones used here are generally dominated by near-cardinal structures (Coppola et al., 1998; Switkes et al., 1978), and these structures which are possibly diagnostic of the image's category are preserved by the filter used here. It is unclear how current findings would generalise to

uncarpentered images, but Goffaux and Dakin (2010) found that people could reliably discriminate between natural scenes (that included some uncarpentered images), when they were filtered to retain near vertical, near horizontal or both near-vertical and near-horizontal orientations only. Notably, they also used Gaussian filters with similar bandwidths to those used here.

Some recent theories of rapid scene classification propose that people utilise properties of the power spectrum that represents the spatial distribution of edges of different orientations and spatial frequencies. This spectral analysis would allow a viewer to obtain a rapid representation of a scene's layout and therefore facilitate scene classification (Brady et al., 2017; Greene & Oliva, 2010). Here we show that the whole spectrum need not be visible for people to reliably classify scenes. With higher image contrasts people could reliably classify orientation-filtered images irrespective of whether they were upright or tilted. It is possible that an analysis of the power spectrum within a diagnostic band of orientations is sufficient for classification since carpentered images like the ones used here contain most of their spectral power near their cardinal orientations. In support of this, Walther et al. (2011) examined the effect of removing contours on classification performance and found that removing long contours impaired classification performance more than removing short contours. They concluded that scene classification survives removal of short contours because long contours are diagnostic of a scene's spatial layout that is essential for rapid scene classification. Therefore, in the current study, classifying carpentered scenes may have survived removing inter-cardinal structures because cardinal structures carry the diagnostic information.

Thirdly, here we find that when images were both filtered and tilted, the increase in threshold is more than would be predicted by the combined effects of filtering and tilting. This suggests that, during image classification, people have difficulty in utilising orientation information presented near the inter-cardinal axes with reference to a gravitational or an egocentric (head-centric or body-centric) reference frame. The critical spatial frame of reference can only be deduced by manipulating the head and/or body orientation of the participant, which is a possible thought for future experiments. Nonetheless, the current findings show that when near-cardinal structures of an image are aligned with the gravitational / egocentric cardinal axes, the image requires relatively lower RMS contrast than when the same structures are aligned with the gravitational / egocentric inter-cardinal axes. Further, the difficulty in utilising orientations near the inter-cardinal axes remained across both stimulus durations tested, showing that the difference in thresholds cannot be attributed to scene classifying mechanisms requiring longer durations to classify scenes that contain limited structural information and/or tilted from canonical orientations.

It is possible that the current findings reveal non-uniformities in orientation sensitivity in the early visual system, specifically in V1. Increased RMS contrast thresholds to detect inter-cardinal orientations could therefore propagate to influence processing in subsequent scene classifying higher-level regions that receive information from regions like V1. However, it has been shown that non-uniformities in contrast sensitivities to different orientations and spatial frequencies are not maintained for suprathreshold stimuli (Georgeson & Sullivan, 1975; Zemon, Conte, & Camisa, 1993). In Experiment 3, we found that the mean threshold for detecting oriented structures near-cardinal and near-inter-cardinal axes are 0.009 and 0.008, respectively. These values are 2.88 (cardinal) and 3.09 (inter-cardinal) times lower than the minimum mean classification

thresholds reported for any condition in Experiment 1 and 2. Given that contrast thresholds for classification were approximately 3 times above detection thresholds, it is unlikely that non-uniformities in contrast sensitivity in early cortical neurons influenced higher-level scene classification.

Another reason why the differences in classification thresholds reported here are probably not simply propagating effects from V1 is that the change in magnitude of the cortical response to stimuli of increasing suprathreshold contrasts in V1 is not proportional to changes in cortical responses in higher-level areas. For instance, Avidan et al. (2002) showed that, although BOLD responses in early cortical regions like V1, V2 and V4 of humans show an increase in their BOLD response for suprathreshold contrasts, the lateral occipital complex (LOC) which is involved in object classification did not show any such increase. In fact, the LOC's BOLD response was saturated at threshold. However, we cannot completely rule-out the possibility that these results reflect propagative effects for two reasons. Firstly, contrast sensitivity has not been examined in scene selective regions like the PPA yet. Secondly, non-uniformities in sensitivity to different orientations are also reported in higher-level regions like the Middle Temporal area (Mannion et al., 2010; Shen, Tao, Zhang, Smith, & Chino, 2014; Xu, Collins, Khaytin, Kaas, & Casagrande, 2006). Therefore, the effect could have propagated from intermediate regions that then feed into scene classifying regions.

Alternatively, the current findings could be attributed to an increased sensitivity of scene classifying brain regions to orientations near the cardinal axes. In fact, Nasr and Tootell's (2012) finding that the PPA responds strongly to cardinal orientations supports this claim. There are also other studies which have shown that BOLD

responses in PPA are stronger to indoor scenes having relatively more cardinal orientations than to outdoor scenes (Bar & Aminoff, 2003; Henderson, Larson, & Zhu, 2007). However, since Nasr and Tootell (2012) also found stronger PPA activation to meaningless stimuli with cardinal orientations, it was unclear to what extent sensitivity to orientations in the PPA is specific to classification *per se*. The current findings strongly suggest that increased sensitivity to cardinal orientations play a significant role in facilitating scene classification. By drawing parallels between orientation statistics of the environment and perceptual performance, it has been shown how low-level feature encoding mechanisms efficiently capture non-uniformities in the distribution of orientations in the environment (Girshick et al., 2011; Hansen et al., 2008; Tomassini et al., 2010). Our findings suggest that, in addition to affecting initial low-level encoding of spatial structure, environmental non-uniformities may also influence classification mechanisms to optimally capture dominant orientation structures in the environment.

5. Chapter 5 - A bias for carpentered images in classification

5.1. Introduction

Chapter 4 addressed the influence of low-level features on image classification. Although the percept of any image arises from an analysis of low-level information, the final percept is not solely determined by this low-level information. For example, our knowledge or expectations about the environment can strongly modify perception (section 1.3). These expectations may have an evolutionary origin leading to permanent changes in the visual system, may develop with years of living in a specific type of environment or could also be learnt in an experimental setting causing temporary changes in the visual system's functioning (Geisler & Diehl, 2003; Geisler & Kersten, 2002; Girshick et al., 2011; Körding & Wolpert, 2004). Using artificial stimuli, it has been shown that the perception of simple features like orientation or motion direction, and more complex features like shape from shading are found to be influenced by prior expectations about the occurrence of these features in the environment (Girshick et al., 2011; Stocker & Simoncelli, 2006; Sun & Perona, 1998; Tomassini et al., 2010). Generally, our perception of these features is biased towards most frequently encountered features in the environment.

We can manipulate stimuli such as Gabors or dot patterns to vary along a single feature dimension like orientation or motion direction, and then measure perceptual biases that are specific to the manipulated feature. Similarly, we can also manipulate naturalistic complex stimuli (e.g., faces) along a single continuum (e.g., gender) and measure biases specific to this manipulated feature based on people's classification of stimuli. Accordingly, a few studies have examined how our perception of complex image properties is influenced by expectations. Biases have been reported for properties like

gaze direction and gender of images depicting faces or bodies (Armann & Bulthoff, 2012; Mareschal et al., 2013; Troje et al., 2006; Watson, Otsuka & Clifford, 2016).

Here, we were interested in examining (a) whether people have perceptual biases for meaningful everyday scenes, and (b) whether these biases result from processes encoding more simple, lower level properties (such as orientation mentioned above). To this end, in Experiment 1, we used a novel, highly versatile method of creating "hybrid" images that allows us to measure biases for complex categorical attributes of natural images while controlling for the visibility of the separate components making-up the hybrid, bypassing confounds that may arise due to differences in sensitivity to low-level visual features (e.g., orientation and spatial frequency). To examine if differences in sensitivity to structural features can account for categorical biases, we conducted Experiment 2 to measure minimum root mean square (RMS) contrast required for participants to detect structures that can be used to classify images from each category.

We investigated whether living in highly carpentered/constructed environments may have altered human perception. We predicted that, our frequent exposure to carpentered images would bias classification of ambiguous hybrids as carpentered rather than uncarpentered (natural). The ambiguous stimuli described below aren't merely ambiguous with respect to orientation content; they are hybrids, whose component images come from two of four different categories, namely "animal", "flower", "house" and "vehicle". The purpose of using these ambiguous hybrids was to maximise the size of the perceptual bias since perceptual biases generally arise when stimuli are highly ambiguous (section 1.3.2). Participants were instructed to report the category of the perceptually dominant component of a series of hybrid images. Their responses proved to be biased toward "house" and "vehicle," categories containing predominantly

cardinal (vertical and horizontal) image structure, even when those orientations were filtered out of the component images. This strongly suggests that the carpentered bias is not the result of the dominance of cardinal structure in carpentered images. Thus, to pre-empt our results, our participants seem to harbour priors favouring the type of images most often seen in their urban environments.

5.2. General Methods

All experimental procedures were approved by the Ethics committee of Queen Mary University of London (QMUL). All participants were members of QMUL and had normal or corrected-to-normal visual acuity. They provided informed consent to participate. All participants have lived in cities with exposure to abundant constructed environments, for at least 10 years prior to the experiment.

Participants were seated in a dimly lit room. A chinrest helped participants to maintain a distance of 0.57 m from the 16" Dell CRT monitor upon which the stimuli were presented. At this distance, each of the screen's 1024 × 768 pixels subtended 1.8 minutes of visual angle. The monitor's refresh rate was 60 Hz. Experimental programs were written in Matlab, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

5.2.1. Experiment 1: filtered hybrids

Participants

Ten participants (3 males; all naïve except AM and IM) took part. This sample size was calculated using a power analysis performed using the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007), based on the results of a pilot experiment

measuring classification biases for hybrids containing image components that are unfiltered for orientation content. According to the pilot's results, a minimum of 8 observers were required to detect a classification bias significantly different from zero, while achieving a statistical power of 80%, a value recommended by Cohen (1992).

Stimuli

From an initial pool of 500 images obtained from the ImageNet database (Deng et al., 2009), we created a 100-image set "C," within which each image was unambiguously recognizable as an animal after application of the cardinal filter described below. Next, we created a 100-image set "I," within which each image was unambiguously recognizable as an animal after application of the intercardinal filter[2]. Some images appeared in both sets. Finally, this process was repeated, creating a set C and a set I for flowers, houses, and vehicles. Example images from all four categories appear in Fig. 5.1.

[2] This was the image selection procedure. Each of the 500 images from each category (animal, flower, house and vehicle; 2000 images in total) was windowed, filtered with a cardinal filter and was presented to participant AM for an unlimited duration, in a random order. Participant AM judged if each image was unambiguously recognizable as an animal, flower, house or vehicle. From the correctly recognized set of images, the first 100 were chosen to create set C for each category. The same procedure was repeated to obtain images for set I, with the exception that instead of a cardinal filter, an intercardinal filter was applied before presenting the image.

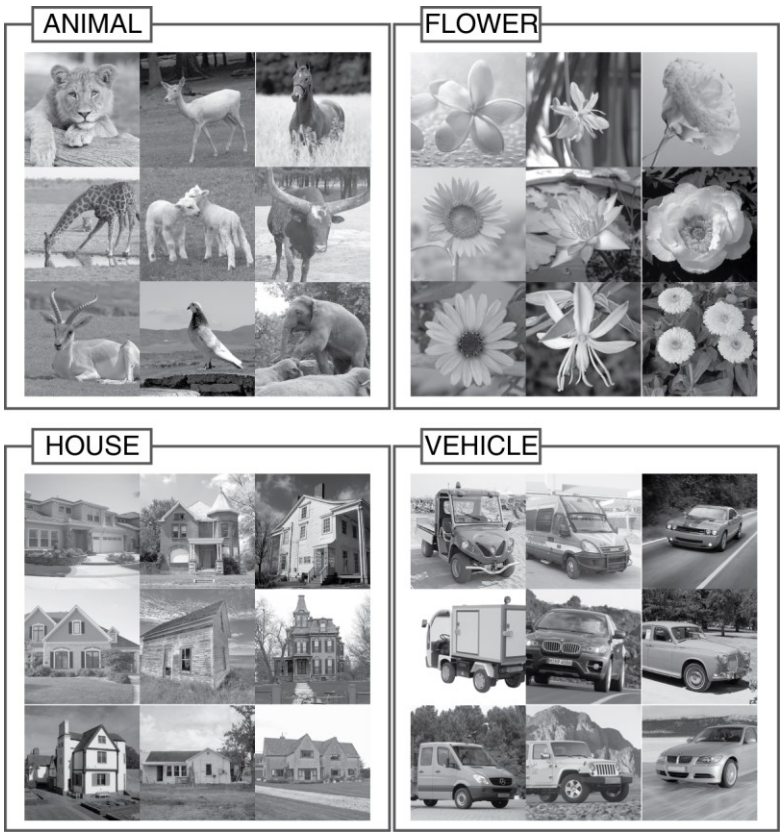


Figure 5.1. Sample images from each category used in filtered hybrids experiment (Experiment 1).

Hybrids were created using a 7-step procedure (steps 3 to 7 are shown in Fig. 5.3a). In step 1, we randomly selected component images from sets C and I in two of the four available categories (e.g., house from set C and flower from set I, as in Fig. 5.3a). In step 2, each component was converted to grayscale by computing the weighted sum of red, green and blue channels of an image ($0.299R + 0.587G + 0.114B$; Hughes et al., 2013). To minimize wrap-around artefacts during Fourier transformation, pixel intensities of each component were multiplied by a 2-dimensional, circularly symmetric, raised cosine window in step 4 (Eq. 5.1). The window has maximum weight in the centre and decreases to zero towards the boundaries of the image. Prior to applying the window (i.e., in step 3), as suggested by van der Schaaf and van Hateren (1996), the weighted mean graylevel was subtracted from each pixel (to prevent leakage

in spectral information from the DC coefficient) and normalized, so that each windowed component would have zero mean intensity (Eq. 5.2).

$$W_{x,y} = \left(0.5 + 0.5cos\left(\frac{r_{x,y}\pi}{R}\right)\right)^{p} \quad (5.1)$$

where W is the window, r is the distance of each pixel from the centre of a 2-dimensional array whose column and row numbers are denoted by x and y, respectively, R is the radius of the window (150 pixels) and p is the power to which the cosine function is raised (0.5).

$$C_{x,y} = \left(\frac{I_{x,y} - \mu}{\mu}\right) W_{x,y} \quad (5.2)$$

Where $C_{x,y}$ is the windowed image, $\mu = \sum_{x,y}(I_{x,y} - W_{x,y})/\sum_{x,y} W_{x,y}$, $I_{x,y}$ is the image to be windowed and $W_{x,y}$ is the cosine window. x and y denote the column and row number of pixels, respectively.

