THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Norms are Not the Norm: Testing Theories of Sensory Encoding using Visual Aftereffects

Katherine Rebecca Storrs

B.A., BPsycSc(Hons I)

A thesis submitted for the degree of Doctor of Philosophy at the University of Queensland in 2015

School of Psychology

ABSTRACT

The goal of visual neuroscience is to explain how patterns of neural activity give rise to visual experiences. Here I use visual aftereffects to explore the computational principles underlying the perception of simple, intermediate, and complex forms (orientations, shapes, and faces). Aftereffects occur when exposure to one stimulus changes the appearance of a subsequent stimulus. In the tilt aftereffect, for example, staring at a left-tilted line can make a vertical line seem briefly to lean to the right. Aftereffects are thought to be caused by neural adaptation — changes in the responsiveness of neurons after prolonged stimulation.

Visual aftereffects allow us to probe sensory encoding, because adaptation within different encoding schemes can predict different patterns of perceptual aftereffect. I focus on two theories of encoding prominent in the face recognition literature. According to norm-based accounts, faces are represented in terms of how they deviate from a unique norm, such as the average of all experienced faces. Similar proposals have been raised in other contexts, for example to encode the aspect ratio of shapes. Norm-based encoding can be implemented by a population in which neurons respond increasingly as a stimulus moves further in the neuron's preferred direction, away from a normative value in stimulus space. Alternatively, exemplar-based accounts propose that faces are encoded in terms of their similarity to a number of previously-experienced exemplars. Exemplar-based encoding can be implemented by a population of non-monotonically tuned neurons that prefer particular *points* in stimulus space, rather than directions.

Norm-based and exemplar-based accounts make distinct predictions for the pattern of aftereffects one should experience following adaptation. Norm-based encoding is associated with renormalisation, in which the adapting stimulus appears more 'neutral' or 'average' after prolonged viewing, and the appearances of other stimuli are altered in the same fashion (e.g. after adapting to a male face, all faces should look more feminine). Exemplar-based encoding is associated with 'local repulsion,' in which the adapting stimulus appears unchanged, but differences between it and subsequent stimuli are exaggerated (e.g. after adapting to a male face, a very masculine face should look even more masculine, while an androgynous or female face should look more feminine).

Over the past fifteen years, norm-based theories of shape and face encoding have gained traction, ostensibly supported by evidence from aftereffects. In Chapter 1, I review this evidence, and relate it to an older debate concerning the role of norms in orientation perception. I conclude that the evidence for renormalisation in form perception is under-

whelming. In Chapter 2 I empirically address the most recent evidence for orientation renormalisation. I show that these data likely arise from an interaction between the task used, and observers' uneven orientation sensitivity. I therefore conclude that orientation aftereffects are best described as purely locally-repulsive.

In order to draw inferences about face and shape representation from aftereffects, it is important to exclude the possibility that aftereffects between complex stimuli are caused entirely by adaptation to lower-level image components, such as local edge orientations. In Chapter 3 I show that shape aspect-ratio aftereffects are determined more strongly by *perceived* shape than retinal shape, and therefore involve computations relatively late in the visual processing hierarchy. In Chapter 4 I show that this higher-level component manifests as a locally-repulsive aftereffect, consistent with an exemplar-based code for aspect ratio.

Having established that tilt and shape aftereffects are more consistent with exemplar-based than norm-based representation, I return my attention to face aftereffects. In Chapter 5 I devise a more diagnostic test than those previously used, by substituting the common binary-classification task for a ternary-classification task. I find that facial gender aftereffects are consistent with local repulsion, while facial distortion aftereffects do not neatly match the predictions of either theory. This raises the previously unsuspected possibility that adaptation along different facial dimensions might involve distinct patterns of perceptual changes.

In Chapter 6 I extend this work by devising a second task, which is more flexible than the first and minimises potential response biases. In a series of experiments I compare appearance between faces presented in differently-adapted retinal conditions. Data show that both facial gender and facial identity aftereffects are consistent with local repulsion, but not renormalisation.

Despite the recent popularity enjoyed by norm-based theories of shape and face encoding, I find that shape and face aftereffects manifest almost exclusively as local repulsions away from the adapted value, consistent with adaptation within an exemplar-based code. In Chapter 7 I discuss these results, and consider their implications for computational models of sensory encoding. The consistency of my results across simple, intermediate, and complex stimuli suggests that the brain may use similar exemplar-based encoding strategies throughout spatial vision.

DECLARATION BY AUTHOR

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

PUBLICATIONS DURING CANDIDATURE

Peer reviewed journal articles

Storrs, K. R. & Arnold, D. H. (2016). Shape adaptation exaggerates shape differences. *Journal of Experimental Psychology: Human Perception and Performance*, *in press*.

Storrs, K. R. (2015b). Facial age aftereffects provide some evidence for local repulsion (but none for re-normalisation). *i-Perception*, *6*(2), 100–103.

Storrs, K. R. & Arnold, D. H. (2015b). Face aftereffects involve local repulsion, not renormalization. *Journal of Vision*, *15*(8), 1–18.

Storrs, K. R. (2015a). Are high-level aftereffects perceptual? *Frontiers in Psychology*, *6*(157), 1–4.

Storrs, K. R. & Arnold, D. H. (2015a). Evidence for tilt normalization can be explained by anisotropic orientation sensitivity. *Journal of Vision*, *15*(1), 1–11.

Spence, M., Storrs, K. R., & Arnold, D. H. (2014). Why the long face? the critical role of vertical configural relations in face 'barcodes' for recognition. *Journal of Vision*, *14*(8), 1–12.

Storrs, K. R. & Arnold, D. H. (2013). Shape aftereffects reflect shape constancy operations: appearance matters. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 616–622.

Storrs, K. R. & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research*, *64*, 7–16.

Conference presentations

Storrs, K R and Arnold, D H. (2015). Enhancing the world with mind: shape adaptation exaggerates shape differences. Poster presented at the European Conference on Visual Perception, Liverpool, UK.

Arnold, D H and Storrs, K R. (2015). Despite what you may have heard, human face and form perception (probably) don't re-normalize. Talk presented at the Asia-Pacific Conference on Vision, Singapore.

Storrs, K R and Arnold, D H. (2015). Faces are repulsive: gender and identity aftereffects involve local repulsion, not re-normalisation. Poster presented at the Vision Sciences Society conference, St Petersburg, Florida.

Storrs, K R. (2014). No evidence from visual aftereffects for norm-based representation of orientations, shapes, or faces. Talk presented at the UQ School of Psychology Centre for Perception and Cognitive Neuroscience workshop, Stradbroke Island, Australia.

Storrs, K R. (2014). A new psychophysical test of norm-based opponent coding in face perception. Poster presented at the European Summer School on Visual Neuroscience, Rauischholzhausen, Germany.

Storrs, K R and Arnold, D H. (2014). Shape aftereffects reflect shape constancy operations: appearance matters. Poster presented at the Vision Sciences Society conference, St Petersburg, Florida.

Spence, M and Storrs, K R and Arnold, D H. (2014). Why the long face? the critical role of vertical configural relations in face 'barcodes' for recognition. Poster presented at the Vision Sciences Society conference, St Petersburg, Florida.

Storrs, K R. (2014). Tilt normalisation may be explained by pre-adaptation to natural orientation statistics. Talk presented at the ModVis workshop, St Petersburg, Florida.

Storrs, K R and Arnold, D H. (2013). Shape aftereffects are more than meets the eye. Talk presented at the AVA Christmas meeting, Leuven, Belgium.

Storrs, K R and Arnold, D H. (2013). Evidence for tilt normalisation may be explained by anisotropic orientation channels. Poster presented at the European Conference on Visual Perception, Bremen, Germany.

Storrs, K R and Arnold, D H. (2013). The shape aftereffect: appearance matters. Talk presented at the Experimental Psychology Conference, Adelaide, Australia.

Storrs, K R and Arnold, D H. (2012). Shape aftereffects reflect a weighted function of retinal and surface slant information. Talk presented at the Australian Cognitive Neuroscience Conference, Brisbane, Australia.

Storrs, K R and Arnold, D H. (2012). Not all face aftereffects are equal. Poster presented at the Vision Sciences Society conference, Naples, Florida.

Storrs, K R and Arnold, D H. (2012). Not all face aftereffects are equal. Talk presented at the Experimental Psychology Conference, Sydney, Australia.

PUBLICATIONS INCLUDED IN THIS THESIS

Included as Chapter 2: Storrs, K. R. & Arnold, D. H. (2015a). Evidence for tilt normalization can be explained by anisotropic orientation sensitivity. *Journal of Vision*, *15*(1), 1–11

Contributor	Statement of contribution
Storrs, K.R.	Designed the experiments (70%) Collected and anlaysed the data (100%) Wrote the paper (70%)
Arnold, D.H.	Designed the experiments (30%) Wrote the paper (30%)

Included as Chapter 3: Storrs, K. R. & Arnold, D. H. (2013). Shape aftereffects reflect shape constancy operations: appearance matters. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 616–622

Contributor	Statement of contribution
Storrs, K.R.	Designed the experiments (70%) Collected and anlaysed the data (100%) Wrote the paper (70%)
Arnold, D.H.	Designed the experiments (30%) Wrote the paper (30%)

Included as Chapter 4: Storrs, K. R. & Arnold, D. H. (2016). Shape adaptation exaggerates shape differences. *Journal of Experimental Psychology: Human Perception and Performance*, *in press*

Contributor	Statement of contribution
Storrs, K.R.	Designed the experiments (70%) Collected and anlaysed the data (90%) Wrote the paper (80%)
Arnold, D.H.	Designed the experiments (30%) Collected and anlaysed the data (10%) Wrote the paper (20%)

Included as Chapter 5: Storrs, K. R. & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research*, *64*, 7–16

Contributor	Statement of contribution
Storrs, K.R.	Designed the experiments (60%) Collected and anlaysed the data (100%) Wrote the paper (60%)
Arnold, D.H.	Designed the experiments (40%) Wrote the paper (40%)

Included as Chapter 6: Storrs, K. R. & Arnold, D. H. (2015b). Face aftereffects involve local repulsion, not renormalization. *Journal of Vision*, *15*(8), 1–18

Contributor	Statement of contribution
Storrs, K.R.	Designed the experiments (100%) Collected and anlaysed the data (100%) Wrote the paper (70%)
Arnold, D.H.	Wrote the paper (30%)

Parts of Chapter 7 modified from: Storrs, K. R. (2015a). Are high-level aftereffects perceptual? *Frontiers in Psychology*, *6*(157), 1–4

Contributor	Statement of contribution
Storrs, K.R.	Wrote the paper (100%)

Included as Appendix 2: Storrs, K. R. (2015b). Facial age aftereffects provide some evidence for local repulsion (but none for re-normalisation). *i-Perception*, *6*(2), 100–103

Contributor	Statement of contribution
Storrs, K.R.	Anlaysed the data (100%) Wrote the paper (95%)
Arnold, D.H.	Wrote the paper (5%)

CONTRIBUTIONS BY OTHERS TO THIS THESIS

My principal advisor, Derek Arnold, contributed to the conception and design of all experiments in this thesis, and was involved in drafting and editing the written work. My secondary advisor, Guy Wallis, gave conceptual guidance throughout the thesis. A number of anonymous reviewers made helpful suggestions regarding each of the empirical chapters.

STATEMENT OF PARTS OF THE THESIS SUBMITTED TO QUALIFY FOR THE AWARD OF ANOTHER DEGREE

No part of this thesis has been submitted towards qualification for another degree.

ACKNOWLEDGEMENTS

My PhD years have enfranchised me. They have unlocked for me a world of freedom, community, and intellectual challenge which I adore. I'm inexpressibly grateful to all the people who helped open the door, but will now try to express those thanks, in something like the order in which you entered my research life:

I owe thanks of course to the University of Queensland Graduate School, for awarding me an APA scholarship which allowed me to do a PhD at all, and later a Travel Award. I'm also very grateful to the School of Psychology, for tangible things like generous conference funding, and for intangible things like being a model of a happy and harmonious academic department.

Thank you to Katie Greenaway and Mike Philipp, for opening the door to the world of research when I barely knew it as a possibility — and for encouraging me exuberantly in those early days.

Later, thank you to the members of the Arnold lab, for making that windowless black room the place I wanted to be more than anywhere. To Sam Pearce and Paul Miller for welcoming me into the family, and sharing years of adventures with me. To Tom Wallis for extending your friendship and generosity to me from across the world. To Morgan Spence and Brendan Keane for keeping the spirit of the Arnold lab alive.

Thank you to the Lunch Committee and the UQ Skeptics — those overlapping collectives that made life at UQ so rich: Will Harrison, Nonie Finlayson, James Retell, Jason Tangen, Matt Thompson, Mel McKenzie, Dustin Venini, Morgan Tear, Zan Saeri and others. Thank you for implacably, exhaustingly, insisting on the application of evidence-based reasoning in every aspect of life, and somehow making that fun. Our lunch table was the most reliable source of great conversation I've yet come across. I feel I was in the department at a very special time — and count myself lucky that many of you now form an international diaspora of friends and collaborators.

When I started travelling to conferences and other labs (which I did wonderfully often, thanks to Derek's encouragement), I met a whole world full of clever and kind people. I found, to my happy surprise, that science is done by a global community of friends (or at least, it is when at its best). The words and experiments in this thesis ended up being worked out not only in my own head, but in conversations and presentations in Brisbane, Perth, Adelaide, Sydney, Florida, Nashville, Boston, London, Cambridge, Leuven and Tübingen. In the process, I found myself driving across Death Valley into mirages; nightswimming in the Gulf of Mexico among bioluminescent dinoflagellates and five hun-

dred vision scientists; eating at Gonville & Caius' high table in Cambridge; getting tattooed in Vegas; and living for two weeks in a German castle. My time as a PhD candidate has been not only intellectually stimulating, but unexpectedly *sensational*.

Back at home in Brisbane, I am deeply grateful to Rohan Kapitany, for adding a whole other dimension to my PhD years — filling them with happiness and affection, and captivating conversations on our couch.

Some of the people I would like to thank were there before I went into research. Thank you to Cameron Anderson, who, to be honest, contributed very little to this thesis, but is wonderful anyway (irrationally, if you like). Thank you also to Mitch Alexander, who has been the proudest and most encouraging friend I can imagine.

Thank you to my parents, Graham and Christine, for all your love and support. Thank you for keeping a place in the world that feels like Home, when research has taken me on three intercontinental moves. Thanks for lending me your beautiful mountaintop while I was writing (you can stop asking whether I've finished my thesis now — look!). I consider myself very fortunate to have parents who so value science and the pursuit of knowledge.

Thanks to Alan Johnston and Niko Kriegeskorte, for tolerating my distraction as I finished my PhD, while I was supposed to be being a Teaching Fellow and a Postdoc for them, respectively.

Thank you to Guy Wallis, my secondary advisor. Although we didn't work closely, you nurtured my interest in computational vision and the problems of object recognition in a way that I expect will shape the rest of my career.

Most of all, thank you to Derek Arnold. Thank you for the years of coffee-sharing and beer-drinking and contemplations at the whiteboard that comprised our time together. Thank you for your endless generosity — for making time to help me think about everything from PsychToolbox code to the philosophy of perception. Thank you for unfailingly putting your students' best interests first. I look forward to debating vision at the Tiki bar of the VSS conference, in the pubs of England, and wherever else we find ourselves over the coming decades. To do a PhD, and with you, were the best decisions I've ever made.

KEYWORDS

visual perception, psychophysics, spatial vision, visual aftereffects, neural encoding, normalisation, face perception, shape perception

AUSTRALIAN AND NEW ZEALAND STANDARD RESEARCH CLASSIFICATIONS (ANZSRC)

ANZSRC code: 170112, Sensory Processes, Perception and Performance 80%
ANZSRC code: 110906, Sensory Systems 20%

FIELDS OF RESEARCH (FOR) CLASSIFICATION

FoR code: 1701, Psychology 80%
FoR code: 1109, Neurosciences 20%

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

2AFC Two-alternative forced-choice

ANOVA Analysis of Variance

BFM Basel Face Model

BOLD Blood Oxygen Level Dependent

CIE *Commission Internationale de l'Eclairage* (International Commission on Illumination)

dva Degrees of Visual Angle

FFA Fusiform Face Area

fMRI Functional Magnetic Resonance Imaging

FWHH Full-Width-at-Half-Height

ISI Inter-Stimulus Interval

JND Just-Noticeable Difference

M Mean

MLE Maximum Likelihood Estimation

N Number of experimental participants

PCA Principal Components Analysis

PSE Point of Subjective Equality

PSS Point of Subjective Stasis

SEM Standard Error of the Mean

UCL University College London

V1 Primary visual cortex

V4 Extrastriate visual area 4

I

NORMS ARE NOT THE NORM: TESTING THEORIES OF SENSORY ENCODING USING VISUAL AFTEREFFECTS

1

CHAPTER 1: INTRODUCTION

Visual neuroscience aims to explain how patterns of neural activity give rise to visual experiences. Within this vast research programme, we can answer some questions in reasonable detail. These include how light is transduced into neural activity, and how the spatial structure of an image on the retina begins to be encoded by the brain. We are only beginning to understand how to answer many other questions, such as how neurons late in the visual processing hierarchy encode an object's identity or a face's expression.

In experiments presented in this thesis, I will explore representations in spatial vision using a powerful psychophysical tool: visual adaptation. An adaptation aftereffect occurs when exposure to one stimulus changes how a subsequent stimulus looks. Figure 1a demonstrates a simple aftereffect. If you look at the dot between the two top patches for ten seconds, then switch your gaze to the dot between the two patches below, you might fleetingly have the impression that the perfectly vertical lines in fact lean toward one another at their tops. This is an example of a tilt aftereffect (Gibson, 1937; Gibson & Radner, 1937; Vernon, 1934). Aftereffects are thought to be caused by neural adaptation — changes in the responsiveness of neurons following prolonged stimulation (Barlow & Hill, 1963; Clifford et al., 2007; Krekelberg, Boynton, & van Wezel, 2006; Solomon & Kohn, 2014; Thompson & Burr, 2009; Wark, Lundstrom, & Fairhall, 2007).

Because encoding schemes involving different arrangements of neurons can predict different patterns of perceptual changes after adaptation, aftereffects can provide insight into how relevant stimulus dimensions are encoded (Clifford, Wenderoth, & Spehar, 2000; Mollon, 1977; Solomon & Kohn, 2014; Webster, 2011). For instance, the study of aftereffects has contributed much to our current knowledge of how the brain encodes colour (see Webster (1996)), motion (see Anstis, Verstraten, and Mather (1998)), orientation (see Clifford, Wyatt, Arnold, Smith, and Wenderoth (2001)), and spatial frequency (see Blakemore and Campbell (1969) and Blakemore, Nachmias, and Sutton (1970)). The relatively recent discovery of aftereffects between far more complex visual properties, including facial characteristics such as expression, gender, and ethnicity (Webster, Kaping, Mizokami, & Duhamel, 2004; Webster & Maclin, 1999) raises the exciting possibility that we might soon gain similar traction in our understanding of how complex objects are represented. Figure 1b shows an example of a face aftereffect.

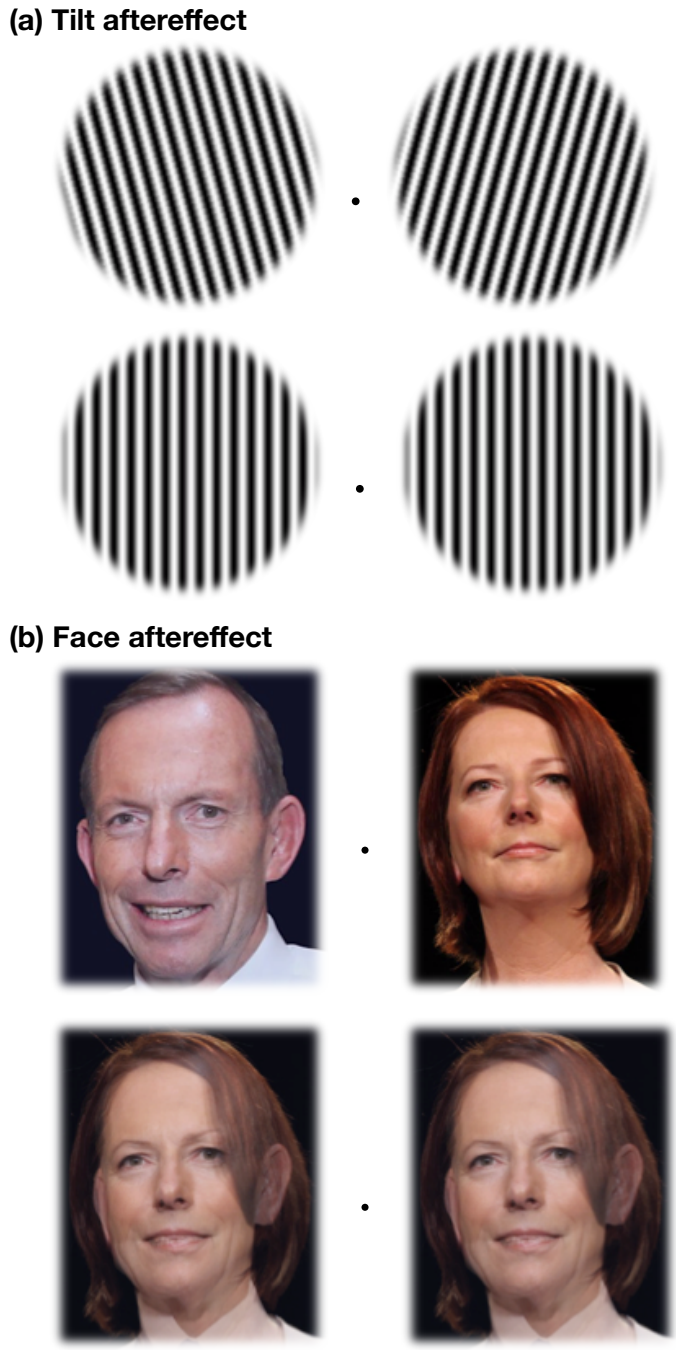


Figure 1. ***(a)*** *A demonstration of a tilt aftereffect. Look at the dot inbetween the upper two patches for about ten seconds, then quickly look at the dot inbetween the lower two patches. You may perceive the lower patches as tilted towards one another at their tops, even though they are physically vertical.* ***(b)*** *A demonstration of a face aftereffect. On the upper left is shown the former prime minister of Australia, Tony Abbott, and on the right, his predecessor Julia Gillard. The two faces at the bottom are identical images created by morphing between the above photographs. Look at the dot inbetween the upper pair of faces for about ten seconds, then quickly look at the dot inbetween the lower pair. You may perceive the identical morphed faces as temporarily taking on the identities of Gillard (on the left) and Abbott (on the right). Aftereffects can also transfer between stimuli of different complexity; if you adapt to the upper stimuli in* ***(a)****, then view the upper stimuli in* ***(b)****, Tony Abbott may lean further to the right.*

1.1 RENORMALISATION AND LOCAL REPULSION: TWO PATTERNS OF VISUAL AFTEREFFECT

Adaptation to different visual properties can produce qualitatively distinct patterns of perceptual aftereffect, which have been linked to distinct representational schemes. I will focus on whether aftereffects provide evidence that the brain represents simple, intermediate, and complex spatial patterns (lines, shapes, and faces) relative to normative values in those domains (for example, to a vertical line, a circular shape, or to an average face). Some aftereffects, such as those induced by tilt adaptation, seem to arise primarily from contrastive processes, in which adaptation exaggerates differences between the adaptor and subsequent other stimuli without changing the adaptor's appearance (Mitchell and Muir (1976); but see Müller, Schillinger, Do, and Leopold (2009) and Chapter 2). Other aftereffects, such as those induced by adaptation to colour saturation, seem to arise from renormalisation processes in which a prolonged adaptor comes to appear more 'neutral' during exposure.

Where renormalising aftereffects are found, they are often interpreted as evidence that neural channels explicitly encode stimuli in terms of how they deviate from a norm (Gibson, 1933; Regan & Hamstra, 1992; Rhodes et al., 2005; Webster & MacLeod, 2011). Supported by apparent evidence for renormalisation, norm-based encoding has been proposed as a common strategy by which shapes, objects, and faces are recognised (Leopold, O'Toole, Vetter, & Blanz, 2001; Panis, Wagemans, & Op de Beeck, 2010; Regan & Hamstra, 1992; Rhodes et al., 2005; Suzuki, 2005; Webster, 2011; Webster & Maclin, 1999). This conjecture has been particularly popular in the field of face perception, where it aligns with a long-standing proposal that the brain recognises new faces by their similarity to abstracted prototypes such as the average of all experienced faces, rather than to particular previously experienced exemplars (Valentine, 1991). However, it remains surprisingly unclear which perceptual dimensions are actually subject to renormalising adaptation, and hence which could be argued to involve explicitly norm-based representations. In the rest of this chapter I will review evidence for renormalisation in the perception of orientation, shape, and face perception, concentrating on the last of these.

1.2 ADAPTATION TO ORIENTATION

Prolonged exposure to a stimulus of a particular orientation has at least two effects on perception. The first is to increase the contrast threshold at which the observer can detect faint stimuli of the same orientation (Blakemore & Campbell, 1969; Boynton & Finney, 2003; Campbell & Kulikowski, 1966; Movshon & Blakemore, 1973). The second is to bias the apparent orientation of differently-oriented test stimuli — the tilt aftereffect (Gibson

& Radner, 1937; Vernon, 1934). The orientation bias is greatest when test stimuli differ in orientation from the adaptor by 10-20 degrees (Campbell & Maffei, 1971; Gibson, 1937; Mitchell & Muir, 1976; Paradiso, Shimojo, & Nakayama, 1989). Tilt aftereffects are greatest when adapting and test stimuli are presented in the same location in the visual field (Gibson, 1937; Knapen, Rolfs, Wexler, & Cavanagh, 2010; Mathôt & Theeuwes, 2013; Muir & Over, 1970) and are of the same spatial frequency (Ware & Mitchell, 1973). If the adaptor is presented to only one eye, and the test only to the other, the inter-ocular tilt aftereffect is as large (Campbell & Maffei, 1971) or almost as large (Gibson, 1937; Movshon, Chambers, & Blakemore, 1972; Paradiso et al., 1989) as that measured when both stimuli are presented to the same eye. This combination of observations suggests that tilt aftereffects are mediated by a neural substrate containing orientation and spatial-frequency selective neurons, sensitive to input from both eyes, with spatially restricted receptive fields, most likely in primary visual cortical area V1 (Campbell & Maffei, 1971; Dragoi, Sharma, & Sur, 2000; Hubel & Wiesel, 1962; Jin, Dragoi, Sur, & Seung, 2005).

1.2.1 *The tilt aftereffect: renormalisation, or local repulsion?*

The tilt aftereffect was originally interpreted by Gibson and others as evidence that orientation is encoded relative to internal norms for vertical and horizontal, which are recalibrated during adaptation toward the current input. In a typical experiment, observers were instructed to adjust a test line until it appeared vertical (or horizontal) before and after viewing a prolonged adapting line (Day & Wade, 1969; Gibson, 1937; Gibson & Radner, 1937; Vernon, 1934). After adapting to a tilted line, tests adjusted to appear subjectively vertical tended to be displaced toward the adapting orientation. According to Gibson's renormalisation theory, this occurs because adaptors come to look more nearly cardinal during adaptation, and lines that had seemed vertical or horizontal look tilted relative to the new norms (Day & Wade, 1969; Gibson, 1937; Gibson & Radner, 1937; Held, 1963; Prentice & Beardslee, 1950; Templeton, 1972). However, these observations are equally consistent with a locally repulsive aftereffect, in which the orientation difference between adapting and test stimuli is exaggerated, with no change in the adaptor's apparent orientation (as noted by Köhler and Wallach (1944)).

1.2.2 *A measurement problem: aftereffects near a putative norm are uninformative*

For several decades, limitations in the methods commonly used to measure tilt aftereffects made it difficult to differentiate between renormalisation and local repulsion. It is straightforward to measure the orientation of the subjective cardinal axes, because "vertical" and "horizontal" are easily communicated to and estimated by observers. Unfortunately, this

constrained experimenters to measuring tilt aftereffects near putative norms, where local repulsion and renormalisation make identical predictions. According to the renormalisation proposal, after adapting to a +15 degree stimulus, the subjective vertical axis is recalibrated to lie closer to 15 degrees, and the previously-vertical axis should appear negatively tilted. The local repulsion proposal makes the same prediction, for a different reason: after adapting to a +15 degree stimulus, all orientation differences between the adaptor and test are exaggerated, so a previously-vertical test should appear negatively tilted. This observation — that measuring the aftereffect at a putative norm is uninformative — will also be relevant to the discussions of shape and face perception later in this chapter.

To differentiate renormalisation from local repulsion, one needs to measure the aftereffect at more diagnostic test values. One diagnostic value is the adapting orientation itself. Renormalisation predicts that after adaptation, the adaptor will appear more 'neutral' (vertical or horizontal), as the perceptual norm is recalibrated to more closely approximate the current input. Local repulsion, since it posits only an exaggeration of *differences* between the adaptor and test, predicts no change in the apparent orientation of the adaptor. Another diagnostic value to measure is one which lies *further from the putative norm* than the adaptor (e.g. a 30 degree line, after adapting to a 15 degree line). According to the renormalisation proposal, this value lies closer to the updated perceptual norm than it had to the unadapted perceptual norm, and should therefore look *less* tilted. Local repulsion predicts the opposite effect: since orientation differences between the adaptor and test are exaggerated after adaptation, a line more tilted than the adaptor should look even *further* tilted.

Changes in perception at diagnostic test values can be measured using a spatial comparison task, in which the observer adapts to an oriented stimulus in one part of the visual field and then compares a reference stimulus presented at the same location to a test stimulus presented elsewhere in the visual field (e.g.Mitchell and Muir (1976)). Because tilt aftereffects are localised to the adapted region of the retina (Gibson, 1937; Knapen et al., 2010; Mathôt & Theeuwes, 2013; Muir & Over, 1970), an aftereffect can create systematic mismatches in perception between an adapted and unadapted location. Any arbitrary orientation may be used as the reference stimulus, freeing experimenters to measure the aftereffect induced between any combination of adapting and test orientations.

By measuring the tuning of the tilt aftereffect for many combinations of adapting and test orientations, it became clear that orientation differences between adaptor and test are exaggerated in both directions — a locally repulsive aftereffect (Mitchell & Muir, 1976). However, some data suggest that there might be a small renormalising component in addition to the predominant local repulsion (Held, 1963; Müller, Schillinger, et al., 2009; Prentice & Beardslee, 1950; Templeton, 1972; Vaitkevicius et al., 2009). In Chapter 2, I address the most compelling recent evidence for tilt renormalisation (Müller, Schillinger, et al., 2009) and show that it can be explained instead by an interaction between the experimental task, and observers' uneven discrimination sensitivity in the orientation domain.

1.2.3 *Multichannel population code models of locally repulsive aftereffects*

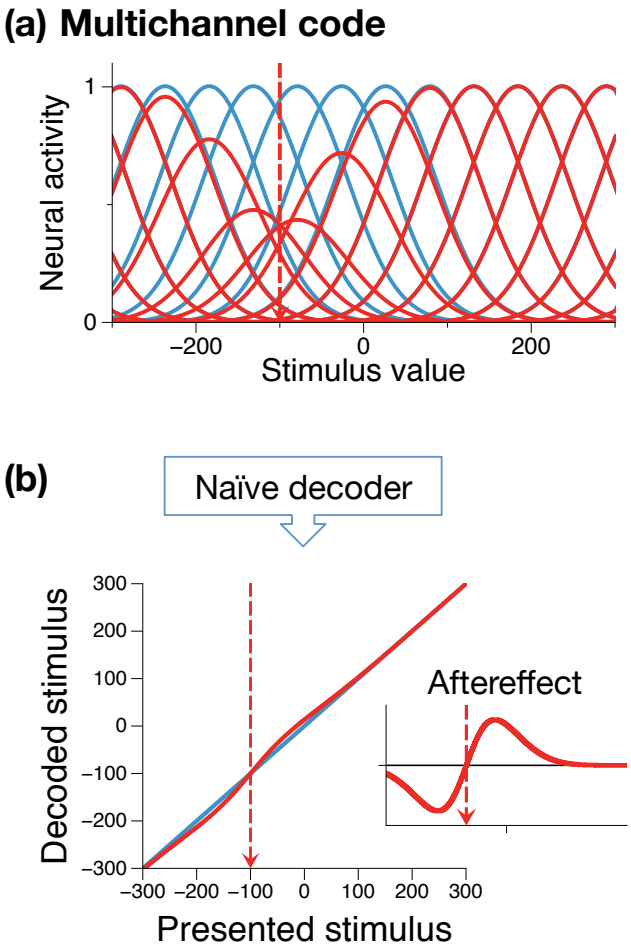


Figure 2. *Schematic illustration of how a locally-repulsive aftereffect can arise from adaptation to a narrow range of stimulus values in a multichannel encoding of a single sensory dimension.* ***(a)*** *Blue curves depict the responsiveness of channels before adaptation; each channel is selective for a relatively narrow range of stimulus values, and the value of zero plays no special role in encoding. After adapting to a stimulus value of -100 (arbitrary units) indicated by a vertical dashed line, the responsiveness of each channel is reduced in proportion to its unadapted response to the adaptor. Post-adaptation response curves are shown in red.* ***(b)*** *A subsequent stage decodes the most probable value of each input stimulus, given the channel activity it elicits, before (blue) and after (red) adaptation. It is assumed that the decoder is "naïve" to the changes in channel responsiveness caused by adaptation. The aftereffect* ***(b, inset)*** *is calculated as the post-adaptation decoded stimulus minus the pre-adaptation decoded stimulus. The aftereffect manifests as a local repulsion of test stimuli away from the adapted value, with no change in the appearance of the adapted value.*

The tilt aftereffect is thought to be due to adaptation in neurons selective for relatively narrow ranges of stimulus orientations (Barlow & Hill, 1963; Bednar & Miikkulainen, 2000; Clifford et al., 2001; Coltheart, 1971; Day, 1962a), likely predominantly in V1 (Campbell & Maffei, 1971; Dragoi et al., 2000; Jin et al., 2005). Tilt aftereffects can be predicted by models in which perceived orientation is determined by activity distributed across a population of orientation-selective channels, wherein each channel is suppressed after adaptation in pro-

portion to its initial response to the adapting orientation (see Figure 2a). Locally-repulsive biases in perceived orientation away from the adapting orientation can occur if one assumes that downstream mechanisms 'decoding' the population activity are unaware of the adaptation-induced changes to the encoding channels' response functions (Ma & Pouget, 2009; Schwartz, Sejnowski, & Dayan, 2009; Seriès, Stocker, & Simoncelli, 2009) — see Figure 2b. A multi-channel model of orientation perception is formalised in Chapter 2, and locally-repulsive aftereffects arising from a similar model are demonstrated in Chapters 5 and 6 and Appendix 1.

1.3 ADAPTATION TO SHAPE

Aftereffects occur not only after prolonged exposure to simple spatial patterns like oriented lines, but also to more complex ones, like two-dimensional shapes. For instance, after looking at a tall ellipse, a perfect circle can appear as a squashed ellipse, and *vice versa* — the aspect ratio aftereffect (Köhler & Wallach, 1944; Regan & Hamstra, 1992). Shape aftereffects can be observed even when adapting and test shapes are shown at widely separated sizes or retinal locations (Regan & Hamstra, 1992; Suzuki & Cavanagh, 1998), suggesting they cannot be explained entirely by early retinally-localised adaptation to the edges constituting a shape. In Chapter 3, I provide further compelling evidence dissociating shape adaptation from retinal contour adaptation by adapting to shapes seen from one viewing angle, and testing with shapes seen from another. These data will show that shape aftereffects depend at least in part on high-level neural substrates, after shape constancy operations have been completed. Shape aftereffects are therefore a useful tool to probe the representation of complex spatial properties.

1.3.1 *A norm-based opponent-channel code for aspect ratio?*

Regan and Hamstra (1992) proposed that aspect ratio is encoded relative to a 'neutral' normative aspect ratio (perceptually neither tall nor squashed). They proposed a model involving two opposing channels, which was later to become influential in models of face representation (Burton, Jeffery, Calder, & Rhodes, 2015; McKone, Jeffery, Boeing, & Clifford, 2014; Pond et al., 2013; Rhodes et al., 2005; Robbins, McKone, & Edwards, 2007; Susilo, McKone, & Edwards, 2010). In the model, one channel is activated by vertically elongated shapes, the other by horizontally elongated shapes (see Figure 3a). During adaptation, the responsiveness of each channel is suppressed in proportion to its original response to the adapting aspect ratio, as in the multichannel model of orientation aftereffects described above. Unlike in the multichannel model, there is a unique norm, signaled by balanced activity across the two channels. Adapting to this shape will not change its apparent as-

pect ratio, but adapting to any other shape will cause the adaptor to seem more nearly 1:1, and bring about a concomitant uniform shift in the appearance of other aspect ratios (see Figure 3b). A norm-based opponent-channel model of stimulus encoding, giving rise to renormalising aftereffects, is formalised in Chapters 5 and 6 and Appendix 1.