In step 5, the C and I components were filtered to retain near-cardinal (horizontal and vertical) and near-intercardinal (45° and 135° clockwise of horizontal) orientations, by multiplying their amplitude spectra with cardinal and intercardinal filters, respectively. The cardinal filter's pass-band was the sum of two wrapped Gaussian functions; one peaking at 0° (horizontal) and the other peaking at 90° (vertical). Each Gaussian had a half-width at half height of 23.6°. The intercardinal filter was rotated 45° but otherwise identical to that of the cardinal filter.

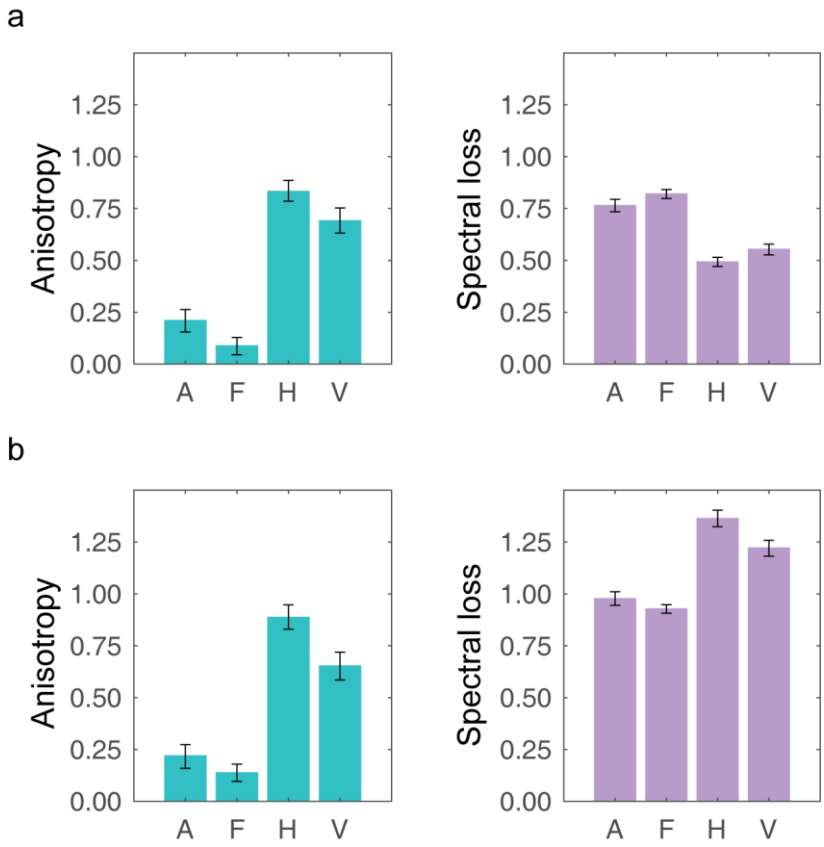


Figure 5.2. Mean orientation anisotropy and mean loss ratio for images: a) Mean orientation anisotropy (left) and mean loss ratio (right) for images in each category (A - animal, F - flower, H - houses, V - vehicle) of set C. b) Mean orientation anisotropy (left) and mean loss ratio (right) for images in each category (A - animal, F - flower, H - houses, V - vehicle) of set I.

To quantify the orientation 'anisotropy' in images, we obtained the log-ratio of total power (integral of power spectrum) after filtering each image cardinally and intercardinally: $A = \ln(P_C - P_I)$, where A is the anisotropy, P_C is the total power after cardinal filtering and P_I is the total power after intercardinal filtering. A positive value represents anisotropy, where there is relatively greater power at near-cardinal than near-intercardinal orientations. Fig 5.2 plots the mean anisotropy across all images from each category (animal, flower, house and vehicle) and each set (C and I). In both sets, houses and vehicles had higher anisotropy while animals and flowers had lower anisotropy. To

quantify how much spectral power is lost in images following orientation filtering, we also calculated the 'spectral loss', by obtaining the log-ratio of total power between an unfiltered image and a cardinally filtered image (for set C images), or between an unfiltered image and an intercardinally filtered image (for set I images):
$S = \ln(P_B - P_A)$, where S is the spectral loss, P_B is the total power before filtering and P_A is the total power after filtering. A positive spectral loss value indicates a reduction in power after filtering. Mean spectral loss across all images from each category and each set are plotted in Fig. 5.2. Cardinal filtering (that attenuates intercardinal orientations) of set C images led to a relatively greater loss of power in animals and flowers, whereas intercardinal filtering (that attenuates cardinal orientations) of set I images resulted in a greater loss of power in houses and vehicles.

In step 6, we uniformly adjusted (reduced or elevated) the amplitude of each component's spatial frequency content, so that the two components would have the desired sum (fixed at 1.33×10^8) and ratio (an independent variable) of notionally visible energies. Notionally visible energy (hereafter "visible energy") is defined as the dot product between an orientation-filtered image's power spectrum and a "window of visibility" that we created, based on Watson and Ahumada (2005). The 'window of visibility' (WV) was the product of two 2-dimensional filters that were the same size as the amplitude spectrum of a component. The first was a 'contrast sensitivity filter' (CSF), whose gain—a truncated log-parabola of spatial frequency (as suggested byLesmes, Lu, Baek, & Albright, 2010; Eq. 5.3.)—was independent of orientation. Three out of four parameters of the truncated log-parabola
($f_{max} = 3.5$ cycles per degree, $\beta = 3.4$ octaves and $\delta = 0.3$ decimal log units below γ_{max}) were those best-fitting the ModelFest dataset (Watson and Ahumada, 2005). The parameter which represents the peak sensitivity (γ_{max}) was set at 1.

The CSF takes the form:

$$S'(f) = \log_{10}\gamma_{max} - K\left(\frac{\log_{10}(f) - \log_{10}(f_{max})}{\beta'/2}\right)^2,$$

$$S(f) = \begin{cases} S'(f), & f \geq f_{max} \\ \log_{10}\gamma_{max} - \delta, & f < f_{max}\ and\ S'(f) < \log_{10}\gamma_{max} - \delta \end{cases} \quad \text{(S3)}$$

where $\gamma_{max}$is the peak sensitivity, f is the spatial frequency, f_{max} is the peak spatial frequency, $\beta' = \log_{10}\beta$ and β is the full-bandwidth at half-height (in octaves), δ is the truncated sensitivity at low spatial frequencies and K is a constant ($K = \log_{10} 2$). $S(f)$ and $S'(f)$ define sensitivity with and without truncation respectively.

The second filter was an 'Oblique Effect filter' (OEF), which models contrast sensitivity as a function of grating orientation and was dependent on spatial frequency (Eq. 5.4; see Watson and Ahumada, 2005).

The OEF takes the form:

$$S(f,\theta) = \begin{cases} 1 - \left(1 - e^{\left(-\frac{f-\gamma}{\lambda}\right)}\right) sin^2(2\theta), & f > \gamma \\ 1, & f \leq \gamma \end{cases} \quad \text{(S4)}$$

where $S(f,\theta)$ defines sensitivity (maximum gain = 1), f is the spatial frequency, γ is the spatial frequency at which sensitivity starts to decline (3.48 cycles per degree), λ is the slope of decline in sensitivity (13.57 cycles per degree) and θ is the orientation.

Combining the CSF with OEF gives the WV, a non-separable filter which models contrast sensitivity as a function of both spatial frequency and orientation of a stimulus. When two hybrid components have equal visible energy (i.e., at a log-ratio of 0), we can assume that the two are equated for low-level visibility, since the WV gives an index of

effective contrast after taking into account non-uniformities in contrast sensitivity to structures of different orientations and spatial frequencies.

In step 7, the filtered, scaled components were back-transformed and combined by adding pixel intensities to create a hybrid. Pixels beyond or below the interval of achievable graylevels were given the maximum or minimum value (i.e., 255 or 0). Although this pixel clipping occurred in 40% of our images, this never affected more than 0.42% of the pixels in the hybrids.

We also created a unique mask for every hybrid image by phase-scrambling the hybrid. This was achieved by adding the phase spectrum of a white noise pattern (300 × 300 pixels with a uniform distribution of pixel intensities between 0 and 1) to the phase spectrum of a hybrid. A unique white noise pattern was generated for each hybrid we created.

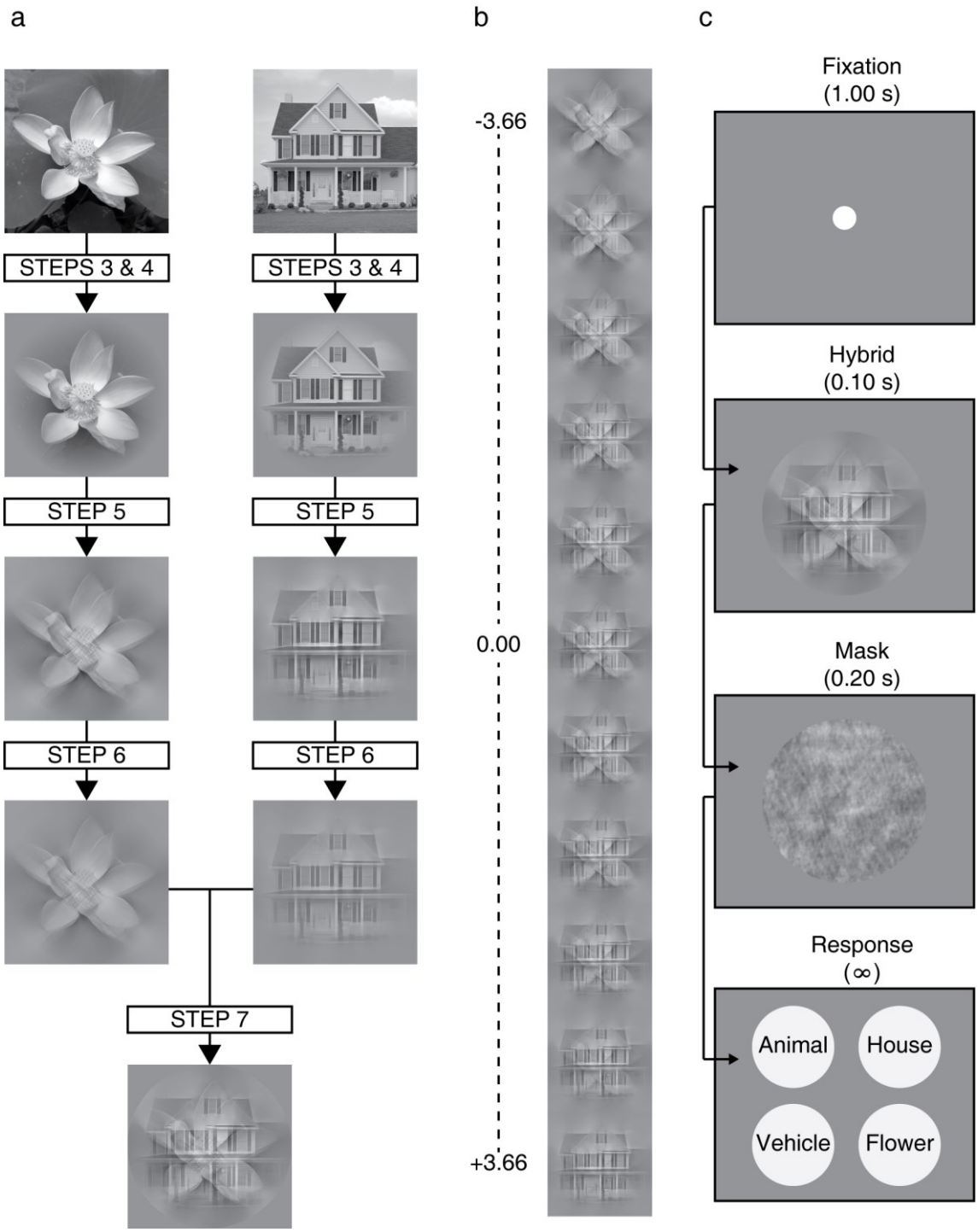


Figure 5.3. Filtered hybrids experiment methods: a) Steps involved in creating a hybrid from two sample grayscale images passed through steps 1 and 2. The house depicted here is the cardinal component and the flower depicts the intercardinal component. b) An example range of hybrid images with different log-ratios (to the left) of visible energy between each hybrid's cardinally and intercadinally filtered components. c) Timeline of a trial from filtered hybrids experiment.

Procedure

There were 8 different conditions, characterized by either the cardinal or the intercardinal component of the hybrid. In 4 conditions, we fixed the cardinal component's category as the animal (CA), flower (CF), house (CH), or vehicle (CV), with the intercardinal component randomly chosen from the remaining 3 categories. In the remaining 4 conditions, we fixed the intercardinal component to be the animal (IA), flower (IF), house (IH), or vehicle (IV), and the cardinal component was randomly chosen from the 3 remaining categories.

Within each condition the log ratio between visible energies of (cardinal and intercardinal) components was selected at random (without replacement) from the set containing 8 copies of these 11 values: {-3.66, –2.20, –1.39, –0.41, –0.20, 0, +0.20, +0.41, +1.39, +2.20, +3.66}. The 8 different conditions were randomly interleaved within each 704-trial session. In every trial, the participant's task was to report the category of the hybrid's most visible component.

The experimental procedure is shown in Fig. 5.3c. Each trial began with the presentation of a white fixation dot (0.3° diameter) centred on a uniform gray background for 1.00 s. This was followed by a hybrid image that was shown for 0.10 s, immediately followed by a mask for 0.20 s. Hybrid and mask were presented in the centre of the screen within a hard-edged circular window (9.4° diameter). After the mask, 4 circular labels (3.8° diameter) of each image category appeared, and the participant responded using one of four keys ('4 – top left', '5 – top right', '1 – bottom left', '2 – bottom right'), which mapped to the screen position of the category label. The position of a given category listed in one of the 4 labels was randomized on every trial.

5.2.2. Experiment 2: detection

Participants

Five participants took part (2 females; all naïve except AM and IM). Three of these participants also took part in Experiment 1. This sample size exceeds that of other studies in which detection thresholds were measured with natural images (Bex & Makous, 2002; Bex et al., 2009).

Stimuli

The image set was expanded to include 555 images per category. Each target and non-target was based on one of these images. To create a target, we started with a Gaussian white-noise pattern of the same size as any image (300 × 300 pixels), having an RMS contrast of 10.00×10^{-2}. Secondly, an image was randomly chosen from one of four available categories (e.g., house) and a circularly symmetric raised cosine window was applied as in Experiment 1. The noise's amplitude spectrum was replaced with the image's amplitude spectrum. Finally, the noise and the image were combined (by adding pixel intensities) to create a target stimulus (Fig. 5.4). The non-target was created in a similar manner except that the image was phase-scrambled before combining with the noise (Fig. 5.4) to preserve the Fourier energy distribution of the image while distorting the higher-order structure.

Procedure

In each trial, we varied the image category used to create target and non-target stimuli and randomly selected two unique images from the same image category. One of the two unique images was superimposed on noise to create the target stimulus and the other was phase-scrambled and superimposed on noise to create the non-target. RMS

contrasts used for the target and non-target were identical and was randomly picked from one of 11 possible values {1.00, 1.26, 1.58, 2.00, 2.51, 3.16, 3.98, 5.01, 6.31, 7.94, 10.00} × 10^{-2}. RMS contrast of the unique noise patterns generated in every trial for the target and non-target was set at 10.00 × 10^{-2}. Each combination of image category and RMS contrast was repeated in 20 trials. A trial began with a white fixation circle (0.3° diameter) on a uniform gray background, shown for 1.00 s. Subsequently, the participant saw the first stimulus followed by the second, each presented for 0.05 s. After each stimulus, a uniform gray screen was presented for 0.30 s. Order of presentation of the target and the non-target was randomized across trials. Participants performed a two-interval-forced-choice task to indicate which stimulus interval contained an image classifiable as an animal, flower, house or vehicle by pressing keys '1' (for first) or '2' (for second).

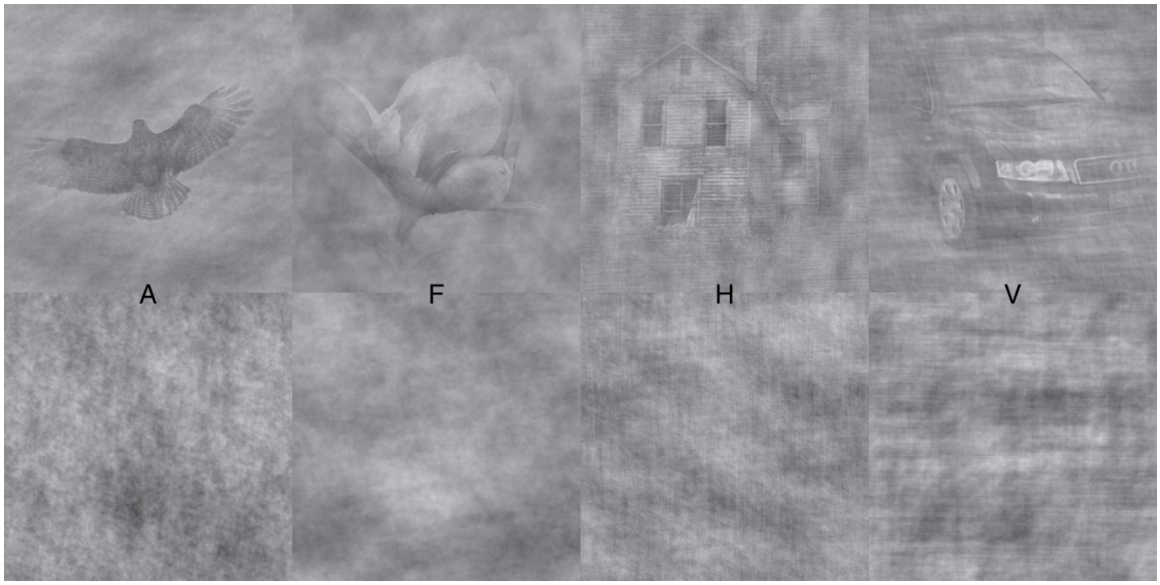


Figure 5.4. Sample images from each category used as target and non-target stimuli in the detection for classification experiment (Experiment 2): top row - unscrambled images superimposed on noise, bottom row – phase-scrambled images superimposed on noise (A - animal, F - flower, H - house and V - vehicle).

5.3. Results

5.3.1. Experiment 1: behavioural results

We obtained estimates of each participant's bias ($-\mu$), in each of the 8 conditions, by maximum-likelihood fitting the two parameters (μ and σ) defining a cumulative Normal distribution to the psychometric function mapping log energy ratio to the proportion of trials on which the cardinal component was selected (Fig. 5.5a). When the two components have equal visible energies (log ratio=0), an unbiased observer should select either component with equal frequency. Positive (negative) biases at log ratio = 0 therefore indicate a tendency for the cardinal (intercardinal) component to dominate perception.

For each estimate of bias, a generalised likelihood-ratio test was performed to evaluate the null hypothesis that the bias does not differ from zero. For this, we fit the data in each condition again with a constrained psychometric function that forced the bias to be zero. We compared the criterion $\alpha = 0.05$ to the value $1 - F(-2 \ln L)$, where F is the cumulative χ^2 distribution with 1 degree of freedom and L is the ratio of likelihood of the constrained fit to the unconstrained fit. If the value is less than α, the bias is significantly different from zero. Table 5.1 shows the number of participants who had positive or negative biases that were significantly different from zero using this likelihood-ratio test. For any given condition, two-tailed one-sample *t*-tests were also conducted to determine if the bias across all participants (mean bias) was significantly different from zero (Table 5.1).

Figure 5.6 (left hand and middle columns) plots the biases from each condition for each participant. It is clear from Table 5.1 and Fig. 5.6 that classification biases were

dependent on the category of images that formed the hybrid's components. In general, when the cardinal component contained an animal or flower the biases were negative, whereas when the intercardinal component contained them, biases were positive. When the cardinal component contained houses or vehicles biases were positive, whereas when the intercardinal component contained them biases were negative.

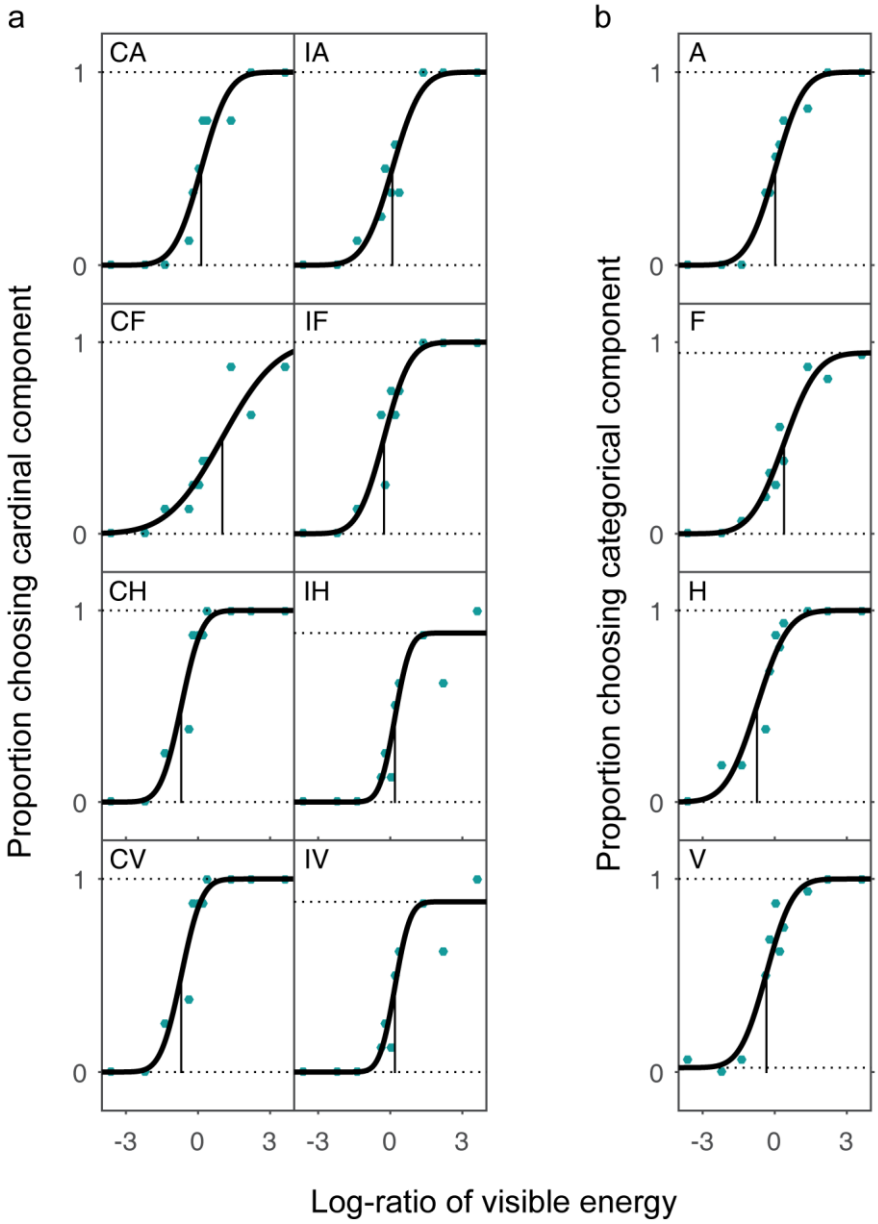


Figure 5.5. Example psychometric functions from participant AM in the filtered hybrid experiment: a) Blue dots plot the proportion of choosing the cardinal component as dominant (ordinate) against the log-ratio of visible energy between cardinal and intercardinal components (abscissa). Each subplot refers to a condition (CA - cardinal

animal, CF - cardinal flower, CH - cardinal house, CV - cardinal vehicle, IA - intercardinal animal, IF - intercardinal flower, IH - intercardinal house and IV - intercardinal vehicle). b) Blue dots plot the proportion of choosing the specific category as dominant (ordinate) against the log-ratio of visible energy between categorical and non-categorical component. Each subplot refers to a category (A - animal, F - flower, H - house and V - vehicle). In all plots (a and b), black curves are best fitting cumulative Normal distribution functions and solid vertical lines denote maximum likelihood estimates of the mean (μ).

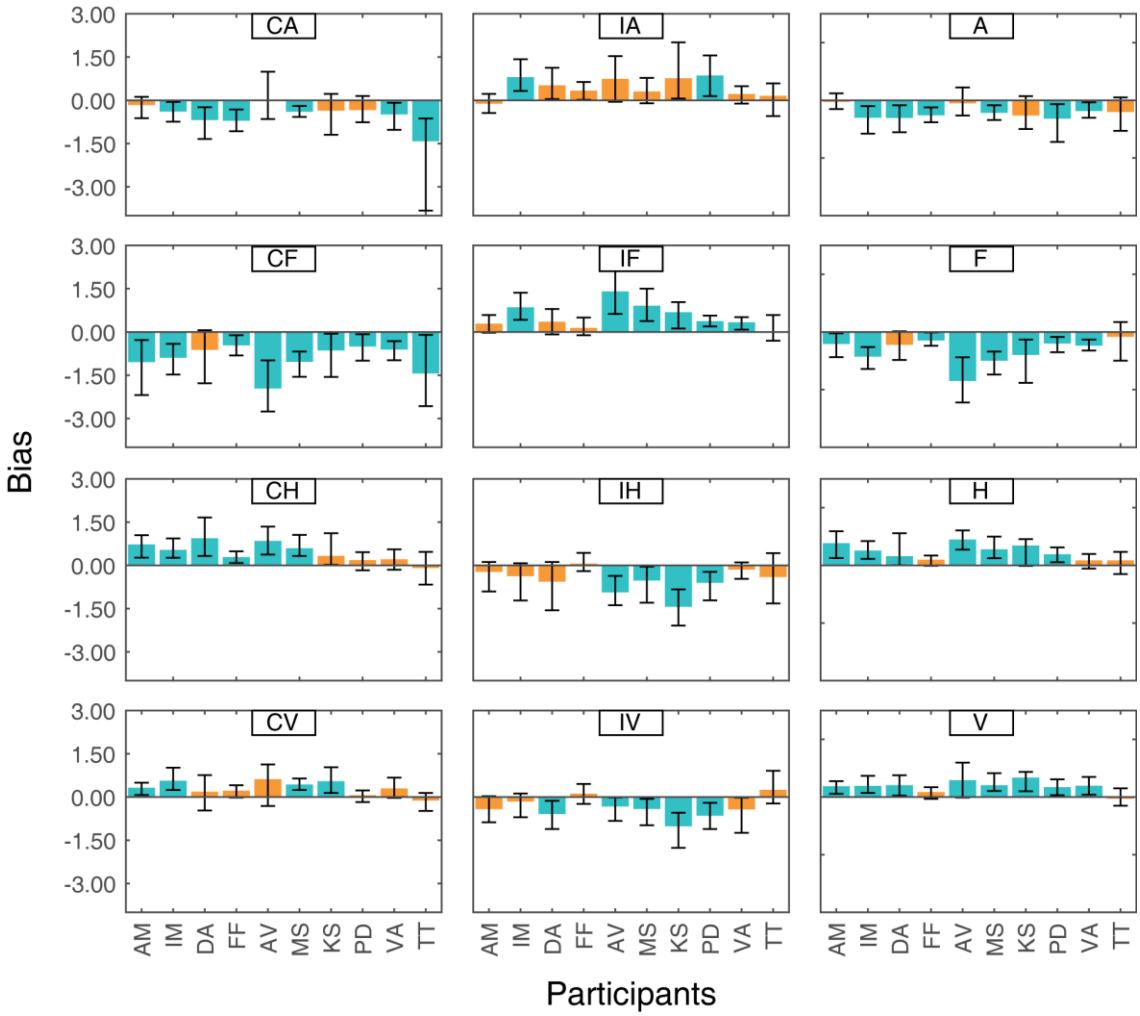


Figure 5.6. Filtered hybrids results: Bar plots showing biases in each condition (left and middle panel: CA - cardinal animal, CF - cardinal flower, CH - cardinal house, CV - cardinal vehicle, IA - intercardinal animal, IF - intercardinal flower, IH - intercardinal house and IV - intercardinal vehicle) and categorical biases (right panel: A - animal, F -

flower, H - house and V - vehicle) for each participant. In each subplot, each bar denotes a single observer (identified by a unique two character initial); blue bars represent biases that significantly differed from zero based on likelihood ratio tests. Error bars represent bias-corrected and accelerated 95% confidence intervals (Efron and Tibshirani, 1994).