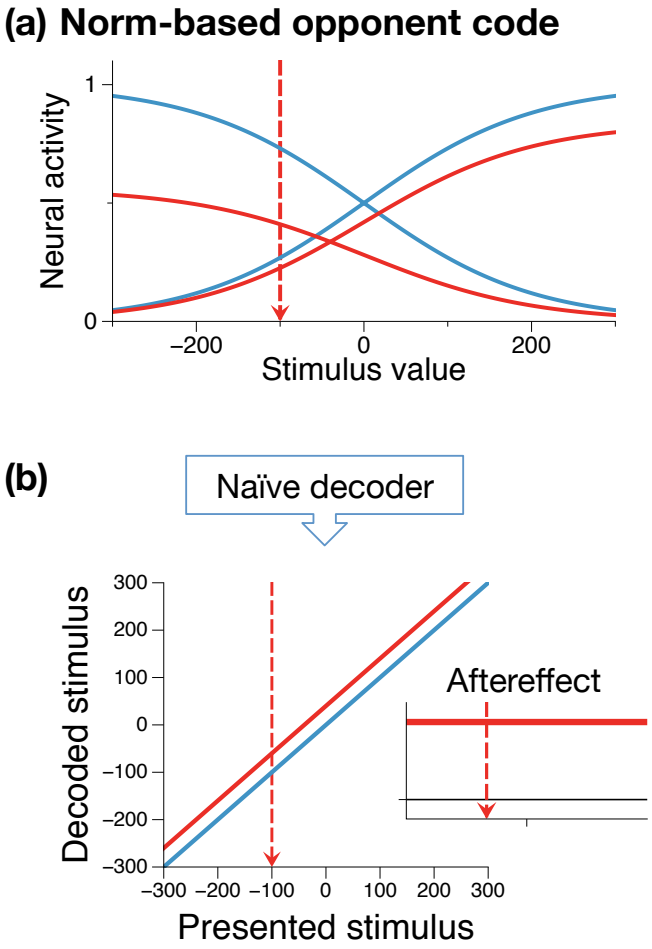


Figure 3. *Schematic illustration of how a locally-repulsive aftereffect might arise from adaptation to a narrow range of stimulus values in a norm-based opponent-channel encoding of a single sensory dimension.* ***(a)*** *Blue curves depict the responsiveness of channels before adaptation; one channel responds increasingly to negative stimulus values, and the other channel responds increasingly to positive stimulus values. The value of zero uniquely elicits balanced activity between the channels, and corresponds to the 'norm,' which can be updated by adaptation. After adapting to a stimulus value of -100 (arbitrary units) indicated by a vertical dashed line, the responsiveness of each channel is reduced (red curves) in proportion to its unadapted response to the adaptor.* ***(b)*** *A subsequent stage decodes the most probable value of each input stimulus, given the channel activity it elicits, before (blue) and after (red) adaptation. The aftereffect* ***(b, inset)*** *manifests as a simple renormalisation, in which the appearance of all test stimuli is shifted in the same direction by the same amount.*

Prior to my thesis, the prediction that aspect ratio adaptation causes a renormalising aftereffect had not been clearly tested. Regan and Hamstra (1992) had measured only changes in the shape perceived as most square or circular after adaptation (i.e. a shift in the perceived 1:1 aspect ratio). As reviewed for orientation aftereffects, measuring changes in appearance near the putative norm does not help dissociate renormalisation from local repulsion. Suzuki and Rivest (1998) used a spatial comparison task to measure aftereffects for many combinations of adapting and test shapes, but used extremely brief 'adaptation' durations (150ms), making it difficult to compare their data to those in the majority of the adaptation literature, and have reported these data only in brief within an edited book chapter. Badcock, Morgan, and Dickinson (2014) reported the tuning of the aspect ratio aftereffect in a conference abstract, but used spatially-overlapping luminance-defined adapting and test stimuli, leaving open the possibility of influence from early retinally-local adaptation. In Chapter 4 I present a spatial comparison task using conventional adaptation durations, and carefully controlling for retinally-local adaptation. In the resulting data, shape aftereffects manifest as a local repulsion away from the adapted aspect ratio, contrary to the predictions of the norm-based opponent-channel model.

1.4 ADAPTATION TO FACES AND OTHER COMPLEX OBJECTS

Face aftereffects were first systematically explored by Webster and Maclin (1999), who showed that after looking at a distorted photograph of a face (e.g. where the features have been 'pinched' together), a subsequent undistorted facial image can seem to be oppositely distorted (e.g. 'expanded' features). Face aftereffects can also be induced by natural differences between faces. For example, after looking at a picture of a male face, an androgynous face can appear feminine (Webster et al., 2004). Analogous effects occur between faces of different expressions (Hsu & Young, 2010), ethnicities (Webster et al., 2004), ages (Schweinberger et al., 2010), eye gaze directions (Jenkins, Beaver, & Calder, 2006; Seyama & Nagayama, 2006), head directions (Fang & He, 2005), and identities (Leopold et al., 2001).
'High-level' aftereffects are by no means unique to faces; they also reportedly occur between distorted images of the same object (Dennett, Edwards, & McKone, 2012; Maclin & Webster, 2001); along morph continua involving arbitrary objects (Daelli, van Rijsbergen, & Treves, 2010), novel objects (Daelli, 2011), and human bodies (Rhodes, Jeffery, Boeing, & Calder, 2013); between objects with different surface material properties (Motoyoshi, 2012) or rendered with different levels of photorealism (Seyama & Nagayama, 2010); and between photographs of landscapes varying along dimensions such as ruralness vs urbanness (Greene & Oliva, 2010).

1.4.1 *Local repulsion or renormalisation in face space?*

Unlike the simpler stimulus properties we have considered so far, which can be fully characterised by a single dimension (orientation or aspect ratio), faces can differ from one another in myriad difficult-to-describe ways. (Valentine, 1991) influentially noted that the perceptual representation of faces could be conceptualised as a multidimensional similarity space. Faces perceived as similar are close to one another in this space, and faces perceived as dissimilar are far apart. Each individual's psychological 'face space' may be thought of as comprising some high number of dimensions, selected and scaled according to personal experience in order to discriminate well amongst the sorts of faces that are most often encountered (Lewis, 2004; Valentine, 1991; Valentine, Lewis, & Hills, 2015). With this metaphor in mind, we could ask how a persons perceptual face space is distorted following adaptation to a particular face. One possibility, illustrated in Figure 4a, is that face aftereffects manifest as a renormalisation, in which the adapted face comes to look more 'average,' with concomitant changes to the appearance of all other faces. Another possibility is that aftereffects manifest as a local repulsion (see Figure 4b), in which the adapted face appears unchanged, but perceived differences between it and other faces are exaggerated.

Rhodes and colleagues (Rhodes & Jeffery, 2006; Rhodes et al., 2005) have proposed that Regan and Hamstra (1992)'s norm-based opponent-channel model of aspect ratio encoding could be extended to encompass multiple dimensions, and thereby provide a substrate for encoding face space. In their proposal, each facial attribute[1] is encoded by two opposing channels: one that responds to higher-than-average values of the attribute, and another which responds to lower-than-average values (Rhodes & Jeffery, 2006; Rhodes et al., 2005). Each pair of channels thereby encodes the position of a face along one attribute dimension. Combined activity across many such channel pairs could specify the location of a face within a multi-dimensional space. The statistical average of all experienced faces would activate all channels equally, uniquely resulting in balanced activity across all opposing pairs, whereas any other face would elicit imbalanced activity in at least one pair of channels. A process of continual adaptation could maintain the balance-point of the opposing channels near the average recently-encountered face, analogous to the manner in which colour adaptation dynamically updates the perceived white/grey point according to the illumination in the prevailing environment (Webster & MacLeod, 2011; Webster & Maclin, 1999). The norm-based opponent-channel proposal thus predicts that face aftereffects will manifest as renormalisation (see Figure 4a).

1 It is unclear what dimensions face-selective mechanisms might be sensitive to. Dimensions may correlate with simple physical differences (e.g. eye height or interocular distance), global properties (e.g. like those captured by a principal components analysis of facial images), or less effable attributes (see Rhodes et al. (2005), Robbins et al. (2007), Valentine (1991), Valentine et al. (2015).

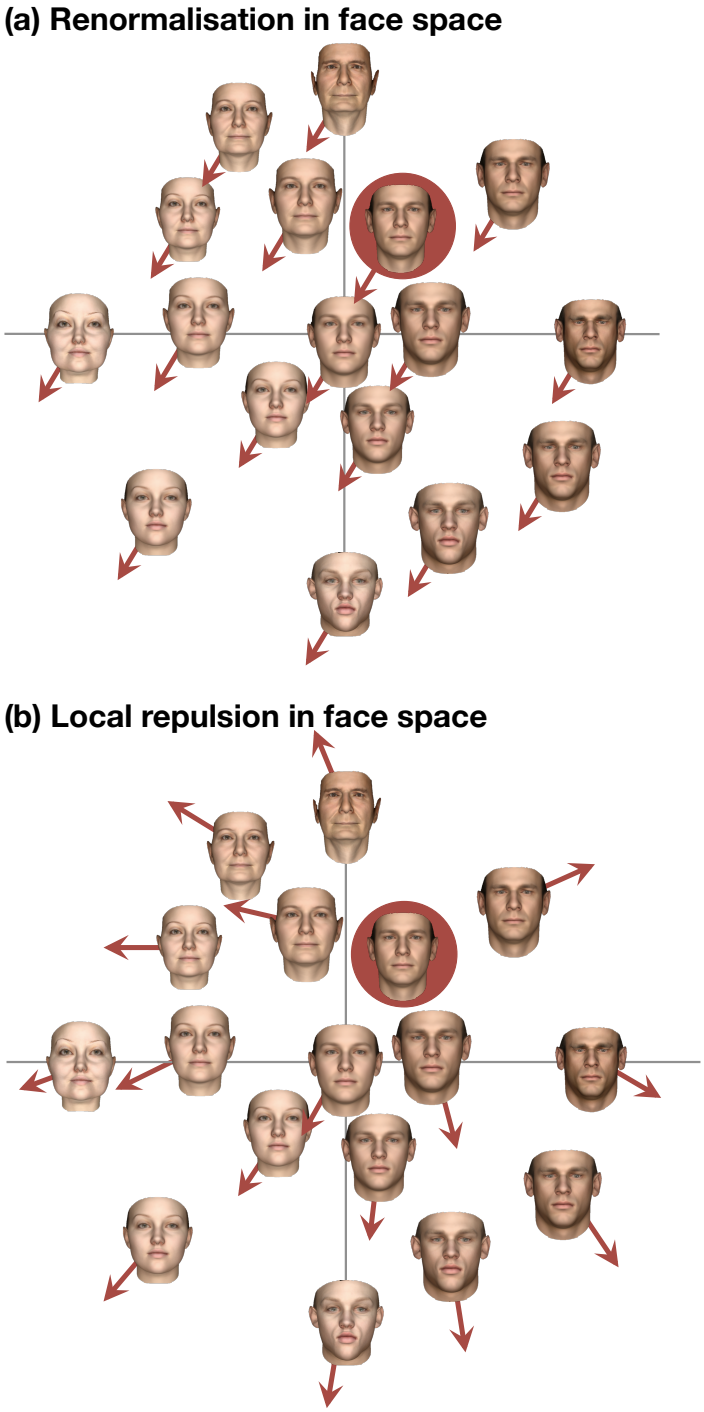


Figure 4. *Schematic depiction of aftereffects within a two-dimensional slice through perceptual 'face space.' Faces were created by sampling from the gender (x-axis) and age (y-axis) vectors within a model based on a principal components analysis of facial structure and texture (Basel Face Model: Paysan, Knothe, Amberg, Romdhani, and Vetter (2009)). At the origin is the average face. After prolonged viewing of a particular face (red circle), adaptation may cause either* ***(a)*** *renormalisation, such that the adapted face appears more neutral, and all other faces change their appearance similarly (red arrows), or* ***(b)*** *local repulsion, such that the differences between the adapting and subsequent other faces appear exaggerated, while the adapting face appears unchanged.*

Alternatively, locations in face space might be encoded by multiple channels, each relatively narrowly-tuned to a sub-region of face space, analogous to the channels thought to underlie the perception of orientation (Giese & Leopold, 2005; Robbins et al., 2007; Ross, Deroche, & Palmeri, 2013; Wallis, 2013; Webster & MacLeod, 2011; Zhao, Seriès, Hancock, & Bednar, 2011). According to this proposal, any dimension of facial variance is encoded by many channels, each responding maximally to some particular value along the dimension and less to values 'higher' or 'lower' than it, with the average face having no unique status. Multichannel encoding predicts that face aftereffects will manifest as a local repulsion (see Figure 4b).

1.4.2 *Evidence from aftereffects for norm-based representation of faces*

A consensus is forming that face aftereffects manifest as renormalisation, and that norm-based opponent-channel models of face representation are therefore supported (Burton et al., 2015; Jeffery et al., 2010, 2011; Leopold et al., 2001; Loffler, Yourganov, Wilkinson, & Wilson, 2005; McKone et al., 2014; Pond et al., 2013; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins et al., 2007; Susilo et al., 2010; Webster & Maclin, 1999). In the remainder of the chapter, I will argue that notwithstanding this consensus, it remains unclear which pattern best describes face aftereffects.

Most face aftereffect experiments are ill-suited to differentiating between local repulsion and renormalisation, because they are designed to measure perceptual changes only near a putative norm. For example, observers are presented with test faces drawn from a continuum spanning male, androgynous and female, and asked to categorise each as either "male" or "female," before and after adapting to a male face (e.g. Webster et al. (2004)). The typical finding in such an experiment is that the boundary at which the observer's responses switch from being predominantly "male" to predominantly "female" is closer to the adapting face after adaptation than before (i.e. after adapting to a male face, they classify more of the gender continuum as "female"). This method tends only to be sensitive to changes in the appearance of stimuli near a category boundary (e.g. androgynous or expressionless faces). Here, faces are more neutral (closer to the origin of the artificial face space) than the adapting face, and both local repulsion and renormalisation predict that neutral faces should take on 'opposite' attributes to the adaptor's. Local repulsion predicts that any differences between the adaptor and test will be exaggerated, so that after viewing a male face, the *relatively* feminine androgynous face should look female. Renormalisation predicts that the perceptual norm will move closer to the adapted face during adaptation, so that after viewing a male face, the new norm will correspond to a somewhat-male face, and the previously androgynous face should appear female. Because shifts in neutral category boundaries cannot discriminate between the two proposals, several alter-

native paradigms have been suggested. I will review five key lines of evidence from these paradigms, that have been taken as support for norm-based encoding of faces.

1.4.2.1 *1. Larger shifts in category boundaries using more extreme adaptors*

It has been suggested that shifts in neutral category boundaries can differentiate locally repulsive from renormalising aftereffects if one compares the magnitudes of shifts induced by different 'strengths' of adaptor (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Jeffery et al., 2010, 2011; McKone et al., 2014; Pond et al., 2013; Skinner & Benton, 2010, 2012; Zhao et al., 2011). A continuum of facial images morphing between different genders (for example) can be thought of as a one-dimensional slice through an artificial face space. The maximally-androgynous face is the norm within this dimension, and increasingly feminine or masculine faces are increasingly high gender strengths. If gender were encoded by two opponent channels (as depicted in Figure 3), then an extremely masculine face should induce a larger renormalisation than a moderately masculine face, and therefore induce a larger change in the appearance of previously-neutral faces, as measured via a binary classification task. Alternatively, if gender were encoded by multiple channels tuned relatively narrowly to particular gender levels (as depicted in Figure 2), then the aftereffect measured near neutral test faces might initially increase as an adaptor becomes more extreme, but then diminish as the distance between adapting and test stimuli exceeds the limited range of local repulsion.

The first studies to test these predictions used only two different levels of adaptor strength, such as a weak and strong version of the same facial expression (Burton et al., 2013; Skinner & Benton, 2010), configural distortion (Jeffery et al., 2010), or facial identity (Jeffery et al., 2011). In all cases the more extreme adaptor induced a larger shift in the category boundary between two perceived expressions, distortions, or identities than the more neutral adaptor. This was argued to be consistent with re-normalisation and not local repulsion. However, a locally-repulsive aftereffect also predicts a gradual increase in aftereffect as the adaptor moves further from the test value, up to some point after which it should begin to decrease (e.g. the tilt aftereffect increases until the adaptor and test are separated by around 15 degrees, after which larger separations decrease the aftereffect; Gibson and Radner (1937), Mitchell and Muir (1976)). Köhler and Wallach (1944) referred to this as the "distance paradox." In a multichannel encoding scheme, the peak of the aftereffect following adaptation is determined by how broadly tuned the underlying channels are, and how locally or broadly adaptation affects population responses. Since these factors are unknown for any putative facial code, larger aftereffects for neutral test stimuli after adapting to 'strong' than 'weak' adaptors are equally consistent with either renormalisation or a broadly-tuned local repulsion.

Zhao and colleagues (2011) measured the magnitude of the facial gender aftereffect after adapting to multiple points across a wider range of stimulus values that extended beyond the natural gender range, into caricatures of male and female faces. Analogously to the previous findings for facial expression, distortion and identity aftereffects, they found increasingly large shifts in the facial gender boundary for adaptors up to and even beyond the average male and female faces, but found that the aftereffect reduced for caricatured adaptors. The tuning of the facial gender aftereffect therefore appeared similar to the tuning of the tilt aftereffect, if a sufficiently broad range of adaptors is considered. Several subsequent studies have also found that facial expression (Skinner & Benton, 2010), gender (Pond et al., 2013), and identity (McKone et al., 2014) aftereffects increase over a broad range of adapting values. These results are inconclusive, since they are consistent with either opponent-channel encoding, or with a scheme involving multiple broadly tuned channels.

1.4.2.2 2. *Larger shifts in category boundaries along stimulus trajectories that span an average face*

In another paradigm, identity aftereffects are compared after adapting and testing along facial identity trajectories that pass through the centre of an artificial face space, as opposed to trajectories that do not (Anderson & Wilson, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006). A uniform renormalisation predicts that the magnitude of an aftereffect will be maximal if one measures changes in perception along an axis connecting the adapting face with the average face, and smaller along any other axis passing through the adapting face. For example, if one adapts to a male face of an average age (e.g. one lying exactly along the positive x-axis in Figure 4a), renormalisation predicts that the adapting face, and all others, will appear less masculine than they had initially, but *unchanged in age*. If one used a binary classification task (”male” or ”female”) to measure changes in perception along the gender axis (which connects the adapting face to the gender-neutral, age-neutral norm), the aftereffect should be maximal. However, if one used an analogous task (”young” or ”old”) to measure changes in perception along the age axis, which is orthogonal to the renormalisation vector, no shift in category boundary is predicted. If one measured the aftereffect along any intermediate axis (e.g. a diagonal axis in Figure 4a’s face space, by asking participants to categorise faces as ”young and male” or ”old and female”), an intermediate shift in the category boundary should be found. Local repulsion predicts instead that a similar magnitude of aftereffect should be measured along any axis that passes through the adapting face, since perception is biased non-selectively in all directions away from the adapted value (see Figure 4b). In the example above, after adapting to a male face of an average age, testing along the age axis should reveal that faces older than the adaptor look exaggeratedly old, and those younger than the adaptor look exaggeratedly

young. A similar pattern of changes should be found regardless of which axis one measures perceptual changes along.

These predictions were tested by Rhodes and Jeffery (2006). An artificial face space was constructed by morphing between and beyond eight identities and an average male face, and participants were trained to recognise four individual faces within this space. For each participant, a target face was chosen (say, "Adam"). In an "opposite" adaptation condition, participants adapted to a face that lay diametrically opposed to Adam on the other side of the average face (say, "Anti-Adam"), and identified test faces drawn from a trajectory passing from Anti-Adam, through the average face, to Adam. In a "non-opposite" condition, participants adapted to a second identity (say, "Bob") and identified test faces drawn from a trajectory passing from Bob, through a non-average face, to Adam. Along the appropriate trajectory, the researchers estimated the proportion of Adams face that needed to be present in the test face in order for participants to reliably identify the face as "Adam," before and after adaptation. Larger shifts in this identity boundary were found after adaptation in the "opposite" than "non-opposite" condition (similar results have been found by other groups using similar designs: Anderson and Wilson (2005), Leopold et al. (2001)).

This appears consistent with a renormalisation interpretation of face aftereffects. However, in this paradigm, the average face plays a special role in how the stimulus space is constructed, and therefore in how participants should divide that space in order to identify individual faces. This raises the possibility that decision-level processes may interact with adaptation-induced perceptual changes, perhaps in non-obvious ways. In a model of face identification decisions, simulations based on both multichannel and opponent-channel encodings of high dimensional face spaces predicted greater shifts in identity boundaries along stimulus trajectories that span an average face than along trajectories that do not (Ross et al., 2013). These simulations suggest that this paradigm is not diagnostic.

1.4.2.3 *3. Boundary shifts in ternary classification tasks after central vs alternating adaptation*

Most face aftereffect experiments involve classification judgements about faces drawn from a single continuum traversing two categories (e.g. male-to-female), yielding a binary classification with a single category boundary. Some facial dimensions lend themselves instead to being divided into three categories — for example, the direction of a face's gaze can easily be classified as being to the left of the viewer, to the right, or directly towards the viewer. This yields a ternary classification task, with two subjective category boundaries (between the lower and middle categories, and between the middle and upper categories), and has the advantage of providing sensitivity to changes in the appearance of stimuli near either of these boundaries.

One paradigm based on a ternary classification task measures changes in both category boundaries after adapting either to a neutral stimulus value, or after adapting to alternating positive and negative stimulus values. For instance, participants might classify faces as "looking left," "looking straight ahead," or "looking right," before and after adapting either to faces that look straight ahead, or to alternating faces that look left and right (Calder, Jenkins, Cassel, & Clifford, 2008). A multichannel code can predict that after consistent neutral-adaptation, the range of stimuli placed in the central category should narrow, and that after alternating negative and positive adaptation, the range of central stimuli should broaden. A simple opponent coding scheme predicts no change in responses after either type of adaptation. When tested on such a task, eye gaze direction (Calder et al., 2008) and head direction aftereffects (Lawson, Clifford, & Calder, 2011) both followed the pattern of changes predicted by a multichannel code.

Burton and colleagues (2015) applied this paradigm to the study of facial expression aftereffects. Adapting to an expressionless face resulted in a robust narrowing of the range of faces participants placed in the middle category, while adapting alternatingly to distinct expressions produced a smaller but significant narrowing of the central category (Burton et al., 2015). Because the shifts were in the same, rather than opposite, directions, this pattern of results was argued to be more consistent with an opponent than multichannel code. However, the predictions of an opponent scheme are unclear, as additional assumptions must be made (e.g. that the response functions of adapted channels steepen or flatten) in order to predict a change in categorisation decisions after either type of adaptation (Burton et al., 2015; Calder et al., 2008; Lawson et al., 2011). In Chapter 5 I present a novel paradigm which tests a more definitive prediction of the renormalising account — whether the appearance of an adapting face changes during adaptation — by measuring shifts in each category boundary in a ternary task after observers adapt to a stimulus that lies at one boundary.

1.4.2.4 *4. No effect of adapting to an average face*

A widely-cited piece of evidence for renormalisation is the claim that adapting to a neutral face is uniquely ineffective at changing the appearance of itself or other faces. This evidence is provided almost exclusively by a single report, from a study of facial distortion aftereffects (Webster & Maclin, 1999). Webster and Maclin (1999) had participants remember a distorted face, and then adjust a test to match this remembered face after adapting either to distorted faces or to an undistorted face. They found that adapting to an undistorted face had no effect on participants' matches of distorted faces. However, there is little evidence that adapting to a neutral value of a *face-specific* attribute (e.g. an adrogynous or expressionless face) does not affect the perception of subsequent non-neutral faces. Instead, adapting to a neutral expression (Burton et al., 2015), or to direct eye gaze (Calder et al., 2008) or

head direction (Lawson et al., 2011), appears to exaggerate the subtle differences in subsequent non-neutral test faces, as indicated by a reduction in "neutral" category responses after adaptation.

1.4.2.5 *5. The adaptor and more extreme faces are rated as more neutral after adaptation*

Perhaps the strongest evidence for normalisation in face aftereffects comes from experiments involving geometrically distorted images. The observation is given anecdotally by Webster and Maclin (1999) that during adaptation a distorted face comes to appear more normal, a key feature of a renormalising aftereffect. This observation has later been supported empirically. Rhodes and colleagues (2003) showed that after adapting to distorted photographs of faces, the adapting faces were rated as more "normal" than they had been before adaptation. Robbins et al. (2007) showed a similar pattern of results using faces in which eye height had been manipulated.

Finally, O'Neil and colleagues (2014) took advantage of the naturally finely-graduated scale of age to explore the pattern of aftereffects induced by facial age adaptation. Age estimates for 80 test faces were collected before adaptation, and after adapting to a sequence of young, middle-aged, or old faces. The authors argue that the aftereffect data are best described as a uniform shift in perceived age across all test values, and therefore indicate a renormalising aftereffect. However, the authors present and analyse only linear fits to their data, which are well-suited to capture the predictions of renormalisation, but not local repulsion. When I re-analysed these data (Storrs (2015b); see Appendix 2), comparing a linear model to a bidirectional model capturing a locally-repulsive aftereffect, I found that the locally-repulsive model explained slightly more variance in the aftereffect data. In a response to my re-analysis, the original authors suggested that the fits of a bidirectional model should be constrained to have the inflection point of the aftereffect fixed at the adapted value (O'Neil, Mac, Rhodes, & Webster, 2015). When such a constraint was added, the linear model performed as well or slightly better than the bidirectional model. However, it is important to note that the best fit of any suggested model explains less than 2% of the variance in aftereffect magnitude. These data therefore do not provide compelling evidence for either local repulsion or renormalisation in age aftereffects.

1.5 THE PRESENT THESIS

Debates over whether spatial aftereffects involve renormalisation, and whether this therefore provides evidence for an explicitly norm-based representation of the underlying dimensions, have recurred several times in the study of spatial vision and are largely unresolved. In the domain of orientation perception, the consensus tips in favour of a multichannel representation of orientation, giving rise to a predominantly locally repulsive tilt

aftereffect, but this remains controversial (see Müller, Schillinger, et al. (2009) and Chapter 2). In the domain of shape perception, a norm-based theory of aspect ratio encoding has been proposed (Regan & Hamstra, 1992; Suzuki, 2005), although there is little relevant evidence from aftereffects to support or falsify this. In the domain of face perception, a consensus has formed in favour of norm-based representation, although this might not be well-supported by the available evidence.

I begin the empirical work of this thesis in Chapter 2, by reassessing recent evidence for renormalisation in orientation perception, before turning to more complex stimulus domains. Chapter 3 demonstrates that shape aftereffects can be attributed to relatively high-level shape representations, and Chapter 4 evaluates whether these high-level shape aftereffects manifest as a local repulsion or a renormalisation. Chapters 5 and 6 focus on face aftereffects. Chapter 5 presents a new paradigm based on a ternary classification task, and applies it to facial distortion and gender aftereffects. Chapter 6 presents a second novel paradigm, involving a spatial comparison task, and applies it to facial identity and gender aftereffects. The data presented in this thesis point to a common encoding strategy throughout spatial vision, in which sensory values are signaled by distributed activity across a population of relatively narrow non-monotonically-tuned channels.

2

CHAPTER 2: EVIDENCE FOR TILT NORMALISATION CAN BE EXPLAINED BY ANISOTROPIC ORIENTATION SENSITIVITY

The order of chapters in this thesis echoes the hierarchy of feedforward processing that is thought to occur in the brain's ventral visual stream, from dots and edges, through shape and texture fragments, to complex objects. It also follows the chronological discovery of different spatial aftereffects, from tilt aftereffects (Gibson & Radner, 1937; Vernon, 1934), through shape aftereffects (Köhler & Wallach, 1944), to face aftereffects (Webster & Maclin, 1999).

It begins, therefore, by examining the tilt aftereffect. In the domain of orientation, the vertical and horizontal axes serve as psychological and linguistic reference points, leading many to propose that these cardinal axes play a unique role in how orientation is encoded neurally (Howard, 1982). Evidence for an explicitly norm-based representation of orientation comes from reports that non-cardinal stimuli 'normalise' toward vertical or horizontal during prolonged viewing (Day & Wade, 1969; Gibson & Radner, 1937; Held, 1963; Prentice & Beardslee, 1950; Vaitkevicius et al., 2009).

The majority of these reports are ambiguous, as they can also be explained by a locally-repulsive tilt aftereffect, or by misperceptions of very briefly presented stimuli (see the Introduction of the present chapter). However, data from a new psychophysical method devised by Müller, Schillinger, et al. (2009) appear to provide clear evidence for normalisation. Müller, Schillinger, et al. (2009) estimated the point of perceptual stasis for slowly-rotating lines, and found that this estimate differed from physical stasis when the initial line orientation was near cardinal. The authors inferred that observers were experienced a slight 'perceptual drift' of near-cardinal orientations toward the cardinal axes — direct evidence for tilt normalisation. My supervisor and I wondered whether these data might instead arise from observers having lower detection thresholds for rotational motion that moved an oriented stimulus toward a cardinal axis than away. Such an asymmetry seemed plausible, given the "oblique effect" (Appelle, 1972) — greater discrimination sensitivity for orientations near the cardinal than oblique axes. The experiments testing this hypothesis appeared as a manuscript in the *Journal of Vision* (Storrs & Arnold, 2015a).

Evidence for tilt normalisation can be explained by anisotropic orientation sensitivity

2.1 ABSTRACT

Some data have been taken as evidence that after prolonged viewing, near-vertical orientations 'normalise' to appear more vertical than they had previously. After almost a century of research the existence of tilt normalisation remains controversial. The most recent evidence for tilt normalisation comes from data suggesting a measurable 'perceptual drift' of near-vertical adaptors toward vertical, which can be nulled by a slight physical rotation away from vertical (Müller, Schillinger, et al., 2009). We argue that biases in estimates of perceptual stasis could, however, result from the anisotropic organisation of orientation-selective neurons in V1, with vertically-selective cells being more narrowly tuned than obliquely-selective cells. We describe a neurophysiologically plausible model that predicts greater sensitivity to orientation displacements *toward* than *away from* vertical. We demonstrate the predicted asymmetric pattern of sensitivity in human observers by determining threshold speeds for detecting rotation direction (Experiment 1), and by determining orientation discrimination thresholds for brief static stimuli (Experiment 2). Results imply that data suggesting a perceptual drift toward vertical instead result from greater discrimination sensitivity around cardinal than oblique orientations (the oblique effect), and thus do not constitute evidence for tilt normalisation.

2.2 INTRODUCTION

Prolonged exposure to a visual stimulus can alter perception in two ways — by changing the appearance of the stimulus itself (e.g. fading colour saturation; see Webster (1996)), or by changing the appearance of other subsequent stimuli (e.g. the motion direction aftereffect; see Anstis et al. (1998)). Patterns of adaptation-induced changes for a particular perceptual dimension can give insight into how that dimension is neurally encoded (see, e.g. Clifford et al. (2007)).

It is well documented that after prolonged exposure to one orientation, the orientation of subsequent stimuli can seem repelled away from the adapted orientation — the tilt aftereffect (Gibson & Radner, 1937; Vernon, 1934). Controversially, it has also been argued that vertical and horizontal (the cardinal orientations) constitute 'norms' for orientation perception (Gibson & Radner, 1937; Howard, 1982; Müller, Schillinger, et al., 2009), and that over time near-cardinally oriented stimuli can undergo a perceptual change, being drawn toward the nearest cardinal axis — tilt normalisation (Day & Wade, 1969; Gibson & Radner, 1937; Held, 1963; Müller, Schillinger, et al., 2009; Prentice & Beardslee, 1950; Vaitkevicius et al., 2009; Vernon, 1934).

Tilt aftereffects are often explained via models in which perceived orientation is determined by activity across a population of orientation-tuned neurons, and can be biased by adaptation-induced changes in the distribution of that activity (Barlow & Hill, 1963; Clifford et al., 2000, 2001; Day, 1962b; Girshick, Landy, & Simoncelli, 2011). However, population models of orientation coding do not strongly predict normalisation (and can even predict the opposite — see below). Some have suggested there might therefore be separate neural processes underlying tilt normalisation and tilt aftereffects (Coltheart, 1971; Morant & Harris, 1965; Morant & Mistovich, 1960; Müller, Schillinger, et al., 2009). An alternative possibility is that data thought to demonstrate tilt normalisation have been misinterpreted. The earliest reports of normalisation were based on the finding that after adapting to a tilted line, lines adjusted to appear vertical tend to be displaced toward the adapting orientation (Day & Wade, 1969; Gibson & Radner, 1937; Vernon, 1934). This can be considered as evidence only for a standard tilt aftereffect. A more direct method to measure normalisation is to adapt observers for a period in one part of their visual field, then briefly present a second stimulus elsewhere, estimating the orientation at which the second stimulus appears parallel with the persistent adaptor. Data from such paradigms suggest that a persistent near-vertical standard is seen as less tilted than a brief comparison to which it is either physically (Prentice & Beardslee, 1950; Vaitkevicius et al., 2009) or perceptually (Held, 1963) matched prior to adaptation (although at least one study failed to replicate this effect — Heinemann and Marill (1954)).

The studies described above provide some evidence that persistent (adapting) and transient (test) stimuli differ in apparent orientation, but it is unclear whether this mismatch is due to normalisation of the prolonged stimulus, or to misperception of the brief stimulus. Andrews (1965, 1967) estimated the orientation at which a brief test appeared parallel to a prolonged standard and found that when the standard was near vertical or horizontal, physically matched tests were judged as more tilted. However, Andrews (1967) showed that the magnitude of this mismatch depended on both the presentation time and spatial extent of the comparison. Larger mismatches were found for shorter lines, and the bias reduced by around two-thirds within the first second of comparison exposure. This dependence on the properties of the comparison suggests the mismatch between brief and prolonged

near-cardinal orientations might be a measurement of misperceptions in briefly-presented stimuli rather than normalisation of prolonged stimuli (Howard (1982); for a similar argument regarding putative curvature normalisation see Coren and Festiniger (1967)). Data from experiments in which the orientation of a persistent 'adaptor' is compared to brief comparators (Andrews, 1965, 1967; Heinemann & Marill, 1954; Held, 1963; Prentice & Beardslee, 1950; Vaitkevicius et al., 2009) are therefore ambiguous.

Recently, Müller, Schillinger, et al. (2009) presented a novel method by which to measure normalisation that eliminates the need for a comparison. They suggested that tilt normalisation involves a detectable 'perceptual drift' of tilted stimuli toward vertical or horizontal, and that this can be quantified by measuring the rate of physical rotation *away from* a cardinal axis required to null the perceptual rotation. The experimenters used an adaptive procedure to estimate points of subjective stasis. For stimuli with an initial orientation of 0°, 90°, or $\pm 45^\circ$ subjective stasis estimates coincided with physical stasis. For all other orientations, subjective stasis estimates corresponded with a physical rotation away from the nearest cardinal axis, with the 'perceptual drift rate' peaking at $\pm 15^\circ$ from cardinal axes. The magnitude of perceptual drift measured in this manner was smaller than that of the tilt aftereffect, followed a different pattern as a function of adapting orientation, and the magnitudes of the two effects were not correlated across observers. These data therefore seemed to present conclusive evidence both that normalisation occurs, and that it arises from a different process than the tilt aftereffect. However, we propose that the measured 'drift' can be more parsimoniously explained as an artefact arising from the oblique effect (i.e. lower detection and discrimination thresholds for orientations near cardinal than oblique axes — Appelle (1972)).

2.3 MODELLING

Here we will show that a neurophysiologically plausible model of orientation coding, which predicts the oblique effect because of anisotropies in the bandwidths of orientation-selective neurons (Girshick et al., 2011; Li, Peterson, & Freeman, 2003; Rose & Blakemore, 1974), also predicts asymmetric just-noticeable-differences (JNDs) for orientation changes from near-vertical and near-horizontal standard orientations. While this type of orientation coding model is well-established, this prediction of such models has not previously been appreciated or formally expressed.

Figure 1a shows an idealised population of 60[1] V1 orientation-selective neurons. Each neuron's tuning curve is represented by a von Mises distribution (Swindale, 1998) drawn on a circular continuum spanning 0 to 180 degrees using the *Circular Statistics Toolbox* for

1 The values predicted by the model for discrimination thresholds (Figure 1c and d) are near-identical across a wide range of numbers of simulated neurons (at least from 20 to 80). We chose the figure of 60 for the sake of consistency with a similar model (Girshick et al., 2011).

Matlab (Berens, 2009). For each tuning curve, the distribution mean indicates the preferred orientation of the simulated neuron, and the distribution variance indicates the bandwidth of the tuning curve (see, e.g. Clifford et al. (2000, 2001), Pouget, Dayan, and Zemel (2000)). The value of a neuron's tuning curve at any given orientation represents its average firing rate (as a proportion of its maximum firing rate) when presented with that orientation.

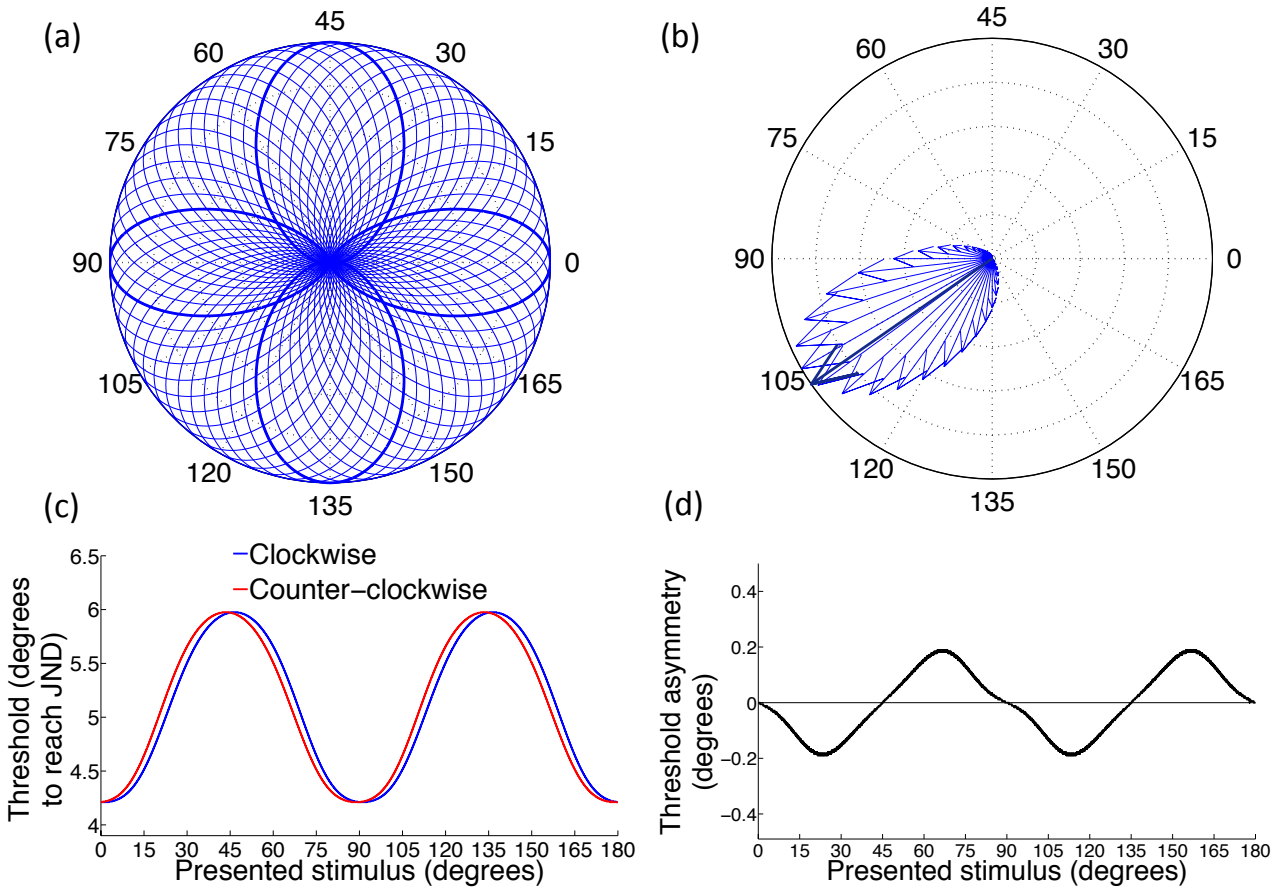


Figure 1. ***(a)*** *The tuning curves of 60 orientation-selective neurons simulated using von Mises distributions, with tuning bandwidths ∼1.4 times wider near the oblique than cardinal axes (two cardinal and two oblique tuning curves are shown in bold to highlight these differences).* ***(b)*** *The response to any presented orientation can be decoded by taking the average (in bold) of the vector responses to that stimulus across the population.* ***(c)*** *The physical orientation difference between two inputs required to reach an arbitrary just-noticeable-difference (JND) of 5° between decoded orientations varies along the orientation continuum. This produces the oblique effect, wherein discrimination sensitivity is higher for cardinal orientations and lower for oblique ones. Thresholds are shown separately for clockwise (blue) and counter-clockwise (red) deviations from each orientation.* ***(d)*** *Clockwise thresholds minus counter-clockwise thresholds. Sensitivity to clockwise and counter-clockwise displacements is equal at the cardinal and oblique axes, but asymmetric elsewhere. In this model, these asymmetries peak at approximately ±20° away from the cardinal axes, where thresholds are lower for detecting displacements toward than away from the nearest cardinal orientation.*

The bandwidth of tuning curves across the population of neurons varies as a rectified sinusoidal function of preferred orientation. Consistent with recorded properties of cat V1 cells (Li et al., 2003; Orban & Kennedy, 1981), the full-width-at-half-height (FWHH) of tuning

curves reaches a maximum of approximately 39° at the oblique axes and a minimum of 28° at cardinal axes. Any oriented input elicits a response from a number of neurons, some reaching close to their maximum firing rate, others responding only slightly. Each neuron's activity can be described by a vector, as illustrated in Figure 1b, where the vector's angle represents the neuron's preferred orientation (the orientation that it is 'voting for'), and the vector's length represents the neuron's firing rate (the 'weight of evidence' provided by the neuron's activity). One simple method to decode the presented orientation from the population activity is by taking the average of all response vectors (Pouget et al., 2000).