We find that at the point of subjective equality, animals and flowers required more energy than the other component of the hybrid, while houses and vehicles required relatively less energy than the other component. Categorical biases were estimated by fitting a cumulative Normal distribution to the function mapping log energy ratio between categorical and non-categorical component to the proportion of trials on which a specific category was selected (i.e., irrespective of filtering; Fig. 5.5b). This involved pooling data from conditions in which a specific category was fixed as either the cardinal or the intercardinal component. For example, data from conditions CA and IA were pooled to plot the proportion of choosing the animal component as dominant against the log-ratio of visible energy between the animal and the non-animal (flower, house or vehicle) components. Individual biases for each image category are given in the right-hand column in Fig. 5.6. As summarized in Table 5.2, group biases were negative and significantly different from zero for animals and flowers, whereas they were significantly positive for houses and vehicles.

Table 5.1. *Individual and group statistics on biases from each hybrid condition.*

Condition	Number of participants		Mean bias	One sample *t*-statistic	*p*-value	Cohen's *d*
	PB	NB				
CA	0	6	–0.46	–3.97*	0.003	–1.25
CF	0	9	–0.89	–5.94*	<0.001	–1.88
CH	6	0	+0.43	+4.21*	0.002	+1.33
CV	4	0	+0.29	+4.26*	0.002	+1.35
IA	2	0	+0.43	+4.08*	0.003	+1.29
IF	6	0	+0.51	+3.81*	0.004	+1.20
IH	0	4	–0.49	–3.77*	0.004	–1.19
IV	0	5	–0.35	–3.31*	0.009	–1.07

Note: 'PB' denotes the number of participants whose bias was positive and significantly different from zero and 'NB' denotes the number of participants with a bias that is negative and significantly different from zero.

Table 5.2. *Individual and group statistics on categorical biases.*

Category	Number of participants		Mean bias	One sample *t*-statistic	*p*-value	Cohen's *d*
	PB	NB				
Animal	0	6	–0.39	–6.06*	<0.001	–1.92
Flower	0	8	–0.62	–4.31*	0.002	–1.36
House	7	0	+0.44	+5.29*	<0.001	+1.67
Vehicle	8	0	+0.34	+5.68*	<0.001	+1.80

Note: 'PB' denotes the number of participants whose bias was positive and significantly different from zero and 'NB' denotes the number of participants with a bias that is negative and significantly different from zero.

We conducted a one-way repeated measures analysis of variance and found a significant difference between mean categorical biases, $F(3, 27) = 25.83$, $p < 0.001$. As summarized in Table 5.3, pairwise comparisons (two-tailed *t*-tests) revealed that mean biases for houses and vehicles were significantly more positive than those for animals and flowers. There was no difference in mean biases between houses and vehicles or between those for animals and flowers.

Table 5.3. *Pairwise comparisons between mean categorical biases.*

Comparison	Mean difference	*p*-value
House – Animal	+0.83	<0.001
House – Flower	+1.06	0.005
House – Vehicle	-0.09	0.826
Vehicle – Animal	+0.74	<0.001
Vehicle – Flower	+0.96	0.004
Animal – Flower	+0.23	1.000

Note: *p*-values displayed are following Bonferroni corrections

5.3.2. Experiment 1: image statistics

Using values of orientation anisotropies computed on unfiltered images (see Methods), we allocated images in both sets C and I to two different groups. One group was named 'ANI' and included images that were highly anisotropic (anisotropy values in the range 0.7 to 1.3). The other group was named 'ISO' which had images with near perfect isotropy (anisotropy values between -0.3 and 0.3). Thirty-four percent of all images did not belong to either of these groups. We found that in the ANI group 95% of images were either houses or vehicles. Consequently, these categories may be considered to contain predominantly "carpentered" images (cf. Coppola et al., 1998; Switkes et al., 1978). In the ISO group 91% were either animals or flowers. Consequently, we consider the images in these categories to be "uncarpentered."

For each condition, we computed the difference in mean anisotropy (DMA) between hybrid components. For example, for condition CA, this corresponds to the difference between the mean anisotropy of all animal images used for the cardinal component and the mean anisotropy of all flower, house and vehicle images used for the intercardinal component. In Fig. 5.7 we plot the relationship between DMA and bias of each condition for each participant. We conducted a Pearson's correlation between bias and DMA of all conditions and found significant correlations ($p < 0.05$) for all 10

participants. There was also a strong correlation between mean bias across participants and DMA, $r(8) = 0.96$, $p < 0.001$, $R^2 = 0.93$. Accordingly, our results indicate that the more an image category is anisotropic, the less energy is required for its component at the point of subjective equality, irrespective of how it was presented in the hybrid.

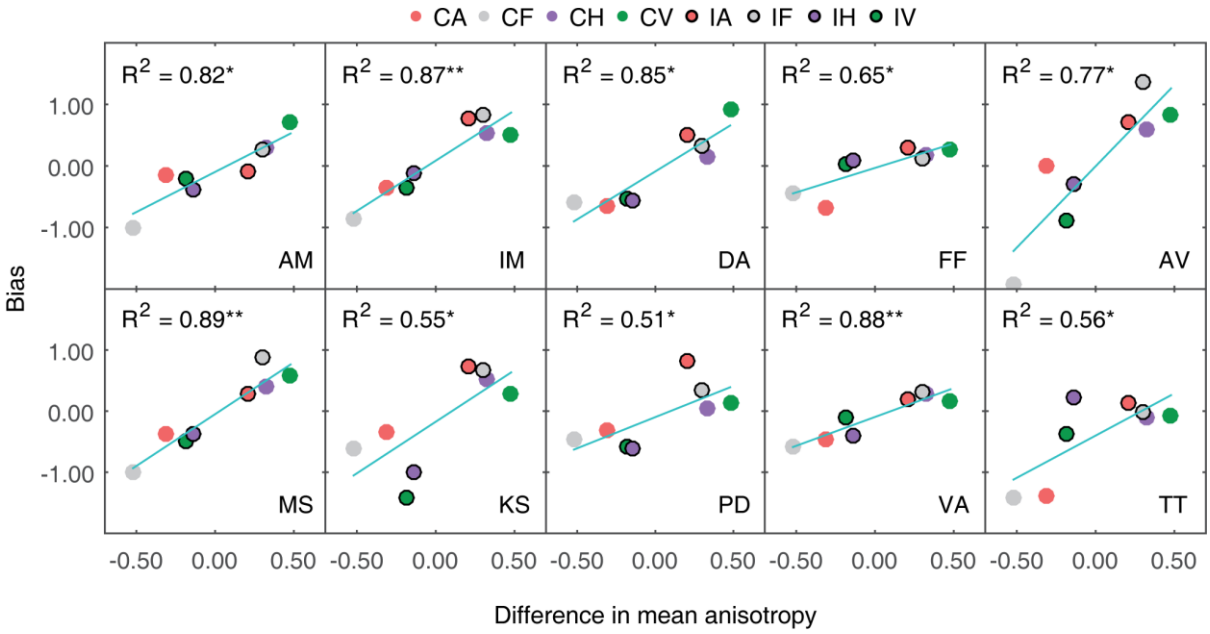


Figure 5.7. Correlation between orientation anisotropy and biases in hybrid conditions: Bias obtained from each condition (CA - cardinal animal, CF - cardinal flower, CH - cardinal house, CV - cardinal vehicle, IA - intercardinal animal, IF - intercardinal flower, IH - intercardinal house and IV - intercardinal vehicle) plotted against difference in mean anisotropy between the cardinal and intercardinal components. Each subplot represents a participant. Each condition is denoted by a uniquely coloured data point (with or without a black border). The coefficient of determination (R^2) of Pearson's correlation for each participant is given at the top left of subplots. Solid blue lines are linear least squares regression fits to the data and asterisks denote significant correlations, *: $p < 0.05$, **: $p < 0.001$.

5.3.3. Experiment 2: detection

We obtained estimates (see Fig. 5.8) of each participant's 63% correct threshold (α; point of inflection of the unscaled sigmoid), for each of the four image categories, by maximum-likelihood fitting a Weibull distribution to the psychometric function mapping log target RMS contrast to the proportion of trials on which the target (rather than the phase-scrambled non-target) was selected. A one-way repeated measures analysis of variance performed on mean thresholds (across participants) revealed no significant difference between image categories, $F(3,12) = 2.84$, $p = 0.08$. Although this p value is approaching significance, none of the (Bonferroni corrected) pairwise comparisons between image categories revealed a significant difference in mean thresholds at the level of $p < 0.05$.

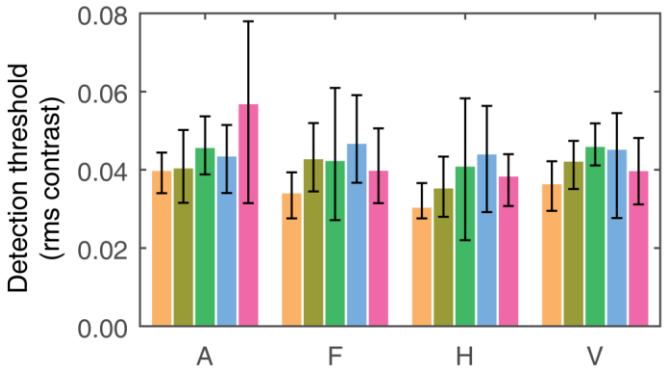


Figure 5.8. Detection thresholds for each image category (A - animal, F - flower, H - house, V - vehicle). Each uniquely coloured bar represents an individual participant. Error bars denote bias-corrected and accelerated 95% confidence intervals.

5.4. Discussion

Here we examined biases in people's categorization of different types of natural images. It was found that, when an ambiguous hybrid image was formed of structures from two different image categories, classification was biased towards the carpentered categories (houses and vehicles) rather than towards the uncarpentered categories (animals and flowers). We propose that this "carpentered bias" is the result of expectations about the world that favor the rapid interpretation of complex scenes as carpentered. Given that the visual diet of our participants is rich in carpentered structures, our results are consistent with a Bayesian formulation of perceptual biases whereby ambiguous stimuli result in biases towards frequently occurring features (Knill et al., 1996).

Schyns and Oliva (1994) report classification biases for the low spatial frequency component in spatial frequency hybrids for short durations only. They interpret this as evidence for a temporal prioritization of low spatial frequency processing during classification, supporting the idea of a "coarse-to-fine" strategy of processing visual information (Breitmeyer, 1975). This strategy is proposed to arise from our experience with natural scenes, whose power spectra display a dominance of low spatial frequency structures (Hughes, Nozawa, & Kitterle, 1996). In the current study, the carpentered bias occurred irrespective of whether the components were filtered to contain near-cardinal or near-intercardinal orientations. Therefore, despite cardinal orientations dominating the structure of our environment (Coppola et al., 1998; Switkes et al., 1978), we find no evidence of an equivalent orientation bias, whereby cardinal orientations would be processed before non-cardinal orientations which would result in a bias to see the image containing cardinal orientations, regardless of its category.

We stress that the carpentered bias is not merely a manifestation of the relative insensitivity to tilted (i.e., neither vertical nor horizontal) contours known as the oblique effect (Appelle, 1972; Berkley et al., 1975). Not only did the participants exhibit biases in favor of houses and vehicles when cardinal orientations had been filtered out of these carpentered components, the detection experiment revealed that houses and vehicles were not any more readily detected than images from the non-carpentered categories. Whereas the oblique effect was established using narrow-band luminance gratings on otherwise uniform backgrounds, it is not expected to influence the perception of broad-band, natural images, such as those used in experiments here. Indeed, if anything, detection thresholds for cardinally oriented structure tend to be higher than those for tilted structure, when those structures are superimposed on broad-band masking stimuli (Essock, DeFord, Hansen, & Sinai, 2003).

We do not claim that intercardinal filtering suffices to remove all easily detectable structures from the images in carpentered categories. Indeed, houses and vehicles almost certainly contain longer straight contours than flowers and animals. However, the results of the detection experiment provide strong ammunition against any sensitivity-based model of the carpentered bias. Whatever structure is contained in unfiltered images of houses and vehicles, on average that structure proved to be no easier to detect than the structure contained in unfiltered images of animals and flowers.

The absence of a difference in sensitivity appears to contradict findings from Crouzet, Joubert, Thorpe, and Fabre-Thorpe (2012), who report that the detection of animals precedes that of vehicles using a saccadic choice task. However comparing contrast sensitivity (detection) to saccadic reaction (decision) is problematic, especially with high contrast stimuli (Carpenter, 2004). Secondly, the difference could be attributed to

the background of images that must be categorized. While Crouzet et al. (2012) controlled contextual masking effects on image category by presenting images occurring in both carpentered and natural contexts, images in the current study were embedded in white noise with the same amplitude spectrum as the image. As Hansen and Loschky (2013) report, the type of mask used (e.g., using a mask sharing only the amplitude spectrum with the image versus one sharing both amplitude and phase information with the image) affects masking strength. It is still unclear which type of masks work best across different image categories (Hansen and Loschky, 2013).

Why might we have a *carpentered bias*? Clifford, Mareschal, Otsuka, and Watson (2015) proposed that certain biases may have potential benefits in daily life, minimizing the cost of errors. For example, in the gender bias, Armann and Bulthoff (2012) suggested that classifying a male face as female could be relatively costlier than the other way around. Given the city context we live in, it might be advantageous to first interpret our surroundings as carpentered. Carpentered environments are better at providing cues for perceptual judgements, notably providing information about the direction of gravity and influencing our judgements of subjective visual vertical (Haji-Khamneh & Harris, 2010). Greene (2013) analysed object frequency in image categories, by measuring the proportion of scenes of a category in which a given object naturally occurring in that category is present. She found that, in general, object frequency was higher for carpentered (indoor and outdoor) scenes than natural scenes. Therefore, our familiarity with carpentered scenes may facilitate identifying things frequently occurring in them. In support of this this view, Remy et al. (2013) found that the 'congruency effect' (i.e., our ability to quickly and more accurately recognise objects when they appear in their naturally occurring contexts) is stronger for carpentered objects in carpentered scenes than for non-carpentered objects in non-

carpentered scenes. When the carpentered nature of environments confers decisional advantages, misclassifying a carpentered scene as non-carpentered may be behaviourally costlier than the other way around.