One can calculate the model's threshold for discriminating any two oriented stimuli by declaring an arbitrary just noticeable difference (JND) value, by which two decoded orientations must differ in order to be discriminable — say five degrees. For each input orientation one can then estimate the amount by which a second input would need to differ in order for the population to signal one JND. Thresholds calculated in this manner are depicted in Figure 1c for both clockwise and counter-clockwise differences from each orientation. Because bandwidths are anisotropic across the orientation continuum, a JND between *decoded* orientations will in some cases be reached when the difference between *physical* inputs is less than the declared JND (e.g. rotations away from 0°), and in other cases a JND will only be reached when the difference between physical inputs is greater than the declared JND (e.g. rotations away from 45°). As depicted in Figure 1c, the model predicts a clear oblique effect for discrimination thresholds, i.e. sensitivity is greater about cardinal orientations and lower about oblique orientations.

As well as predicting the well-known oblique effect, the pattern of anisotropies captured in this model predicts asymmetric sensitivity to clockwise vs counter-clockwise changes from certain orientations. In Figure 1d, the discrimination thresholds for counter-clockwise differences have been subtracted from the thresholds for clockwise differences, to reveal orientations at which sensitivity is asymmetric. In this model, threshold asymmetries peak at approximately $\pm 20°$ from cardinal orientations, with greater sensitivity predicted for orientation changes toward the cardinal axes than away.

We believe this asymmetric sensitivity can explain data that have been taken as evidence for tilt normalisation. Müller, Schillinger, et al. (2009) propose that tilt normalisation involves a measurable 'perceptual drift' of tilted stimuli toward vertical or horizontal, which can be nulled by a slight physical rotation away from the cardinal axis. Rotation speeds required to null this putative drift were estimated by an adaptive procedure (a single staircase, see Cornsweet (1962)). For 0°, 90° and $\pm 45°$ initial orientations, the procedure converged on physical stasis, but for other orientations converged on a slight rotation away from the nearest cardinal axis.

Müller, Schillinger, et al. (2009)'s method of estimating perceptual stasis assumes a single point of stasis lying mid-point between the speeds of rotation that result in chance performance when judging whether a stimulus is rotating clockwise or counterclockwise. Judg-

ing direction of rotation involves two perceptual thresholds, one for detecting that stimuli are rotating clockwise, another for detecting that stimuli are rotating counter-clockwise. If these two thresholds are equal in magnitude, and physically static inputs are seen as static, an adaptive procedure searching for a single estimate of perceptual stasis should converge on a physically static stimulus. If, however, thresholds are asymmetric, with observers being more sensitive to orientation changes toward, rather than away from, cardinal axes, then an adaptive procedure will likely converge on a stimulus that is slowly rotating away from a cardinal orientation, even if physically static inputs are seen as static.
In Experiment 1 we seek to replicate the effect reported by Müller, Schillinger, et al. (2009) by measuring the apparent rotation direction of oriented stimuli, to which varying amounts of physical rotation have been applied via adaptive staircase procedures. Unlike in the previous report, we determine independent clockwise and counterclockwise rotation speed thresholds for tests originating at a range of orientations.

2.4 EXPERIMENT 1

2.4.1 *Method*

2.4.1.1 *Participants*

There were 12 participants, comprising the first author and 11 experienced psychophysical observers who were naïve to the hypotheses of the experiment. All experimental procedures were approved by the School of Psychology's Ethical Review Board at the University of Queensland.

2.4.2 *Stimuli and apparatus*

Stimuli were presented on a gamma-corrected 19" Sony Trinitron Multiscan CPD-G520 monitor, with a resolution of 1024 x 768 pixels and a refresh rate of 120Hz. Matlab software was used to drive a VSG 2/3 stimulus generator from Cambridge Research Systems. Participants viewed stimuli from a distance of 52cm, with their head restrained by a chin rest. Stimuli were viewed through a black cylindrical cardboard tube 16cm in diameter and 52cm in length (see Figure 2a). The surface of the monitor could be viewed through a circular aperture cut into a piece of black cardboard attached to the end of the viewing tube, which subtended 17.5 degrees of visual angle (dva). This minimised the influence of potential reference orientations, such as those provided by the edge of the monitor screen or by peripheral objects in the room.
On each trial a single Gabor was presented (see Figure 2b), with a red 1-pixel fixation dot at its centre (approximately 0.03 dva). The spatial Gaussian envelope for Gabors subtended

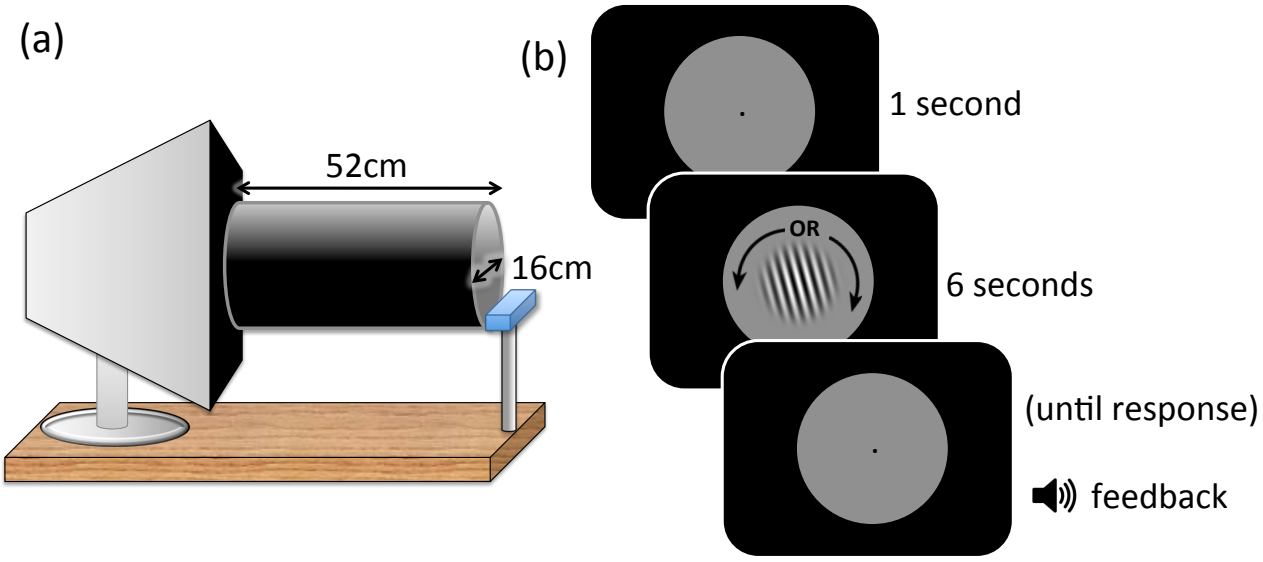


Figure 2. ***(a)*** *Apparatus used in Experiments 1 and 2. The participant viewed stimuli through a black cardboard tube, with a circular aperture cut into a sheet of black cardboard covering the screen.* ***(b)*** *Trial sequence in Experiment 1. After a 1 second fixation, a single Gabor appeared, rotating for 6 seconds at a speed and direction determined by an adaptive procedure. At the end of this period the participant reports if the stimulus had rotated clockwise or counter-clockwise. They were subsequently given auditory feedback.*

9.9 dva, with a standard deviation of 1.7 dva. The Michelson luminance contrast of Gabors was 0.50 and they had a spatial frequency of 5 cycles/dva. The phase of the Gabor waveform was randomised on a trial-by-trial basis, and the display background was grey (CIE chromaticity co-ordinates x = 0.27, y = 0.30, Y = 45). The average luminance of test stimuli was matched to the background.

2.4.3 *Procedure*

On each trial the test Gabor rotated at a speed determined by an adaptive procedure for a period of 6 seconds (see below). After the test presentation, the display was filled with static white noise (individual elements subtended 0.03 dva x 0.03 dva), which persisted until the participant indicated in which direction they felt the Gabor had rotated, by pressing one of two mouse buttons. Feedback was then provided in the form of a high-pitched (correct) or low-pitched (incorrect) beep.

Speed thresholds for detecting the direction of rotational motion were measured at 8 orientations. These consisted of the two cardinal orientations (0° and 90°), the two obliques (45° and 135°), and four orientations located ±15° to either side of the cardinal orientations (i.e. 15°, 75°, 105°, and 165°). Here and throughout, orientations are labeled according to the geometric convention, with 0° indicating a horizontal line and 90° a vertical line.

Adaptive 'staircase' procedures (Cornsweet, 1962) were used to determine speed thresholds for detecting rotation direction, independently for clockwise and counter-clockwise rotations from stimuli initialised at each of 8 test orientations. Each of the 16 staircase procedures began by presenting a clearly rotating stimulus, at a speed of 0.8 angular degrees/second. Test speeds were then adjusted according to a 'two-down, one-up' procedure, to derive an estimate of the 71% threshold (i.e. if the participant got one response wrong, speed was increased for that staircase, and if they got two successive responses correct within a staircase procedure, speed was decreased — see Levitt (1970)). Whenever the direction of adjustment (increasing vs. decreasing speed) differed from the direction of the previous adjustment within a staircase, a 'reversal' was recorded for that staircase.
As all staircases began with clearly rotating stimuli, stimulus speed was decreased by 0.16°/s on each trial until the first reversal was recorded within a given staircase procedure (i.e. until the participant's first incorrect response), after which speed was adjusted by 0.04°/s for that staircase. If such an adjustment would generate a negative speed for that staircase, stimulus speed was set to 0°/s, ensuring that staircase procedures searching for clockwise rotation thresholds did not contain any counter-clockwise rotating stimuli, and *vice versa*. Each staircase terminated after 6 reversals, and the mean of the stimulus speeds at the last 3 reversals was taken as an estimate of the threshold for identifying directional rotation from the initial test orientation.
To prevent fatigue, the 16 staircase procedures were completed in four blocks of trials, each taking approximately 12 minutes to complete. Within a block of trials, individual trials were drawn randomly from one of four interleaved staircases. In block A, staircases were conducted to determined speed thresholds for clockwise and counter-clockwise rotations from initial test orientations of 0° and 90°. In block B initial test orientations were 75° and 165°, in block C initial test orientations were 45° and 135°, and in block D initial test orientations were 15° and 105°. Half of the participants completed blocks in the order A, B, C, D, and the other half completed blocks in the order D, C, A, B.

2.4.4 *Results*

We determined thresholds for each of the 8 test orientations, independent of rotation direction, by averaging clockwise and counter-clockwise threshold estimates for each test orientation for each participant. This revealed a clear oblique effect (see Figure 3a), with highest thresholds for oblique test orientations (45° and 135°) and lowest thresholds for cardinal orientations (0° and 90°).
To estimate thresholds for test stimuli rotating toward the nearest cardinal axis we averaged individual data from the 15° clockwise, 75° counter-clockwise, 105° clockwise, and 165° counter-clockwise conditions. To estimate thresholds for test stimuli rotating away

from the nearest cardinal axis we averaged individual data from the 15° counter-clockwise, 75° clockwise, 105° counter-clockwise, and 165° clockwise conditions. Analyses revealed that speed thresholds for detecting rotation *toward* the nearest cardinal axis (Mean 0.22 ± SEM 0.03 degrees/second) were lower than thresholds for detecting rotation *away from* the nearest cardinal axis (0.33 ±0.03; two-tailed paired t_{11} = -2.58, p = .026; see Figure 3b). One interpretation of this result is that sensitivity is symmetric, but that the point of subjective stasis (PSS) deviates from physical stasis. Assuming symmetric sensitivity, the PSS could be estimated by taking the signed average of the toward- and away-from-cardinal axis thresholds for each participant, yielding a mean PSS of a 0.05°/s (±0.02) away from the nearest cardinal axis. These data are broadly consistent with Müller, Schillinger, et al. (2009)'s key results (see their Figure 3b), which were taken as evidence for a perceptual drift of near-cardinal orientations.

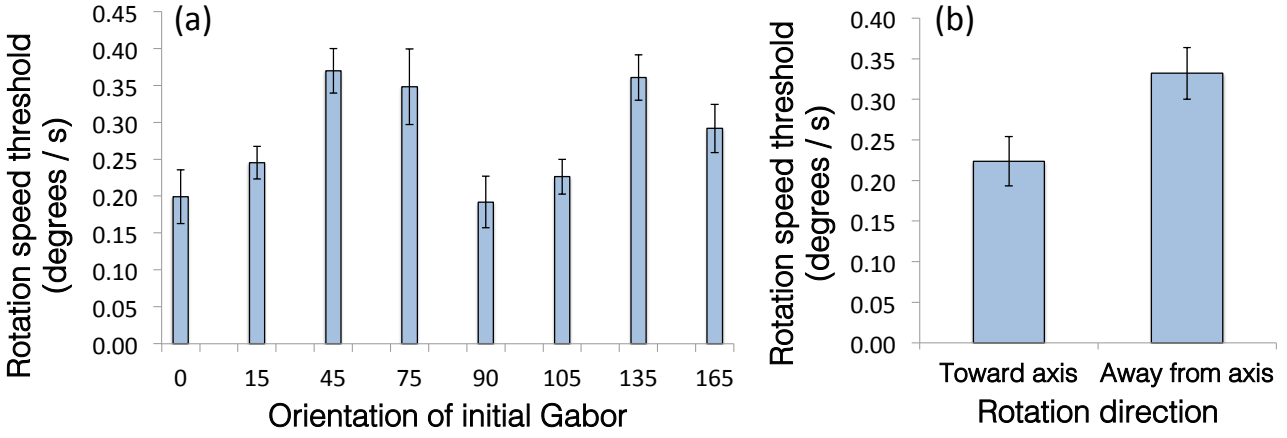


Figure 3. *Rotation speed thresholds (averaged across clockwise and counter-clockwise conditions) for detecting rotation direction in a stimulus displayed for 6 seconds, starting at various initial test orientations. The oblique effect is evidenced by lower thresholds for detecting rotation direction from cardinal orientations, and higher thresholds for oblique orientations. (b) Rotation speed thresholds for detecting rotation direction for tests with an initial orientation ±15° from a cardinal axis. These data were grouped, for each participant, by whether the stimulus rotated toward or away from the nearest cardinal axis. All data show threshold estimates averaged across participants. Error bars show ±1 SEM.*

However, we do not believe our data provide unambiguous evidence that an observer's PSS differs from physical stasis. Another interpretation is that subjective stasis corresponds with physical stasis, but sensitivity is higher for toward-cardinal than away-from-cardinal axis rotations. This is consistent with the predictions of the population model of orientation coding outlined in the previous section. If sensitivity to rotational motion of an oriented stimulus is limited by orientation coding sensitivity, or if observers rely on judging absolute orientation at the end of a test presentation (as opposed to actually detecting movement), participants would be more sensitive to rotations that change orientation toward a cardinal axis, as opposed to rotations that change orientation toward an oblique axis. A third possi-

bility is that asymmetric thresholds measured using a method of single stimuli result from a systematic response bias, for example a tendency to report that stimuli are rotating away from the nearest cardinal axis when in doubt (Morgan, Dillenburger, Raphael, & Solomon, 2012). In Experiment 2 we disambiguate these possibilities by using brief static tests, and a bias-minimising measure of sensitivity.

2.5 EXPERIMENT 2

If asymmetric thresholds for detecting rotations toward and away from cardinal orientations measured in Experiment 1 result from perceptual drift, one would expect that a prolonged stimulus exposure would be necessary to measure this evidence. If, on the other hand, orientation sensitivity differences due to anisotropic orientation tuning (Li et al., 2003) are responsible for asymmetric thresholds, we should find asymmetric sensitivity when people try to detect orientation differences among brief static tests. In Experiment 2 we tested this proposition using a bias-minimising measure of orientation sensitivity — a three-interval forced-choice 'odd one out' task.

2.5.1 *Method*

Details were as for Experiment 1, with the following exceptions.

2.5.1.1 *Participants*

There were 6 participants, comprising the first author and 5 experienced observers who were naïve to experimental hypotheses.

2.5.2 *Stimuli and apparatus*

Each trial involved three sequential presentations of static Gabors, with a spatial frequency of 3 cycles/dva, a randomised waveform phase and a Michelson contrast of 1.0 (see Figure 4). These were shown at fixation for 300ms, each separated by 200ms intervals filled by dynamic noise (individual elements subtending 0.03 dva x 0.03 dva, updated at the monitor refresh rate). Two of the Gabors were of a standard orientation, and the orientation of the third 'odd-ball' was varied on a trial-by-trial basis according to adaptive staircase procedures (see below). Order of presentation was randomized on a trial-by-trial basis. After the final dynamic noise sequence, a static white noise field was displayed until the participant indicated, by pressing one of three mouse buttons, if the first, second, or third Gabor had had a different orientation from the other two. Auditory feedback was given, and the next trial began after 1 second.

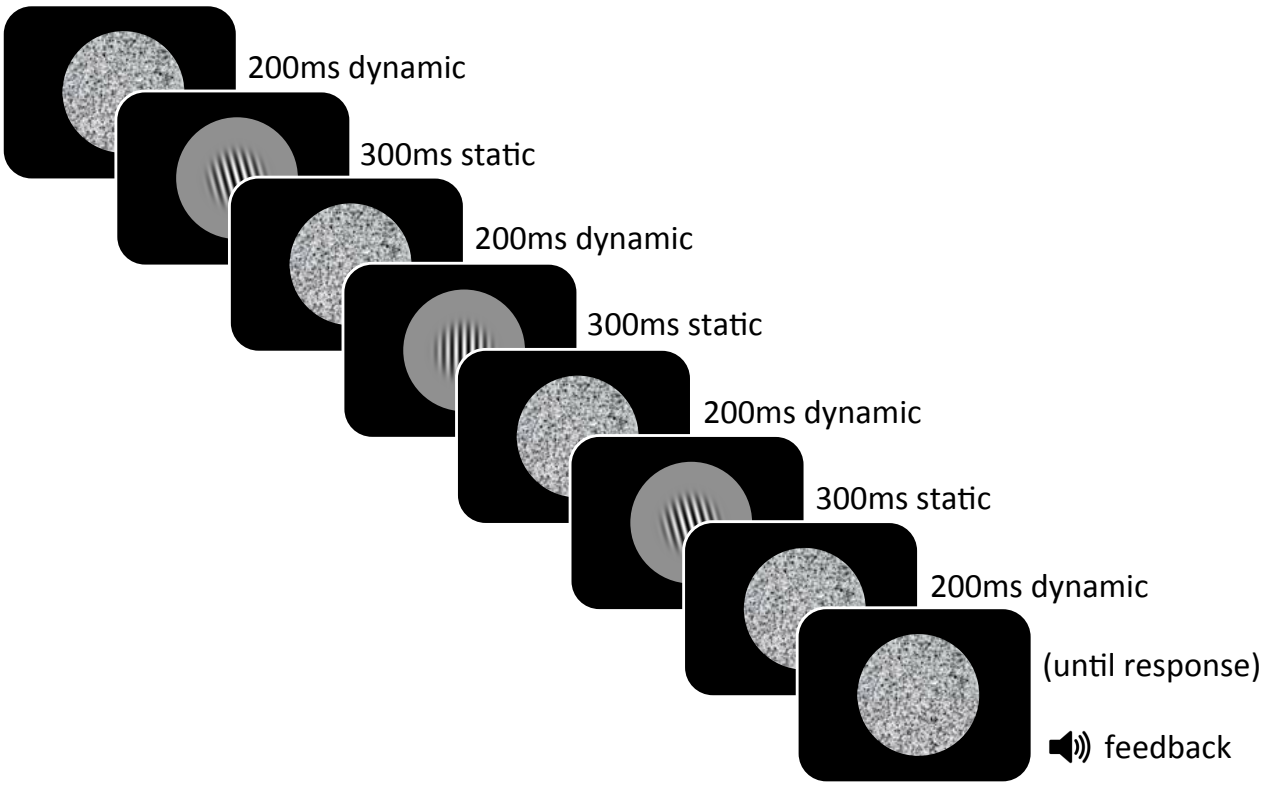


Figure 4. *Example trial sequence in Experiment 2. Three static Gabors are shown sequentially for 300ms each, preceded and followed by 200ms bursts of dynamic white noise. Two of the Gabors had a standard orientation, and one deviated by an amount determined by an adaptive procedure (here the odd-ball is presented in the second interval). At the end of this sequence, participants reported in which interval the odd-ball had appeared, followed by auditory feedback.*

Orientation difference thresholds for correctly detecting the 'odd-ball' were measured at the same 8 orientation points tested in Experiment 1, using four separate staircase procedures for each point. Two of the staircases sampled odd-balls tilted clockwise from standards, one initiated at a large orientation difference (20°, a 'shrinking' staircase) and another at no orientation difference (0°, a 'growing' staircase). A complementary pair of staircases sampled odd-balls tilted counter-clockwise from standards. Odd-ball orientations were adjusted by 1° within each staircase procedure until the first 6 reversals were recorded, after which orientation was adjusted in steps of 0.5° for that staircase. Staircase values had a set minima of 0°, ensuring that staircase procedures searching for clockwise thresholds did not contain any counter-clockwise tilted stimuli, and vice versa. Each staircase terminated after 12 reversals.

Orientation values corresponding to the last 6 reversals in each of the paired shrinking and growing staircases were averaged to estimate independent orientation difference thresholds for each participant for odd-balls tilted clockwise and counter-clockwise from each of the 8 standard orientations.

Participants completed two blocks of trials, each of which sampled tests drawn randomly from one of 16 interleaved staircase procedures. One block of trials tested thresholds

around cardinal and oblique orientations (i.e. four staircases centred around each of 0°, 45°, 90° and 135°), while the other tested thresholds around near-cardinal orientations (i.e. 15°, 75°, 105°, and 165°). These groupings ensured that participants were not adapting to an average orientation during any block of trials. Half of the participants completed the block containing cardinal and oblique tests first, the other half completed the block containing near-cardinal orientations first.

2.5.3 *Results*

Odd-ball detection thresholds, expressed as a function of standard orientation (see Figure 5a), display a clear oblique effect, with largest orientation difference thresholds for oblique standards (45° or 135°), and smallest for cardinal standards (0° or 90°).
Discrimination thresholds for near-cardinal standards (15°, 75°, 105°, and 165°), grouped and averaged according to whether odd-balls were tilted toward or away from the nearest cardinal axis, revealed that larger physical differences were required to detect a brief static odd-ball tilted *away from* a cardinal axis (10.1 ±1.2°) than *toward* a cardinal axis (6.3 ±0.5°; two-tailed paired t_5 = -4.81, p = .005; see Figure 5b).

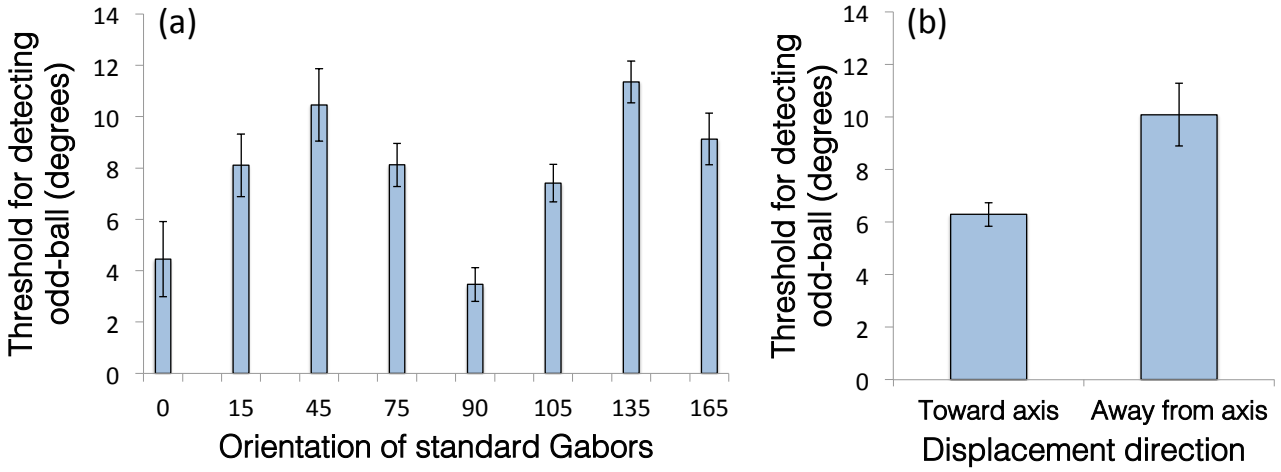


Figure 5. ***(a)*** *Mean thresholds for detecting odd-balls tilted relative to various standard orientations. Note the oblique effect.* ***(b)*** *Average thresholds at near-cardinal test points, for detecting an odd-ball tilted either toward or away from the nearest cardinal axis. Error bars show ±1 SEM.*

2.6 DISCUSSION

Our data demonstrate greater sensitivity to changes in orientation toward than away from a nearby cardinal axis. This was true both when changes in orientation were gradual over time (rotation-direction detection thresholds in Experiment 1) and when changes were between brief static stimuli (Experiment 2). An anisotropic model of orientation encoding,

in which neurons selective for cardinal orientations have narrower tuning curves than those selective for oblique orientations, predicts these asymmetric discrimination thresholds (see Modelling section), in addition to the well-known oblique effect (Girshick et al., 2011; Li et al., 2003; Rose & Blakemore, 1974). We hold that asymmetric sensitivity to changes toward vs away from cardinal axes provides a more parsimonious explanation for Müller, Schillinger, et al. (2009)'s data than does perceptual drift caused by tilt normalisation.

Müller, Schillinger, et al. (2009) reported that the magnitude of the perceptual drift suggested by their data was smaller, by a factor of five, than the magnitude of the tilt aftereffect. They also reported that there was no correlation between the magnitudes of these two effects. Our account explains this dissociation, as it posits that the tilt aftereffect arises from adaptation-induced changes in the responsiveness of orientation-selective neurons, whereas the bias measured when nulling the rotational motion of oriented stimuli results from asymmetric sensitivity to orientation changes which is present before adaptation.

We have not yet addressed Müller, Schillinger, et al. (2009)'s curvature data, but suspect they are amenable to a similar explanation. Greater sensitivity to *decrements* than to *increments* in curvature would produce data consistent with the 'perceptual uncurling' of slightly curved stimuli posited by Müller, Schillinger, et al. (2009). One possible reason for such asymmetric sensitivity is that when a slightly-curved near-vertical line segment (as in Müller, Schillinger, et al. (2009)) is straightened, the local orientation changes are toward vertical, whereas when the segment is further curved the local orientation changes are toward oblique axes. Enhanced sensitivity for toward-cardinal compared to toward-oblique changes would therefore predict enhanced sensitivity to decrements compared to increments in curvature for such stimuli. This would explain why Müller, Schillinger, et al. (2009)'s 'perceptual uncurling' rate estimates are tied to the aspect ratio of the stimulus rather than to absolute curvature, as the local orientations comprising a stimulus do not change with stimulus size. If this explanation is correct, a reversed pattern of results should ensue for obliquely-oriented curved line segments, which when straightened would contain local orientations *further* from the cardinal axes.

While the predictions of our model agree qualitatively with our data both in terms of the shape of the oblique effect (compare Figure 1c to Figures 3a and 5a), and the direction of the near-cardinal discrimination threshold asymmetries, our model makes no firm predictions regarding the magnitudes of these effects. The threshold values predicted by the model depend on an arbitrarily declared average JND (5°), and the size of any asymmetry predicted would vary substantially for different JND values. Further, we have not made any assumptions regarding the temporal evolution or integration of neural orientation signals. In Experiment 1 participants were able to detect a change in orientation from vertical of approximately 1.2° during 6-second presentations. In Experiment 2 they were able to detect a difference in orientation of approximately 3.5° between successive static inputs, each presented for just 300ms. This discrepancy suggests an evolution in signal to noise, possibly

due to reiterative processes. Since there is insufficient background information to justify assumptions about these processes, our modelling makes no quantitative predictions for JND values in different tasks.

2.6.1 *If tilt normalisation did happen, how could it arise?*

Although evidence for tilt normalisation remains underwhelming, our data do not rule it out as a possibility. There are various ways in which normalisation could be predicted. Vaitkevicius et al. (2009) recently proposed a population-coding model to account for both tilt aftereffects and tilt normalisation. In their model narrowband orientation selectivity in cortex is derived from sub-cortical mechanisms that have broad selectivity for the two cardinal orientations, and normalisation occurs as a flow-on effect of adaptation within these cardinal channels. However, this model seems to fail on empirical grounds, erroneously predicting that adaptation to 0°, 90° and $\pm 45^\circ$ should not induce a tilt aftereffect (which is inconsistent with data in Templeton, Howard, and Easting (1965)), and that the appearance of stimuli oriented $\pm 22.5^\circ$ and $\pm 67.5^\circ$ from vertical should not be affected by adaptation (which is inconsistent with data, e.g. Mitchell and Muir (1976)).
Models similar to that described here could be modified to predict either normalisation or anti-normalisation, depending on the average channel bandwidth and the degree of bandwidth variance. Given that there are no firm estimates from investigations of human visual cortex to constrain assumptions regarding these parameters, the fact that any particular instantiation of a model might, or might not, predict normalisation cannot be regarded as firm evidence for the existence or absence of this process. Another possibility is that tilt normalisation happens because a population code, like that described here, is pre-adapted to an environment containing a preponderance of cardinal, compared to oblique, orientations. Arguably, this could be expected from adaptation to natural scenes (Girshick et al., 2011; Hansen & Essock, 2004). Data pertaining to blur normalisation has been similarly explained, by assuming pre-adaptation to a $1/f$ spatial frequency distribution (Elliott, Georgeson, & Webster, 2011). It remains to be seen, however, if such explanations are necessary to explain tilt normalisation, as it is unclear if tilt normalisation exists; our data suggests it doesn't.

2.6.2 *A general methodological point*

We believe our data highlight an important methodological issue. Sensitivity is non-uniform along most, if not all, perceptual continua. For magnitude continua (luminance, weight, loudness, etc), Weber's law dictates that sensitivity to a decrement will be superior to an increment of the same magnitude. Along other continua, sensitivity is often maximal

around null points, such as zero binocular disparity (Stevenson, Cormack, Schor, & Tyler, 1992) a white/grey hue (Webster, 1996), or at subjective boundaries, such as those between phonemes, colours (Goldstone & Hendrickson, 2009) and complex object categories (Beale & Keil, 1995; Newell & Bülthoff, 2002). Most methods for establishing points of subjective equality will systematically err in all of these circumstances. Staircase procedures converge on the centre of a region of uncertainty, which will coincide with the point of subjective equality only when discrimination thresholds are symmetric. Symmetric thresholds are also assumed when fitting data with symmetric psychometric functions. If this assumption is violated, the estimate of central response tendency will deviate from the point of subjective equality. As uneven sensitivity to stimulus changes is common, procedures and analyses should generally be preferred that avoid the assumptions of symmetric sensitivity and or symmetric criteria placement (see Yarrow, Jahn, Durant, and Arnold (2011)).

2.6.3 *Conclusion*

The existence of tilt normalisation remains under question. Recent data taken as evidence for tilt normalisation might instead result from anisotropic sensitivity to orientation differences — the oblique effect.

3

CHAPTER 3: SHAPE AFTEREFFECTS REFLECT SHAPE CONSTANCY OPERATIONS: APPEARANCE MATTERS

The remaining chapters concern perception of increasingly complex stimulus properties, such as the aspect ratio of a shape, or the gender of a face. A fundamental question regarding such aftereffects is whether they arise from adaptation within substrates explicitly encoding complex attributes, or whether they can be explained entirely by the aggregation of aftereffects between simpler image elements such as colour, contrast, and edge orientation (see, e.g. Dickinson, Almeida, Bell, and Badcock (2010) and Dickinson and Badcock (2013)).

Aspect ratio aftereffects provide an interesting test case. When adapting and test shapes are presented at the same size and retinal location, any resulting aftereffect could be attributed either to *shape* adaptation, or to adaptation to the orientations of local edges constituting the shape. Tilt aftereffects are tightly retinotopically localised (Gibson, 1937; Knapen et al., 2010), and can therefore be mitigated by changing the size or position of stimuli between adaptation and test. Previous work shows that shape aftereffects can survive substantial changes in size and position between adaptation and test (Badcock et al., 2014; Köhler & Wallach, 1944; Regan & Hamstra, 1992; Suzuki & Cavanagh, 1998), indicating a relatively late locus of adaptation, where neurons have large receptive fields.

However, the conscious perception of shape is not determined solely by the shape of images cast on the retina, but depends on further stages of 'shape constancy' computations. These partially compensate for perspective foreshortening and other distortions, to more faithfully signal the shapes of distal objects (Epstein & Park, 1963; Vishwanath, Girshick, & Banks, 2005). We wondered whether it was possible to adapt shape representations that lay not only after some degree of spatial invariance had been achieved, but also after shape constancy calculations had been completed. The resulting manuscript appeared in the *Journal of Experimental Psychology: Human Perception and Performance* (Storrs & Arnold, 2013).

Shape aftereffects reflect shape constancy operations: appearance matters

3.1 ABSTRACT

One of the oldest known visual aftereffects is the shape aftereffect, wherein looking at a particular shape can make subsequent shapes seem distorted in an 'opposite' direction. After viewing a narrow ellipse, for example, a perfect circle can look like a broad ellipse. It is thought that shape aftereffects are determined by the dimensions of successive retinal images. However, *perceived* shape is invariant for large retinal image changes resulting from different viewing angles; current understanding suggests that shape aftereffects should not be impacted by the operations responsible for this viewpoint invariance. By viewing adaptors from an angle, with subsequent fronto-parallel tests, we establish that shape aftereffects are not solely determined by the dimensions of successive retinal images. Moreover, by comparing performance with and without stereo surface-slant cues, we show that shape aftereffects reflect a weighted function of retinal image shape and surface slant information, a hallmark of shape constancy operations. Thus our data establish that shape aftereffects can be influenced by perceived shape, as determined by constancy operations, and must therefore involve higher-level neural substrates than previously thought.

3.2 INTRODUCTION

Sensory aftereffects are ubiquitous in human vision, and are generally attributed to neural adaptation (Barlow & Hill, 1963; Clifford et al., 2007; Wark et al., 2007). One of the oldest examples is the shape aftereffect, wherein exposure to a particular shape makes subsequent shapes seem 'oppositely' distorted (Köhler & Wallach, 1944). After exposure to a narrow rectangle, for instance, a square can appear as a wide rectangle. Shape aftereffects have been linked to V4 activity (Gheorghiu & Kingdom, 2007; Müller, Wilke, & Leopold, 2009; Suzuki, 2003), wherein shape may be encoded by the position and magnitude of points of maximal local curvature (Pasupathy & Connor, 2002). Models of shape perception based on this information can achieve size and position invariance (Poirier & Wilson, 2006), both established properties of the shape aftereffect (Regan & Hamstra, 1992; Suzuki & Cavanagh,

1998). Note, however, that in such models the description of a shape depends solely on the shape of its retinal image.

A single retinal image can, however, result in the perception of many different shapes. Conversely, a single object can result in many different retinal images, with different local curvatures (see Figure 1). This dissociation between retinal image shape and perceived shape is overcome by shape constancy calculations, which enable more accurate shape perception by discounting the retinal image changes that result from viewing objects from different perspectives. These operations are not included in our current understanding of shape aftereffects (Gheorghiu & Kingdom, 2007; Poirier & Wilson, 2006).

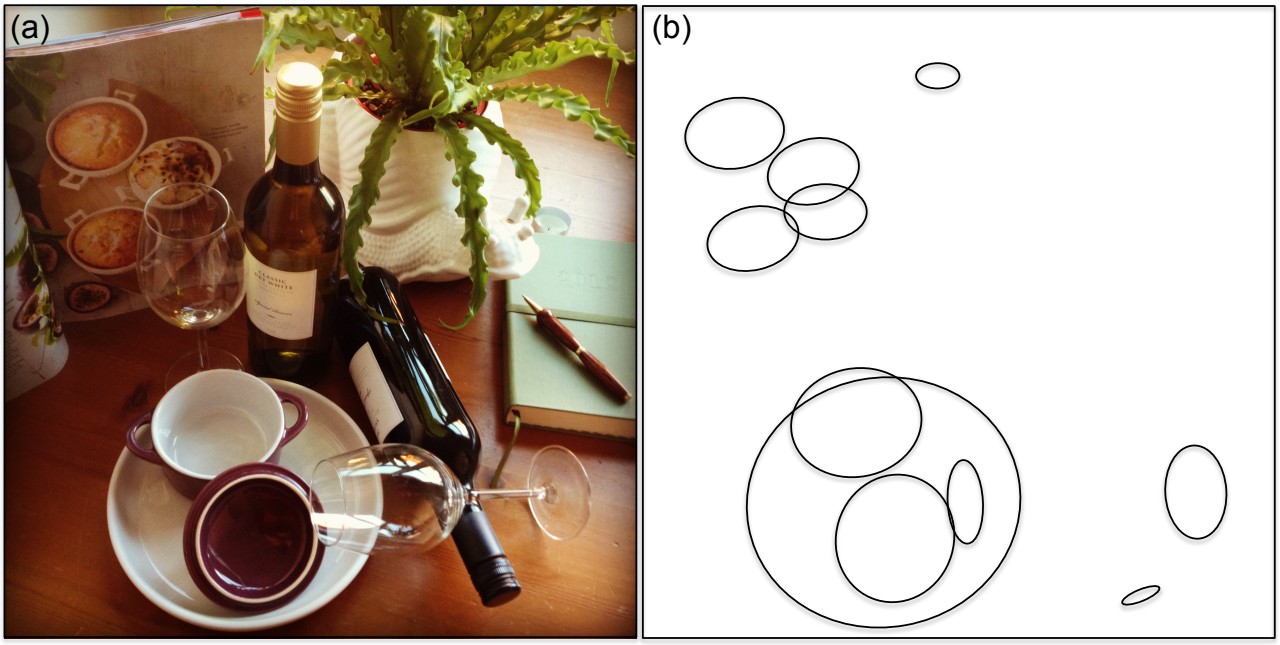


Figure 1. *Shape constancy demonstration. The scene on the left* ***(a)*** *seems to contain many circular shapes, although on the image surface plane and on the retina, each corresponds to a different non-circular ellipse* ***(b)****. In some cases the geometry of apparently circular shapes is inconsistent with that of the rest of the scene (i.e. the images in the magazine — see Vishwanath, Girshick, and Banks (2005).*

For two-dimensional ellipses, distal shape cannot be recovered solely from the shape of the retinal image, as any distal ellipse can project an infinite number of retinal ellipses, depending on the viewer's perspective (Pizlo, 1994). Accurate perception of an ellipse instead depends on recovering the slant of the surface on which the image lies (Pizlo, 1994; Thouless, 1931). Local surface slant (S_{local}) can be estimated using binocular disparity and contextual information, such as that provided by perspective and texture cues (Vishwanath et al., 2005). One can then infer distal ellipse aspect ratio (A/B) from the retinal image ratio (α/β):

$$A/B = \frac{\alpha/\beta}{cos(S_{\text{local}})}$$

Neural responses in human vision don't exhibit full shape constancy until areas beyond V4, a current candidate substrate for shape aftereffects Janssen, Vogels, and Orban, 1999;

Kourtzi and Kanwisher, 2001. Here we test whether shape aftereffects reflect shape constancy calculations.

3.3 EXPERIMENT 1: INFLUENCE OF SURFACE SLANT ON SHAPE PERCEPTION

First, we quantified the impact of viewing angle and stereoptic surface slant information on perceived shape.

3.3.1 *Method*

3.3.1.1 *Participants*

There were eight participants, comprising six experienced psychophysical observers who were naïve to the hypotheses of the experiment, and the two authors.

3.3.1.2 *Stimuli and apparatus*

Stimuli were presented on a gamma-corrected 19″ Sony Trinitron Multiscan G520 monitor (resolution 1024 x 768 pixels and a refresh rate of 120 Hz). Matlab software was used to drive a VSG 2/3 stimulus generator from Cambridge Research Systems. The monitor was fixed in the centre of a circular table, with seven viewing positions marked around the edge of the table: at viewing angles of 0° (fronto-parallel), 10°, 20°, 30°, 40°, 50°, and 60° clockwise from fronto-parallel (see Figure 2a, apparatus). Participants viewed stimuli at each position from a distance of 77cm with their heads restrained by a chinrest in normal (well lit) lighting conditions.

Stimuli consisted of ellipses rendered in white noise, with an element size subtending 0.11 degrees of visual angle (dva) shown against a grey background (CIE chromaticity coordinates x = 0.27, y = 0.30, Y = 47), such that the ellipse and background had the same average luminance. Stimuli were presented in the centre of the display, with a fixed height subtending 6.2 dva. Stimulus width was adjusted by the participant during the trial (see below). Two strips of white noise were added to the top and bottom of the display, each 5.7 dva in height and vertically centred 7.45 dva from fixation (see Figure 2a, display). The presence of these textured strips, coupled with the lighting conditions, ensured that there were ample perspective and texture cues regarding the slant of the test display relative to the observer.

3.3.1.3 *Adjustment procedure*

On each trial an ellipse was presented in the centre of the screen with a randomly determined initial aspect ratio of either 0.6 (distinctly narrow) or 1.7 (distinctly wide). Partici-

pants adjusted this shape using mouse buttons to reduce or increase its width, until they felt it looked like a circle. Twenty such adjustments were completed from a single viewing angle during each block of trials. Participants completed 14 blocks of trials: one from each of 7 viewing angles, once under normal binocular viewing conditions, and once with the participant's right eye covered by an eye patch (monocular condition). Participants completed the blocks in a pre-determined pseudo-random order (0°, 40°, 20°, 60°, 50°, 10°, 30°), with monocular and binocular runs conducted in separate testing sessions. Four participants completed all binocular blocks of trials first, then all monocular blocks. This order was reversed for the remaining four participants.

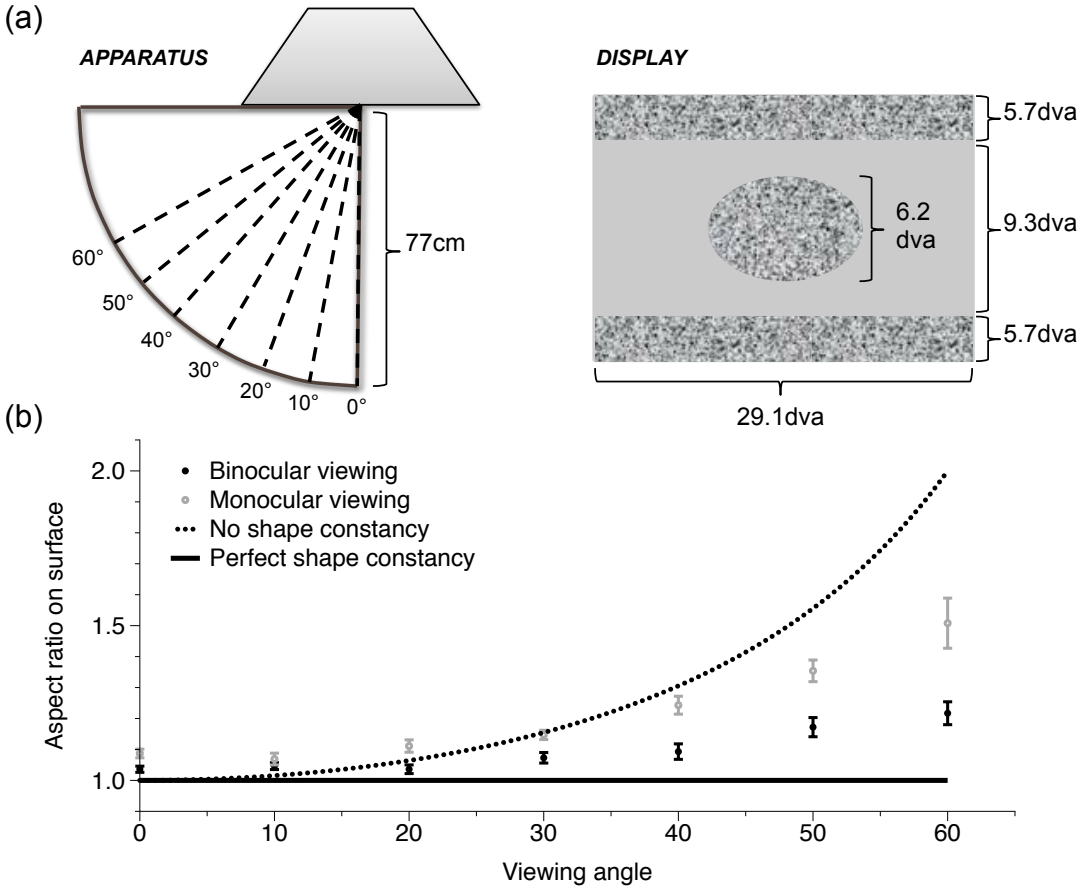


Figure 2. ***(a)*** *Schematic illustrations of the apparatus and stimuli. A monitor was placed in the centre of a section of circular table, and chinrests were attached at various angles to allow oblique viewing at a fixed distance of 77cm. In all three experiments, stimuli consisted of ellipses drawn in white noise on a grey background, with bars of white noise at top and bottom of the screen.* ***(b)*** *Average "circle" settings from different viewing angles. Solid line shows predicted settings if perceived shape were perfectly viewpoint invariant. Dotted line shows predicted settings if perceived shape matched the dimensions of retinal images. Error bars show ±1 SEM.*

3.3.2 *Results*

As found previously Epstein and Park, 1963; Thouless, 1931; Vishwanath et al., 2005, apparent shape was not given by the dimensions of retinal images, but nor was there perfect compensation for retinal image foreshortening as viewing angle increased (see Figure 2b). Instead, apparent shape reflected a weighted function of retinal image dimensions and surface slant estimates (with a slight vertical-horizontal illusion noticeable at smaller viewing angles, e.g. Kunnapas (1955)). Perceived shape was markedly closer to retinal shape when viewing stimuli with one eye, rather than two, revealing a contribution of stereo cues to surface slant estimates.

3.4 EXPERIMENT 2A: ADAPTING TO RETINAL OR PHYSICAL CIRCLES

To answer our central question, observers adapted to ellipses from a viewing angle of 50°, then moved to judge the aspect ratio of test shapes from directly in front.

3.4.1 *Method*

Details for Experiment 2a were as for Experiment 1, with the following exceptions.

3.4.1.1 *Participants*

There were twelve participants, including the authors and ten additional experienced psychophysical observers who were naïve to the experimental hypotheses. Eight of the twelve had also participated in Experiment 1.

3.4.1.2 *Stimuli*

Test stimuli were ellipses consisting of static white noise, as in Experiment 1. Adapting stimuli were ellipses consisting of white noise that was dynamically updated at a rate of 0.5Hz, in order to mitigate the build-up of retinal afterimages. Two adaptor conditions were used to dissociate the effects of retinal from perceived adaptor shape on the appearance of tests. In the first condition people adapted to a physical circle on the surface of the display screen, which resembled a narrow ellipse from the adapted viewing angle (due to a partial failure of shape constancy, see Figure 2b). In the second condition people adapted to an ellipse that projected a circular retinal image (surface aspect ratio = $1/cos(50^\circ)$). To participants, this resembled a wide ellipse during adaptation. Both adaptor conditions were completed under both binocular and monocular viewing.

3.4.1.3 *Procedure*

Participants were instructed to fixate a small red fixation point in the centre of each adapting and test stimulus throughout its presentation. All test stimuli were viewed while seated at 0°, and all adaptors were viewed from 50°. Each participant completed two preliminary baseline blocks of trials (one under binocular viewing and another under monocular viewing conditions), during which tests were viewed without pre-exposure to an adapting stimulus.

During all blocks of trials the test ellipse aspect ratio was varied according to the method of constant stimuli, sampling ratios of 0.90, 0.95, 0.98, 0.99, 1.00, 1.01, 1.02, 1.05, and 1.10. Each ratio was presented 12 times (during baseline blocks of trials) or 6 times (during adaptation blocks of trials) in a pseudo-random order, yielding a total of 108 (baseline) or 54 (adaptation) individual trials. After each test presentation the participant classified the test as being narrower or wider than a perfect circle, by pressing one of two mouse buttons.

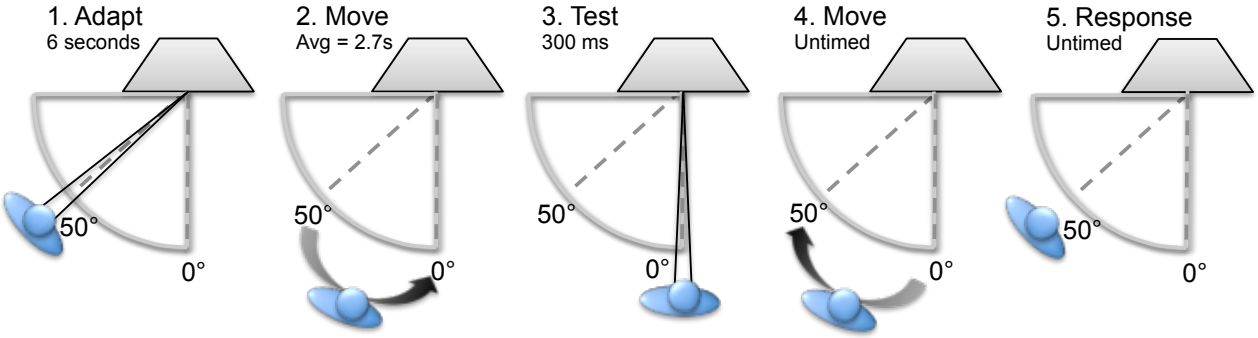


Figure 3. *Trial sequence during adaptation blocks in Experiments 2a and b.* ***(1)*** *Observer views an adaptor for 6 seconds from 50°,* ***(2)*** *then moves to a chin rest at 0° and* ***(3)*** *triggers a 300ms test display.* ***(4)*** *The observer then moves back to 50° and indicates if the test had been narrower or wider than a circle* ***(5)****.*

During adaptation blocks of trials, participants performed the same task at 0° after viewing one of the two adaptors, for 30 seconds on the first trial or for 6 seconds on subsequent trials (see Figure 3). The inter-stimulus interval (ISI) between adaptation and test presentations was dictated by how long it took the participant to first move between the 50° and 0° viewing positions and then to indicate with a button press that he or she was ready to view a test stimulus. This value was recorded (M = 2.74s, SEM = 0.20s).

Six participants completed the experimental conditions in the following order: binocular baseline, surface-circle adaptation, then retinal-circle adaptation in one testing session, followed in a separate session by monocular baseline, surface-circle adaptation, and retinal-circle adaptation. The order for the remaining six participants was: monocular baseline, retinal-circle adaptation, and surface-circle adaptation in the first session, followed by binocular baseline, retinal-circle adaptation, and surface-circle adaptation in the second session.

Experimental sessions were minimally separated by an hour, and often took place on separate days.
Each block of trials provided a distribution of perceived aspect ratio as a function of physical aspect ratio. Logistic functions were fitted to upper responses ("wider than a circle") using the *psignifit* toolbox version 2.5.6 for Matlab (Wichmann & Hill, 2001), and 50% points were taken as estimates of the ellipse aspect ratio that resembled a circle. Aftereffect magnitudes were calculated by taking the difference between PSE estimates from baseline and adaptation runs of trials. Aftereffects during monocular and binocular adaptation conditions were respectively calculated relative to the participant's monocular and binocular baselines.

3.4.2 *Results*

After adaptation to a shape that formed a circle on the surface of the screen, frontally viewed ellipses tended to look wider than they had before adaptation (binocular viewing paired-samples t_{11} = -5.50, p $<$ .001; monocular t_{11} = -5.14, p $<$ 0.001, see Figure 4). It is important to note that these aspect ratios are asymmetric about 1. For example, a shape with an aspect ratio (width/height) of 0.5 is more horizontally stretched than a shape with aspect ratio 1.5 is vertically stretched. These data can be made symmetric by taking logarithmic transforms of raw aspect ratios, i.e. $\log_{10}$(width/height). Transforming the data in this way does not substantively change the above result (binocular viewing t_{11} = -5.35, p $<$ .001; monocular t_{11} = -5.11, p $<$ .001). Throughout the rest of the text, analyses performed on log-transformed data are reported after those performed on raw aspect ratios.
In the 'surface circle' condition just described, the adapting ellipse is both retinally and subjectively narrower than a circle, so this result is consistent with the shape aftereffect being tuned to either retinal or perceived shape, and does not speak to our primary question. Critically, though, after adaptation to a retinal circle frontally viewed ellipses tended to look *narrower* than they had prior to adaptation (binocular raw data t_{11} = 6.05, p $<$ 0.001; log transformed t_{11} = 6.16, p $<$ 0.001; monocular raw data t_{10} = 9.47, p $<$ 0.001; log transformed t_{10} = 9.44, p $<$ .001; see Figure 4). This shows that shape aftereffects cannot be determined *solely* by retinal image aspect ratios, otherwise no shape aftereffect should have ensued. Instead, shape aftereffects appear to be influenced by perceived shape, as derived by shape constancy operations.

3.5 EXPERIMENT 2B: ADAPTING TO PERCEIVED CIRCLES

Experiment 2a demonstrated that, in the absence of any *retinal* deviation from circularity, a shape aftereffect still ensues, presumably driven by a contribution from *perceived* shape.

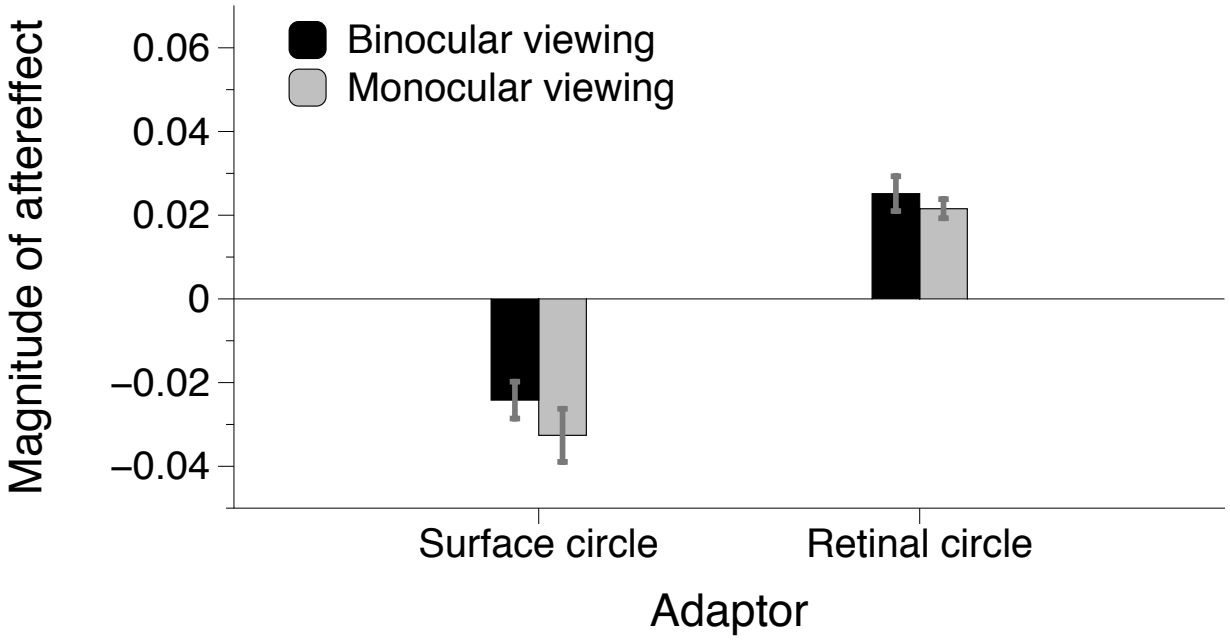


Figure 4. *Bar graph showing shape aftereffects after adaptation to a surface circle (surface aspect ratio = 1.00; retinal aspect ratio approximately = 0.64), and a retinal circle (surface = 1.56; retinal = 1.00), both viewed from 50°. Aftereffect magnitudes represent differences in the point of subjective circularity between baseline and adaptation blocks of trials.*

Experiment 2b tests whether the shape aftereffect is tuned entirely for perceived shape, or if a retinal contribution remains. If shape aftereffects are governed predominantly, or entirely, by perceived shape, then adapting to an apparent circle should produce little or no aftereffect in this context.

3.5.1 *Method*

A subset of participants from Experiment 2a (N = 8, including the authors) also completed Experiment 2b, which was identical except for involving two different adapting shapes. These were determined by the average subjective "circle" settings from Experiment 1 during binocular and monocular viewing from 50° (these had retinal aspect ratios of approximately 0.75 and 0.87). Participants adapted to each from a viewing angle of 50°, under both binocular and monocular viewing conditions, and also completed binocular and monocular no-adaptation baseline blocks of trials at the beginning of the respective testing session. This yielded eight possible orders in which the four adaptation conditions could be conducted in a 2 x 2 x 2 (order of viewing condition x order of binocular adaptors x order of monocular adaptors) permutation. Each of the eight participants was assigned a unique order. The ISI between adaptation and test presentations was again recorded (M_{ISI} = 3.03s, SEM = 0.12s).

3.5.2 *Results*

Experimental conditions were either congruent, in that the adaptors *looked* circular (i.e. binocular adaptation to aspect ratio 0.75 and monocular adaptation to aspect ratio 0.87), or incongruent, in that adaptors did not look circular (binocular adaptation to aspect ratio 0.87 and monocular adaptation to aspect ratio 0.75).

The impact of perceived shape is again evident in the results of incongruent conditions. In these conditions, monocular adaptation to a binocular circle made tests appear *wider* than binocular adaptation to the same physical adaptor (raw data paired-samples t_7 = 2.83, p = .026; log transformed t_7 = -2.80, p = .027; see Figure 5). Note that the retinal adaptor shapes were identical in these two conditions, but the adaptor was *perceived* as wider when stereo slant cues were available during binocular adaptation. Similarly, binocular adaptation to a monocular circle made test ellipses appear *narrower* than monocular adaptation to the same retinal image shape (raw data t_7 = 2.72, p = .030; log-transformed t_7 = 2.72, p = .030). Stereo surface slant cues, which contribute to shape constancy operations, can thus modulate shape aftereffects.

For the congruent conditions, there was no significant difference between aftereffects induced by binocular (aspect ratio 0.75) and monocular (aspect ratio 0.87) adaptation (paired-samples t_7 = 0.61, p = 0.56; log transformed t_7 = 0.62, p = 0.55). We therefore collapsed these data into an average aftereffect estimate for each participant, by taking the mean of the two congruent aftereffects for each participant. Analysis of these data revealed a slight tendency for test ellipses to appear *wider* after adapting to a subjective circle ($M_{PSE\ shift}$ -0.007 ±SEM 0.003; raw data one-sample t_7 = 2.28, p = .056; log transformed t_7 = 2.30, p = .055) These data suggest that adapting to images that look circular (but project narrowed retinal images) can still produce a small aftereffect when judging tests that both look circular and have circular retinal images. Hence these data, in conjunction with the results of Experiment 2a, show that shape aftereffects involve contributions from *both* retinal image shape and perceived shape.

It is important to note that there were no significant differences between ISIs (which reflect how long it took to move between adapting and test positions) during binocular and monocular viewing conditions; binocularly viewed 'binocular circle' M_{ISI} = 2.97 ±SEM 0.38 seconds; monocularly M_{ISI} = 3.04 ±0.29s (paired t_{11} = 0.31, p = .77), binocularly viewed 'monocular circle' M_{ISI} = 2.97 ±0.34s; monocularly M_{ISI} = 3.13 ±0.28 (paired t_{11} = 0.78, p = .46). Hence, differences between aftereffect magnitudes under monocular and binocular viewing cannot be attributed to a different length of time having elapsed between adaptation and test.

Finally, we also examined sensitivity changes induced by shape adaptation by comparing just noticeable differences (calculated as the difference between the 80% and 50% points on individual psychometric functions) before and after adaptation during Experiments 2a

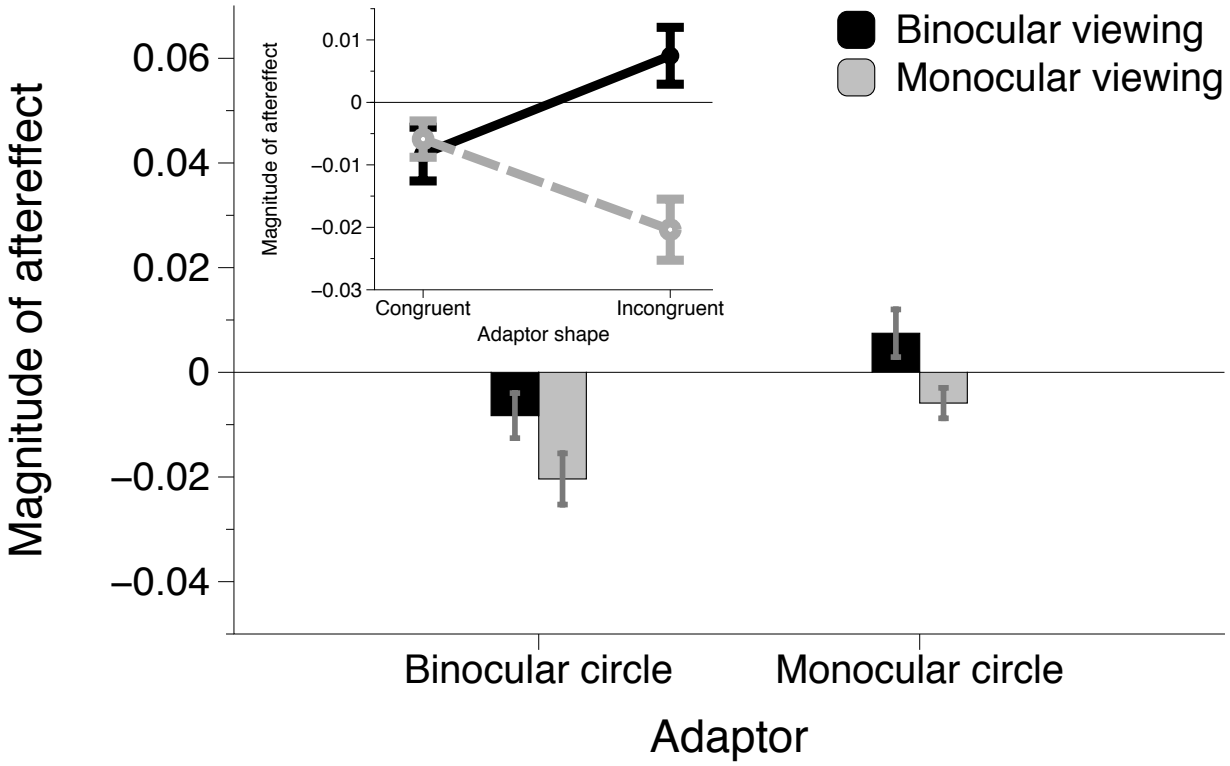


Figure 5. *Aftereffects induced by binocular and monocular adaptation to ellipses that looked circular when viewed binocularly (surface aspect ratio = 1.35; retinal = 0.87) or monocularly (surface = 1.17; retinal = 0.75). Inset: Same data expressed as aftereffect magnitudes after adapting to ellipses that looked circular (congruent) or non-circular (incongruent), illustrating the interaction between adaptor shape and stereoptic cues. All error bars show ±1 SEM.*

and 2b. Overall, thresholds were lower after adaptation than they were before (change in discriminant threshold between baseline and adaptation blocks in both experiments combined $M_{sensitivity}$ = -0.003 ±SEM 0.001, one-sample t_{11} = 2.77, p = .018). This improvement in sensitivity was not, however, well tuned for either retinal image shape or perceived shape, because a 2 x 4 repeated measures ANOVA revealed no main effects or interaction of viewing condition or adaptor shape. This suggests that a functional benefit, in terms of sensitivity to differences in shape, ensued from all adaptation conditions, and any variance in the magnitude of this effect across our adaptation conditions was too subtle to detect.

3.6 DISCUSSION

The shape aftereffect is known to be somewhat tolerant to changes in size (Regan & Hamstra, 1992) and translations across a fronto-parallel surface between adaptation and test (Suzuki & Cavanagh, 1998). However, these transformations both preserve[1] the relative aspect ratios and local curvatures of retinal images, and so provide no information about

1 Or almost preserve them, within the range of translations tested. The maximum local surface slant relative to observers in Suzuki and Cavanagh (1998) was slight, at approximately 6°.

the interplay between shape aftereffects and shape constancy operations. Our data reveal that shape aftereffects can be influenced by perceived shape, as well as by retinal image shape, and thus that shape aftereffects reflect the weighted calculations necessary for (partial) viewpoint invariance.

Precautions taken to minimize the impact of retinotopic adaptation, and associated afterimages, might have been critical to our discovery of a view-invariant contribution to shape aftereffects. In most circumstances perceptual aftereffects, even those for supposedly complex stimuli (such as shapes and faces), will at least partially reflect the effects of adaptation to low-level image properties (Dennett et al., 2012; Dickinson et al., 2010; Xu, Dayan, Lipkin, & Qian, 2008). We minimized such contributions by using dynamic, texture-defined white noise as an adaptor. The time-averaged luminance of these adaptors was equal to that of the grey background, thereby minimizing the impact of adaptation to localized luminance changes. If we had used static, luminance-defined stimuli, viewpoint invariant contributions might have been obscured by a greater contribution from retinotopic aftereffects.

Several previous studies have indicated that high-level visual aftereffects are broadly retinotopic, rather than viewpoint invariant. For example, face aftereffects can transfer between adaptors and tests depicting three-dimensional faces viewed from different perspectives, but they diminish with increasing angular distance (Fang & He, 2005; Jeffery, Rhodes, & Busey, 2007). Similarly, the facial gender aftereffect is greatest when adapting and test images are presented in the same retinal, rather than the same physical, coordinates (Afraz & Cavanagh, 2009). However, these investigations are not directly comparable to ours, as there is no analogous dissociation between (for example) 'perceived facial gender' and 'retinal facial gender' at different viewing angles, as there is with perceived and retinal aspect ratio.

We are only aware of one previous attempt to examine the impact of perceived shape on shape aftereffects. Bell, Dickinson, and Badcock (2008) showed that radial frequency aftereffects can transfer partially between adaptors and tests with different horizontal compressions, consistent with a transfer between shapes viewed from different perspectives. Critically, the aftereffect induced by a horizontally compressed adaptor was larger when disparity cues were added (via stereo goggles), resulting in an adapting image that was consistent with viewing an uncompressed shape from an angle (Bell et al., 2008). This result is consistent with our data, although the studies are only weakly analogous. The previous experiment quantified aftereffects by contrast threshold elevation, rather than by shape appearance, and was exploratory, including only two observers, and only found an effect of stereo cues at the most extreme of four horizontal compressions (Bell et al., 2008).

There are, however, complementary observations concerning the primacy of apparent properties over retinal image properties from other contexts. For instance, size and spatial frequency aftereffects can reflect perceived, rather than physical, properties (Bennett & Cortese, 1996; Burbeck, 1987; Parker, 1981; Sutherland, 1961), and illusory size changes can

influence tilt aftereffects and sensitivity to orientation changes (Arnold, Birt, & Wallis, 2008; Schindel & Arnold, 2010).

Until now, our knowledge of shape aftereffects had been consistent with a neural substrate in which shapes are represented by the magnitudes and positions of points of maximal local curvature in the retinal image, indicating V4 as a plausible substrate (Bell et al., 2008; Gheorghiu & Kingdom, 2007; Pasupathy & Connor, 2002; Poirier & Wilson, 2006; Suzuki, 2003). The influence of shape constancy operations are not, however, evident until later stages in the visual hierarchy (Janssen et al., 1999; Kourtzi & Kanwisher, 2001; Tanaka, 1996, 2003). Our data imply that these higher-level substrates contribute to the determination of shape aftereffects, consistent with Suzuki and Cavanagh, 1998's suggestion that shape representations in inferotemporal cortex (and / or in the superior temporal sulcus) might be involved in generating shape aftereffects. At this point we cannot specify the precise nature of this contribution. It is possible that the contribution is toward recalibrating a sensory code. Equally, it is possible that the contribution has more to do with re-setting decisional criteria. Teasing apart these possibilities will be the focus of future research in our lab.

4

CHAPTER 4: SHAPE ADAPTATION EXAGGERATES SHAPE DIFFERENCES

The results in the previous chapter suggest that shape aftereffects involve shape representations relatively late in the visual processing hierarchy. We can now ask whether these representations are likely to involve a norm-based representation of aspect ratio, as has been proposed (Regan & Hamstra, 1992; Suzuki, 2005). In the present chapter I test whether adaptation to a particular aspect ratio induces renormalisation or local repulsion within the aspect ratio domain.

In order to differentiate renormalisation from local repulsion, I used a spatial comparison method, with a bias-minimising forced-choice task, similar to that used in Chapter 6 (chronologically, the experiments reported in the present chapter were the last conducted during my candidature). The pattern of perceptual changes associated with the adaptation of late-stage shape representations was isolated by using a spatially-jittering adaptation protocol. At the time of writing, the resulting manuscript is in press at the *Journal of Experimental Psychology: Human Perception and Performance* (Storrs & Arnold, 2016).

Shape adaptation exaggerates shape differences

4.1 ABSTRACT

Adaptation to different visual properties can produce distinct patterns of perceptual aftereffect. Some, like those following adaptation to colour, seem to arise from recalibrative processes. These are associated with a re-appraisal of which physical input constitutes a normative value in the environment — in this case what appears 'colourless', and what 'colourful'. Recalibrative aftereffects can arise from coding schemes in which inputs are referenced against malleable norm values. Other aftereffects seem to arise from contrastive processes. These exaggerate differences between the adaptor and other inputs, without changing the adaptor's appearance. There has been conjecture over which process best describes adaptation-induced distortions of spatial vision, such as of apparent shape or facial identity. In three experiments, we determined whether recalibrative or contrastive processes underlie the shape aspect ratio aftereffect. We found that adapting to a moderately elongated shape compressed the appearance of narrower shapes and further elongated the appearance of more elongated shapes (Experiment 1). Adaptation did not change the perceived aspect ratio of the adaptor itself (Experiment 2), and adapting to a circle induced similar bidirectional aftereffects on shapes narrower or wider than circular (Experiment 3). Results could not be explained by adaptation to retinotopically local edge orientation, or single linear dimensions of shapes. We conclude that aspect ratio aftereffects are determined by contrastive processes that can exaggerate differences between successive inputs, inconsistent with a norm-referenced representation of aspect ratio. Adaptation might enhance the salience of novel stimuli, rather than recalibrate our sense of what constitutes a 'normal' shape.

4.2 INTRODUCTION

Visual aftereffects along different sensory dimensions might involve qualitatively distinct processes. Some appear to arise from recalibrative processes that reference the appearance of inputs relative to malleable normative values (Anstis et al., 1998; Webster, 2011; Webster

& Leonard, 2008). For instance, one might have the impression that a particular hue and saturation is a neutral grey, and that all other combinations of hue and saturation are colourful. One might then update one's impression of what appears grey, thereby changing the apparent colour of all points in colour space (Webster, 1996).

Other aftereffects appear to arise from contrastive processes that exaggerate differences between adapting and other inputs. The tilt aftereffect (Gibson, 1937; Vernon, 1934), for instance, is induced by prolonged exposure to a stimulus of a particular orientation. Afterwards, differences between this and other similar orientations tend to be exaggerated, but the apparent orientation of the adaptor itself seems unchanged (Mitchell & Muir, 1976). Similar 'locally repulsive' aftereffects are found following adaptation to spatial frequency (Blakemore & Sutton, 1969) or to a particular direction of motion (Clifford, 2002; Mather, 1980).

It has been suggested that shapes, faces, and other complex spatial stimuli are encoded relative to perceptual norms (Freiwald, Tsao, & Livingstone, 2009; Kayaert, Biederman, Op de Beeck, & Vogels, 2005; Leopold, Bondar, & Giese, 2006, 2001; Loffler et al., 2005; Panis et al., 2010; Rhodes et al., 2005; Webster & Maclin, 1999). According to these proposals, the appearance of complex forms can be distorted via adaptation-induced re-appraisals of what constitutes a normative input (McKone et al., 2014; O'Neil, Mac, Rhodes, & Webster, 2014; Pond et al., 2013; Susilo et al., 2010; Webster & MacLeod, 2011; Webster & Maclin, 1999). For example, after adapting to an unusual face, ones impression of what constitutes a normal face might be updated to more closely resemble the unusual face. This would impact on the appearance of all faces. Both the adapting face and more 'extreme' versions of it, for instance, would appear more normal after adaptation (e.g. Susilo et al. (2010)).

While facial appearance is popularly thought to be subject to recalibrative perceptual aftereffects, there is some contention on this point, with some suggesting that these perceptual distortions are better explained as contrastive aftereffects (Ross et al., 2013; Storrs, 2015b; Storrs & Arnold, 2012, 2015b; Zhao et al., 2011). From a conceptual perspective, this would mean that adaptation-induced distortions of complex form, such as facial appearance, are qualitatively similar to distortions of attributes often described as low-level, such as spatial frequency and orientation (Blakemore & Sutton, 1969; Mitchell & Muir, 1976). We therefore felt it would be interesting to closely examine an intermediately-complex spatial attribute — the aspect ratio of a two-dimensional shape.