6. Chapter 6 - Summary

In this thesis, we addressed three different aspects of higher-level natural image representations in the human visual system. We first examined if we possess neural mechanisms that selectively encode higher-level properties of meaningful natural images, and found a mechanism that selectively encodes image "uprightness". Next, we examined if we prioritise specific low-level properties of edges when encoding higher-level representations of images. We found that neural mechanisms underlying the classification of meaningful natural images are highly sensitive to edges near cardinal (vertical and horizontal) orientations that dominate natural scenes we mostly experience. Finally, we examined if our semantic classification of images is influenced by our expectations about the environment, over and above the visual system's differences in sensitivity to low-level properties of edges that construct the image. We found that, in cases of high ambiguity people are biased to classify natural images as "carpentered" (man-made), possibly because we live in an urban environment. The rest of this chapter will summarise these major findings and discuss their implications.

6.1. Selectivity to a higher-level natural image property

Adaptation after-effects have been widely used among psychophysicists to reveal the sensitivity of the mechanisms involved in image processing. However, the majority of the studies have identified selective mechanisms for very basic low-level features of localised edges in images such as luminance contrast, orientation and spatial frequency. In general these studies find that prolonged exposure to a specific feature value (the "adaptor"), repulsively biases the perception of a subsequently viewed feature value (the "test"). These repulsive after-effects are used as evidence to suggest the existence

of neural mechanisms that are tuned to specific feature values. In most cases, selective encoding of low-level features has been attributed to neurons in the early stages of the visual system, such as V1 (sections 1.2.1. and 1.2.2.). Some studies have also revealed selectivity of the visual system to complex properties of natural images. When measuring selectivity to natural image properties, it is very important to demonstrate that after-effects to these properties: 1) do not simply reflect adaptation to low-level features and 2) do not reflect adaptation to a complex property that is not unique to natural images *per se*. By doing so, one can infer the existence of neural mechanisms specifically encoding natural image properties, while also proposing most likely candidate cortical regions that would be involved in this selectivity. Accordingly, several studies have demonstrated selectivity to properties of natural images, most often using faces as stimuli (section 1.2.4.1).

A few studies have examined higher-level properties of natural images other than faces and provided evidence for the existence of mechanisms selectively encoding properties like viewpoint of objects and naturalness of scenes (Fang and He, 2005; Greene & Oliva, 2010; Kaping et al., 2007). However, most studies on adaptation after-effects, either using artificial or natural stimuli, have relied on using the method of single stimuli (MOSS), where a viewer adapts to a single image and is also tested on a single image. There is no way to be certain whether after-effects obtained using the MOSS reflect perceptual biases or non-perceptual biases such as response biases (Morgan et al., 2013). Therefore, experimental designs have recently been developed to measure after-effects largely uncontaminated from non-perceptual biases (Jogan & Stocker, 2014; Morgan, 2014; Morgan, Grant, Melmoth, & Solomon, 2015), but so far, these designs have only been applied to measure adaptations to low-level features.

The aim of chapters 1 and 2 was to examine after-effects to the "uprightness" or the global orientation of images of houses. With carefully designed experiments we showed that: 1) we have neural mechanisms selectively encoding uprightness, 2) this mechanism is distinct from those low-level mechanisms encoding the orientation of localised edges, 3) uprightness is a unique property encoded from stimuli that contain higher-order structure which conveys semantic meaning and 4) uprightness after-effects weren't a result of non-perceptual sources of bias.

The first two claims are supported by findings showing that adapting to house images tilted in the fronto-parallel plane repulsively biased the appearance of subsequently viewed test houses. After-effects to uprightness survived manipulations of spatial position and spatial frequency overlap between adaptors and tests. However, contrary to many findings in the literature (sections 2.1.2 and 2.1.3), after-effects to the local orientation of gratings also survived these manipulations. The major finding that distinguished after-effects to uprightness and local orientations is that, when the difference in mean orientation content between adaptor and tests was 90°, after-effects to uprightness survived whereas the other did not. Together, these findings point to a higher-level neural mechanism that encodes uprightness that is invariably responsive to changes in spatial position, spatial frequency and orientation of local edges present in the image.

After-effects to uprightness were unique to house images retaining higher-order structure and were not produced by adapting to images whose higher-order structure is fully distorted. Therefore, higher-order structure that conveys the semantic meaning of a scene (Andrews, Clarke, Pell, & Hartley, 2010; Coggan, Liu, Baker, & Andrews, 2016; Loschky et al., 2010) appears to be a prerequisite for encoding uprightness. However, it

must be noted that we have only examined selectivity to uprightness using a single category of images. Future works must therefore examine if this generalises to other categories of images, while making sure to avoid using images that people are poor at judging for uprightness as is the case with uncarpentered natural scenes (Haji-Khamneh & Harris, 2010). Given that the magnitude of the after-effects we report are comparatively smaller than tilt after-effects reported for reduced stimuli like gratings, it is possible that poor uprightness judgements could dilute after-effects of selectivity to uprightness. Nonetheless, this is the first time a psychophysical method immune to non-perceptual biases like response and decisional biases have been exploited to examine adaptation after-effects to higher-level natural image properties. We stress that this psychophysical method qualifies to examine after-effects to more natural image properties with strong immunity against non-perceptual biases.

A mechanism to encode the global orientation of a scene is advantageous because it may play a role in several other perceptual components of vision. Global scene orientation influences perception by: 1) informing the viewer of the direction of gravity (Howard & Childerson, 1994), 2) affecting judgements of subjective vertical of localised edges within a scene (Haji-Khamneh & Harris, 2010) and 3) facilitating semantic classification, since uprightness in most cases reflect the canonical orientation in which we are used to seeing images in the real world (Loschky et al., 2015; also see chapter 4). In addition, global scene orientation is also known to affect behavioural components of vision, for instance, showing its influence on the pattern of eye movements during the initial exposure to a scene (Foulsham et al., 2008). Therefore, it is possible that uprightness mechanisms exert feedback modulations on mechanisms encoding other image attributes and mechanisms mediating associated oculomotor behaviour. Further work is required to elucidate the nature of these interactions.

Studying global properties like the uprightness of an image is not only important to unravel attribute-selective neural mechanisms, but can also reveal how the visual system in different populations functions. For example, there is ample evidence to show that some individuals with autism demonstrate enhanced processing of fine local information within an image while also lacking a sense of its global properties (Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). Adaptation after-effects can be used to examine if neural mechanisms responsible for encoding global image properties are present in autistic individuals who are known to have impaired global perception.

6.2. Low-level features of edges and scene classification

Vision has evolved to efficiently capture the information typically present in the environments in which we live. Most studies have examined the relationship between the anisotropic distribution of local edge features in the environment and the properties of early cortical mechanisms encoding these features (Bex et al., 2009; Girshick et al., 2011; Hansen et al., 2008). Recently, some findings have revealed how our visual system prioritises some low-level features typically experienced in everyday life to facilitate higher-level meaningful attributes we perceive from images. For example, structures near the horizontal axes are found to be relatively more informative when perceiving meaningful attributes of faces and structures near low spatial frequencies are processed more rapidly when passively classifying faces and scenes (Dakin & Watt, 2009; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Oliva & Schyns, 1997; Schyns & Oliva, 1994).

In a similar vein, we found that people are relatively more sensitive to structures near the cardinal axes when classifying natural scenes. Most importantly, this did not occur as a result of non-uniformities in sensitivity of V1 neurons in detecting oriented structures. The most plausible explanation for the findings in chapter 4 is that higher-level scene classifying mechanisms prioritise cardinally oriented structures. Since natural scenes we often encounter are characterised by a dominance of cardinal structures (Coppola et al., 1998; Switkes et al., 1978), it is perhaps advantageous for scene classifying mechanisms to prioritise this information as they might contain most of the structures diagnostic of a scene's semantic category.

In a hierarchical framework of scene perception, early cortical regions such as V1 are believed to encode low-level properties of scenes like orientation and spatial frequency, and neurons down the hierarchy encode more complex attributes of scenes (Andrews, Watson, Rice, & Hartley, 2015; Felleman & Van Essen, 1991; Hochstein & Ahissar, 2002). Generally, it is believed that higher-level regions such as the PPA encode higher-dimensional attributes like spatial layout (e.g., openness) and semantic category (e.g., carpentered) of images. This is because response properties of higher-level regions are specific to the semantic category and spatial layout of images and the two properties can also be used in turn to predict response patterns in scene selective regions (Huth et al., 2012; Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011; Park, Konkle, & Oliva, 2014; Stansbury et al., 2013; Walther et al., 2009; Walther et al., 2011). The invariable behavioural judgements of higher-level image attributes on stimuli with large manipulations of low-level features also support this claim (section 4.1.2.). However, some recent findings challenge this and show that higher-level regions may in fact be sensitive to low-level properties like orientation and spatial frequency of images that are otherwise identical in semantic attributes (Nasr & Tootell,

2012; Watson, Hartley, & Andrews, 2014, 2017; Watson et al., 2016). These studies also propose that scene selective regions can distinguish between semantic categories based on low-level properties that covary with an image's category.

The afore-mentioned studies (Nasr & Tootell, 2012; Watson et al., 2014, 2017; Watson et al., 2016) have generally relied on measuring differences in BOLD response patterns in higher-level areas to stimuli created by filtering natural images to retain some feature values, but they did not find behavioural judgements of higher-level scene attributes that vary alongside differential representations of low-level features in higher-level areas. This raises a question as to what is the use of differential representations if we do not know how they influence the ultimate perceptual judgements of an observer.

The findings from chapter 4 establish a possible link between non-uniform representations of low-level features in higher-level areas and resultant behavioural judgements of image category. We suggest that experiments measuring behavioural judgements must be incorporated into experiments measuring higher-level cortical response patterns to unravel interdependencies between their sensitivity to low-level properties and categorical perception. Furthermore, although some studies have examined the visual system's sensitivity to other low-level properties like spatial frequency of edges or more complex properties like $1/f^{\alpha}$ spectral slope of naturalistic stimuli (Bex et al., 2009; Tadmor & Tolhurst, 1994), they haven't really explored how it affects scene classification. Since these are properties we might be using to judge the semantic category of a scene, it is important to understand if a visual system optimized to encode specific features is in fact facilitating classification using those features.

6.3. Priors for meaningful natural image properties

Visual perception is a product of two components which form the crux of the Bayesian formulation of vision - a sensory representation of a stimulus combined with our expectations about the stimulus. While chapter 4 demonstrated how image classification is affected by non-uniformities in our sensory representation of an image, chapter 5 showed that expectations might play a role in influencing image classification, beyond non-uniformities in sensory representations. Here we exploited the Bayesian prediction that when the stimulus is highly degraded, people demonstrate biases reflecting our expectations about the environment. Accordingly, when natural image components from carpentered and uncarpentered categories were combined to create ambiguous hybrids, people's classifications of hybrids were biased towards carpentered categories. These hybrid components were matched for low-level visibility by using a "window of visibility" that was created based on existing models of sensitivity to low-level features like orientation and spatial frequency of edges. Accordingly, we are confident that biases were not the result of increased sensitivity to low-level structures that could be prevalent in carpentered images.

Structures of carpentered images contain a dozen other attributes that are prevalent in carpentered scenes such as long and straight contours and rectilinear structures and neurons in any cortical area that is involved in scene perception could be more sensitive to these structures. In fact, a recent study had shown that scene selective regions are highly responsive to rectilinear compared to curved or non-rectangular structures (Nasr, Echavarria, & Tootell, 2014). However, minimum contrasts required to detect structures from images that can be used to classify them did not differ between the image categories we have used to create hybrids. This clarified that biases weren't the result of

the visual system's increased sensitivity to attributes of structures that could be dominant in carpentered images.

Chapter 5 has only measured biases at a high level of stimulus uncertainty. A more comprehensive assessment of perceptual biases in the context of Bayesian perception would therefore require a manipulation of ambiguity in hybrid images to measure how biases change with different levels of ambiguity. When doing so, it is important to be careful of the source of ambiguity being manipulated. For example, in chapter 5, the high ambiguity in hybrids was a result of combining two images while also orientation filtering them, where the latter was aimed at examining if biases to carpentered images are a result of a dominance of cardinal orientations typical of scenes.

Nonetheless, chapter 5 has two major implications for research on natural image perception. Firstly, the current work has successfully implemented an experimental design that can measure perceptual biases for complex meaningful attributes of natural images, while also controlling for possible confounds arising from sensitivity to different attributes of structures that make-up the image. This method could be applied to examine biases for semantic properties of natural images like faces, objects and scenes, and also for complex spatial layout properties of scenes like openness. Secondly, given the possibility that expectations of people extensively experiencing urban views throughout their life may have resulted in carpentered biases we report, it raises the intriguing question of whether people (e.g., indigenous communities) living in different, non-urban environments display such biases. In fact, a few decades back, studies had shown how people living in different environments (e.g., city dwellers versus tribal communities) are differently susceptible to visual illusions like the Müller-

Lyer illusion which is believed to arise from how people make inferences based on their expectations of the environment (Jahoda, 1966; Segall, Campbell, & Herskovits, 1963).

Unfortunately, this idea of studying how expectations of people living in different environments may influence visual perception has not gained much popularity in vision research. For the most part influence of prior expectations on image perception has been studied by reverse engineering the shape of the prior common to a group of participants after measuring how different levels of stimulus uncertainty affect perceptual biases. Alternatively, Powell, Meredith, McMillin, and Freeman (2016) have modelled individual differences in the shape of the prior by comparing how perceptual biases to attributes like perceived speed of dot patterns in autistic and non-autistic individuals vary. They found that autistic individuals have a flatter prior for perceived speed compared to non-autistic individuals with a prior peaking at low-speeds. However, since these groups share the same environment, differences in priors cannot be attributed to the environment. Therefore, it is important that priors in people living in different environments are measured to reveal how perception has evolved in tandem with our environment.