The apparent aspect ratio of a shape can be distorted via adaptation. A circle can appear vertically elongated after adapting to a horizontally-elongated ellipse, and can appear horizontally elongated after adapting to a vertically-elongated ellipse (Köhler & Wallach, 1944; Regan & Hamstra, 1992; Suzuki & Cavanagh, 1998). These changes in what looks 'circular' could be explained by a recalibrative aftereffect, which updates one's impression of what constitutes a normal circular shape (Regan & Hamstra, 1992; Suzuki, 2005). They

could be equally well explained by a contrastive aftereffect, which exaggerates aspect ratio differences between the adapted and other shapes (Badcock et al. (2014); see Figure 1b-c).
Aspect ratio is an interesting case in which to test for recalibrative vs contrastive aftereffects, for several reasons. First, the human visual system appears to have dedicated mechanisms for encoding aspect ratio (Badcock et al., 2014; Nachmias, 2011; Regan & Hamstra, 1992; Stankiewicz, 2002). Second, aspect ratio aftereffects likely arise from adaptation at a reasonably late stage of processing, as they reflect *perceived* differences between the adapted and test values, as opposed to differences in the physical aspect ratios of retinal images (see Chapter 3). Finally, several groups have proposed that aspect ratio mechanisms encode shapes relative to a 'neutral' norm corresponding to an apparent 1:1 aspect ratio Kayaert et al., 2005; Regan and Hamstra, 1992; Suzuki, 2003, 2005; Suzuki and Rivest, 1998.
As adaptation to spatial patterns likely occurs at multiple levels of visual processing (Dhruv & Carandini, 2014; Dickinson et al., 2010, 2012; Xu et al., 2008), and we wanted to investigate shape-specific adaptation, it was important to mitigate, as far as possible, the effects of adaptation in mechanisms encoding local contrast or orientation. For example, when adapting and test shapes overlap, any observed effect could be strongly impacted by contour repulsion or tilt aftereffects, likely produced in primary visual cortex rather than later shape-specific stages of processing (Dragoi et al., 2000; Jin et al., 2005). To overcome this we used an adapting stimulus that was intermittently re-positioned about the physical test location, such that the external contours of the adaptor traced a symmetrical outline about the test location (see Figure 3c). Moreover, we diminished the influence of pre-cortical adaptation by adapting and testing with shapes rendered in dynamic white noise, which had the same average luminance over time as the grey display background. With these two precautions, adapting stimuli had no persistent luminance-defined contour that could induce adaptation at the initial tightly-retinotopic stages of processing.
To distinguish between the possibilities that shape aspect-ratio aftereffects involve recalibrative or contrastive processes we had observers adapt to moderately horizontally-elongated ellipses, and examined the effect this had on the appearance of both more horizontally-elongated shapes, and less elongated shapes (circles). If shape aftereffects involve the recalibration of a norm for aspect ratio, both types of test should look *less* horizontally elongated, and more vertically elongated, after adaptation. If shape aftereffects involve a contrastive process, circular tests should look less horizontally elongated (i.e. vertically elongated) whereas the horizontal elongation of more elongated tests should be further exaggerated — that is, we should find opposite distortions for the two types of tests (see Figure 1).

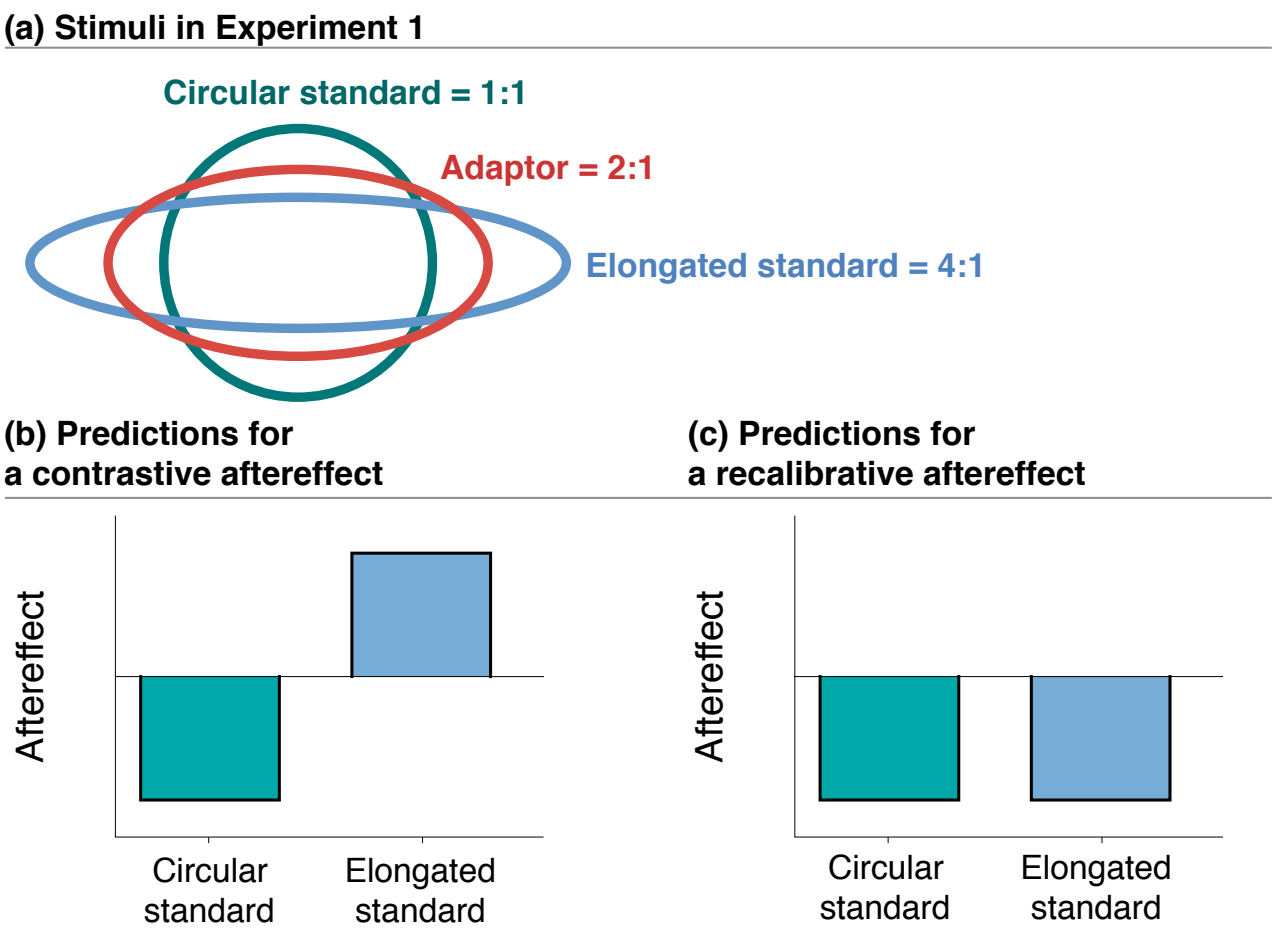


Figure 1. ***(a)*** *Aspect ratios of adapting and standard test shapes used in the experiment.* ***(b)*** *Schematic predictions for how the perceived aspect ratio of each Standard test should change after adaptation, in a contrastive aspect ratio aftereffect, and* ***(c)*** *in a recalibrative aspect ratio aftereffect.*

4.3 EXPERIMENT 1: ASPECT RATIO ADAPTATION

4.3.1 *Method*

4.3.1.1 *Participants*

Ten observers participated, comprising the two authors and eight additional experienced psychophysical observers naïve to the research hypotheses. Experiment 1 was approved by the School of Psychology Ethics Committee at the University of Queensland.

4.3.1.2 *Stimuli and apparatus*

Stimuli were presented on either a 19″ Samsung SyncMaster 950SL, a 19″ Samsung SyncMaster 950p+, or a 19″ Dell Trinitron monitor, all set to a 1280 x 1024 pixel resolution and a refresh rate of 75Hz. Stimuli were generated using the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). Participants viewed stimuli from a distance of 57cm, using a chinrest to stabilise their heads.

Stimuli were elliptical patches of dynamic white noise updated every 10ms, rendered on a grey background. By defining stimuli using dynamic noise textures we minimise adaptation in pre-cortical sites, and in those V1 neurons that act as linear luminance filters (Baker & Mareschal, 2001). In runs of trials involving adaptation, the adapting stimulus had an aspect ratio of 2 (i.e. a 2:1 width-to-height ratio). Two standard test stimuli were used: a "Circular Standard" with an aspect ratio of 1, and an "Elongated Standard" with an aspect ratio of 4 (see Figure 1a). The area of adapting and test shapes was held constant at 14,400 pixels^2 — approximately 7.8 square degrees of visual angle (dva). The Circular Standard therefore subtended 2.8 (width) x 2.8 (height) dva, the Elongated Standard subtended 5.6 x 1.4 dva, and the Adaptor subtended 4.0 by 2.0 dva. Adapting and test stimuli were centred 4.6 dva above or below a central fixation cross, which subtended 0.5dva (see Figure 2).

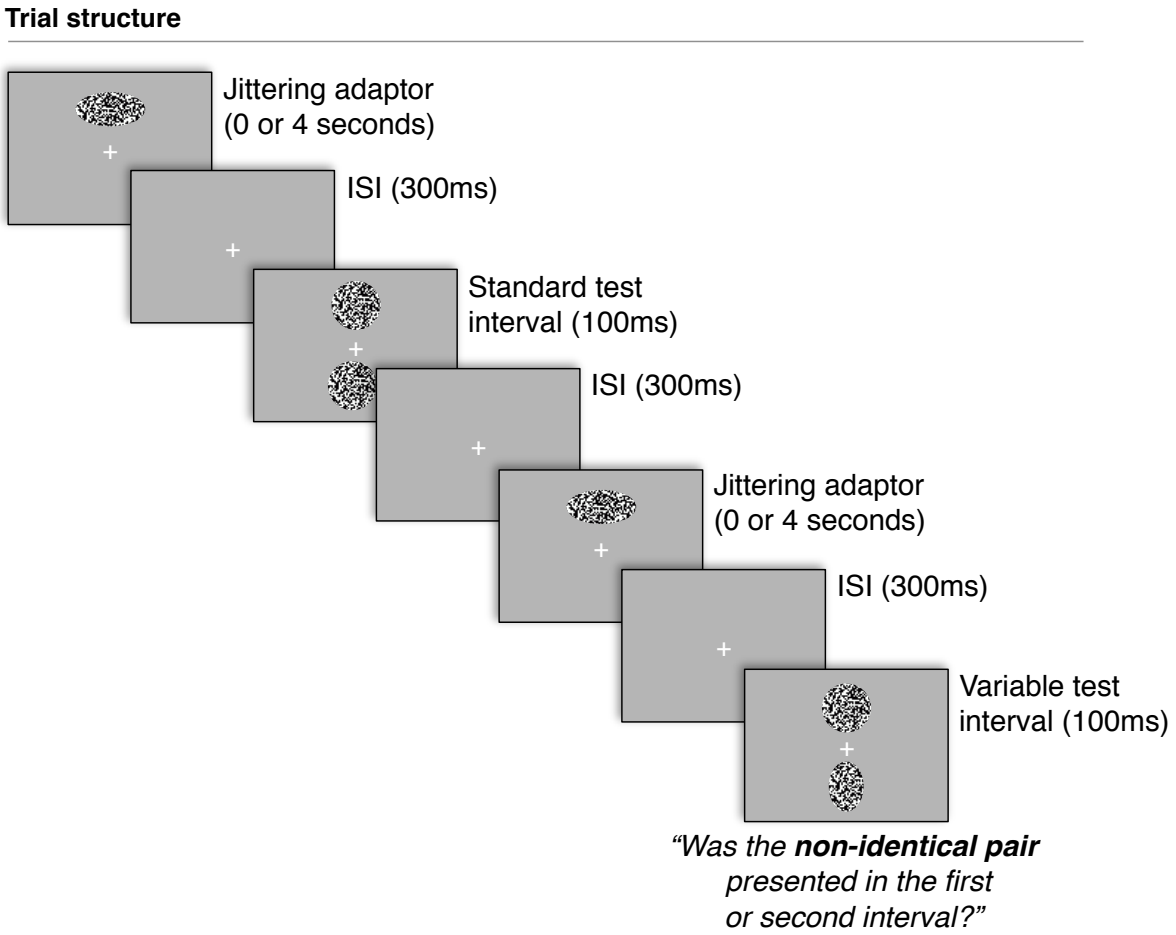


Figure 2. *Example trial structure. Adapting stimuli 'jittered,' appearing at a different location every 100ms (see Figure 3 and main text for details), while test stimuli appeared in a fixed location. Both adapting and test stimuli were rendered in white noise that updated every 10ms. The interval in which the variable test stimulus appeared was randomly chosen on each trial.*

The spatial location of the adapting stimulus jittered randomly within an allowable region (see Figure 3). By ensuring no systematic retinotopic overlap between the contours of adapting and test shapes, location jitter minimises contributions from 'contour repulsion' and tilt adaptation, both of which can be driven by channels sensitive to second-order stimuli such

as ours (Whitaker, McGraw, and Levi (1997) and Larsson, Landy, and Heeger (2006), respectively). The spatial jitter was implemented by randomly selecting a new adaptor location every 100ms, from within a vertically elongated ellipse with an aspect ratio (1:2) opposite to that of the adaptor. The allowable adaptor region was equal in size to the adaptor, and was centred on the test location (see Figure 3a). This resulted in the contours of adapting stimuli tracing an approximately circular region across a block of trials, which importantly had a 1:1 aspect ratio (with a width and height of 5.93 dva; see Figure 3c). Adaptors were presented above fixation for five participants and below fixation for the other five.

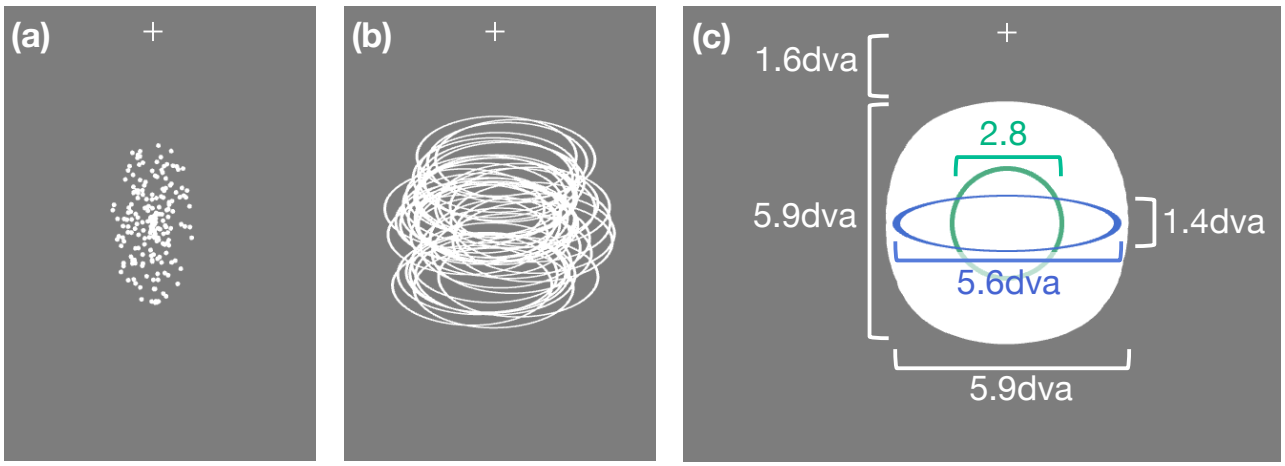


Figure 3. *Illustration of how adaptor positions were jittered.* ***(a)*** *Every 100ms during adaptation the location of the centre of the adapting shape was resampled from within a vertically elongated ellipse region (aspect ratio 1:2) centred on the test location. Dots depict a random sample of 200 possible adaptor locations. Locations were uniformly sampled in polar coordinates, with the result that locations near the test location occurred more frequently than locations further away.* ***(b)*** *White ellipses show the contours of a random sample of 40 adaptor locations, representative of the range of adaptor locations seen by an observer during a four-second adaptation period (note that in the experiment adaptors were rendered in dynamic white noise).* ***(c)*** *Using this sampling method, the region within which adaptor edges could appear approximated a circular shape with a 1:1 aspect ratio (shown as a solid white region). For reference, the outlines of the Circular (green) and Elongated (blue) Standard test shapes are also shown. The sizes of each test shape, and the size and distance from fixation of the adapted region, are indicated in degrees of visual angle (dva).*

4.3.1.3 *Procedure*

Each participant completed one run of trials without adaptation, followed immediately by a run of trials with adaptation. During adaptation runs the adapting shape was displayed for four seconds at the start of each trial. Test stimuli were presented using a dual-pair task (Kaplan, Macmillan, & Creelman, 1978; Rousseau & Ennis, 2001), in which four test stimuli were presented in two sequential pairs (see Figure 2). On each trial three of the four stimuli were of a standard test aspect ratio (either Circular or Elongated, selected pseudo-randomly on each trial), while the fourth varied according to a method of constant stimuli (described below). Each test pair was presented for 100ms. The variable test was always presented in

the unadapted location, in an interval chosen randomly on each trial. There was a blank inter-stimulus interval (ISI) of 300ms between each test pair, and between the adaptor and tests.

The observer's task was to indicate via a keypress whether the non-identical pair (i.e. the pair containing the variable test) had appeared first or second. Observers were instructed to pick the interval containing the larger difference, if both intervals appeared to contain non-identical shapes. This is expected to be the case in many trials after adaptation, unless the variable test shape exactly matches the adaptation-induced distortion of the standard test shape. The point of subjective equality (PSE) between adapted and unadapted locations occurs at the value for which the adaptation-induced distortion of the variable test compensates for its physical difference from the Standard. Near this value, the adaptation-induced difference between the physically identical standards appears larger than that between the variable test and standard, and observers should respond systematically incorrectly. The peak in the proportion of incorrect responses can therefore be taken as an estimate of the observer's PSE.

On each trial the standard test was pseudo-randomly selected to be of either the Circular or Elongated Standard. The variable test was selected pseudo-randomly according to a method of constant stimuli from one of seven aspect ratios, centred logarithmically about the respective standard stimulus. For trials involving Circular Standards (aspect ratio 1), variable tests were selected from aspect ratios of 0.59, 0.71, 0.84, 1.00, 1.19, 1.41, or 1.68. For trials involving Elongated Standards (aspect ratio 4) tests were selected from aspect ratios of 2.38, 2.83, 3.36, 4.00, 4.76, 5.66, or 6.73. Variable tests had the same area as standard tests and the adaptor. Within a run of trials eight samples of each variable test aspect ratio were presented for each of the two types of standard test, yielding a total of 112 trials.

4.3.2 *Results*

Trials involving Circular and Elongated Standard test stimuli were analysed separately. For each, data were expressed as the proportion of trials on which the observer had incorrectly reported on the order of the variable test interval. A Gaussian function was fitted to the proportion of incorrect responses as a function of variable test aspect ratios. The peak of the fitted function was taken as an estimate of the point of subjective equality (PSE) in terms of aspect ratio between the adapted and unadapted locations. A proportional aftereffect score was calculated by dividing the log(PSE estimate derived from adaptation runs of trials) by the log(PSE estimate derived from baseline runs of trials).

After adapting to an aspect ratio of 2, Circular Standard stimuli were matched to more contracted ellipses relative to baseline trials (mean proportional aftereffect -0.12 $\pm$Standard Error of the Mean (SEM) 0.001; t_9 = -9.08, $p < .001$; see Figure 4a). An oppositely-directioned

aftereffect was observed for tests more elongated than the adaptor (0.12 ±0.03; t_9 = 4.75, p = .001). This bi-directional pattern of perceptual changes was found for each observer (see Figure 4b).

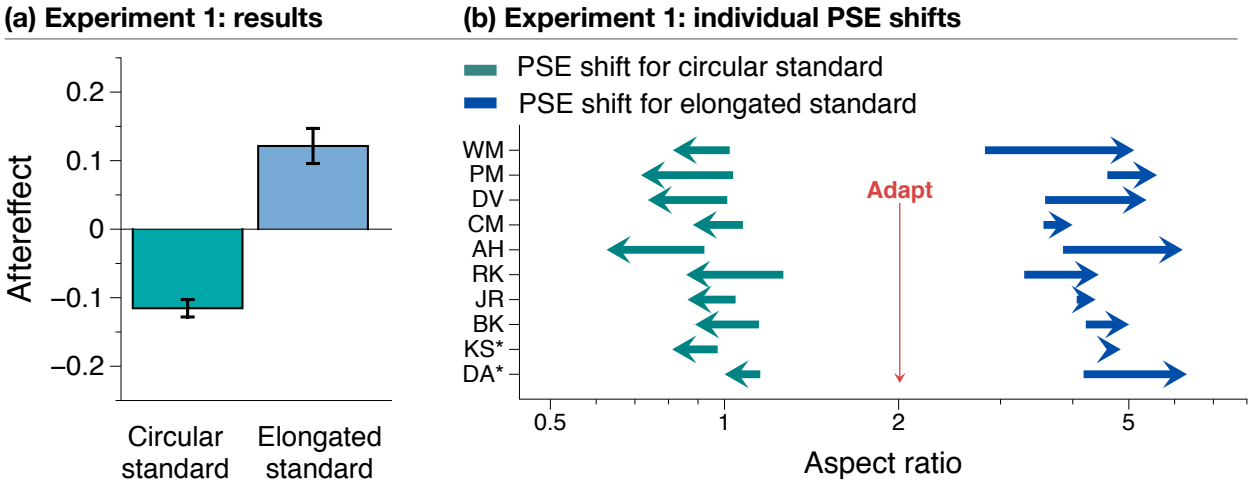


Figure 4. ***(a)*** *Mean proportional aftereffect on trials involving Circular (aspect ratio 1:1) and Elongated (aspect ratio 4:1) Standard test stimuli. Error bars indicate ±1 standard error of the mean (SEM).* ***(b)*** *The same data displayed as individual PSE shifts for each participant for each standard stimulus. The tail of each arrow indicates the baseline PSE estimated during trials involving Circular (green) or Elongated (blue) Standard tests, and arrowheads indicate the corresponding post-adaptation PSE estimate. The aspect ratio of the adapting stimulus is indicated by a red arrow. Note that the abscissa is in log units. Authors' data are indicated by asterisks.*

These results demonstrate that the aspect ratio aftereffect manifests predominantly as a contrast between successive shapes. If it is *entirely* mediated by contrastive mechanisms, two additional results are predicted: first, when the test shape is identical to the adapting shape, there should be no change in the appearance of the test. Alternatively, a contribution from recalibrative processes predicts that the adapted aspect ratio will appear closer to circular after adaptation ('renormalisation'). Experiment 2 tests this prediction. Second, in a contrastive aftereffect, adapting to a 1:1 aspect ratio should induce bi-directional aftereffects on test aspect ratios smaller or larger than 1:1. Recalibration predicts that adapting to a 1:1 shape should be uniquely ineffective in inducing aftereffects, since it is the norm for aspect ratio perception under this hypothesis. Experiment 3 tests this prediction.

4.4 EXPERIMENT 2: TESTING FOR RENORMALISATION OF THE ADAPTED SHAPE

4.4.1 *Method*

Details were as for Experiment 1, with the following exceptions.

4.4.1.1 *Participants*

Ten observers participated, comprising the first author, four experienced psychophysical observers naïve to hypotheses, and five inexperienced observers recruited from the MRC Cognition and Brain Sciences Unit volunteer panel, who were compensated with £9 for their time. Experiments 2 and 3 were approved by the Cambridge Psychology Research Ethics Committee.

4.4.1.2 *Stimuli and apparatus*

Stimuli were presented on a 17" Dell P791, set to a 1024x768 pixel resolution and a refresh rate of 75Hz. In runs of trials involving adaptation, the adapting stimulus had an aspect ratio of 2. Three standard test stimuli were used: a "Narrower Standard" with an aspect ratio of 1.5, an "Identical Standard" with aspect ratio 2, and a "Wider Standard" with aspect ratio 2.67. Stimuli were centred 6.8 dva above or below a central fixation cross, which subtended 0.7 dva. The area of adapting and test shapes was held constant at 28,800 pixels^2 — approximately 33.3 dva^2. The Narrower Standard therefore subtended approximately 7.1 x 4.7 dva, the Identical Standard / Adaptor subtended 8.2 x 4.1 dva, and the Wider Standard subtended 9.4 x 3.5 dva. Each pair of test stimuli was presented for 200ms.

4.4.1.3 *Procedure*

On each trial a standard test was pseudo-randomly selected from among the Narrower, Identical, and Wider aspect ratios. The variable test was selected pseudo-randomly according to a method of constant stimuli from one of seven aspect ratios, centred logarithmically about the respective standard stimulus. For trials involving Narrower Standards, variable tests were selected from aspect ratios of 1.06, 1.26, 1.40, 1.50, 1.61, 1.78, and 2.12. For trials involving Identical Standards tests were selected from aspect ratios of 1.41, 1.68, 1.87, 2.00, 2.14, 2.38, and 2.83, and for trials involving Wider Standards tests were selected from aspect ratios of 1.89, 2.42, 2.49, 2.67, 2.89, 3.17, and 3.77. Eight samples of each variable test aspect ratio were presented for each Standard test within a run of trials, yielding a total of 168 trials. On runs of trials involving adaptation, the adapting stimulus was presented above fixation for five participants and below fixation for the other five.

4.4.2 *Results*

Trials involving each of the three Standard test stimuli were analysed separately. After adapting to an aspect ratio of 2, Narrower Standard stimuli were matched to more contracted ellipses relative to baseline trials (proportional aftereffect -0.03 $\pm$0.01, t_9 = -3.64, p

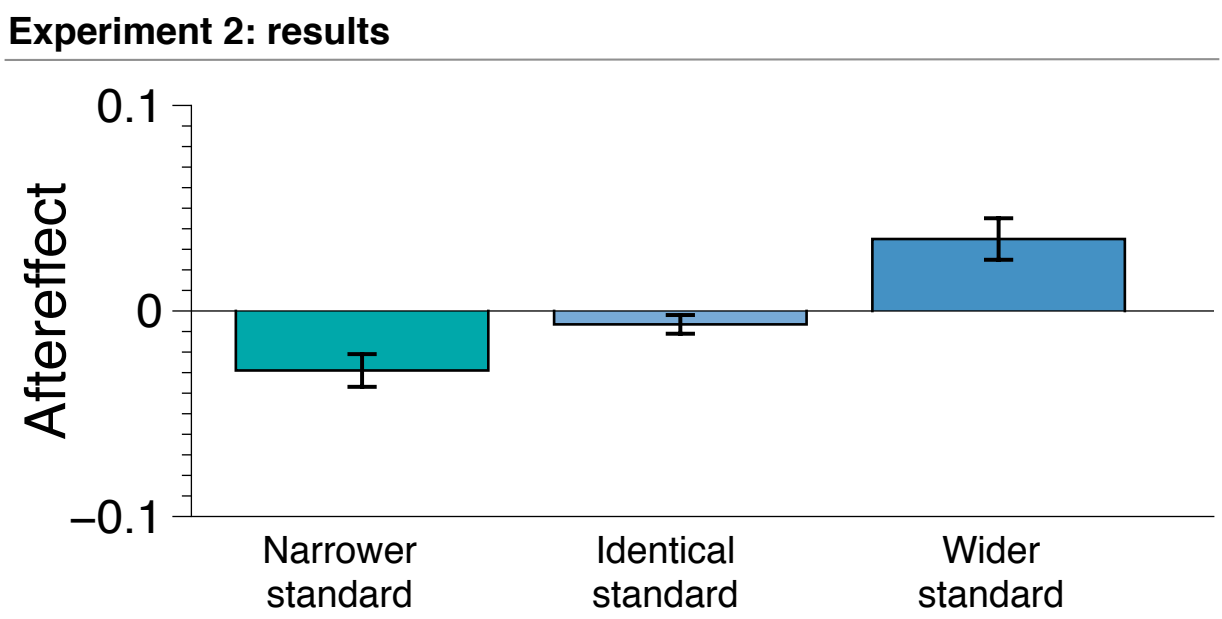


Figure 5. *Mean proportional aftereffects in Experiment 2, following adaptation to an ellipse with aspect ratio 2:1, for trials involving Standard test stimuli of a narrower (1.5:1), identical, or wider (2.67:1) aspect ratio. Error bars indicate ±1 standard error of the mean (SEM).*

= .005; see Figure 5). After adaptation, Wider Standards were matched to more elongated ellipses relative to baseline (0.04 ±0.01, t_9 = 3.47, p = .007). In the critical condition, in which standard stimuli had the same aspect ratio as the adaptor, adaptation had no significant effect on aspect ratio perception (-0.01 ±0.005, t_9 = -1.43, p = .19).

Results suggest that aspect ratio aftereffects can be characterised as contrastive, exaggerating differences between adapting and test aspect ratios without changing the appearance of the adapted aspect ratio. In Experiment 3, we test a final point of difference between the contrastive and recalibrative hypotheses: whether adaptation to the putative norm (a 1:1 aspect ratio) induces aftereffects. In addition, we introduce a size change between the adapting and test stimuli, to assess the possibility that observers are adapting to width or height alone, rather than aspect ratio.

4.5 EXPERIMENT 3: EFFECT OF ADAPTING TO A 1:1 ASPECT RATIO

4.5.1 *Method*

Details were as for Experiment 2, with the following exceptions.

4.5.1.1 *Participants*

Fifteen observers participated, comprising the first author, two experienced psychophysical observers naïve to hypotheses, and twelve inexperienced paid observers.

4.5.1.2 *Stimuli and apparatus*

Adapting shapes had an area of 14,400 pixels² — approximately 16.6 dva². Stimuli were centred 6.8 dva above or below a central fixation cross, which subtended 0.7 dva. In runs of trials involving adaptation, the adapting stimulus had an aspect ratio of 1 (i.e. circular), and subtended 4.1 x 4.1 dva. Four standard test stimuli were used: "small" and "large" narrower standards, with an aspect ratio of 0.8, and "small" and "large" wider standards, with an aspect ratio of 1.25. The square root of the area of "small" test stimuli was set to $\frac{3}{4}$ the square root area of the adaptor, and the square root area of "large" test stimuli was set to $\frac{4}{3}$ the square root area of the adaptor (see Figure 6a). The spatial location of the adapting stimulus jittered randomly within a circular region, such that, across a block of trials, its contours inscribed a circle subtending 8.7 x 8.7 dva. Standard stimuli subtended 2.7 x 3.4 dva (Small Narrower Standard), 3.4 x 2.7 dva (Small Wider Standard), 4.9 x 6.1 (Large Narrower Standard), or 6.1 x 4.9 dva (Large Wider Standard). Importantly, small test stimuli were smaller in both width and height than adapting stimuli, while large test stimuli were larger in both width and height than adaptors. Any aftereffect induced by adaptation to uni-dimensional width or height should therefore distort the aspect ratio of both Narrower and Wider standards in the same direction (within a size condition), contrary to the predictions of a locally-repulsive aspect ratio aftereffect. For example, if adapting to width, both narrower and wider Large Test stimuli should appear more elongated in aspect ratio, since both have a larger width than the adaptor. Adaptors were presented above fixation for eight participants and below fixation for the other seven.

4.5.1.3 *Procedure*

On each trial the variable test was presented with the same area as the Standard stimulus selected for that trial. For trials involving Narrower Standards (aspect ratio 0.8), variable tests were chosen from aspect ratios of 0.57, 0.67, 0.75, 0.80, 0.86, 0.95, or 1.13. For trials involving Wider Standards (aspect ratio 1.25) tests were chosen from aspect ratios of 0.88, 1.05, 1.17, 1.25, 1.34, 1.49, or 1.77. A run of trials consisted of 224 trials.

4.5.1.4 *Results*

Trials involving each of the four Standard test stimuli were analysed separately. For large test stimuli, after adapting to a circle, Narrower Standard stimuli were matched to more contracted ellipses relative to baseline trials (proportional aftereffect -0.02 $\pm$0.01, t_{14} = -2.38, p = .032; see Figure 5b), and Wider Standards were matched to more elongated ellipses relative to baseline (0.01 $\pm$0.01, t_{14} = 2.93, p = .011). For small test stimuli, Narrower Standards were matched to narrower ellipses than during baseline trials (-0.01 $\pm$0.01, t_{14} = -2.80, p = .014). Wider Standard stimuli were matched to slightly wider shapes than during baseline, but this difference was not significant (0.01 $\pm$0.01; t_{14} = 1.05, p = .313).

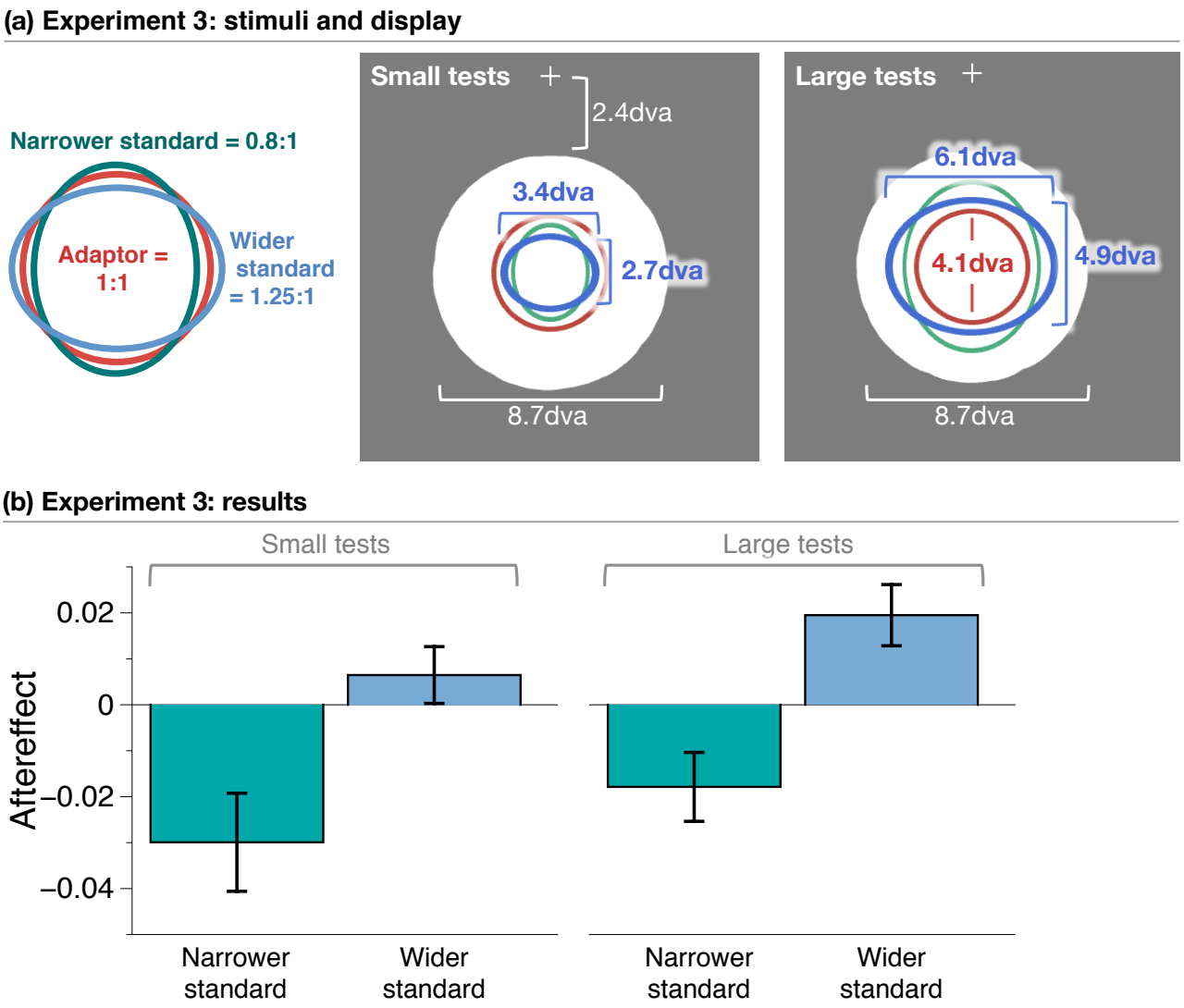


Figure 6. ***(a)*** *Left: aspect ratios of adapting and standard test shapes used in Experiment 2. Middle and Right: schematic illustrations of stimulus displays, showing the relative sizes of adapting and test stimuli in Small and Large test conditions in Experiment 2, respectively. The red circle indicates the size of the adaptor, while the white disc indicates the region within which it could appear. Outlines of the Narrower (green) and Wider (blue) Standard tests are also shown.* ***(b)*** *Mean proportional aftereffects in Experiment 2. Error bars indicate ±1 SEM.*

4.6 GENERAL DISCUSSION

Our data imply that aspect ratio aftereffects arise from a contrastive process, which perceptually exaggerates differences between adapting and test shapes. We are confident that these data reflect a perceptual effect, as we have used a signal detection task (a two-alternative forced-choice task, in which observers pick the interval containing non-identical tests) in order to minimise the potential impact of decisional response bias. We are also confident that these data cannot be accounted for by positional repulsion between contours, tilt adaptation to local edges, or by adaptation to a single shape dimension (width or height). We used stimuli rendered in dynamic white noise to minimise the contribution of processing in pre-cortical and some V1 mechanisms (Baker & Mareschal, 2001), which adapt well

to luminance-defined stimuli, but not dynamic white noise. Moreover, we minimised the influence of tightly-retinotopic visual processes by using a novel "jittering adaptor" display (inspired by Baker and Meese (2012)), in which the adaptor's external contours did not systematically overlap with those of the test stimuli. In Experiment 3 we also introduced a size change, so that both the width and height of test stimuli were smaller or larger than those of the adaptor. If observers adapted only to the width or height of the adaptor, both narrower and wider test stimuli should be distorted in the same direction, within a size condition. Instead, we found opposite aftereffects for narrower vs wider test shapes in the Large Test condition (aftereffects in the Small Test condition did not reach significance).

Our observations are inconsistent with a norm-based representation of aspect ratio, in which adaptation recalibrates the putative norm (Regan & Hamstra, 1992; Suzuki, 2003, 2005; Suzuki & Rivest, 1998). Norm-based aspect ratio coding predicts uni-directional recalibration after adapting to a non-normative aspect ratio, rather than the bi-directional contrastive aftereffects we observed in Experiment 1. Additionally, it predicts perceptual renormalisation of an adapted shape toward a neutral aspect ratio, which we did not observe in Experiment 2. Finally, norm-based aspect ratio encoding predicts no aftereffect following adaptation to a 1:1 aspect ratio (the putative norm), whereas we found in Experiment 3 that adapting to a circle exaggerates the aspect ratios of tests away from circular.

The present results suggest the existence of multiple aspect ratio channels, each tuned to a different aspect ratio, which signal shape via the distribution of activity across the population of channels (Badcock et al., 2014; Storrs & Arnold, 2012, 2015b; Webster & MacLeod, 2011). Similar encoding schemes underlie the perception of orientation and spatial frequency (Blakemore & Campbell, 1969; Clifford et al., 2001; Goris, Putzeys, Wagemans, & Wichmann, 2013; Pouget et al., 2000). Channels encoding aspect ratio are likely retinotopically localised, as we were able to induce and measure spatially-contingent aftereffects. However, the relevant channels must have broad spatial receptive fields, since a shape jittering over approximately 6 degrees of visual angle was an effective adaptor.