7. Chapter 7 - References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057-1058. doi:10.1038/nn1312

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision, 9*(10). doi:10.1167/9.10.10

Aguirre, G. K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. *Neuroreport, 10*(1), 189-194. doi:10.1097/00001756-199901180-00036

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21*(2), 373-383. doi:10.1016/s0896-6273(00)80546-2

Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). SPATIAL CONTRAST ADAPTATION CHARACTERISTICS OF NEURONS RECORDED IN THE CATS VISUAL-CORTEX. *Journal of Physiology-London, 347*(FEB), 713-739.

Anderson, N. D., Habak, C., Wilkinson, F., & Wilson, H. R. (2007). Evaluating shape after-effects with radial frequency patterns. *Vision Research, 47*(3), 298-308. doi:10.1016/j.visres.2006.02.013

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage, 49*(1), 703-711.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision, 15*(7), 3-3.

Appelle, S. (1972). PERCEPTION AND DISCRIMINATION AS A FUNCTION OF STIMULUS ORIENTATION - OBLIQUE EFFECT IN MAN AND ANIMALS. *Psychological Bulletin, 78*(4), 266-&. doi:10.1037/h0033117

Armann, R., & Bulthoff, I. (2012). Male and female faces are only perceived categorically when linked to familiar identities - And when in doubt, he is a male. *Vision Research, 63*, 69-80. doi:10.1016/j.visres.2012.05.005

Asch, S. E., & Witkin, H. A. (1948). Studies in space orientation: I. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology, 38*(3), 325.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87*(6), 3102-3116.

Bachatene, L., Bharmauria, V., Cattan, S., Rouat, J., & Molotchnikoff, S. (2015). Reprogramming of orientation columns in visual cortex: a domino effect. *Scientific Reports, 5*. doi:10.1038/srep09436

Baddeley, R. J., & Hancock, P. J. (1991). A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings of the Royal Society of London B: Biological Sciences, 246*(1317), 219-223.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38*(2), 347-358.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *The Journal of physiology, 141*(2), 337-350.

Barlow, H. B., & Hill, R. M. (1963). Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature, 200*(4913), 1345.

Barton, J. J. S., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform, face area impair perception of facial configuration in prosopagnosia. *Neurology, 58*(1), 71-78.

Bednar, J. A., & Miikkulainen, R. (2000). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Computation, 12*(7), 1721-1740. doi:10.1162/089976600300015321

Berkley, M. A., Kitterle, F., & Watkins, D. W. (1975). GRATING VISIBILITY AS A FUNCTION OF ORIENTATION AND RETINAL ECCENTRICITY. *Vision Research, 15*(2), 239-244. doi:10.1016/0042-6989(75)90213-8

Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, A. C., & Welling, L. L. M. (2010). Face aftereffects suggest interdependent processing of expression and sex and of expression and race. *Visual Cognition, 18*(2), 255-274. doi:10.1080/13506280802708024

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America a-Optics Image Science and Vision, 19*(6), 1096-1106. doi:10.1364/josaa.19.001096

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision, 9*(10). doi:10.1167/9.10.1

Biederman, I. (1987). RECOGNITION-BY-COMPONENTS - A THEORY OF HUMAN IMAGE UNDERSTANDING. *Psychological Review, 94*(2), 115-147. doi:10.1037//0033-295x.94.2.115

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*(17), 2885-2899.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*(1), 38-64.

Blakemore, C., & Campbell, F. W. (1969). ON THE EXISTENCE OF NEURONS IN THE HUMAN VISUAL SYSTEM SELECTIVELY SENSITIVE TO THE ORIENTATION AND SIZE OF RETINAL IMAGES. *Journal of Physiology (Cambridge), 203*(1), 237-260.

Blakemore, C., Carpenter, R. H., & Georgeson, M. A. (1970). LATERAL INHIBITION BETWEEN ORIENTATION DETECTORS IN HUMAN VISUAL SYSTEM. *Nature, 228*(5266), 37-&. doi:10.1038/228037a0

Blakemore, C., Nachmias, J., & Sutton, P. (1970). The perceived spatial frequency shift: evidence for frequency-selective neurones in the human brain. *The Journal of physiology, 210*(3), 727-750.

Blakemore, C., & Tobin, E. A. (1972). LATERAL INHIBITION BETWEEN ORIENTATION DETECTORS IN CATS VISUAL-CORTEX. *Experimental Brain Research, 15*(4), 439-+.

Boi, M., Ogmen, H., & Herzog, M. H. (2011). Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus-Pikler display. *Journal of Vision, 11*(3). doi:10.1167/11.3.7

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39*(2), 257-269.

Boynton, G. M., & Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *Journal of Neuroscience, 23*(25), 8781-8787.

Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global Ensemble Texture Representations are Critical to Rapid Scene Perception. *Journal of Experimental Psychology-Human Perception and Performance, 43*(6), 1160-1176. doi:10.1037/xhp0000399

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897x00357

Breitmeyer, B. G. (1975). SIMPLE REACTION-TIME AS A MEASURE OF TEMPORAL RESPONSE PROPERTIES OF TRANSIENT AND SUSTAINED CHANNELS. *Vision Research, 15*(12), 1411-1412. doi:10.1016/0042-6989(75)90200-x

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7*(8), 880-886. doi:10.1038/nn1278

Caelli, T., Brettel, H., Rentschler, I., & Hilz, R. (1983). Discrimination thresholds in the two-dimensional spatial frequency domain. *Vision Research, 23*(2), 129-133.

Campbell, F. W., & Kulikowski, J. J. (1972). The visual evoked potential as a function of contrast of a grating pattern. *The Journal of physiology, 222*(2), 345-356.

Campbell, F. W., Kulikowski, J. J., & Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *The Journal of physiology, 187*(2), 427-436.

Campbell, F. W., & Maffei, L. (1971). The tilt after-effect: a fresh look. *Vision research*, *11*(8), 833-840.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of physiology, 197*(3), 551-566.

Carandini, M., & Heeger, D. J. (1994). SUMMATION AND DIVISION BY NEURONS IN PRIMATE VISUAL-CORTEX. *Science, 264*(5163), 1333-1336. doi:10.1126/science.8191289

Carpenter, R. H., & Blakemore, C. (1973). INTERACTIONS BETWEEN ORIENTATIONS IN HUMAN VISION. *Experimental Brain Research, 18*(3), 287-303.

Carpenter, R. H. S. (2004). Contrast, probability, and saccadic latency: Evidence for independence of detection and decision. *Current Biology, 14*(17), 1576-1580. doi:10.1016/j.cub.2004.08.068

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*(10), 913-919.

Churches, O., Nicholls, M., Thiessen, M., Kohler, M., & Keage, H. (2014). Emoticons in mind: An event-related potential study. *Social neuroscience, 9*(2), 196-202.

Clifford, C. W. G., Mareschal, I., Otsuka, Y., & Watson, T. L. (2015). A Bayesian approach to person perception. *Consciousness and Cognition, 36*, 406-413. doi:10.1016/j.concog.2015.03.015

Clifford, C. W. G., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society B-Biological Sciences, 267*(1454), 1705-1710.

Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2016). Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. *Neuroimage, 135*, 107-114.

Cohen, J. (1992). A POWER PRIMER. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Coltheart, M. (1971). Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review, 78*(2), 114.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real word. *Proceedings of the National Academy of Sciences of the United States of America, 95*(7), 4002-4006. doi:10.1073/pnas.95.7.4002

Coren, S. (1970). LATERAL INHIBITION AND GEOMETRIC ILLUSIONS. *Quarterly Journal of Experimental Psychology, 22*, 274-&. doi:10.1080/00335557043000212

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal Detection Precedes Access to Scene Category. *Plos One, 7*(12). doi:10.1371/journal.pone.0051471

Dakin, S. C., & Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *Journal of the Optical Society of America a-Optics Image Science and Vision, 15*(6), 1486-1499. doi:10.1364/josaa.15.001486

Dakin, S. C., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision, 9*(4), 2-2.

Daniel, P., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology, 159*(2), 203-221.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). SPATIAL-FREQUENCY SELECTIVITY OF CELLS IN MACAQUE VISUAL-CORTEX. *Vision Research, 22*(5), 545-559. doi:10.1016/0042-6989(82)90113-4

Dekel, R., & Sagi, D. (2015). Tilt aftereffect due to adaptation to natural stimuli. *Vision Research, 117*, 91-99. doi:10.1016/j.visres.2015.10.014

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Li, F. F., & Ieee. (2009, Jun 20-25). *ImageNet: A Large-Scale Hierarchical Image Database.* Paper presented at the IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops, Miami Beach, FL.

Desimone, R., & Schein, S. J. (1987). VISUAL PROPERTIES OF NEURONS IN AREA V4 OF THE MACAQUE - SENSITIVITY TO STIMULUS FORM. *Journal of Neurophysiology, 57*(3), 835-868.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? *Neuron, 73*(3), 415-434. doi:10.1016/j.neuron.2012.01.010

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision, 3*(10), 586-598. doi:10.1167/3.10.1

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron, 28*(1), 287-298. doi:10.1016/s0896-6273(00)00103-3

Duong, T., & Freeman, R. D. (2007). Spatial frequency-specific contrast adaptation originates in the primary visual cortex. *Journal of Neurophysiology, 98*(1), 187-195. doi:10.1152/jn.01364.2006

Dyde, R. T., Jenkin, M. R., & Harris, L. R. (2006). The subjective visual vertical and the perceptual upright. *Experimental Brain Research, 173*(4), 612-622. doi:10.1007/s00221-006-0405-y

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the bootstrap*. CRC press

Engel, S. A., & Furmanski, C. S. (2001). Selective adaptation to color contrast in human primary visual cortex. *Journal of Neuroscience, 21*(11), 3949-3954.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex, 7*(2), 181-192. doi:10.1093/cercor/7.2.181

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601. doi:10.1038/33402

Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research, 43*(12), 1329-1335.

Fang, F., & He, S. (2005). Viewer-centred object representation in the human visual system revealed by viewpoint after-effects. *Neuron, 45*(5), 793-800

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:10.3758/bf03193146

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12*(6), 893-924.

Felleman, D. J., & Kaas, J. H. (1984). RECEPTIVE-FIELD PROPERTIES OF NEURONS IN MIDDLE TEMPORAL VISUAL AREA (MT) OF OWL MONKEYS. *Journal of Neurophysiology, 52*(3), 488-513.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*(1), 1-47. doi:10.1093/cercor/1.1.1

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience, 8*(12), 1643-1646.

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience, 23*, 441-471. doi:10.1146/annurev.neuro.23.1.441

Field, D. J., Hayes, A., & Hess, R. F. (1993). CONTOUR INTEGRATION BY THE HUMAN VISUAL-SYSTEM - EVIDENCE FOR A LOCAL ASSOCIATION FIELD. *Vision Research, 33*(2), 173-193. doi:10.1016/0042-6989(93)90156-q

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research, 50*(8), 779-795. doi:10.1016/j.visres.2010.01.019

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research, 48*(17), 1777-1790. doi:10.1016/j.visres.2008.05.018

Freeman, R., & Thibos, L. (1975). Visual evoked responses in humans with abnormal visual experience. *The Journal of physiology, 247*(3), 711-724.

Freeman, W. T. (1994). THE GENERIC VIEWPOINT ASSUMPTION IN A FRAMEWORK FOR VISUAL-PERCEPTION. *Nature, 368*(6471), 542-545. doi:10.1038/368542a0

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of experimental psychology: General, 108*(3), 316.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience, 3*(6), 535-536. doi:10.1038/75702

Gattass, R., Sousa, A. P. B., & Gross, C. G. (1988). VISUOTOPIC ORGANIZATION AND EXTENT OF V3 AND V4 OF THE MACAQUE. *Journal of Neuroscience, 8*(6), 1831-1845.

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), 568-573. doi:10.1038/9224

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12*(3), 495-504. doi:10.1162/089892900562165

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science, 27*(3), 379-402. doi:10.1016/s0364-0213(03)00009-0

Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience, 5*(6), 508-510. doi:10.1038/nn0602-508

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*(6), 711-724. doi:10.1016/s0042-6989(00)00277-7

Georgeson, M. A., & Blakemore, C. (1973). Apparent Depth and the Müller—Lyer Illusion. *Perception, 2*(2), 225-234.

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of physiology, 252*(3), 627-656.

Gerardin, P., Kourtzi, Z., & Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 107*(37), 16309-16314. doi:10.1073/pnas.1006285107

Gibson, J. J. (1937). Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology, 20*(6), 553.

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology, 20*(5), 453.

Gilbert, C. D., & Wiesel, T. N. (1979). MORPHOLOGY AND INTRA-CORTICAL PROJECTIONS OF FUNCTIONALLY CHARACTERIZED NEURONS IN THE CAT VISUAL-CORTEX. *Nature, 280*(5718), 120-125. doi:10.1038/280120a0

Gilbert, C. D., & Wiesel, T. N. (1989). COLUMNAR SPECIFICITY OF INTRINSIC HORIZONTAL AND CORTICOCORTICAL CONNECTIONS IN CAT VISUAL-CORTEX. *Journal of Neuroscience, 9*(7), 2432-2442.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*(7), 926-U156. doi:10.1038/nn.2831

Gizzi, M. S., Katz, E., Schumer, R. A., & Movshon, J. A. (1990). SELECTIVITY FOR ORIENTATION AND DIRECTION OF MOTION OF SINGLE NEURONS IN CAT STRIATE AND EXTRASTRIATE VISUAL-CORTEX. *Journal of Neurophysiology, 63*(6), 1529-1543.

Goddard, E., Clifford, C. W. G., & Solomon, S. G. (2008). Centre-surround effects on perceived orientation in complex images. *Vision Research, 48*(12), 1374-1382. doi:10.1016/j.visres.2008.02.023

Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology, 1*.

Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00777

Gorea, A. (2015). A refresher of the original Bloch's Law paper (bloch, july 1885). *i-Perception, 6*(4), 2041669515593043.

Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science, 20*(4), 464-472. doi:10.1111/j.1467-9280.2009.02316.x

Greene, M. R., & Oliva, A. (2010). High-Level Aftereffects to Global Scene Properties. *Journal of Experimental Psychology-Human Perception and Performance, 36*(6), 1430-1442. doi:10.1037/a0019058

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159-166. doi:10.1016/s0959-4388(03)00040-0

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555-562. doi:10.1038/nn1224

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping, 6*(4), 316-328. doi:10.1002/(sici)1097-0193(1998)6:4<316::aid-hbm9>3.0.co;2-6

Haji-Khamneh, B., & Harris, L. R. (2010). How different types of scenes affect the Subjective Visual Vertical (SVV) and the Perceptual Upright (PU). *Vision Research, 50*(17), 1720-1727. doi:10.1016/j.visres.2010.05.027

Hansen, B. C., & Essock, E. A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision, 4*(12), 5-5.

Hansen, B. C., Haun, A. M., & Essock, E. A. (2008). The ‘‘horizontal effect’’: A perceptual anisotropy in visual processing of naturalistic broadband stimuli. *Visual cortex: New research. New York, NY: Nova Science Publishers*.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi:10.1126/science.1063736

Heeger, D. J. (1992). NORMALIZATION OF CELL RESPONSES IN CAT STRIATE CORTEX. *Visual Neuroscience, 9*(2), 181-197.

Helmholtz, H. V. (1925). Treatise on Physiological Optics (Southall, J. P. C., Trans). Electronic edition (2001): University of Pennsylvania. Retrieved from http//psych.upenn.edu/backuslab/helmholtz

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: an fMRI study. *Experimental Brain Research, 179*(1), 75-84.

Hess, R. F., Dakin, S. C., & Field, D. J. (1998). The role of "contrast enhancement" in the detection and appearance of visual contours. *Vision Research, 38*(6), 783-787. doi:10.1016/s0042-6989(97)00333-7

Hess, R. F., Hayes, A., & Field, D. J. (2003). Contour integration and cortical processing. *Journal of Physiology-Paris, 97*(2-3), 105-119. doi:10.1016/j.jphysparis.2003.09.013

Hochberg, J., & Galper, R. E. (1967). Recognition of faces: I. An exploratory study. *Psychonomic Science, 9*(12), 619-620.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804. doi:10.1016/s0896-6273(02)01091-7

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature, 436*(7047), 71-77. doi:10.1038/nature03689

Howard, I. P., & Childerson, L. (1994). THE CONTRIBUTION OF MOTION, THE VISUAL FRAME, AND VISUAL POLARITY TO SENSATIONS OF BODY TILT. *Perception, 23*(7), 753-762. doi:10.1068/p230753

Hsu, S. M., & Young, A. W. (2004). Adaptation effects in facial expression recognition. *Visual Cognition, 11*(7), 871-899. doi:10.1080/13506280444000030

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591.

Hubel, D. H., & Wiesel, T. N. (1961). Integrative action in the cat's lateral geniculate body. *The Journal of physiology, 155*(2), 385-398.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology, 160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology, 195*(1), 215-243.

Hubel, D. H., & Wiesel, T. N. (1974). UNIFORMITY OF MONKEY STRIATE CORTEX - PARALLEL RELATIONSHIP BETWEEN FIELD SIZE, SCATTER, AND MAGNIFICATION FACTOR. *Journal of Comparative Neurology, 158*(3), 295-306. doi:10.1002/cne.901580305

Huettel, S. A., & McCarthy, G. (2000). Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. *Neuroimage, 11*(5), 547-553. doi:10.1006/nimg.2000.0553

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience, 8*(3), 197-230. doi:10.1162/jocn.1996.8.3.197

Hughes, J. F., Van Dam, A., Mcguire, M., Sklar, D. F., Foley, J. D., Feiner, S. K., & Akeley, K. (2013). *Computer Graphics: Principle and Practice* (3rd ed.). Ohio: Addison-Wesley.

Hunt, J. J., Bosking, W. H., & Goodhill, G. J. (2011). Statistical structure of lateral connections in the primary visual cortex. *Neural systems & circuits, 1*(1), 3.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron, 76*(6), 1210-1224. doi:10.1016/j.neuron.2012.10.014

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). SIZE AND POSITION INVARIANCE OF NEURONAL RESPONSES IN MONKEY INFEROTEMPORAL CORTEX. *Journal of Neurophysiology, 73*(1), 218-226.

Jahoda, G. (1966). Geometric illusions and environment: A study in Ghana. *British Journal of Psychology, 57*(1-2), 193-199.

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me - Direction-specific aftereffects in gaze perception. *Psychological Science, 17*(6), 506-513. doi:10.1111/j.1467-9280.2006.01736.x

Jiang, F., Blanz, V., & O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation - A view from the other side of the mean. *Psychological Science, 17*(6), 493-500. doi:10.1111/j.1467-9280.2006.01734.x

Jin, D. Z., Dragoi, V., Sur, M., & Seung, H. S. (2005). Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex. *Journal of Neurophysiology, 94*(6), 4038-4050. doi:10.1152/jn.00571.2004

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision, 14*(3). doi:10.1167/14.3.20

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289-303.

Jolicoeur, P., & Milliken, B. (1989). IDENTIFICATION OF DISORIENTED OBJECTS - EFFECTS OF CONTEXT OF PRIOR PRESENTATION. *Journal of Experimental Psychology-Learning Memory and Cognition, 15*(2), 200-210. doi:10.1037/0278-7393.15.2.200

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience, 9*(1), 133-142. doi:10.1162/jocn.1997.9.1.133

Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition, 15*(1), 12-19. doi:10.1080/13506280600856660

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research, 107*, 49-57.

Kelly, D., & Savoie, R. (1978). Theory of flicker and transient responses. III. An essential nonlinearity. *JOSA, 68*(11), 1481-1490.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304. doi:10.1146/annurev.psych.55.090902.142005

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762-1776.

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision, 10*(1). doi:10.1167/10.1.8

Knill, D. C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. *Perception as Bayesian inference*, 1-21.

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*(3), 856-867.

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology, 97*(5), 3155-3164. doi:10.1152/jn.00086.2007

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971), 244-247. doi:10.1038/nature02169

Kovács, G., Zimmer, M., Banko, E., Harza, I., Antal, A., & Vidnyanszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cerebral Cortex, 16*(5), 742-753. doi:10.1093/cercor/bhj020

Kovacs, I., & Julesz, B. (1993). A CLOSED CURVE IS MUCH MORE THAN AN INCOMPLETE ONE - EFFECT OF CLOSURE IN FIGURE GROUND SEGMENTATION. *Proceedings of the National Academy of Sciences of the United States of America, 90*(16), 7495-7497. doi:10.1073/pnas.90.16.7495

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *Journal of Neuroscience, 31*(20), 7322-7333.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*(1), 37-68.

Kurtenbach, W., & Magnussen, S. (1981). INHIBITION, DISINHIBITION, AND SUMMATION AMONG ORIENTATION DETECTORS IN HUMAN-VISION. *Experimental Brain Research, 43*(2), 193-198.

Ledgeway, T., Hess, R. F., & Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of "association field" models of contour integration. *Vision Research, 45*(19), 2511-2522. doi: 10.1016/j.visres.2005.04.002

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science, 11*(5), 379-385. doi:10.1111/1467-9280.00274

Lee, T. S. (2002). Top-down influence in early visual processing: a Bayesian perspective. *Physiology & Behavior, 77*(4-5), 645-650. doi:10.1016/s0031-9384(02)00903-4

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), 1434-1448. doi:10.1364/josaa.20.001434

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience, 4*(1), 89-94. doi:10.1038/82947

Lesmes, L. A., Lu, Z. L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision, 10*(3). doi:10.1167/10.3.17

Li, B., Peterson, M., & Freeman, R. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology, 90*(1), 204-217. doi:10.1152/jn.00954.2002

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, 99*(14), 9596-9601.

Liu, T., & Hou, Y. (2011). Global feature-based attention to orientation. *Journal of Vision, 11*(10):8, 1-8. doi:10.1167/11.10.8

Liu, T. T., Jiang, Y., Sun, X. H., & He, S. (2009). Reduction of the Crowding Effect in Spatially Adjacent but Cortically Remote Visual Stimuli. *Current Biology, 19*(2), 127-132. doi:10.1016/j.cub.2008.11.065

Liu, T. S., & Mance, I. (2011). Constant spread of feature-based attention across the visual field. *Vision Research, 51*(1), 26-33. doi:10.1016/j.visres.2010.09.023

Livingstone, M. S., & Hubel, D. H. (1984). SPECIFICITY OF INTRINSIC CONNECTIONS IN PRIMATE PRIMARY VISUAL-CORTEX. *Journal of Neuroscience, 4*(11), 2830-&.

Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research, 48*(20), 2106-2127. doi:10.1016/j.visres.2008.03.006

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8*(10), 1386-1390. doi:10.1038/nn1538

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). SHAPE REPRESENTATION IN THE INFERIOR TEMPORAL CORTEX OF MONKEYS. *Current Biology, 5*(5), 552-563. doi:10.1016/s0960-9822(95)00108-4

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics, 72*(2), 427-444.

Loschky, L. C., Ringer, R. V., Ellis, K., & Hansen, B. C. (2015). Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist. *Journal of Vision, 15*(6). doi:10.1167/15.6.11

Magnussen, S., & Johnsen, T. (1986). TEMPORAL ASPECTS OF SPATIAL ADAPTATION - A STUDY OF THE TILT AFTEREFFECT. *Vision Research, 26*(4), 661-672. doi:10.1016/0042-6989(86)90014-3

Magnussen, S., & Kurtenbach, W. (1980). Adapting to two orientations: disinhibition in a visual aftereffect. *Science, 207*(4433), 908-909.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., . . . Tootell, R. B. H. (1995). OBJECT-RELATED ACTIVITY REVEALED BY FUNCTIONAL MAGNETIC-RESONANCE-IMAGING IN HUMAN

OCCIPITAL CORTEX. *Proceedings of the National Academy of Sciences of the United States of America, 92*(18), 8135-8139. doi:10.1073/pnas.92.18.8135

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81*(1), B1-B9. doi:10.1016/s0010-0277(01)00116-0

Mannion, D. J., McDonald, J. S., & Clifford, C. W. (2010). Orientation anisotropies in human visual cortex. *Journal of Neurophysiology, 103*(6), 3465-3471.

Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2013). Humans Have an Expectation That Gaze Is Directed Toward Them. *Current Biology, 23*(8), 717-721. doi:10.1016/j.cub.2013.03.030

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649-652. doi:10.1038/379649a0

Mathot, S., & Theeuwes, J. (2013). A reinvestigation of the reference frame of the tilt-adaptation aftereffect. *Scientific Reports, 3*. doi:10.1038/srep01152

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*(1), 8-15. doi:10.1016/j.tics.2006.11.002

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology-Human Perception and Performance, 27*(3), 573-599. doi:10.1037//0096-1523.27.3.573

Meese, T. S., Baker, D. H., & Summers, R. J. (2017). Perception of global image contrast involves transparent spatial filtering and the integration and suppression of local contrasts (not RMS contrast). *Royal Society open science, 4*(9), 170285.

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology, 15*(19), 1745-1748. doi:10.1016/j.cub.2005.08.044

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience, 10*(7), 903-907. doi:10.1038/nn1917

Merigan, W. H., & Pham, H. A. (1998). V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Visual Neuroscience, 15*(2), 359-367.

Mitchell, D. E., & Muir, D. W. (1976). DOES TILT AFTEREFFECT OCCUR IN OBLIQUE MERIDIAN. *Vision Research, 16*(6), 609-613. doi:10.1016/0042-6989(76)90007-9

Mitchell, D. E., & Ware, C. (1974). INTEROCULAR TRANSFER OF A VISUAL AFTEREFFECT IN NORMAL AND STEREOBLIND HUMANS. *Journal of Physiology-London, 236*(3), 707-721. doi:10.1113/jphysiol.1974.sp010461

Mollon, J. (1977). Neural analysis. *The perceptual world*, 71-97.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill

Montaser-Kouhsari, L., Landy, M. S., Heeger, D. J., & Larsson, J. (2007). Orientation-selective adaptation to illusory contours in human visual cortex. *Journal of Neuroscience, 27*(9), 2186-2195. doi:10.1523/jneurosci.4173-06.2007

Morgan, M. J. (2014). A bias-free measure of retinotopic tilt adaptation. *Journal of Vision, 14*(1), 7-7.

Morgan, M. J., Grant, S., Melmoth, D., & Solomon, J. A. (2015). Tilted frames of reference have similar effects on the perception of gravitational vertical and the planning of vertical saccadic eye movements. *Experimental Brain Research, 233*(7), 2115-2125. doi:10.1007/s00221-015-4282-0

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience, 30*(5-6), 197-206. doi:10.1017/s095252381300045x

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of autism and developmental disorders, 36*(1), 27-43.

Movshon, J. A., & Lennie, P. (1979). PATTERN-SELECTIVE ADAPTATION IN VISUAL CORTICAL-NEURONS. *Nature, 278*(5707), 850-852. doi:10.1038/278850a0

Muir, D., & Over, R. (1970). TILT AFTEREFFECTS IN CENTRAL AND PERIPHERAL VISION. *Journal of Experimental Psychology, 85*(2), 165-&. doi:10.1037/h0029509

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. *Journal of Neuroscience, 34*(20), 6721-6735. doi:10.1523/jneurosci.4802-13.2014

Nasr, S., & Tootell, R. B. H. (2012). A Cardinal Orientation Bias in Scene-Selective Visual Cortex. *Journal of Neuroscience, 32*(43), 14921-14926. doi:10.1523/jneurosci.2036-12.2012

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology, 14*(4), 474-480.

Ng, M., Ciaramitaro, V. M., Anstis, S., Boynton, G. M., & Fine, I. (2006). Selectivity for the configural cues that identify the gender, ethnicity, and identity of faces in human cortex. *Proceedings of the National Academy of Sciences of the United States of America, 103*(51), 19552-19557. doi:10.1073/pnas.0605358104

Nowak, L., Munk, M., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience, 12*(2), 371-384.

O'Leary, A., & McMahon, M. (1991). ADAPTATION TO FORM DISTORTION OF A FAMILIAR SHAPE. *Perception & Psychophysics, 49*(4), 328-332. doi:10.3758/bf03205988

O'Neil, S. F., & Webster, M. A. (2011). Adaptation and the perception of facial age. *Visual Cognition, 19*(4), 534-550. doi:10.1080/13506285.2011.561262

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72-107.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175. doi:10.1023/a:1011139631724

Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Research, 44*(7), 669-683.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1? *Neural Computation, 17*(8), 1665-1699.