Because our method measured differences in perceived aspect ratio between an adapted an unadapted location, it would be insensitive to any retinotopically-global adaptation. It therefore remains possible that there exists a location-tolerate stage of aspect ratio that conforms to recalibration, although deconfounding such non-localised adaptation from decisional bias is problematic (see e.g. Morgan (2014)).

4.6.1 *Previous attempts to characterise the tuning of shape aftereffects*

Regan and Hamstra (1992) were the first to propose a norm-based encoding of aspect ratio, based on the observation that discrimination sensitivity is highest near a 1:1 aspect ratio (the putative norm). They presented a model comprising two channels oppositely tuned

to high and low aspect ratios. The responsiveness of both channels as a function of input aspect ratio changed most steeply around an aspect ratio of 1:1, and thereby predicted enhanced discrimination sensitivity about this value. This observation is, however, also compatible with a multiple-channel code with an uneven distribution of channels along the dimension of aspect ratio. Channels tuned to neutral aspect ratios might be more numerous, and/or more narrowly tuned, than those preferring more extreme aspect ratios. This would be analogous to the anisotropies found in orientation-tuned channels, that are thought to underlie our superior discrimination ability about vertical and horizontal relative to oblique orientations (Girshick et al., 2011; Li et al., 2003; Mannion, McDonald, & Clifford, 2010; Storrs & Arnold, 2015a).

Badcock et al. (2014) recently communicated via a conference abstract that they had measured spatially-contingent aspect ratio aftereffects, and also concluded that aspect ratio adaptation involves a local repulsion. They tested the influence of a range of adapting shapes on a test stimulus with a fixed aspect ratio of 2:1, and found a classic locally-repulsive aftereffect tuning with a minimum when adapting and test aspect ratios were identical. However, in these experiments stimuli were defined by first-order luminance information and the spatial overlap of adapting and test shapes was consistent. This left open the possibility that the data in question reflected adaptation to local contours rather than to shape aspect ratio. Our data show that locally-repulsive aftereffect tuning for shape aspect ratio persists when this possibility is eliminated.

The only other previous attempt to assess the tuning of aspect ratio aftereffects, communicated via an edited book chapter and a conference abstract (Suzuki, 2005; Suzuki & Rivest, 1998), found evidence of recalibration. However, those authors used extremely brief adaptation (150ms) and test (60ms) periods, making the results difficult to compare to those of the present experiments.

4.6.2 *What computations underlie global shape perception?*

The precise nature of computations underlying global shape perception are, as yet, unclear. It is, however, entirely possible that these will involve the activity of channels with retinally localised receptive fields that are attuned to different curvatures (Badcock et al., 2014). Related experiments have revealed aftereffects induced by the curvature of shapes which cannot be explained entirely by local tilt adaptation (Gheorghiu & Kingdom, 2007, 2008). These manifest as contrastive aftereffects, rather than as a recalibration of curvature perception (Gheorghiu & Kingdom, 2007, 2008), and they have been successfully modelled by adaptation within a multichannel code (Gheorghiu, Kingdom, Bell, & Gurnsey, 2011). Our results might be mediated by the influence that sequential adaptation of such channels has on computations underlying global shape perception. Alternatively, the locus of

adaptation might be more directly within channels that encode shape from broader retinal expanses. In either case, our data suggest that the product of shape adaptation is a contrastive aftereffect, rather than a renormalisation of shape perception.

4.6.3 *Adaptation may enhance the salience of novel stimuli*

Both multiple-channel and norm-based encoding schemes can predict improved discrimination sensitivity about adapted test values post adaptation, but these have not been found reliably. For instance, after adaptation to spatial patterns, Clifford et al. (2001), Regan and Beverley (1985) and Oruç and Barton (2011) reported enhanced performance in psychophysical spatial discrimination tasks, whereas Barlow, Macleod, and Van Meeteren (1976), Rhodes, Maloney, Turner, and Ewing (2007) and Westheimer and Gee (2002) reported no such advantage.

An intriguing recent suggestion is that perceptual distortions created by contrastive aftereffects might be behaviourally useful even in the *absence* of improved discrimination in a psychophysical task. According to this hypothesis, the perceptual distortions enhance the salience of novel inputs (McDermott, Malkoc, Mulligan, & Webster, 2010; Ranganath & Rainer, 2003). By perceptually exaggerating differences between the adapted environment and subsequent inputs, adaptation might ensure that scenes and objects that are statistically unusual in the context of the recent past capture attention. Adaptation might therefore rapidly and implicitly update our knowledge of which stimuli are typical in the current context (e.g. Kayaert, Op de Beeck, and Wagemans (2011)), and provide a host of behavioural advantages for atypical stimuli, such as faster and more accurate detection in cluttered environments (Kompaniez-Dunigan, Abbey, Boone, & Webster, 2015; McDermott et al., 2010; Wissig, Patterson, & Kohn, 2013). Our demonstration of a contrastive shape aftereffect is entirely consistent with this conjecture.

4.6.4 *Norms might not be the norm in spatial vision*

The theory that the brain represents spatial patterns relative to perceptual norms has gained traction over the past couple of decades (Freiwald et al., 2009; Kayaert et al., 2005; Leopold et al., 2006, 2001; Loffler et al., 2005; Panis et al., 2010; Rhodes et al., 2005; Webster & Maclin, 1999). Norm-based opponent-channel coding proposals have been particularly prevalent in the face perception literature (Giese & Leopold, 2005; Leopold et al., 2001; McKone et al., 2014; O'Neil et al., 2014; Pond et al., 2013; Rhodes et al., 2005; Susilo et al., 2010; Webster & MacLeod, 2011; Webster & Maclin, 1999). There is, however, mounting evidence questioning accounts of facial aftereffects that rely on norm-based encoding. Specifically, several studies have found that facial aftereffect patterns are not well-described

by opponent-channel based hypotheses, but can be better explained by multiple-channel models (Ross et al., 2013; Storrs, 2015a; Storrs & Arnold, 2012, 2015b; Zhao et al., 2011). The present data, and other recent results (Badcock et al., 2014), similarly suggest that a norm-based opponent-channel hypothesis does not well explain shape aspect ratio after-effects. Instead, our data suggest a continuity in spatial vision, such that adaptation to a moderately complex spatial property (aspect ratio) creates a classical contrastive aftereffect similar to those found for simpler spatial properties, such as spatial frequency (Blakemore & Sutton, 1969), orientation (Mitchell & Muir, 1976), and curvature (Gheorghiu & Kingdom, 2008). This complements findings suggesting that adaptation to still more complicated spatial properties, such as facial gender and identity (Ross et al., 2013; Storrs, 2015a; Storrs & Arnold, 2012, 2015b; Zhao et al., 2011) also produce contrastive aftereffects. It would seem, therefore, that local repulsion, rather than re-normalisation, might be the norm for aftereffects in spatial vision. In all contexts aftereffects might serve to enhance the salience of changes to the spatial properties of the environment, whether those properties signal changes in objects, scenes or faces.

5

CHAPTER 5: NOT ALL FACE AFTEREFFECTS ARE EQUAL

Having established that shape aftereffects are consistent with a multi-channel representation of aspect ratio in mid-level vision, I turn my attention in the last two chapters to the neural representation of a far more complex class of stimuli — faces. As reviewed in Chapter 1, several lines of evidence suggest that face aftereffects manifest as a renormalisation within a perceptual 'face space,' consistent with a norm-based representation of facial attributes (Burton et al., 2015; Jeffery et al., 2010, 2011; Leopold et al., 2001; Loffler et al., 2005; McKone et al., 2014; Pond et al., 2013; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins et al., 2007; Susilo et al., 2010; Webster & Maclin, 1999). However, as was also reviewed, there is some contention over the strength this evidence, and the discovery of more conclusive data has been hindered by a reliance on binary classification tasks.

I therefore wanted to devise a new experimental paradigm that could more decisively test whether face aftereffects manifest as a renormalisation or a local repulsion along particular facial attribute dimensions. In this chapter we show that a simple extension of the standard binary-classification task to a ternary-classification task, combined with adaptation to one of the category boundaries, generates qualitatively distinct predictions for the two proposals. The new paradigm is validated using colour saturation aftereffects, before being applied to facial distortion and facial gender aftereffects. The resulting manuscript appeared in *Vision Research* (Storrs & Arnold, 2012).

The introduction to this chapter necessarily covers some of the same material as the Introduction to the thesis in Chapter 1.

Not all face aftereffects are equal

5.1 ABSTRACT

After prolonged exposure to a female face, faces that had previously seemed androgynous are more likely to be judged as male. Similarly, after prolonged exposure to a face with expanded features, faces that had previously seemed normal are more likely to be judged as having contracted features. These facial aftereffects have both been attributed to the impact of adaptation upon a norm-based opponent code, akin to low-level analyses of colour. While a good deal of evidence is consistent with this, some recent data are contradictory, motivating a more rigorous test. In behaviourally matched tasks we compared the characteristics of aftereffects generated by adapting to colour, to expanded or contracted faces, and to male or female faces. In our experiments opponent coding predicted that the appearance of the adapting image should change and that adaptation should induce symmetric shifts of two category boundaries. This combination of predictions was firmly supported for colour adaptation, somewhat supported for facial distortion aftereffects, but not supported for facial gender aftereffects. Interestingly, the two face aftereffects we tested generated different patterns of response shifts. Our data suggest that superficially similar aftereffects can ensue from mechanisms that differ qualitatively, and therefore that not all high-level categorical face aftereffects can be attributed to a common coding strategy.

5.2 INTRODUCTION

Just as judgements regarding simple visual attributes like orientation (Gibson & Radner, 1937) or motion (Adams, 1834) can be changed by prior exposure to an 'adapting' stimulus, our experience of more complex images can depend on what we have recently seen. High-level categorical aftereffects were first reported for facial images that had been subjected to a configural distortion. Exposure to a face with contracted features, for instance, tended to make subsequent faces seem to have more expanded features (Webster & Maclin, 1999). A large literature has subsequently developed on face aftereffects, demonstrating that they not only impact apparent facial configuration, but also the apparent race (Webster et al.,

2004), age (Schweinberger et al., 2010), gender (Rhodes et al., 2004), emotional expression (Hsu & Young, 2010), gaze direction (Seyama & Nagayama, 2006), head direction (Fang & He, 2005), and identity (Leopold et al., 2001) of a face.

Face aftereffects have often been attributed to the impact of adaptation upon a two-channel norm-based opponent code (Anderson & Wilson, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006; Rhodes & Leopold, 2011; Rhodes et al., 2005; Susilo et al., 2010; Tsao & Freiwald, 2006). In a hypothetical two-channel opponent code, pools of neurons are arranged in complementary pairs, with each pool responding maximally to stimuli at one extreme of the encoded dimension, similar to the red-green and blue-yellow channels that exist at early stages of colour coding (Derrington, Krauskopf, & Lennie, 1984; Hurvich & Jameson, 1957). When both pools of opponent neurons are equally activated, the presence of a perceptual mid-point can be signalled. For opponent colour coding this may correspond to an 'achromatic' white or grey stimulus, which lies at the midpoints of both the blue-yellow and red-green axes (Hurvich & Jameson, 1957; Webster, 1996). After prolonged exposure to a stimulus coloured along one of these axes, one of the two opposing pools (the one that was more responsive to the stimulus initially) can become less responsive[1]. Consequently a more saturated colour along that axis will be needed induce an equivalent level of activation across the two pools of neurons. For example, after prolonged exposure to a red stimulus, a stimulus that had previously looked grey can seem to have a greenish hue, and one that had previously seemed reddish can seem grey.

It is thought that faces might be subject to a similar coding scheme, with facial appearance being signalled by ratios of activity across multiple opponent mechanisms (Giese & Leopold, 2005; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Webster & MacLeod, 2011). Adaptation to an individual face would disproportionately reduce activity in those channels that it most stimulated, causing faces that had previously seemed 'average' to take on an appearance 'opposite' to that of the adaptor (see Figure 1a for an illustration of adaptation in an opponent code).

An alternative strategy is to encode a stimulus property by the pattern of responses across a population consisting of multiple channels. In a multichannel coding scheme, each channel is tuned to a different stimulus value, as in the case of the relatively narrowly-tuned orientation-selective cells found in primary visual cortex (Hubel & Wiesel, 1962). When a given channel is active it can be thought of as 'voting' for the stimulus value for which it is tuned. Perception might be determined by the average value signalled by active channels, with the averaging process weighted by how active each channel is (Barlow & Hill, 1963; Georgopoulos, Schwartz, & Kettner, 1986; Pouget et al., 2000).

1 Note that the neural adaptation that results in reduced responsiveness need not occur at the level of the opponent mechanism, but rather within a preliminary cell. Colour adaptation occurs primarily within photoreceptor cells, before information is passed to an opponent coding mechanism (Webster & MacLeod, 2011).

In both coding schemes, it is proposed that adaptation reduces the responsiveness of a subset of channels (e.g. Webster (2011)). In a multichannel code, this effect is maximal for the channel(s) most responsive to the adapting stimulus, with progressively lesser effects for channels tuned to progressively dissimilar stimulus values. Perceptual aftereffects can then occur because the reduced responsiveness of channels tuned to values near the adaptor can bias the population average toward values signalled by now disproportionately active channels (see Figure 1b for an illustration of adaptation in a multichannel code).

5.2.1 *Evidence for two-channel norm-based opponent coding*

Since face aftereffects were first reported, much work has been done to evaluate which coding scheme best accounts for the observations. Substantial evidence in favour of a two-channel norm-based opponent scheme has been amassed. The first test of opponent coding predictions was conducted in the initial report, using faces with different configural distortions (Webster & Maclin, 1999). In an opponent code, unlike multichannel coding schemes, there is a unique stimulus value at the equilibrium point between the two opposing channels, adaptation to which should not induce any aftereffect (Webster, 2011; Webster & Leonard, 2008). Exposure to this stimulus value adapts both of the opposing channels equally, and so does not result in a perceptual shift, which only ensues when the *proportional* influence of the two channels for a given input is changed by adaptation. For example, in colour perception there exists a unique hue and saturation, corresponding roughly to a subjective "white" or "grey," to which adaptation creates no perceptual aftereffect (Webster, 1996; Webster & Leonard, 2008). Analogously, Webster and Maclin (1999) showed that although adapting to various strengths of facial distortion all yielded an aftereffect, no measurable perceptual changes ensued from adapting to a normal-looking undistorted face.

Another crucial prediction of norm-based opponent coding is that aftereffects will tend to bias appearance in a consistent direction along a stimulus dimension (given by a vector that points from the adaptor to the unadapted norm value in the perceptual space), rather than indiscriminately exaggerating any differences between successive stimuli (Benton & Burgess, 2008; Leopold et al., 2001; Rhodes et al., 2005). Consistent with this prediction, Leopold et al. (2001) showed that facial aftereffects were greater when tested along a dimension linking faces that could be described as computational opposites (in that they differed from the average face in geometrically opposite ways), relative to a dimension linking non-opposite faces. However, the non-opposite faces in that experiment were perceptually more similar than the opposite faces, so smaller aftereffects might simply have reflected a smaller contrast effect (Anderson & Wilson, 2005; Rhodes & Jeffery, 2006). Rhodes and Jeffery (2006) therefore performed a conceptually similar experiment, using test dimensions

matched in terms of perceptual similarity, and still found that testing along a dimension linking non-opposite pairs resulted in weaker aftereffects.

Further evidence for the opponent coding account of face aftereffects comes from changes in the perceived normality or attractiveness of adapting faces (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Robbins et al., 2007; Webster & Maclin, 1999). In an opponent code, adaptation shifts the equilibrium point (at which responses of the opposing channels are in balance) toward the adapting stimulus. Because a balanced response signals the presence of the perceptual mid-point along a given stimulus dimension, after prolonged exposure a distorted face should look more normal than it did previously. Likewise, a happy face should seem to have a more neutral expression, a male face should seem more androgynous, and so on. Evidence for this has been found after adapting to configural facial distortions, both for faces with expanded or contracted features (Maclin & Webster, 2001; Watson & Clifford, 2003; Webster & Maclin, 1999) and for faces with high- or low-set eyes (Robbins et al., 2007).

5.2.2 *Reasons to question the norm-based opponent coding account of face aftereffects*

Although there are several observations that support an opponent coding explanation for high-level categorical face aftereffects, there are also some observations that encourage further investigation.

First, while a norm-based opponent coding scheme is thought to provide a good account for facial identity aftereffects, including facial gender aftereffects, it is not considered a general explanation for all aftereffects that involve faces (Susilo et al., 2010). Aftereffects induced by the direction of eye gaze (Seyama & Nagayama, 2006) and head viewpoint (Fang & He, 2005), for instance, seem to be better accounted for by a multichannel coding scheme Calder et al., 2008; Lawson et al., 2011.

Recent measurements of aftereffect magnitude as a function of adaptor value have been interpreted as inconsistent with opponent facial coding (Zhao et al., 2011). In an opponent code, provided that adapting and test stimuli lie along and within the confines of an encoded dimension, aftereffect magnitudes will tend to scale with the difference between the adapting image and the centre of the encoded dimension, and should not become mitigated for extreme differences (Jeffery et al., 2010, 2011; Robbins et al., 2007; Susilo et al., 2010; Zhao et al., 2011).[2] Multichannel coding schemes, however, predict that aftereffect magnitudes should increase over a certain range of distances between adaptor and test, and then decrease for larger distances. Facial configuration aftereffects have been reported to increase monotonically for increasingly extreme adaptors (Robbins et al., 2007; Susilo

2 Note, however, that this suggestion is contingent on the two opponent channels either being maximally responsive to inputs lying at the extremities of an encoded dimension, or having saturating responses that do not diminish for extreme values.

et al., 2010), consistent with the predictions of an opponent code. More recently, however, it has been reported that facial gender aftereffects increase in magnitude up to a certain point, and then decrease, consistent with the predictions of a multichannel code (Zhao et al., 2011).

Finally, it is worth noting the recent proliferation of studies reporting high-level categorical aftereffects for diverse non-face stimuli. Categorical aftereffects have been reported for images depicting more and less cluttered landscapes (Greene & Oliva, 2010), for test arrays linking images of disparate objects (such as a mushroom and a lightbulb, see Daelli et al. (2010)), and for test arrays linking images of unnatural three-dimensional shapes (Daelli, 2011). While each of these aftereffects has been attributed to the influence of a norm-based opponent coding scheme (Daelli, 2011; Daelli et al., 2010; Greene & Oliva, 2010), the fact that such a diversity of aftereffects exists, coupled with the observation that not all face aftereffects are well accounted for by an opponent coding scheme Calder et al., 2008; Lawson et al., 2011; Zhao et al., 2011, suggests that there could be a diversity of causal mechanisms underlying high-level categorical aftereffects.

5.2.3 *A discriminant test for two-channel opponent vs multi-channel coding*

Here we will propose a new protocol to test if a given categorical aftereffect primarily reflects renormalisation within a two-channel opponent coding scheme, or a contrastive aftereffect arising from a multichannel code. Schematic models of an opponent code and a multichannel code are shown in Figure 1. The top left panel (a) depicts two unadapted opponent channels, intersecting at the subjective neutral stimulus value (here a value of zero). The top right panel (b) depicts an unadapted multichannel code consisting of multiple equally responsive channels each tuned to a different stimulus value. In both coding schemes adaptation reduces channel responsiveness, as shown in the middle panels. Here we have depicted the response profiles of (a) an opponent and (b) a multichannel code before (pale dashed lines) and after (bold solid lines) adapting to a stimulus value slightly offset from neutral (an arbitrary value of +20 units). Response profiles represent the magnitude of expected channel responses (e.g. average firing rate) upon exposure to different stimulus values. The reduction in responsiveness post-adaptation has been calculated as a divisive change in response amplitude, proportional to the initial response of each channel to the adapting stimulus value.

There are several types of perceptual judgement an observer can make about a stimulus. For instance, an observer can be asked to perform a binary classification task, such as judging if a face looks male or female. This example could involve a single criterion separating two categories — the point at which facial images look androgynous. Note, however, that a criterion need not be placed in the middle of a stimulus range, nor be restricted to dividing

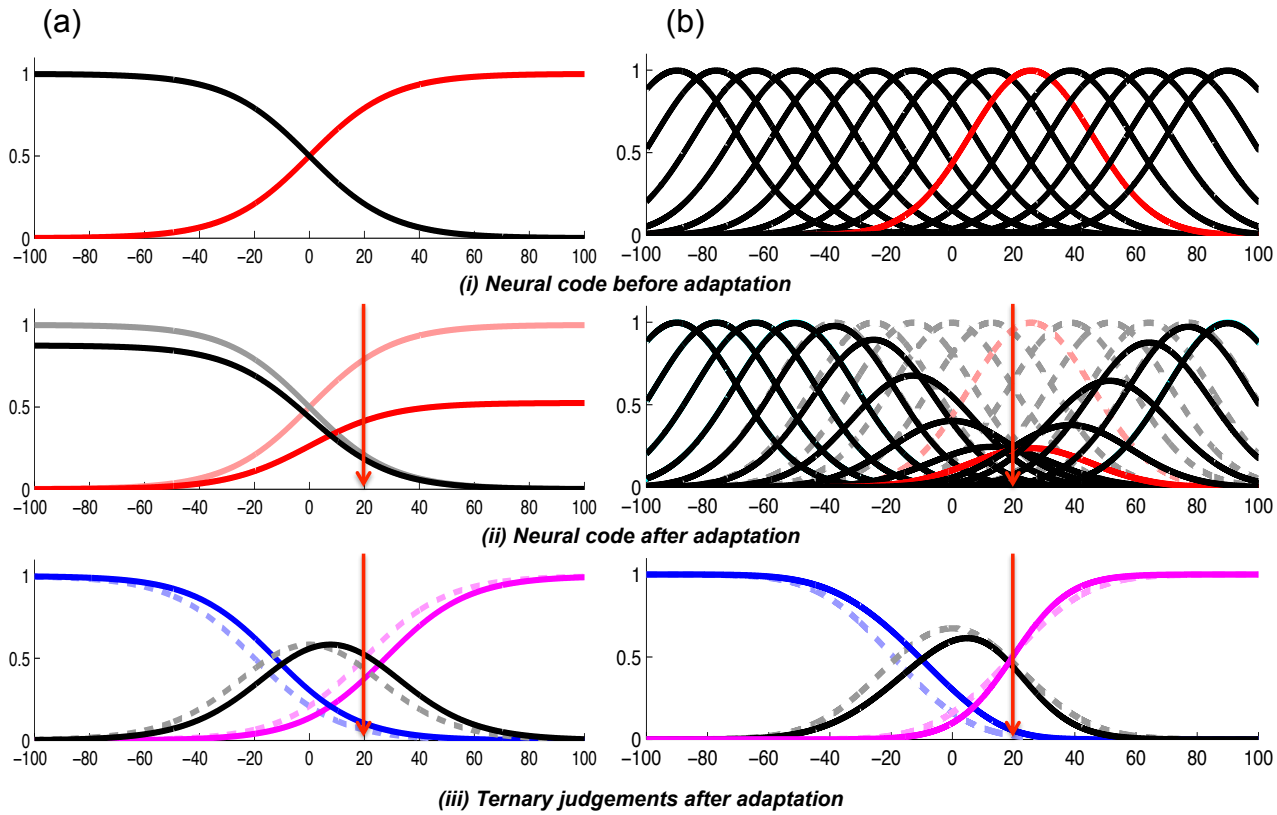


Figure 1. *Graphic depiction of adaptation in* ***(a)*** *an opponent code, and* ***(b)*** *a multichannel code. The top row (i) shows channel activity to different stimulus values before adaptation. The dark curves in the second row (ii) show channel activity after adapting to a stimulus value of +20 (units are arbitrary), as indicated by the red arrow. The faded, dashed curves show unadapted responses for comparison. The bottom row shows perceptual judgements predicted by the neural codes above for a three-category classification task, assuming a fixed set of decisional criteria. The dark curves in row (iii) show the predicted likelihood of classifying each stimulus value as belonging to the lower category (blue), as being neutral (black), or as belonging to the upper category (magenta) after adapting to the stimulus value indicated by the red arrow. The faded, dashed curves represent the same probabilities before adaptation. Note first that the upper and lower category boundaries (i.e. their intersections with the middle curve) shift equally and symmetrically in an opponent code but not a multichannel code. Note second that the classification of the adapting stimulus itself (at the site of the red arrow) changes in an opponent but not a multichannel code.*

the continuum into only two categories. In a ternary classification task, an observer could decide if a facial image appears female, androgynous, or male. These judgments reflect at least two criteria (see Allan (1975) and Ulrich (1987)): one for deciding that faces are sufficiently feminine to be classified "female," and another for deciding that faces are sufficiently masculine to be classified "male". "Androgynous" judgments would indicate that neither criterion had been met. While binary and ternary classification tasks differ in terms of the number of categories participants are allowed to use, the two tasks may tap the same number of subjective criteria. In a binary classification task, for instance, participants might perform the task by deciding if a given face is sufficiently female-looking to be judged as female, sufficiently male-looking to be judged as male, and if neither criterion is met (such as when a face looks perfectly androgynous), they might simply guess at random.

A binary classification task limits the experimenter to measuring shifts in a single category boundary, centred on a neutral putative 'norm' value (e.g. androgyny), while a ternary tasks allows one to measure shifts in two category boundaries, neither centred on the putative norm. Crucially, a norm-based opponent code predicts that adaptation to any non-normative value will cause that stimulus to appear more neutral after adaptation, while a multichannel code predicts that, for any adaptor value, the appearance of the adaptor itself will not be changed during adaptation. To use orientation as an example: if an observer adapts to a line of an arbitrary orientation and subsequently matches standard lines to tests presented in an unadapted retinal location, a 'repulsive' tilt aftereffect will be observed for standards to either side of the adaptor, but there will be no change in the apparent orientation of standards matched to the adaptor's orientation (Mitchell & Muir, 1976). This is not the case in an opponent code, where there is only one stimulus value along a perceptual dimension which will remain unchanged after adaptation. For example any adapted colour, other than grey or white, will tend to look less saturated after adaptation (e.g. Webster and Mollon (1995)). Here, using ternary classification tasks, we will develop a behaviourally matched protocol to test whether facial configuration, facial gender and coloured aftereffects reflect similar computational processes.

5.2.4 *Model predictions for ternary classifications*

The bottom left panel in Figure 1 shows response probabilities for a range of inputs in a ternary classification task, predicted from activity in an opponent code shown in the panel above. The placement of the boundaries between the three categories will depend on criteria. If the observer is unwilling to classify a colour as "colourless," or a face as "androgynous," there would be no separation of the upper and lower category boundaries. Any willingness to classify inputs in the mid category will result in a lateral separation of the lower and upper category boundaries. We have modelled this for opponent coding by differentially weighting the upper and lower opponent channels when calculating response probability curves, to reflect a critical ratio of activity which must be exceeded before an upper or lower category response is made. This results in a lateral shift of the response probability curves, here by ± 20 (arbitrary units). We can then estimate a distribution for mid-category classifications by taking the summed probability of lower and upper category classifications at each stimulus value from 1 (solid black / dashed grey lines). The lower and upper category boundaries are given by the intersection of the mid-category functions and the lower (blue) and upper (pink) category classification functions. Note that this scheme predicts a *symmetric* rightward shift of the lower and upper category boundaries post-adaptation (solid bold lines compared to dashed pale lines), and a matched shift of the perceptual midpoint (here given by the peak of the black/grey mid-category function).

Note that because the adapting value is not at the equilibrium point, the appearance of the adaptor itself will also change. After adaptation the adaptor is more likely to be placed in the middle category than it was prior to adaptation.

We have also depicted response probabilities for a ternary classification task based on a multichannel code (Figure 1, bottom right). Here the probability of making an upper category response at each stimulus value is calculated as the total sum of activity in channels tuned for stimulus values *greater than* +20, divided by activity summed across all active channels (pink lines). The probability of a lower category response is given by the summed response of active channels tuned to values *less than* -20, again divided by activity summed across all active channels (blue lines). Mid-category responses are predicted by taking the summed responses of the lower and upper category functions at each stimulus value from 1 (black/grey lines) and category boundaries are given by the intersection of this and the other category functions. Note that this scheme also predicts a rightward shift of the perceptual midpoint post-adaptation, but whereas the lower category boundary is similarly shifted, the upper category boundary is unaffected. This is because the upper category boundary lies at the adapted value, and within a multichannel code the appearance of the adaptor itself is unchanged (Mitchell & Muir, 1976; Webster, 2011).

We have thus derived qualitatively different predictions for how ternary classifications should be affected by adaptation based on a two-channel opponent code, compared to a multichannel coding scheme. When adapting to a value corresponding to an upper category boundary, adaptation within an opponent code should result in (a) altered adaptor appearance, and (b) symmetrical shifts of the upper and lower category boundaries. In contrast, adaptation within a multichannel code should result in (a) unchanged adaptor appearance, and (b) asymmetrical upper and lower categorical boundary shifts (with upper category boundary shifts being minimal or non-existent). These predictions, and the conceptual models upon which they are based, are in line with existing accounts of high-level categorical aftereffects (e.g. (Calder et al., 2008; Lawson et al., 2011; Rhodes et al., 2005; Robbins et al., 2007; Susilo et al., 2010; Webster & MacLeod, 2011). Different predictions could be derived given different assumptions regarding the number, distribution and / or structure of underlying channels, so the following experiments should be regarded as a test of the validity of these specific proposals.

To test our predictions we had participants make ternary classifications for colour, facial distortion, and facial gender. Participants judged stimuli as either belonging to a lower category (cyan hue / contracted face / female face), to an upper category (magenta hue / expanded face / male face), or as belonging to a mid-category (grey / undistorted face / androgynous face). To ensure these conditions were behaviourally matched our protocol involved a crucial initial procedure to identify unadapted category boundaries and perceptual midpoints. We then conducted targeted baselines, during which participants adapted to the image at their perceptual midpoint, and finally had people adapt to an image near

their upper category boundary. We measured the impact this had on (a) the appearance of the adaptor, and (b) the placement of their two category boundaries.

5.3 METHOD

5.3.1 *Participants*

There were six participants in each condition, comprising two of the authors and four experienced psychophysical observers who were naïve to the hypotheses of the experiment. All participants conducted one run of baseline trials and one run of adaptation trials per condition, except for the authors, who conducted two runs of baseline trials and two runs of adaptation trials.

5.3.2 *Stimuli*

Stimuli were presented on a gamma-corrected 21" Samsung SyncMaster 1100p+ or on a 19" Sony Trinitron Multiscan G420 monitor (resolution 1024 x 768 pixels and a refresh rate of 120Hz). Matlab software was used to drive ViSaGe stimulus generators from Cambridge Research Systems. The white points for colour space calculations were set at CIE chromaticity coordinates of x = 0.28, y = 0.29, Y = 60 cd/m² and at x = 0.29, y = 0.34, Y = 66 cd/m² respectively for the two monitors.

Example stimuli are shown in Figure 2a. In the colour condition we generated square coloured patches of 100 different hues ranging from cyan, through grey, to magenta, from which we selected a subset of 7 test stimuli (see below for details). For the facial gender condition, we used Abrosoft FantaMorph software to create an array of 200 images, morphing between a male and a female face. For the facial configuration condition, we applied the spherize distortion filter in Adobe Photoshop (following Rhodes et al. (2003)) to a neutral-expression female face to generate an array of 80 images, each with a different filter level between -40 and +40. In all conditions the most extreme points in the test ranges were assigned values of -100% (cyan, expanded face, female face) or +100% (magenta, pinched face, male face). These values corresponded to a position within each array, such that a value of 0% demarked the mid-point between the two categories (a grey patch, an undistorted face, an androgynous face).

Except for the colour patches, all images were greyscale and equated for luminance using the *SHINE* toolbox for Matlab (Willenbockel et al., 2010). Stimuli were presented with a mean luminance of 66cd/m² and shown against a black background. In order to reduce the contribution of low-level aftereffects, test images were presented at a different size to the adapting images. Adapting images subtended 2.94 degrees of visual angle (dva), and

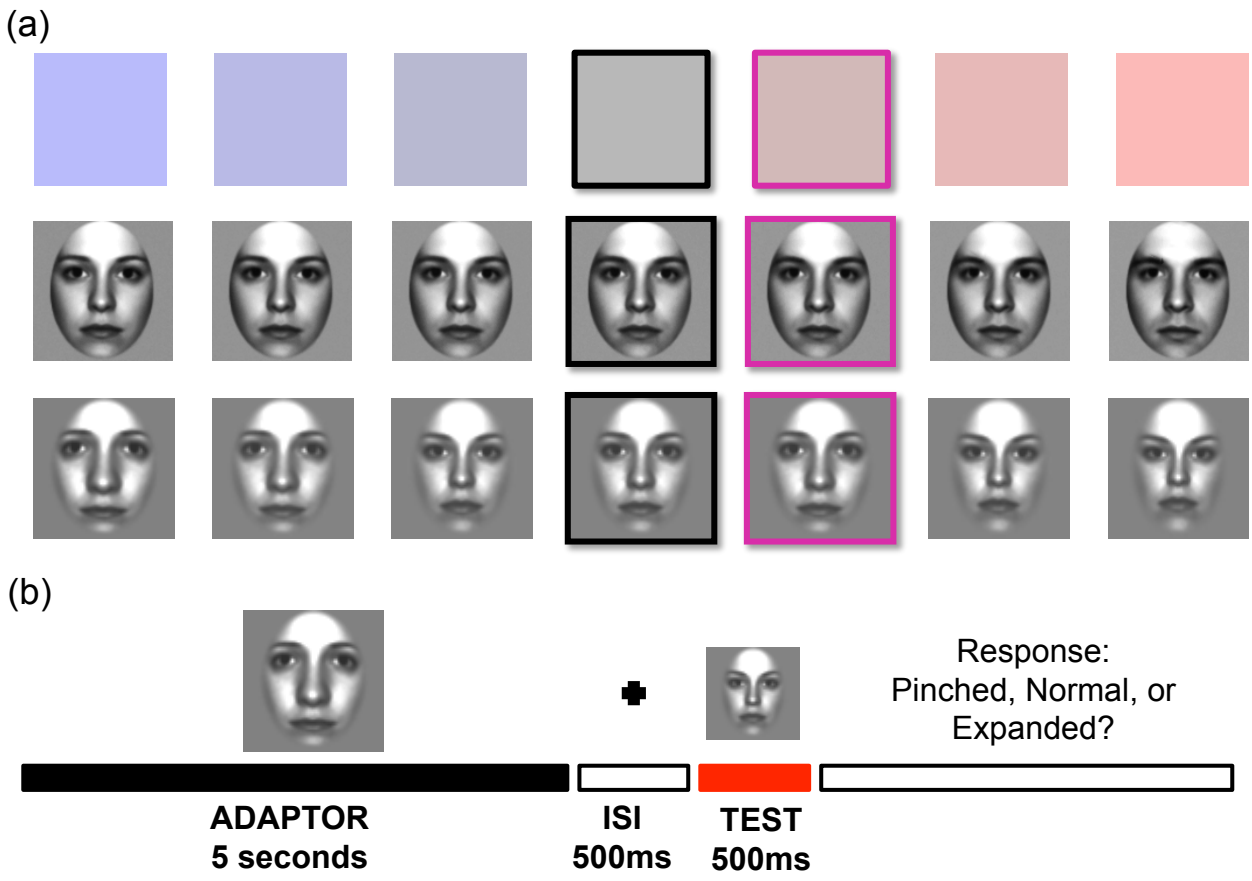


Figure 2. *Stimuli and trial sequence. From top to bottom,* ***(a)*** *shows an example of a possible test range in the colour, facial gender, and face configuration conditions. Images with black borders indicate those closest to the peak of a Gaussian fitted to "neutral" responses in each image range, during a pre-test calibration run of trials (see main text for details). These then served as adaptors during baseline runs of trials. Images with magenta borders indicate those close to the upper category boundary, identified during the pre-test calibration. These served as adaptors during the adaptation runs of trials.* ***(b)*** *Graphic depiction of a trial sequence.*

test images subtended 2.1 dva, except for colour patches, which were all presented at the smaller size. Participants viewed the images from a distance of 114cm with their head restrained by a chinrest.

5.3.3 *Pre-test procedure*

Critical to our protocol was the use of ranges of test stimuli that were equated for each participant, on each stimulus dimension, in terms of their subjective discriminability. Each participant therefore conducted a pre-test calibration run of trials to estimate their subjective neutral point and category boundaries within each stimulus range. These values were used to create a tailored range of test images for that participant, along that stimulus dimension, in the following manner.

For the colour condition, we first determined each participant's equiluminant LM (red-green) colour axis using a minimum motion technique (Anstis & Cavanagh, 1983). Based

on this procedure, 9 colour levels spanning -50% to +50% of the range between the participant's subjectively isoluminant cyan and magenta values were selected as test stimuli for the pre-test run of trials, and sampled 8 times each. In the face image conditions, 21 stimulus levels spanning -100% to +100% of the image ranges were sampled 6 times each. In each trial, a pseudo-randomly selected image was displayed for 500ms, after which it was replaced by a fixation cross until the participant responded, at which point the next trial began. Participants were instructed to classify each stimulus as belonging to either the upper category, the lower category, or the mid category. For example, in the colour condition participants were asked to press the left mouse button if the square looked cyan, the right if it looked magenta, and the middle button if it looked grey or white.
A Gaussian function was then fitted to the participant's "neutral" responses and, based on the mean and spread of this function, seven images were selected from the full image range to be used during subsequent experimental runs of trials. Specifically, the seven test images were determined as those most closely matched to values for the mean, the mean $\pm$0.4 x the full-width-at-half-height (FWHH) of the fitted function, the mean $\pm$2.4 x FWHH, and the mean $\pm$4.4 x FWHH. Thus, although we did not attempt to physically equate the varied range of test images across different experimental conditions, stimuli were carefully equated in terms of their subjective discriminability.

5.3.4 *Test procedure*

The test procedure consisted of a baseline run of trials, during which participants adapted to the middle image in their 7-image test range (their subjective neutral point), and an adaptation run of trials during which they adapted to the fifth image in their range (mean + 0.4 x FWHH). Each trial commenced with an adapting image presentation, for 30 seconds on the first trial and for 5 seconds on subsequent trials. Participants were instructed to fixate the centre of adapting images throughout their presentation. After the adaptation period there was a 500ms inter-stimulus-interval, during which a central fixation point was presented, followed by a test image presentation of 500ms. Participants then completed a ternary classification task. During a run of trials each of 7 test images was presented 20 times in a pseudo-random order according to the method of constant stimuli, yielding a total of 140 individual trials.
Each run of trials provided a distribution of apparent image identity as a function of test image value. Logistic functions were fitted to "lower" and "upper" category responses using *psignifit* toolbox version 2.5.6 for Matlab (Wichmann & Hill, 2001), and 50% points were taken as estimates of the two category boundaries, at which an image was equally likely to be classified as belonging to that category, or to the mid category. A point of subjective equality (PSE), which can be regarded as an estimate of the perceptual midpoint

for the tested dimension, was estimated by taking the midpoint between the two category boundaries. Aftereffect magnitudes were calculated by taking the difference between PSE estimates from targeted baseline and adaptation runs of trials. Raw response data for each participant in each condition are shown in Figure 3, along with fitted logistic and Gaussian functions.

5.4 RESULTS

5.4.1 *Aftereffects*

As depicted in Figure 4, each of our experimental conditions generated a perceptual aftereffect, as indicated by differences in PSEs from targeted baseline and adaptation runs of trials. In all cases average PSE values after adapting to a magenta coloured patch, to a pinched face, or to a male face were shifted toward the adapted value relative to baseline (Colour paired t_5 = -7.00, p = .001; Facial distortion t_5 = -5.27, p = .003; Facial gender t_5 = -3.11, p = .026).

5.4.2 *Appearance of the adaptor*

Grey bars in Figure 5 show the change in the proportion of trials that the adapting image was judged as being in the mid-category after adaptation, relative to baseline. In the colour condition the adaptor was rated as neutral more often after adaptation than before (mean proportion of "grey" responses pre-adaptation = 0.23, ±SEM 0.10; post-adaptation = 0.68 ±0.09, paired t_5 = -4.56, p = .006). This was also the case in the configural facial distortion condition (mean proportion of "undistorted" responses pre-adaptation = 0.40, ±0.12; post-adaptation = 0.69 ±0.11, t_5 = -9.10, $p < .0001$). However, facial gender adaptation did not result in the adaptor being rated as androgynous more often relative to baseline (mean proportion of "androgynous" responses pre-adaptation = 0.16 ±0.07; post-adaptation = 0.13 ±0.05, t_5 = 1.89, p = .118).

5.4.3 *Category boundary shifts*

Coloured bars in Figure 5 show shift estimates for lower (blue) and upper (magenta) category boundaries. For colour, shifts in the upper and lower category boundaries were matched in magnitude (repeated measures t_5 = -0.58, p = .586). This was not the case for either of the face conditions. For configural face adaptation the shift of the lower "expanded" face category boundary was smaller than the shift of the upper "pinched" category boundary (repeated measures t_5 = -2.62, p = .047). In contrast, for facial gender adaptation, the

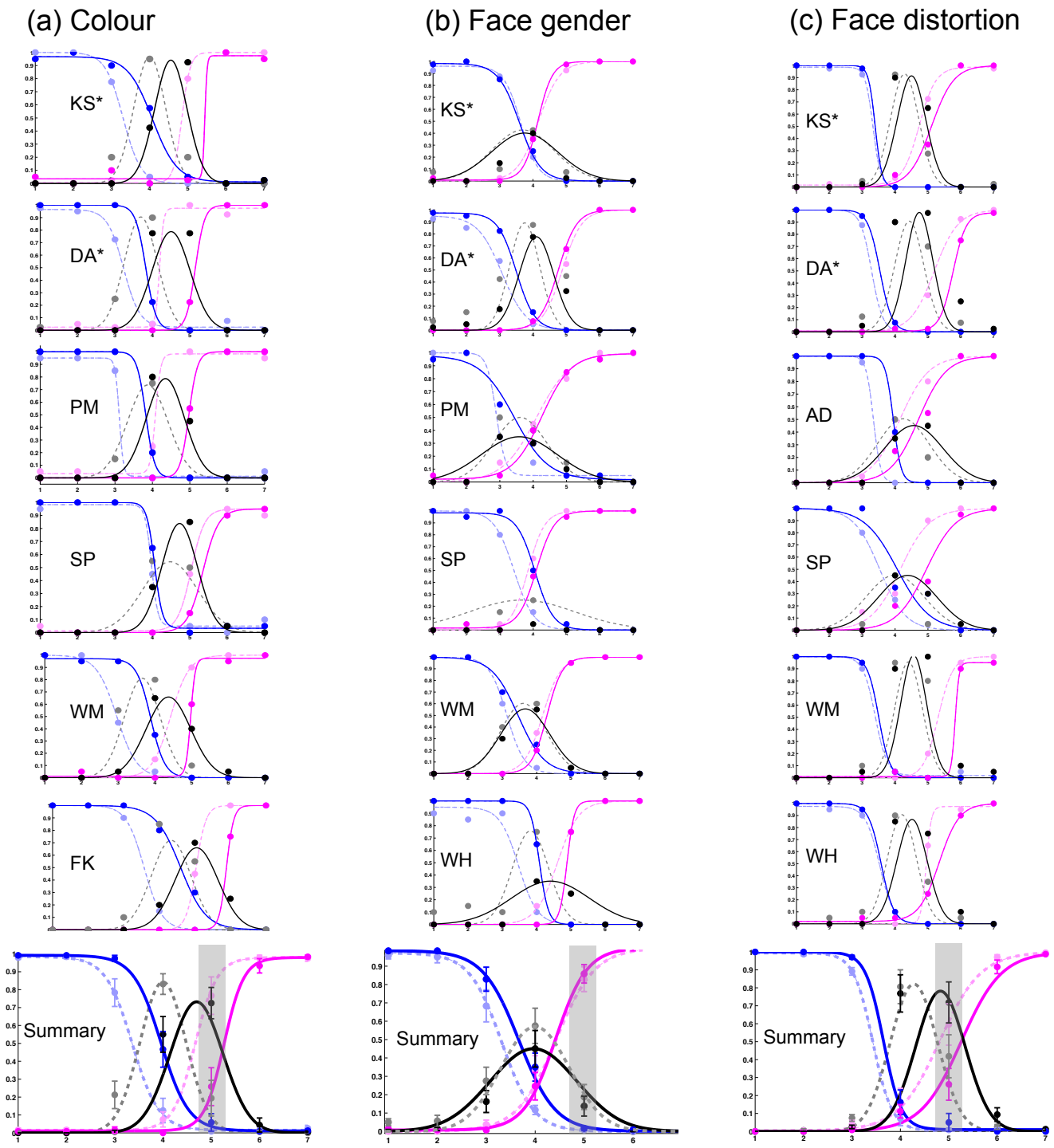


Figure 3. *Logistic and Gaussian curves fitted to the lower category (blue), neutral category (black) and upper category (magenta) responses for each participant in the* ***(a)*** *colour,* ***(b)*** *facial gender, and* ***(c)*** *facial distortion adaptation conditions. Dashed lines indicate responses before adaptation, and solid lines indicate responses after adapting to stimulus 5 in the 7-stimulus test range. Authors' data are indicated by an asterisk. Summary plots at the bottom of each panel show aggregate responses, with error bars showing ±1 SEM between individual responses, and grey columns indicating the position of the adaptor.*

shift of the lower "female" category boundary was greater than the shift of the upper "male" category boundary (repeated measures $t_5 = 3.24$, $p = .023$). Thus while category boundary

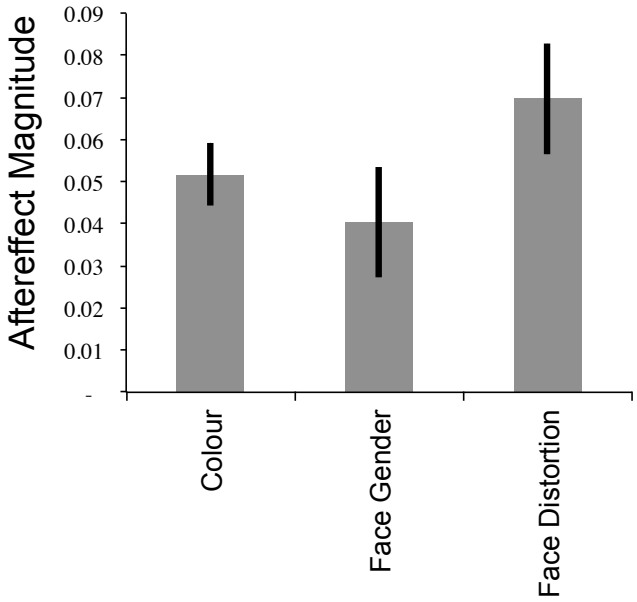


Figure 4. *Bar graph showing aftereffects in each condition. Aftereffect magnitudes are given by taking the shift in PSE between baseline and adaptation conditions as a proportion of the total image range. Error bars show ±1 SEM between individual PSE shift estimates in each condition.*

shifts after colour adaptation were symmetrical, boundary shifts after face adaptation were asymmetric.

5.5 DISCUSSION

We have used behaviourally matched tasks to compare perceptual aftereffects resulting from colour adaptation, from exposure to expanded or contracted faces, and from exposure to male or female faces. In our protocol adaptation within a two-channel opponent coding scheme predicted that the adapting image should appear more neutral after adaptation, and that shifts of the lower and upper category boundaries should be equal. This combination of predictions was confirmed for colour adaptation, and was broadly consistent with the impact of adaptation on distorted facial images. However, results for facial gender adaptation were qualitatively different, in that there was no change in the appearance of the adaptor, and shifts of the two category boundaries were asymmetric. In our protocol, this pattern of results matched the predictions for an aftereffect mediated via a multichannel coding strategy (see Figure 1b for a graphic depiction).

Results from adapting to distorted facial images were broadly consistent with opponent coding predictions, in that the adapting image was more often rated as 'normal' after adaptation, and although shifts of the upper and lower category boundaries were asymmetric, both shifted in the direction predicted by opponent coding. It is interesting to note that the 'normalisation' effect evident in these data, (i.e. a change in the appearance of the adapted

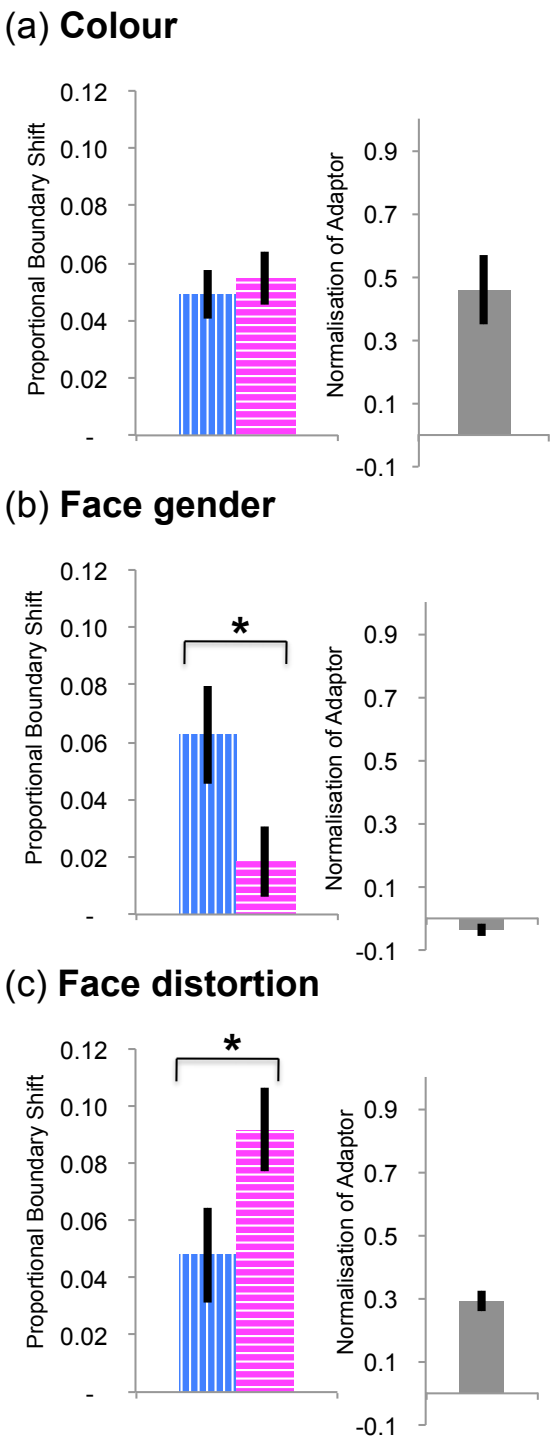


Figure 5. *Three panels showing the proportional shift in lower (blue vertically striped bars) and upper (magenta horizontally striped bars) category boundaries, as well as the change in the proportion of adaptor images which were classified as 'neutral' after adaptation, relative to during baseline (grey solid bars), for* ***(a)*** *colour,* ***(b)*** *facial gender,* ***(c)*** *and facial configuration. Error bars show ±1 SEM between individual measurements.*

image) has only previously been reported after adaptation to configurally distorted facial images (Rhodes et al., 2003; Robbins et al., 2007; Webster & Maclin, 1999). Our data thus replicate these previous findings, but suggest that the same normalisation effect does not happen after adaptation to facial gender.

One potential concern is that the diversity of results we have obtained was caused not by a diversity of underlying mechanisms, but by a failure to adequately match stimuli either in terms of their subjective discriminability, or in terms of adaptation magnitude. An inspection of individual and summary data plots in Figures 3 and 4 reveals that our stimuli were in fact well matched in terms of discriminability. For instance, in all three conditions the adaptor (stimulus 5, highlighted with a grey shaded column) was somewhat ambiguous during baseline trials, ensuring that our protocol was sensitive to even slight changes in the adaptor's appearance and avoided ceiling effects. Other test values along baseline discriminant functions were similarly well matched across different stimulus conditions.
Our protocol does not equate the magnitudes of adaptation that might occur in different contexts, but this does not pose a problem for our paradigm. The strength of our protocol is that it predicts different *patterns* of aftereffect, rather than different *magnitudes*. Note that in all conditions a robust aftereffect was generated by the adapting stimulus (see Figure 4). In two conditions this was accompanied by shifts in each of the two category boundaries, as predicted by our opponent coding model. For colour these shifts were symmetric, precisely as predicted by opponent coding, whereas for facial configuration they were asymmetric (see following paragraph for further discussion of this result). For facial gender there was a shift of the lower category boundary, but not of the upper — just as predicted by our multichannel code. Larger aftereffect magnitudes should only exaggerate these qualitative differences, whereas smaller aftereffects could eliminate them. This does not appear to be a problem in our experiments, as measurable aftereffects occurred in all conditions, indexed by a shift in the stimulus that appeared most neutral in all three conditions. One could argue that all aftereffects were in fact driven by an opponent code, and that any asymmetry was due to measurement error. We think this is unlikely, due to the marked asymmetry between the lower and upper category boundary shifts for facial gender (see Figure 5b).

5.5.1 *Adaptation of a sensory code, or decisional criterion shifts?*

The predictions we derived for a two-channel opponent coding strategy, and for a multichannel code, assume a direct mapping between a sensory code and perceptual decisions (see Figure 1, also see Calder et al. (2008), Leopold et al. (2001), Rhodes et al. (2005), Susilo et al. (2010)). Perceptual aftereffects predicted on this basis are presumed to have been caused by changes to a sensory code (Seriès et al., 2009). However, subjectively similar perceptual experiences, and consequent datasets, could ensue from an unchanged sensory code if decisional criteria were modified instead.
In order to classify an image, an observer must apply decisional criteria. For example, in our facial distortion condition participants had to apply criteria to decide wither facial images should be classified as expanded, contracted, or undistorted. If the sensory evidence

exceeded the "expanded" criterion threshold, the stimulus would be classified as expanded. Alternatively, if sensory evidence exceeded the "contracted" threshold, the opposite classification would be made. The placement of these two criteria need not be bound; one could be altered independent of the other. Yarrow et al. (2011) recently made a similar point with respect to timing judgments, suggesting that repeated exposure to particular temporal relationships might independently alter the placement of multiple threshold criteria. Thus even if the sensory information available for decision making is unchanged, a perceptual aftereffect could ensue from either selectively, or disproportionately, changing just one of two decision-making criteria (Yarrow et al., 2011). In this case, after prolonged exposure to a slightly contracted face the criterion for judging subsequent faces as contracted might be relaxed, resulting in slightly contracted faces being classified as more normal than they had been prior to adaptation. The criterion for judging faces as expanded, however, might be relatively unchanged. This would result in the shifted and expanded distribution of "undistorted" judgments post-adaptation which we observed, rather than the simple lateral shift predicted by opponent coding.

We would like to emphasise that shifts in decisional criteria could not be differentiated from adaptation of a sensory code on the basis of subjective experience. Criteria must underlie perceptual decision processes at all stages of coding. Many of these processes are likely to be rapid and unconscious, meaning that changes to the sensory code would be subjectively indistinguishable from changes to decisional criteria. At this point we think it is somewhat ambiguous as to what extent face aftereffects reflect sensory coding changes, or altered decisional criteria.

5.5.2 *Adaptation of a sensory code, or a perceptual contrast effect?*

A final point we would like to make is that although we have found that facial gender aftereffects can successfully be predicted by a multichannel population based coding scheme, this does not dictate that these effects necessarily ensue from sensory adaptation. The tuning of the tilt aftereffect can successfully be modelled on the basis of a similar coding strategy, but so too can the tilt illusion (Clifford et al., 2000; Schwartz et al., 2009). The tilt illusion can occur when a central vertical test is surrounded by an annulus containing an oblique orientation. The central stimulus typically appears tilted away from the orientation of the annulus (Gibson & Radner, 1937). The tilt illusion is better conceptualised as a contrast effect, as opposed to an adaptation effect, as it occurs for simultaneously presented tests and annuli. Moreover, it can happen for very briefly presented stimuli (Westheimer, 1990).

The general implication of the preceding point is that competitive interactions between clearly distinct inputs can both shape perception and be well described by the dynamics of

a multichannel coding scheme (Clifford et al., 2000; Webster, 2011). The specific implication for high-level categorical aftereffects is that some of these might ensue due to a contrast effect between successive images, as opposed to an renormalisation process that recalibrates a neural code. The fact that some aftereffects can ensue after very brief 'adaptation' periods (of just 1 second: see Leopold, Rhodes, Muller, and Jeffery (2005) and Rhodes, Jeffery, Clifford, and Leopold (2007)) encourages us to entertain this possibility.

5.5.3 *Conclusion*

We have presented a novel psychophysical protocol, and demonstrated that it is capable of revealing qualitative differences between apparently similar visual aftereffects. While colour aftereffects met the predictions of a two-channel opponent code, face gender aftereffects were better predicted by a multichannel coding strategy and facial distortion aftereffects could only be said to be broadly consistent with opponent coding. These data suggest that high-level categorical aftereffects are diverse in nature and may be caused by computational processes that differ qualitatively.

6

CHAPTER 6: FACE AFTEREFFECTS INVOLVE LOCAL REPULSION, NOT RENORMALISATION

The previous chapter presented a novel ternary-classification paradigm that makes distinct predictions for renormalising vs locally repulsive aftereffects, unlike the more commonly-used binary classification task. Resulting data suggested that facial gender aftereffects are better described as locally repulsive perceptual aftereffects, whereas facial distortion aftereffects may be better described as renormalising ones. However, this interpretation assumes that changes in classification behaviour between baseline and adaptation trials arise from changes to perceptual encoding, rather than changes to decision-level criteria. This is not unique to our paradigm, but is a general weakness of method-of-single-stimulus classification tasks (Gescheider, Herman, & Phillips, 1970; Green & Swets, 1966; Morgan, Melmoth, & Solomon, 2013; Morgan et al., 2012; Storrs, 2015a; Yarrow et al., 2011).

In this final chapter, I present a second novel paradigm, which leverages the spatial-contingency of face aftereffects to compare appearance between standard and test faces presented in two differently-adapted retinal locations. By interleaving multiple standard stimuli in a forced-choice task we are able to measure perceptual aftereffects with as little influence from decision-level biases as is reasonable in psychophysics. The method has the additional advantage that by selecting any standard stimulus, the aftereffect can be measured at any test point in which the experimenter is interested. The resulting manuscript appeared in the *Journal of Vision* (Storrs & Arnold, 2015b).

This chapter again necessarily covers some of the same material as the preceding Chapter 5, and the Introduction to the thesis in Chapter 1 — I apologise for the repetition.

Face aftereffects involve local repulsion, not renormalisation

6.1 ABSTRACT

After looking at a photograph of someone for a protracted period (adaptation), a previously neutral-looking face can take on an 'opposite' appearance in terms of gender, identity, and other attributes — but what happens to the appearance of other faces? Face aftereffects have repeatedly been ascribed to perceptual re-normalisation. Renormalisation predicts that the adapting face and more extreme versions of it should appear more *neutral* after adaptation (e.g. if the adaptor was male, it and hyper-masculine faces should look more feminine). Other aftereffects, such as tilt and spatial frequency, are locally-repulsive, exaggerating differences between adapting and test stimuli. This predicts that the adapting face should be little-changed in appearance after adaptation, while more extreme versions of it should look even more *extreme* (e.g. if the adaptor was male, it should look unchanged, while hyper-masculine faces should look even more masculine). Existing reports do not provide clear evidence for either pattern. We overcame this by using a spatial comparison task to measure the appearance of stimuli presented in differently-adapted retinal locations. In behaviourally-matched experiments we compared aftereffect patterns after adapting to tilt, facial identity, and facial gender. In all three experiments data matched the predictions of a locally-repulsive, but not a renormalising, aftereffect. These data are consistent with the existence of similar encoding strategies for tilt, facial identity, and facial gender.

6.2 INTRODUCTION

After looking at a photograph of someone, a previously neutral-looking face can take on an 'opposite' appearance in terms of its gender (Rhodes et al., 2004; Webster et al., 2004), identity (Leopold et al., 2001), emotional expression (Hsu & Young, 2010), ethnicity (Webster et al., 2004), age (Schweinberger et al., 2010), eye gaze and head direction (Calder et al., 2008; Fang & He, 2005; Lawson et al., 2011). It is less clear how non-neutral faces are affected.

For example, does the appearance of the adapting face change during adaptation? Would a caricatured version of the original identity look more or less caricatured after adaptation? The answer to these questions may yield insights into how faces are represented neurally. Face aftereffects have repeatedly been attributed to a form of perceptual re-normalisation (Burton et al., 2015; Jeffery et al., 2010, 2011; Leopold et al., 2001; McKone et al., 2014; Pond et al., 2013; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins et al., 2007; Susilo et al., 2010; Webster & Maclin, 1999). According to this account, faces are represented in terms of how they differ from a perceptual norm. The perceptual norm is constantly updated according to recent experience to minimise differences between the norm and the prevailing average stimulus in the environment (Rhodes & Jeffery, 2006; Rhodes et al., 2005; Webster et al., 2004). This process can change the appearance of a given face by increasing or decreasing its distance from the norm, explaining the face aftereffect.

The perceptual renormalisation hypothesis predicts that as the norm is updated to lie closer to the adaptor, the adapting face should appear more 'neutral'. More 'extreme' versions of the adaptor should also look more neutral after adaptation, as the distance between these and the updated norm, relative to the unadapted norm, is also decreased. Previously neutral-looking faces should take on an appearance 'opposite' to the adaptor, as the distance between these and the updated norm will be increased relative to the unadapted norm (Robbins et al., 2007; Webster & MacLeod, 2011; Webster & Maclin, 1999). This pattern constitutes a uni-directional aftereffect, in which the appearance of all stimuli is shifted in the same direction to re-centre around an updated norm (see Figure 1b).

Adaptation to simple spatial patterns, like lines and shapes, instead results in a bi-directional 'locally-repulsive' pattern of aftereffects (see Figure 1c). In the tilt aftereffect (Gibson & Radner, 1937; Vernon, 1934) adaptation to any oriented stimulus can exaggerate differences in orientation between adapting and test stimuli in both directions (Mitchell & Muir, 1976), with little change in the appearance of the adapted orientation itself (Köhler & Wallach, 1944; Mitchell & Muir, 1976; Storrs & Arnold, 2015a). Locally-repulsive aftereffects have also been reported following adaptation to spatial frequency (Blakemore et al., 1970) and aspect-ratio (Badcock et al., 2014). If face aftereffects follow a similar locally-repulsive pattern, the appearance of the adapting face should be unchanged by adaptation, while differences between the adaptor and subsequent test faces should be exaggerated. Taking facial gender as an example, a locally-repulsive aftereffect uniquely predicts that adapting to a masculine face should make less masculine faces seem more feminine and more masculine faces seem more masculine (Robbins et al., 2007; Storrs & Arnold, 2012; Webster, 2011; Zhao & Chubb, 2001). As yet, it is unclear whether face aftereffects follow the pattern predicted by renormalisation, or that predicted by local-repulsion.

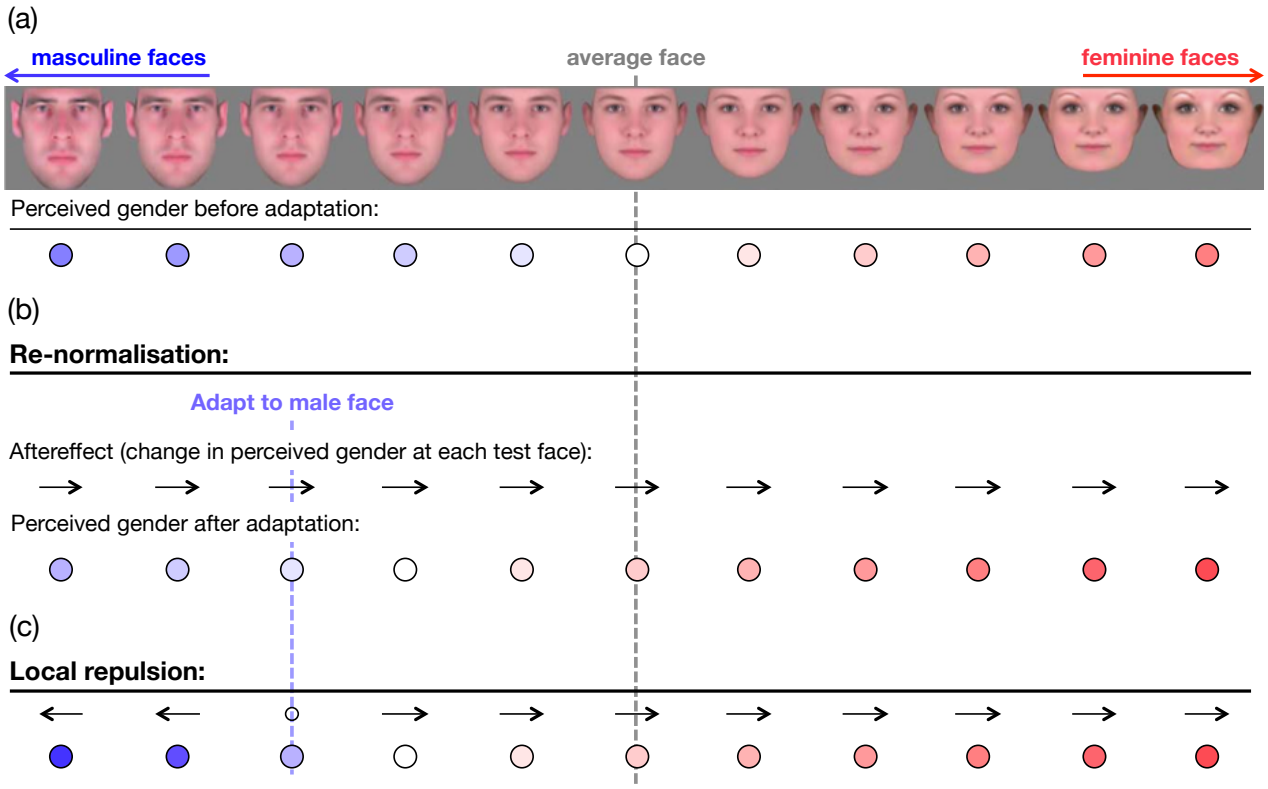


Figure 1. *A depiction of how a re-normalising and a locally-repulsive aftereffect might manifest along a dimension of facial gender.* ***(a)*** *A physical stimulus continuum, which ranges from caricatured male faces, through the gender-average face, to caricatured female faces. Prior to adaptation, the observer perceives a different strength of masculinity (blue) or femininity (red) in each face.* ***(b)*** *According to the re-normalisation account, facial gender is represented in terms of how each face differs from a perceptually gender-neutral norm (indicated by white circle). After adapting to a male face, the gender-neutral norm is recalibrated to lie closer to the adapting face, and all other faces in the encoded range change in appearance in the same direction.* ***(c)*** *If the facial gender aftereffect follows a locally-repulsive pattern, then after adapting to a male face, there should be no change in the appearance of the adapting face, but the differences between adapting and test faces should be exaggerated. Faces that had previously appeared more masculine than the adaptor should be exaggerated in their masculinity, while faces that had appeared less masculine than the adaptor should appear more neutral or feminine than they had previously. Note that the two proposals predict similar perceptual changes for faces more neutral than the adapting face, but make different predictions for the adapting face itself and for more extreme versions of the adapting face.*

6.2.1 *Evidence that face aftereffects result from renormalisation is ambiguous*

The method most often used to quantify face aftereffects a single stimulus binary-classification task, has limitations that make it difficult to determine what precise pattern a given after-effect follows. On each trial in such an experiment, a participant is shown a single test stimulus, usually selected from a continuum (e.g. a morph of facial genders, from masculine to feminine). The participant is then asked to classify the test as belonging to one of two categories (e.g. ”male” or ”female”). An estimate of the point along the continuum at which classifications switch from being predominantly of one category to predominantly

the other is taken as an estimate of the observer's category boundary. A shift in this boundary after adaptation is taken as a measure of the perceptual aftereffect.

This method has two weaknesses. The first is that a shift in the category boundary is equally consistent with either a change in the appearance of stimuli, or with a change in criteria used to apply category labels (Gescheider et al., 1970; Green & Swets, 1966; Morgan, 2014; Morgan et al., 2013; Morgan et al., 2012; Storrs, 2015a; Yarrow et al., 2011). A second weakness is that binary classifications tend only to be sensitive to changes in the appearance of stimuli near a category boundary. Here, a small change in appearance can produce a large change in categorisation behaviour. For stimuli far away from the boundary, however, even a large perceptual change might be insufficient to change how stimuli are categorised. A very masculine-looking face, for instance, might be made to look less masculine, but still look sufficiently male to be so categorised.

Unfortunately the subjective appearance of stimuli near a category boundary, as measured in a standard binary-classification task, is usually unhelpful in differentiating renormalising from locally-repulsive aftereffects. After adapting to a non-neutral stimulus, both proposals predict a shift in the appearance of neutral stimuli away from the adapted value (see Figure 1). Also problematic is that after adapting to a neutral stimulus neither proposal predicts a shift in the appearance of that stimulus.

It has been suggested that renormalising and locally-repulsive aftereffects can be dissociated by determining aftereffect magnitudes at the category boundary as a function of increasingly non-neutral adapting stimuli (Jeffery et al., 2010, 2011; McKone et al., 2014; Pond et al., 2013; Zhao et al., 2011). The rationale is that, if an aftereffect entails a local repulsion, then the aftereffect for stimuli near the category boundary might initially increase, but should then diminish as the distance between adapting and test stimuli exceeds the limited range of the local repulsion. On the other hand, if an aftereffect involves renormalisation of the whole perceptual dimension toward the adapting stimulus, the extent of recalibration should only increase for more extreme adaptors. There is, however, an important caveat that prevents this protocol from being diagnostic. Facial dimensions are thought to be finite, so increasingly non-neutral stimuli will eventually escape the confines of the dimension, and begin to look unnatural. An increasingly male-looking face, for instance, might begin to look non-human — ogre-ish. A gradual increase, and eventual reduction, in aftereffect magnitude for increasingly non-neutral adaptors (now repeatedly reported, see McKone et al. (2014), Zhao et al. (2011)), is therefore equally consistent with either a locally repulsive aftereffect, or with a renormalisation process that is not strongly activated by unrealistic faces (Pond et al., 2013).

6.2.2 *A novel spatial-comparison task*

To test the renormalisation and local-repulsion hypotheses, one needs a task for which these aftereffect patterns make distinct predictions. In a previous report (Storrs & Arnold, 2012), we showed that by using a three-category classification task (e.g. "male, androgynous, or female?"), and adapting to one of the two category boundaries, one can measure an aftereffect at a particularly diagnostic point: the adapted value itself. A renormalising aftereffect will result in a previously non-neutral adaptor looking more neutral after adaptation, while a locally-repulsive aftereffect predicts no change in the adaptor's appearance. For facial gender adaptation, we found no change in how the adapted value was categorised, consistent with a locally-repulsive aftereffect. After adapting to a configurally distorted face, however, there was an increased tendency to place the adaptor in the "undistorted" category, consistent with renormalisation. While this suggested multiple facial encoding strategies, caution must be exercised when interpreting these data, due to the subjective nature of the categorical tasks on which the protocol depended. In the following experiments we adopt another protocol that allows one to measure aftereffects at *any* point along a test continuum, and to avoid relying on subjective categorical decisions.

Our protocol makes use of two characteristics of face aftereffects. First, face aftereffects are strongest when adapting and test stimuli are matched in retinal position (Afraz & Cavanagh, 2009; Afraz & Cavanagh, 2008) although some tolerance is displayed to positional variance, (Yamashita, Hardy, De Valois, & Webster, 2005; Zhao & Chubb, 2001). Second, when different adaptors are shown simultaneously in different retinal locations, distinct spatially-contingent aftereffects can ensue (Afraz & Cavanagh, 2008). This allows for a powerful psychophysical method: a spatial comparison task. One can present a standard test stimulus in one location, and find which stimulus value in a second location appears to match it (Elliott et al., 2011; Farell & Pelli, 1999; Jakel & Wichmann, 2006; Kompaniez, Abbey, Boone, & Webster, 2013). At baseline, perceptually-matched stimuli are likely also to be physically matched (for an exception, see Afraz, Pashkam, and Cavanagh (2010)). If adaptation impacts perception, a standard stimulus presented in an adapted location should appear to match a physically different stimulus presented in an unadapted location. Aftereffect magnitudes can therefore be calculated as the difference between matches to the same standard stimulus pre- and post-adaptation.

In the experiments reported here, we use a forced-choice 'double-pair task' (Kaplan et al., 1978; Rousseau & Ennis, 2001). Two pairs of peripheral stimuli are shown at test. One contains an identical pair of standard stimuli, whereas the other contains one standard and a test that differs from the standard by a variable amount. Participants are asked to indicate which interval contains the *identical* pair. If the proportion of times that a participant selects the incorrect interval is plotted as a function of the test value, one should find a bell-shaped

distribution peaking at the value that perceptually matches the standard. The magnitude of any aftereffect is signalled by the shift of this peak, from pre- to post-adaptation.

Rather than compare appearance between an adapted and an unadapted location, we instead adopt a double-adaptor paradigm, comparing appearance between two *differently* adapted locations. We do this for two reasons. First, previous research suggests adaptation to a single peripheral facial image can affect classifications of spatially distant tests, so it may be difficult to find a measurable difference between 'adapted' and 'unadapted' locations (Afraz & Cavanagh, 2008). In the same study, however, spatially-contingent aftereffects were induced when multiple adaptors were presented, highlighting the possibility of differentially adapting different locations. Second, using two adaptors minimises the potential that asymmetric spatial attention might bias subsequent perceptual decisions. If there were just one adaptor, people might attend more to stimuli in that location (attention has been shown to modulate face aftereffects measured via a binary classification task — Rhodes, Jeffery, Evangelista, Ewing, and Peters (2011)), or be tempted to compare stimuli to the single adaptor. With two adaptors, spatial attention is more likely balanced, and it becomes harder to compare the four test images to the two different adaptors.

If face aftereffects reflect a renormalisation of all stimuli about a new neutral point (Rhodes et al., 2005; Webster & Maclin, 1999), aftereffects measured at different positions along a test continuum will tend to be constant (see Figure 2d-f). As we will be measuring the *relative* aftereffect induced between two differently-adapted locations at different standard test values, re-normalisation predicts a constant magnitude of relative aftereffect for all test values, equal to the difference between the aftereffects induced in each location. If, on the other hand, face aftereffects involve local repulsion, we should find that the relative aftereffect is largest for standard test values midway between the two adaptors (see Figure 2a-c). Such tests should be 'pushed' in opposite directions by opposite local repulsions, whereas aftereffects measured at either of the adapted values should be smaller, as they will reflect the influence of just one adaptor.

In the experiments that follow, we test these predictions for tilt adaptation (Experiment 1; to validate our protocol in a context repeatedly linked to locally-repulsive aftereffects), for facial identity adaptation (Experiment 2) and for facial gender adaptation (Experiment 3). We find that data in all three contexts are consistent with locally-repulsive, but not renormalising, patterns of aftereffect.