Pantle, A., & Sekuler, R. (1968). Size-detecting mechanisms in human vision. *Science, 162*(3858), 1146-1148.

Paradiso, M. A., Shimojo, S., & Nakayama, K. (1989). SUBJECTIVE CONTOURS, TILT AFTEREFFECTS, AND VISUAL CORTICAL ORGANIZATION. *Vision Research, 29*(9), 1205-1213. doi:10.1016/0042-6989(89)90066-7

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*(4), 1333-1340.

Park, S., Konkle, T., & Oliva, A. (2014). Parametric coding of the size and clutter of natural scenes in the human brain. *Cerebral cortex*, *25*(7), 1792-1805.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86*(5), 2505-2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*(12), 1332-1338. doi:10.1038/nn972

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442. doi:10.1163/156856897x00366

Pelli, D. G., & Zhang, L. (1991). ACCURATE CONTROL OF CONTRAST ON MICROCOMPUTER DISPLAYS. *Vision Research, 31*(7-8), 1337-1350. doi:10.1016/0042-6989(91)90055-a

Pizlo, Z., & Salachgolyska, M. (1995). 3-D SHAPE PERCEPTION. *Perception & Psychophysics, 57*(5), 692-714. doi:10.3758/bf03213274

Potter, M. C. (1975). MEANING IN VISUAL SEARCH. *Science, 187*(4180), 965-966. doi:10.1126/science.1145183

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*(1), 10-15. doi:10.1037/h0027470

Powell, G., Meredith, Z., McMillin, R., & Freeman, T. C. (2016). Bayesian models of individual differences: combining autistic traits and sensory thresholds to predict motion perception. *Psychological Science, 27*(12), 1562-1572.

Ramachandran, V. S., & Gregory, R. L. (1991). PERCEPTUAL FILLING IN OF ARTIFICIALLY INDUCED SCOTOMAS IN HUMAN VISION. *Nature, 350*(6320), 699-702. doi:10.1038/350699a0

Regan, D., & Hamstra, S. (1992). Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research, 32*(10), 1845-1864.

Remy, F., Saint-Aubert, L., Bacon-Mace, N., Vayssiere, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: A life-span study. *Vision Research, 91*, 36-44. doi:10.1016/j.visres.2013.07.006

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14*(23), 2119-2123. doi:10.1016/j.cub.2004.11.053

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: Diversity and Laminar dependence. *Journal of Neuroscience, 22*(13), 5639-5651.

Roach, N. W., & Webb, B. S. (2013). Adaptation to implied tilt: extensive spatial extrapolation of orientation gradients. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00438

Roach, N. W., Webb, B. S., & McGraw, P. V. (2008). Adaptation to global structure induces spatially remote distortions of perceived orientation. *Journal of Vision, 8*(3). doi:10.1167/8.3.31

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology-Human Perception and Performance, 33*(3), 570-592. doi:10.1037/0096-1523.33.3.570

Rolls, E. T., & Cowey, A. (1970). Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. *Experimental Brain Research, 10*(3), 298-310.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and cognitive neuroscience reviews, 1*(1), 63-75.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*(1), 107.

Roufs, J. (1974). Dynamic properties of vision—IV: Thresholds of decremental flashes, incremental flashes and doublets in relation to flicker fusion. *Vision Research, 14*(9), 831-851.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8*(12), 1647-1650.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). SPATIAL-FREQUENCY CHANNELS IN HUMAN VISION. *Journal of the Optical Society of America, 61*(9), 1176-&. doi:10.1364/josa.61.001176

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8*(5), 374-378. doi:10.1111/j.1467-9280.1997.tb00428.x

Schneider, K. A., Richter, M. C., & Kastner, S. (2004). Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: A high-resolution functional magnetic resonance imaging study. *Journal of Neuroscience, 24*(41), 8975-8985. doi:10.1523/jneurosci.2413-04.2004

Scholl, B. J. (2005). Innateness and (Bayesian) visual perception. *The innate mind: Structure and contents, 1*, 34.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research, 122*, 105-123.

Schyns, P. G., & Oliva, A. (1994). FROM BLOBS TO BOUNDARY EDGES - EVIDENCE FOR TIME-SCALE-DEPENDENT AND SPATIAL-SCALE-DEPENDENT SCENE RECOGNITION. *Psychological Science, 5*(4), 195-200. doi:10.1111/j.1467-9280.1994.tb00500.x

Sclar, G., & Freeman, R. D. (1982). ORIENTATION SELECTIVITY IN THE CATS STRIATE CORTEX IS INVARIANT WITH STIMULUS CONTRAST. *Experimental Brain Research, 46*(3), 457-461.

Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., . . . Schomer, D. L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport, 8*(12), 2749-2754.

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1963). Cultural differences in the perception of geometric illusions. *Science, 139*(3556), 769-771.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. H. (1995). BORDERS OF MULTIPLE VISUAL AREAS IN HUMANS REVEALED BY FUNCTIONAL MAGNETIC-RESONANCE-IMAGING. *Science, 268*(5212), 889-893. doi:10.1126/science.7754376

Shen, G., Tao, X., Zhang, B., Smith, E. L., & Chino, Y. M. (2014). Oblique effect in visual area 2 of macaque monkeys. *Journal of Vision, 14*(2), 3-3.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193-1216. doi:10.1146/annurev.neuro.24.1.1193

Singh, K. D., Smith, A. T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas. *Neuroimage, 12*(5), 550-564. doi: 10.1006/nimg.2000.0642

Smith, A., & Over, R. (1975). TILT AFTEREFFECTS WITH SUBJECTIVE CONTOURS. *Nature, 257*(5527), 581-582. doi:10.1038/257581a0

Smith, A. T., Scott-Samuel, N. E., & Singh, K. D. (2000). Global motion adaptation. *Vision Research, 40*(9), 1069-1075. doi:10.1016/s0042-6989(00)00014-6

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex, 11*(12), 1182-1190. doi:10.1093/cercor/11.12.1182

Snowden, R. J., & Milne, A. B. (1996). The effects of adapting to complex motions: position invariance and tuning to sprial motions. *Journal of Cognitive Neuroscience, 8*(5), 435-452. doi: 10.1162/jocn.1996.8.5.435

Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex. *Neuron, 79*(5), 1025-1034. doi:10.1016/j.neuron.2013.06.034

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578-585. doi:10.1038/nn1669

Stone, J. V., Kerrigan, I. S., & Porrill, J. (2009). Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings of the Royal Society B-Biological Sciences, 276*(1663), 1797-1804. doi:10.1098/rspb.2008.1635

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00157

Storrs, K. R., & Arnold, D. H. (2015). Face aftereffects involve local repulsion, not renormalization. *Journal of Vision, 15*(8). doi:10.1167/15.8.1

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience, 1*(3), 183-184. doi:10.1038/630

Sutherland, N. (1961). Figural after-effects and apparent size. *Quarterly Journal of Experimental Psychology, 13*(4), 222-228.

Suzuki, S. (2001). Attention-dependent brief adaptation to contour orientation: a high-level aftereffect for convexity? *Vision Research, 41*(28), 3883-3902. doi:10.1016/s0042-6989(01)00249-8

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). SPATIAL-FREQUENCY ANALYSIS OF VISUAL ENVIRONMENT - ANISOTROPY AND CARPENTERED ENVIRONMENT HYPOTHESIS. *Vision Research, 18*(10), 1393-&.

Tadmor, Y., & Tolhurst, D. (1994). Discrimination of changes in the second-order statistics of natural and synthetic images. *Vision Research, 34*(4), 541-554.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*(8), 764-769. doi:10.1038/77666

Tarr, M. J., & Pinker, S. (1989). MENTAL ROTATION AND ORIENTATION-DEPENDENCE IN SHAPE-RECOGNITION. *Cognitive Psychology, 21*(2), 233-282. doi:10.1016/0010-0285(89)90009-1

Taylor, G., Hipp, D., Moser, A., Dickerson, K., & Gerhardstein, P. (2014). The development of contour processing: evidence from physiology and psychophysics. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00719

Taylor, L. M., & Mitchell, P. (1997). Judgments of apparent shape contaminated by knowledge of reality: Viewing circles obliquely. *British Journal of Psychology, 88*, 653-670.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522. doi:10.1038/381520a0

Thouless, R. H. (1931). Phenomenal regression to the 'real'object. II. *British Journal of Psychology, 22*(1), 1-30.

Tolhurst, D. J., & Thompson, P. G. (1975). ORIENTATION ILLUSIONS AND AFTEREFFECTS - INHIBITION BETWEEN CHANNELS. *Vision Research, 15*(8-9), 967-972. doi:10.1016/0042-6989(75)90238-2

Tomassini, A., Morgan, M. J., & Solomon, J. A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision Research, 50*(5), 541-547. doi:10.1016/j.visres.2009.12.005

Tripathy, S. P., & Levi, D. M. (1994). LONG-RANGE DICHOPTIC INTERACTIONS IN THE HUMAN VISUAL-CORTEX IN THE REGION CORRESPONDING TO THE BLIND SPOT. *Vision Research, 34*(9), 1127-1138. doi:10.1016/0042-6989(94)90295-x

Troje, N. F., Sadr, J., Geyer, H., & Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *Journal of Vision, 6*(8), 850-857. doi:10.1167/6.8.7

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6*(4), 1160-1170.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci., 31*, 411-437.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience, 24*(46), 10440-10453.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*(4), 471-491.

Valentine, T., & Endo, M. (1992). TOWARDS AN EXEMPLAR MODEL OF FACE PROCESSING - THE EFFECTS OF RACE AND DISTINCTIVENESS. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 44*(4), 671-703.

Van der Schaaf, v. A., & van Hateren, J. v. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research, 36*(17), 2759-2770.

Van Der Zwan, R., & Wenderoth, P. (1995). MECHANISMS OF PURELY SUBJECTIVE CONTOUR TILT AFTEREFFECTS. *Vision Research, 35*(18), 2547-2557. doi:10.1016/0042-6989(95)00012-o

Wallace, G. (1969). The critical distance of interaction in the Zöllner illusion. *Perception & Psychophysics, 5*(5), 261-264.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience, 29*(34), 10573-10581. doi:10.1523/jneurosci.0559-09.2009

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences of the United States of America, 108*(23), 9661-9666. doi:10.1073/pnas.1015666108

Ware, C., & Mitchell, D. E. (1974). SPATIAL SELECTIVITY OF TILT AFTEREFFECT. *Vision Research, 14*(8), 735-737. doi:10.1016/0042-6989(74)90072-8

Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of vision*, *5*(9), 6-6.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *Neuroimage, 99*, 402-410.

Watson, D. M., Hartley, T., & Andrews, T. J. (2017). Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex. *Cortex, 92*, 162-174.

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage, 124*, 107-117. doi:10.1016/j.neuroimage.2015.08.058

Watson, T. L., & Clifford, C. W. G. (2003). Pulling faces: An investigation of the face-distortion aftereffect. *Perception, 32*(9), 1109-1116. doi:10.1068/p5082

Watson, T. L., Otsuka, Y., & Clifford, C. W. G. (2016). Who are you expecting? Biases in face perception reveal prior expectations for sex and age. *Journal of Vision, 16*(3). doi:10.1167/16.3.5

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*(6982), 557-561. doi:10.1038/nature02420

Webster, M. A., & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America a-Optics Image Science and Vision, 25*(11), 2817-2825. doi:10.1364/josaa.25.002817

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1571), 1702-1725. doi:10.1098/rstb.2010.0360

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6*(4), 647-653. doi:10.3758/bf03212974

Weigelt, S., Limbach, K., Singer, W., & Kohler, A. (2012). Orientation-selective functional magnetic resonance imaging adaptation in primary visual cortex revisited. *Human Brain Mapping, 33*(3), 707-714. doi:10.1002/hbm.21244

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598-604. doi:10.1038/nn858

Welchman, A. E., Lam, J. M., & Bulthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences of the United States of America, 105*(33), 12087-12092. doi:10.1073/pnas.0804378105

Weliky, M., Kandler, K., Fitzpatrick, D., & Katz, L. C. (1995). PATTERNS OF EXCITATION AND INHIBITION EVOKED BY HORIZONTAL CONNECTIONS IN VISUAL-CORTEX SHARE A COMMON RELATIONSHIP TO ORIENTATION COLUMNS. *Neuron, 15*(3), 541-552. doi:10.1016/0896-6273(95)90143-4

Wenderoth, P., & Johnstone, S. (1988). THE DIFFERENT MECHANISMS OF THE DIRECT AND INDIRECT TILT ILLUSIONS. *Vision Research, 28*(2), 301-312. doi:10.1016/0042-6989(88)90158-7

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American journal of ophthalmology, 46*(1), 102-113.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). *Sun database: Large-scale scene recognition from abbey to zoo.* Paper presented at the Computer vision and pattern recognition (CVPR), 2010 IEEE conference on.

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the cortical hierarchy: Low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience, 28*(13), 3374-3383. doi:10.1523/jneurosci.0182-08.2008

Xu, X., Collins, C. E., Khaytin, I., Kaas, J. H., & Casagrande, V. A. (2006). Unequal representation of cardinal vs. oblique orientations in the middle temporal visual area. *Proceedings of the National Academy of Sciences, 103*(46), 17490-17495.

Yarrow, K., Martin, S. E., Di Costa, S., Solomon, J. A., & Arnold, D. H. (2016). A roving dual-presentation simultaneity-judgment task to estimate the point of subjective simultaneity. *Frontiers in Psychology, 7*, 416.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian Decision Theory and Psychophysics. *Perception as Bayesian Inference*. Cambridge University Press.

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception, 42*(11), 1166-1178.

Zeki, S. M. (1978). UNIFORMITY AND DIVERSITY OF STRUCTURE AND FUNCTION IN RHESUS-MONKEY PRESTRIATE VISUAL-CORTEX. *Journal of Physiology-London, 277*(APR), 273-290.

Zemon, V., Conte, M. M., & Camisa, J. (1993). Stimulus orientation and contrast constancy. *International journal of neuroscience, 69*(1-4), 143-148.

Zimmermann, E., Morrone, M. C., Fink, G. R., & Burr, D. (2013). Spatiotopic neural representations develop slowly across saccades. *Current Biology, 23*(5), R193-R194. doi:10.1016/j.cub.2013.01.065

الحمد لله