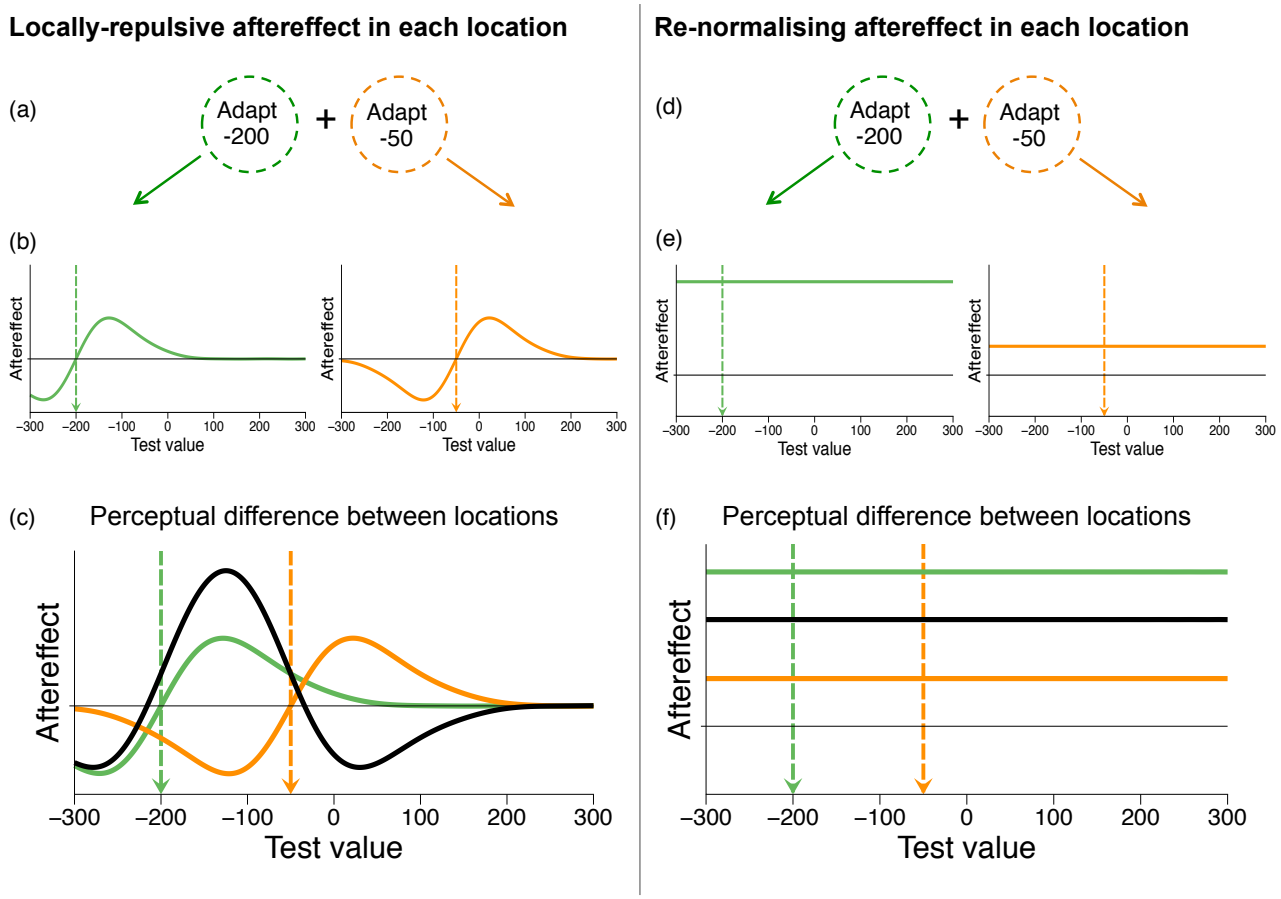


Figure 2. *Patterns of results predicted by a locally-repulsive (left), and a re-normalising aftereffect (right) in our spatial-comparison task.* ***(a)*** *After adapting in two different retinal locations to stimulus values of -200 (left; arbitrary units) and -50 (right), an aftereffect is induced in each location which manifests* ***(b)*** *as a local repulsion of nearby values away from the adapted value, with no change in the appearance of the adapted value.* ***(c)*** *The predicted mismatch in appearance between stimuli presented at the two differently-adapted locations is given by the difference between the two aftereffect functions (black). The shape of the difference function varies depending on the shape and separation of the aftereffects in each location, but the qualitative pattern remains constant across a wide range of possible values: maximal aftereffects are predicted for test values inbetween the two adaptors, and lesser aftereffects are predicted for tests at either of the two adapted values.* ***(d)*** *If, after adapting to these same stimulus values, aftereffects manifest as* ***(e)*** *a uniform renormalisation of all stimulus values, then the difference in appearance between two differently-adapted locations* ***(f)*** *will also be uniform across all encoded test values.*

6.3 EXPERIMENT 1: TILT AFTEREFFECTS

6.3.1 *Method*

6.3.1.1 *Participants*

Ten observers participated in Experiment 1, including the first author, three experienced psychophysical observers naïve to the research hypotheses, and six inexperienced observers recruited from the University College London (UCL) psychology participation pool, who

were compensated with £6 for their time. The Experimental Psychology Ethics Committee at UCL approved all experiments.

6.3.1.2 *Stimuli and apparatus*

Stimuli were presented on a 22" Mitsubishi Diamond Plus 230SB monitor (resolution 1280 x 1024 pixels; refresh rate 85Hz; not Gamma corrected), using the *Psychophysics Toolbox for Matlab* (Brainard, 1997; Pelli, 1997). Participants viewed stimuli from a distance of 57cm using a chinrest. During both a preliminary sensitivity measure and the main experimental task, pairs of stimuli were presented simultaneously, centred 4.5 degrees of visual angle (dva) to either side of a central white fixation point (which had a diameter subtending 0.24 dva). Participants were instructed to fixate the central fixation point throughout all experiments.

Test and adapting stimuli were pairs of Gabors with a Michelson contrast of 1 and a spatial frequency of 2.5 cycles/dva, presented within Gaussian spatial envelopes with standard deviations subtending 1.1 dva. The phase of each Gabor waveform was randomised on a trial-by-trial basis. The display background was grey, and matched the average luminance of test stimuli.

6.3.1.3 *Procedure*

In the main experiment, we measured the relative aftereffect between two differently-adapted retinal locations, at three different Standard test values. The three Standard test values were determined independently for all participants via a preliminary procedure that measured orientation discrimination sensitivity.

Preliminary procedure

Stimuli were presented in a dual-pair task (Rousseau & Ennis, 2001), also known as a '4-interval AX' task (Kaplan et al., 1978) — see Figure 3. Two pairs of static Gabors were presented sequentially for 300ms, separated by a 300ms blank inter-stimulus interval (ISI). On each trial, three of the test stimuli were set to a Standard value of -45° (which will serve as the Central of the three Standard test values in the main experiment). The orientation of the fourth varied via an adaptive procedure described below. The interval and position (left or right of fixation) of the variable test was randomised on a trial-by-trial basis. The participant indicated which of the two intervals had contained identical stimuli by clicking one of two mouse buttons. Feedback was given, with the fixation dot turning green for correct decisions and red for incorrect decisions, for 500ms prior to the next trial.

The orientation difference between the variable test and the Standard test value was varied on a trial-by-trial basis according to adaptive 'staircase' procedures (Cornsweet, 1962). Four

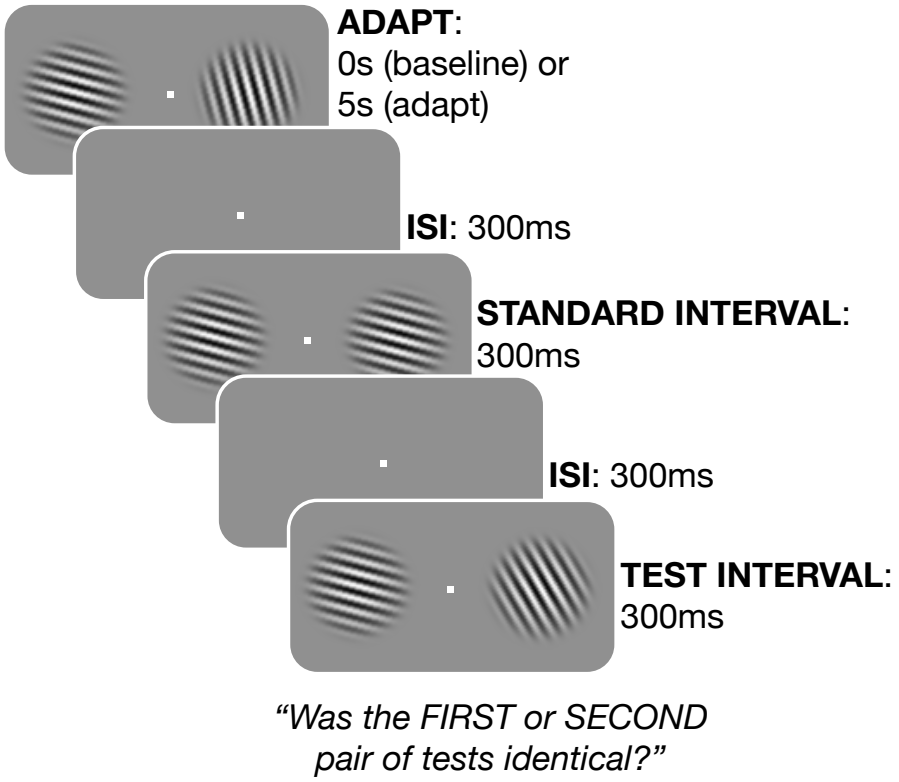


Figure 3. *Trial structure for Experiment 1. Two pairs of test stimuli were shown for 300ms each, sequentially with a 300ms inter-stimulus-interval. Three of the four tests were set to an identical standard test value (-45°), while the orientation of the fourth was varied according to an adaptive procedure (see main text for description). Tests were either shown in isolation (during the preliminary sensitivity measure, and during baseline trials of the main experiment) or after exposure to a pair of adapting stimuli for 5 seconds. Participants were asked to indicate which of the two test intervals had contained identical stimuli.*

staircases were interleaved, with two sampling values rotated counter-clockwise from the Standard and two sampling values rotated clockwise from the Standard. For both counter-clockwise and clockwise sampling procedures, one of the two staircases was initiated at an orientation difference of 0° (i.e. an initial orientation of -45°), and the other was initiated at an orientation difference of 10° (i.e. initial orientations of -55° for clockwise-sampling staircases, and -35° for counter-clockwise-sampling staircases).

Test orientation differences were adjusted according to a 'three-down, one-up' decision rule, wherein three consecutive *correct* responses resulted in a decrease in the magnitude of orientation differences, and any incorrect response resulted in an increase in orientation difference magnitude. This decision rule converges on 79% correct performance (Levitt, 1970). For each staircase, a 'reversal' was recorded when the direction of adjustment (increasing vs decreasing the orientation difference between the -45° Standard and the variable test) differed from the direction of the last adjustment for that staircase. Test orientation differences were adjusted in steps of 2° until the first 3 reversals for that staircase, after which adjustments were made in steps of 0.5°. Minimum orientation differences of 0° were enforced,

by repeatedly sampling this value if repeated correct judgments made this necessary. Each staircase terminated after 6 reversals. Orientation difference values corresponding with the last 3 reversals in each staircase were averaged across the four staircases to produce a single just-noticeable difference (JND) estimate from the Standard test value. The average JND was 9.8° (minimum = 6.1°, maximum = 14.7°).

Main experimental procedure

The Central Standard value for the main experiment was set to -45° for all participants, and Lower and Upper Standard test values were set to multiples of the participant's JND from the Central Standard (see Figure 4a). Lower Standard test values were -45° minus two JNDs, and Upper Standard test values were -45° plus two JNDs. In adaptation trials, the adaptor shown to the left of fixation was of the Lower Standard value, and the adaptor shown to the right of fixation was of the Upper Standard value. The average Lower Standard test value was -64.5° (±2.0° standard error of the mean (SEM)) and the average Upper Standard test value was -25.5° (±2.0°).

All participants first completed a baseline run of trials, during which they viewed sequential tests in the absence of adaptation. They then completed an adaptation run of trials, in which they viewed sequential tests after either 30 (first trial) or 5 (all subsequent trials) seconds of exposure to the adapting stimuli (see Figure 3).

On each trial, during both baseline and adaptation runs of trials, three of the four test stimuli were set to a Standard test value, pseudo-randomly selected on each trial from either the Lower, Central or Upper Standard test values. The value of the fourth, variable, test stimulus was determined according to an adaptive procedure described below. The test stimulus was always shown on the right hand side of fixation, in a random interval. Observers indicated which of the two intervals had contained identical stimuli. No performance feedback was given.

The value of the variable test stimulus was determined on each trial by a Pólya urn adaptive sampling procedure (Rosenberger & Grill, 1997; Yarrow, Sverdrup-Stueland, Roseboom, & Arnold, 2013). The range of possible test values was -135° to +45°, sampled in increments of 0.25° (i.e. 360 possible test values). Three independent sampling procedures were maintained, one for each of the three Standard test values.

On each trial, a Standard test value was pseudo-randomly selected, and a variable test value was selected according to the relevant sampling procedure. The initial probability distribution for each of the three sampling procedures was uniform, with a nominal sampling probability value of 1 assigned to each test value falling within a range ±3 JNDs from the relevant Standard, and zero to values falling outside this range. The actual probability of sampling any given test value on a trial was the current nominal probability of sampling that test value, divided by the sum of the current nominal probability values across all test

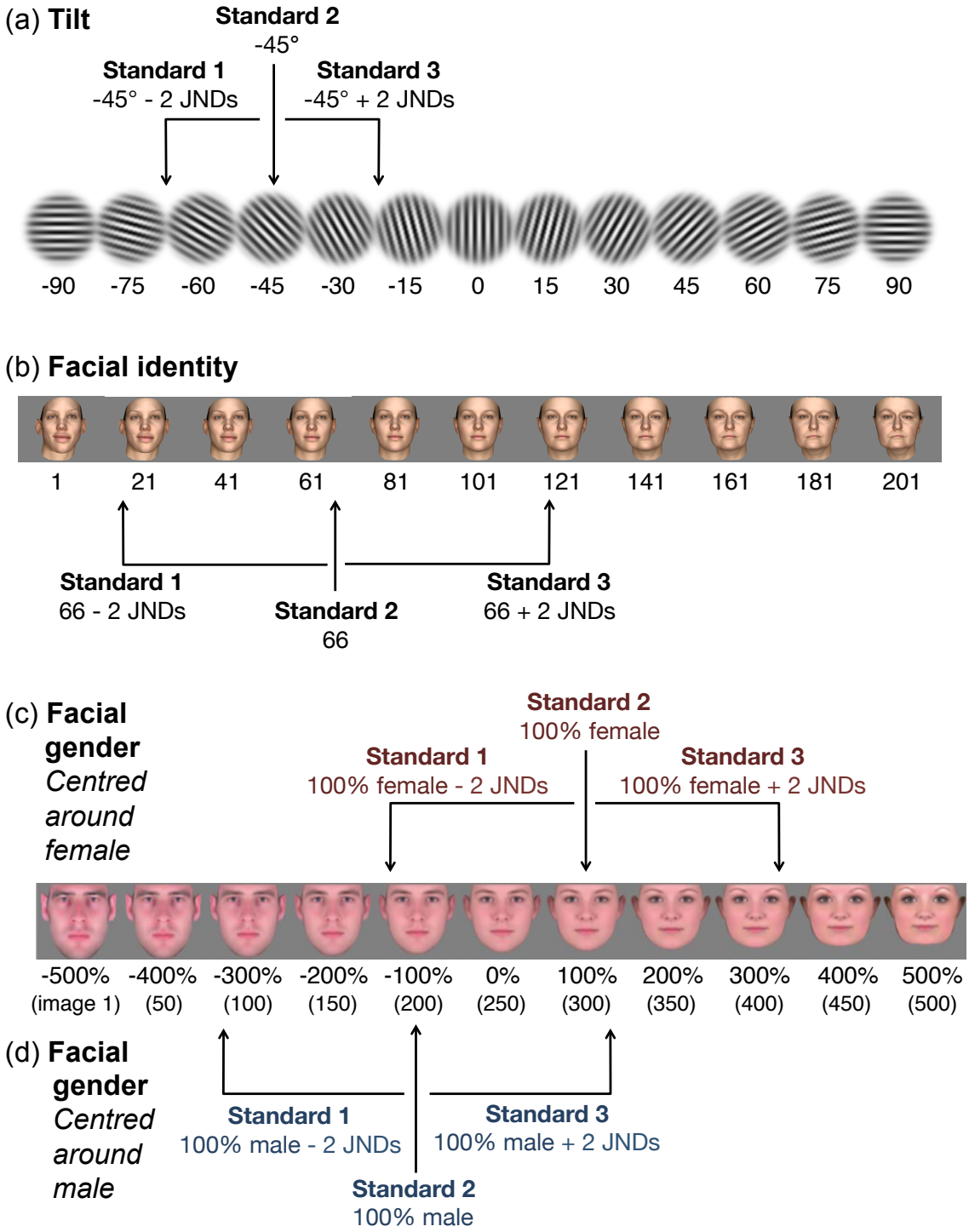


Figure 4. *Selection of Standard test stimuli for* ***(a)*** *tilt,* ***(b)*** *facial identity, and* ***(c, d)*** *facial gender experiments. The selection procedure was identical in all cases. A Central Standard test value was chosen (-45° for tilt adaptation; image 66 for facial identity adaptation; image 300 for female-centred facial gender adaptation, and image 200 for male-centred facial gender adaptation). Estimates of the participant's just-noticeable-difference (JND) from Central Standard test values were determined in a preliminary sensitivity measure (see main text for details). Lower and Upper Standard test values were then set respectively to -2 JNDs and to +2 JNDs from Central Standard test values. Arrows indicate approximate average Lower, Central and Upper Standard test values in each experiment.*

values. On each trial, a particular value was chosen for the variable test via the generation of a random number used to index ordered sampling probability values. Whenever a

participant incorrectly selected the interval containing the variable test as having contained identical stimuli, nominal sampling probability values associated with test values +/-0.5° and ±2° from the test were increased by 2 and 1 respectively. These adaptive procedures ensured disproportionate sampling of tests perceptually similar to each Standard stimulus. Each of the three adaptive procedures was sampled for 50 trials, yielding a total run of 150 trials (taking approximately 10 minutes in baseline runs of trials, and 20 minutes in adaptation runs of trials).

6.3.2 *Results*

A run of trials provided three distributions, one for each Standard stimulus, describing the proportion of times each variable test value had been presented on which it was mistaken as being identical to the relevant Standard stimulus (see Figure 5a). Gaussian distributions were fit to these data, using a least-squares regression weighted by the number of observations at each variable test value. Peaks of the fitted Gaussians were taken as estimates of the participant's point of subjective equality (PSE) between tests presented in the right retinal location and each Standard stimulus presented in the left retinal location. Relative aftereffects between the two differently-adapted locations were calculated as the post-adaptation PSE for a given Standard minus the corresponding baseline PSE.
Re-normalisation predicts that the magnitude and direction of relative aftereffects should be constant for our three Standards (see Figure 2f). Local repulsion predicts that relative aftereffects will be greater for the Central Standard than for Lower and Upper Standards (see Figure 2c). We confirmed that our pattern of results conformed to the latter prediction by conducting a repeated-measures ANOVA. This revealed a significant quadratic trend for data relating to our three Standards ($F_{1,9}$ = 10.16, p = .011), with the greatest aftereffect magnitude at the Central Standard — as predicted by a locally-repulsive aftereffect (see Figure 5b).

6.4 EXPERIMENT 2: FACIAL IDENTITY AFTEREFFECTS

Having validated our protocol for tilt adaptation, we then applied it to facial identity adaptation. Details were as for Experiment 1, with the following exceptions.

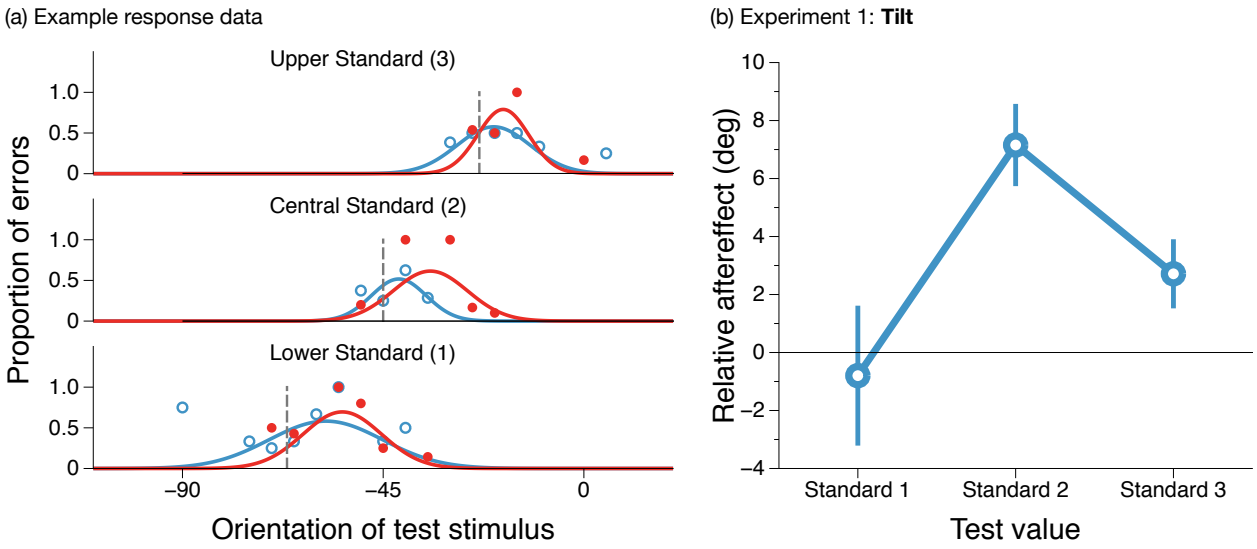


Figure 5. ***(a)*** *Example data from one participant in the tilt adaptation experiment. Open blue dots represent the proportion of incorrect responses for each variable test value during baseline trials on which the Upper (top panel), Central (middle) or Lower (bottom) Standard test stimulus was shown; closed red dots represent the same during adaptation trials. Data are binned into 5° bins for illustration purposes. Dotted grey lines indicate the physical value of each of the three Standard test stimuli. Curves show Gaussians fit to response data before (blue) and after (red) adaptation. The peaks of the fitted Gaussians indicate the participant's PSE between the two retinal locations for each Standard stimulus. The relative aftereffect between the two locations is calculated independently for each Standard stimulus by subtracting the baseline PSE from the post-adaptation PSE.* ***(b)*** *Relative aftereffects at each of the three Standard test values in the tilt adaptation experiment. Error bars indicate ±1 standard error of the mean (SEM) between individual estimates.*

6.4.1 *Methods*

6.4.1.1 *Participants*

Fifteen observers participated in the facial identity experiment, including the first author, three experienced psychophysical observers who were naïve to the experimental hypotheses, and 11 naïve paid volunteers.

6.4.1.2 *Stimuli and apparatus*

Test and adapting stimuli were pairs of face images. Test images subtended 5.8 dva in width and 7.7 dva in height whereas adaptors subtended 7.3 by 9.6 dva. The size difference between adaptors and tests was intended to mitigate local retinotopic adaptation, a precaution widely used in face aftereffect research (see Zhao and Chubb (2001)).

Face stimuli were colour images generated from the Basel Face Model (BFM[1]; Paysan, Knothe, Amberg, Romdhani, and Vetter (2009)), a generative model created by performing principal components analysis (PCA) on the three-dimensional head scans of 100 male

1 Available from http://faces.cs.unibas.ch/bfm/

and 100 female, predominantly-Caucasian, individuals. Images generated by the model are cropped around the ears, necks, and forehead to remove hair (see Figure 4b).
In the BFM two sets of 199 principal components have been derived, independently for facial structure and facial texture (capturing variations due to skin colour, eye colour, lighting, etc). This allowed us to minimise colour and luminance artefacts by generating faces that varied structurally but not texturally. Data accompanying the BFM specify directions in face-structure and face-texture spaces along which variance is greatest for individuals of different genders, heights, weights, and ages. This allows these attributes to be manipulated in a naturalistic manner.
We generated a unique continuum of face images for each participant, to ensure our data were not dependent on idiosyncrasies of any single stimulus set. Each continuum consisted of 201 images drawn from a trajectory in face-structure space linking two different same-gender identities, centred on an average male or female face (7 male continua, 8 female). For each continuum, image 1 was a novel male (or female) identity chosen by adding a random proportion and direction of the height, weight, and age vectors to an average male (or female) face, along with a small random amount of each principal component. Image 101 was the average male (or female) face, and image 201 was an 'anti-identity' face, that differed from the male (or female) average in an opposite manner relative to image 1. Other images were then generated from the BFM, traversing in equal steps the distance between images 1 and 201. Selected images from one stimulus continuum are shown in Figure 4b, and from all 15 continua in Appendix 3.

6.4.1.3 *Procedure*

Preliminary procedure

Sequential pairs of face images were presented for 500ms each, separated by 300ms ISIs. The Central Standard test value was set to image 66 of each individual face continuum.
Four independent, randomly interleaved staircase procedures were used to estimate each participant's JND from the Central Standard face. Two procedures sampled test faces with lower image numbers than the Central Standard, and two sampled higher image numbers. Within each pair of sampling procedures, one staircase was initiated at an image number difference of 0 (i.e. image number 66), while the other was initiated at either image number 26 (for procedures sampling lower image numbers) or 106 (for procedures sampling higher image numbers). The average JND estimate was 26 image units (minimum = 14, maximum = 36).

Main experimental procedure

Two participants had JNDs that would have led to the Lower Standard (Central Standard minus 2 JNDs) being outside the range of available test faces (JNDs = 34 and 36). For these observers, an arbitrary JND of 32 was used to select Standard test images.
Whenever a participant incorrectly selected the test interval as having contained identical faces, nominal sampling probability values associated with test images numbered ±10 and ±20 from the test were increased by 2 and 1 respectively.

6.4.2 *Results*

Analyses of data were as for Experiment 1, unless specified.
To assess whether aftereffects were greatest for the Central Standard, relative to Lower and Upper Standards, we conducted a repeated-measures ANOVA. This revealed a significant quadratic trend for data relating to our three Standards ($F_{1,14} = 7.14$, $p = .018$; see Figure 6), with a maximal aftereffect for the Central Standard — as predicted by a locally-repulsive aftereffect.

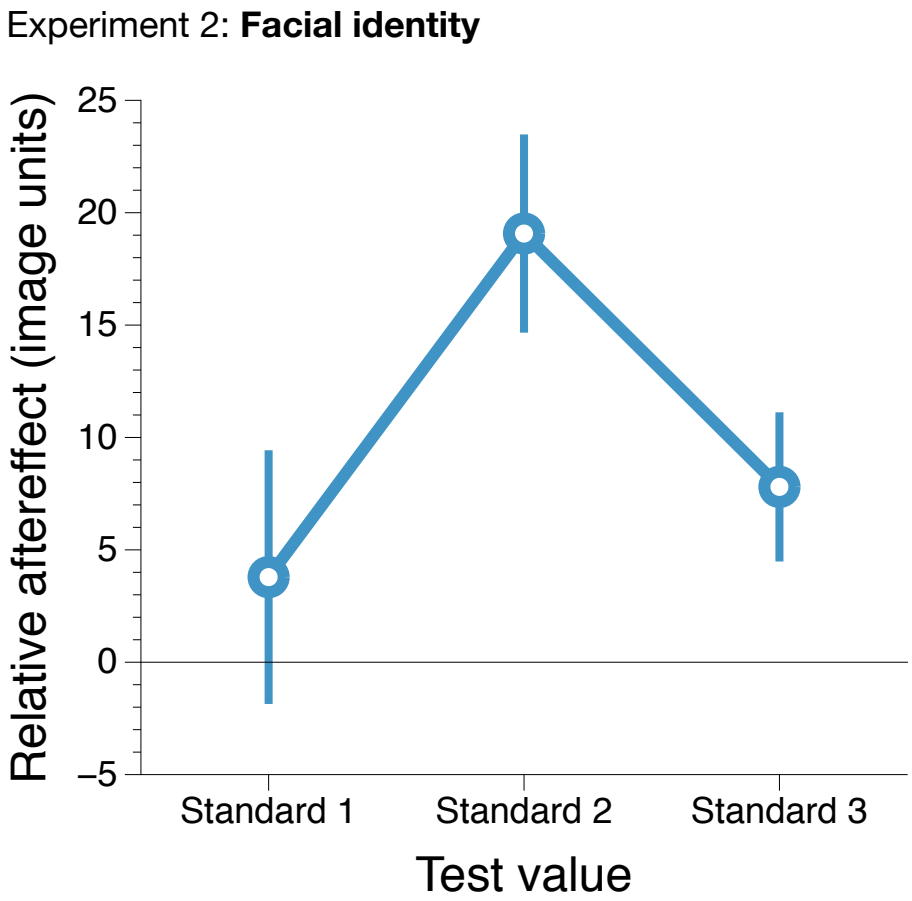


Figure 6. *Relative aftereffect at each of the three standard test values in the facial identity aftereffect experiment. Details as for Figure 5.*

6.5 EXPERIMENT 3: FACIAL GENDER AFTEREFFECTS

Details for Experiment 3 were as for Experiment 2, with the following exceptions.

6.5.1 *Method*

6.5.1.1 *Participants*

Fifteen observers participated in Experiments 3a (male-face-centred experiment), including the first author, five experienced psychophysical observers naïve to the experimental hypotheses, and nine naïve paid volunteers. In Experiment 3b (female-face-centred experiment), participants included the first author, six experienced observers, and eight paid volunteers. Note that some people participated in both experiments, with a total of 19 unique observers (five male, fourteen female). When the same person participated in both experiments, these were conducted on separate days.

6.5.1.2 *Stimuli and apparatus*

Test images subtended 5.9 dva in width and 6.5 dva in height. Adaptors subtended 7.8 by 8.7 dva.

Face stimuli were 500 colour images[2] previously used in Zhao et al. (2011) and Pond et al. (2013). These were generated by morphing between and beyond an average male face (-100% gender; image number 200) and an average female face (+100% gender; image 300), to create a continuum spanning from a masculine caricature (-500% gender; image 1) through androgynous (0% gender; image 250) to a feminine caricature (+500% gender; image 500. For further details see those describing the 'twenty' continuum in Zhao et al. (2011). Face images were presented against a grey background, inside a grey oval frame of the same luminance, which occluded the ears and side of the head. All participants saw the same stimulus continuum.

6.5.1.3 *Procedure*

Experiment 3a had a Central Standard stimulus of 200, Experiment 3b a Central Standard stimulus of 300.

Preliminary procedure

In both Experiments, four staircase procedures were used to locate each participant's JND for differences from the Central Standard, two sampling lower image numbers and two

2 Available from http://homepages.inf.ed.ac.uk/jbednar/stimuli/

sampling higher image numbers. Within each pair of staircase procedures, one was initiated at an image number difference of 0 (i.e. initiated at image numbers 200 and 300 respectively for Experiments 3a and 3b), while the other was initiated at an image number difference of 100 (i.e. lower-sampling procedures were initiated at image 100 in Experiment 3a and 200 in Experiment 3b; higher-sampling procedures were initiated at image 300 in Experiment 3a and 400 in Experiment 3b). Average JNDs were 54 image units (minimum = 28, maximum = 92) in Experiment 3a, and 57 image units (minimum = 35, maximum = 97) in Experiment 3b.

Main experimental procedure

In Experiment 3a, variable test stimuli were shown on the right hand side of fixation, in the same location as the Upper adaptor, as in the previous experiments. In Experiment 3b, variable test stimuli were shown to the left of fixation, in the same location as the Lower adaptor. This meant that aftereffects should predominantly manifest as participants matching standard stimuli to *lower* variable test values after adaptation, rather than *higher* values as in the previous experiments. We took this precaution to avoid a possible ceiling effect when estimating the aftereffect at the Upper Standard, which was for some participants near the end of the image continuum (see Figure 4c).
In both Experiments, whenever a participant incorrectly selected the test interval as having contained identical faces, nominal sampling probability values associated with test images numbered ± 10, ± 25 and ± 50 from the variable test value were increased by 3, 2 and 1 respectively.

6.5.2 *Results*

For both experiments, data for each Standard were grouped into 10-image bins, then expressed in terms of the proportion of times, for variable test values within each bin, that the participant had responded incorrectly. Gaussian functions were fit to these data, and their peaks were taken as PSE estimates for each Standard stimulus, before and after adaptation. To assess whether aftereffects were greater for Central Standards than Lower and Upper Standards, we conducted two repeated-measures ANOVAs. For Experiment 3b, this revealed a significant quadratic trend for data relating to our three Standards ($F_{1,14} = 6.33$, $p = .025$; see Figure 7), with a maximal aftereffect at the Central Standard — as predicted by a locally-repulsive aftereffect. In Experiment 3a, aftereffects were evident (a two-tailed one-sample t-test performed on the average aftereffect across all three Standards for each participant revealed a mean relative aftereffect of 42 image units, $t_{14} = 4.67$, $p < .001$), but there was no significant quadratic trend between data relating to the three Standards ($F_{1,14}$

= 1.82, p = .200). The reader should note, however, that the pattern of results in this Experiment conformed with the pattern of results in all other experiments we have reported here, in that the largest aftereffect observed was for the Central standard, whereas aftereffects for the Lower and Upper Standards were smaller, albeit not statistically different in this case.

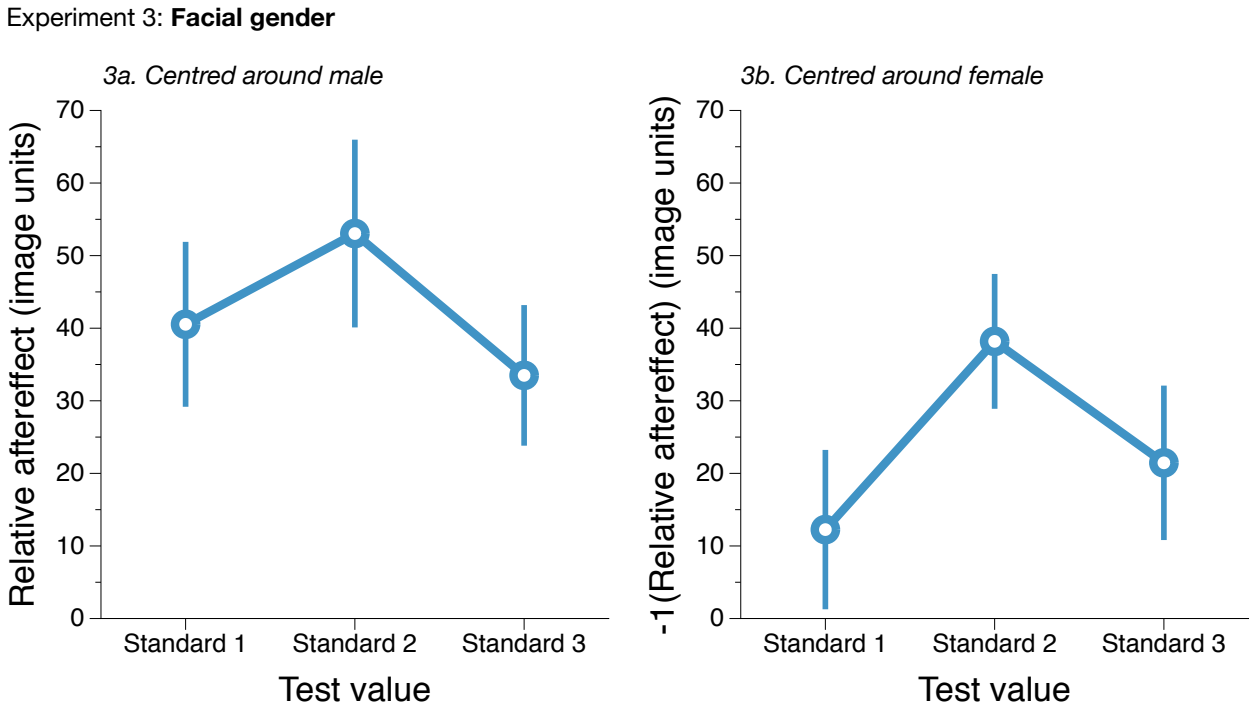


Figure 7. *Relative aftereffect at each of the three standard test values in (left) the male-face-centred and (right) the female-face-centred facial gender adaptation experiments. Because the variable test was shown in the same location as the Lower adaptor for Experiment 3b (but that of the Upper adaptor for other Experiments), the aftereffects manifested as a negative rather than a positive shift, relative to baseline. For ease of comparison, aftereffects are shown here with their sign flipped. Other details as for Figures 5 and 6.*

6.6 DISCUSSION

We measured aftereffects following tilt, facial identity, and facial gender adaptation, and found strikingly similar patterns of results, which were in excellent agreement with the predictions of a locally-repulsive aftereffect. These data were, however, inconsistent with the predictions of a re-normalising aftereffect.

Locally-repulsive aftereffects provide evidence for multichannel encoding Perceptual aftereffects are often thought to result from neural adaptation — changes in the responsiveness of neurons after prolonged stimulation (Clifford et al., 2007; Wark et al., 2007; Webster, 2011). Two broad classes of encoding scheme have been proposed. Locally-repulsive aftereffect patterns have been associated with encoding schemes in which multiple channels are narrowly tuned for different input values, with no particular stimulus playing a special role (Blakemore et al., 1970; Clifford et al., 2000; Pouget et al., 2000; Seriès et al., 2009). In

Figure 8a we illustrate a simulated multichannel code before adaptation (blue curves), consisting of Gaussian channels with peaks uniformly spaced along the stimulus dimension (see Appendix 1 for details of this model).

Adaptation to a given stimulus is proposed to cause a temporary reduction in the excitability of a subset of channels, in proportion to their response to the adaptor (red curves in Figure 8a). Perceptual experience is presumably determined by the pattern of activity across the population of channels. When this 'decoding' process is simulated, via a technique such as maximum-likelihood estimation (MLE), systematic differences can arise between estimates based on pre- and post-adaptation activity (Ma & Pouget, 2009; Pouget et al., 2000). These take the form of a locally-repulsive pattern of biases (red curve, and inset, in Figure 8b), in which the decoded value of the adaptor is unchanged, but differences in decoded values between the adaptor and nearby stimuli are exaggerated. Crucially, a multichannel encoding scheme accurately predicts the patterns of perceptual bias seen after adapting to orientation or spatial frequency (Blakemore et al., 1970; Clifford et al., 2000; Goris et al., 2013).

Renormalising aftereffect patterns have been associated with opponent-channel encoding schemes, in which two channels respond increasingly as stimulus values deviate from a perceptual norm; one preferring low values along the encoded dimension and the other high values (Giese & Leopold, 2005; Regan & Hamstra, 1992; Rhodes et al., 2005; Susilo et al., 2010). In Figure 8d we illustrate a one-dimensional simulation of such an encoding scheme, in which two monotonic channels intersect at stimulus value zero (blue curves) (see Appendix 1 for details). The stimulus corresponding to the 'norm' value (here, zero) activates both channels equally, while any other stimulus elicits an imbalanced response from the two channels. Again, the perceived stimulus is determined by the particular combination of channel activities.

As in the multichannel scheme, adaptation is thought to bring about a temporary reduction in the excitability of one or both channels, in proportion to their response to the adaptor (red curves, Figure 8d). When one channel is disproportionately suppressed, the input that elicits a balanced response across the two channels shifts toward the adapted value, thereby updating the 'norm' to more closely resemble recent experience. This recalibration can produce systematic changes in the perception of the same physical stimulus before and after adaptation. The opponent-channel code predicts a uniform re-normalisation, in which the adaptor appears more neutral after adaptation, and the appearance of all other stimuli is altered in the same direction by the same amount (red line, and inset, in Figure 8e). The magnitude and direction of this constant shift depends on the adapting value, with adaptors further from neutral producing larger shifts, and adaptation to zero uniquely producing no aftereffect.

A uniform shift in appearance of *all* test stimuli is perhaps the simplest possible prediction that can be derived from an opponent-channel code. One can also plausibly predict

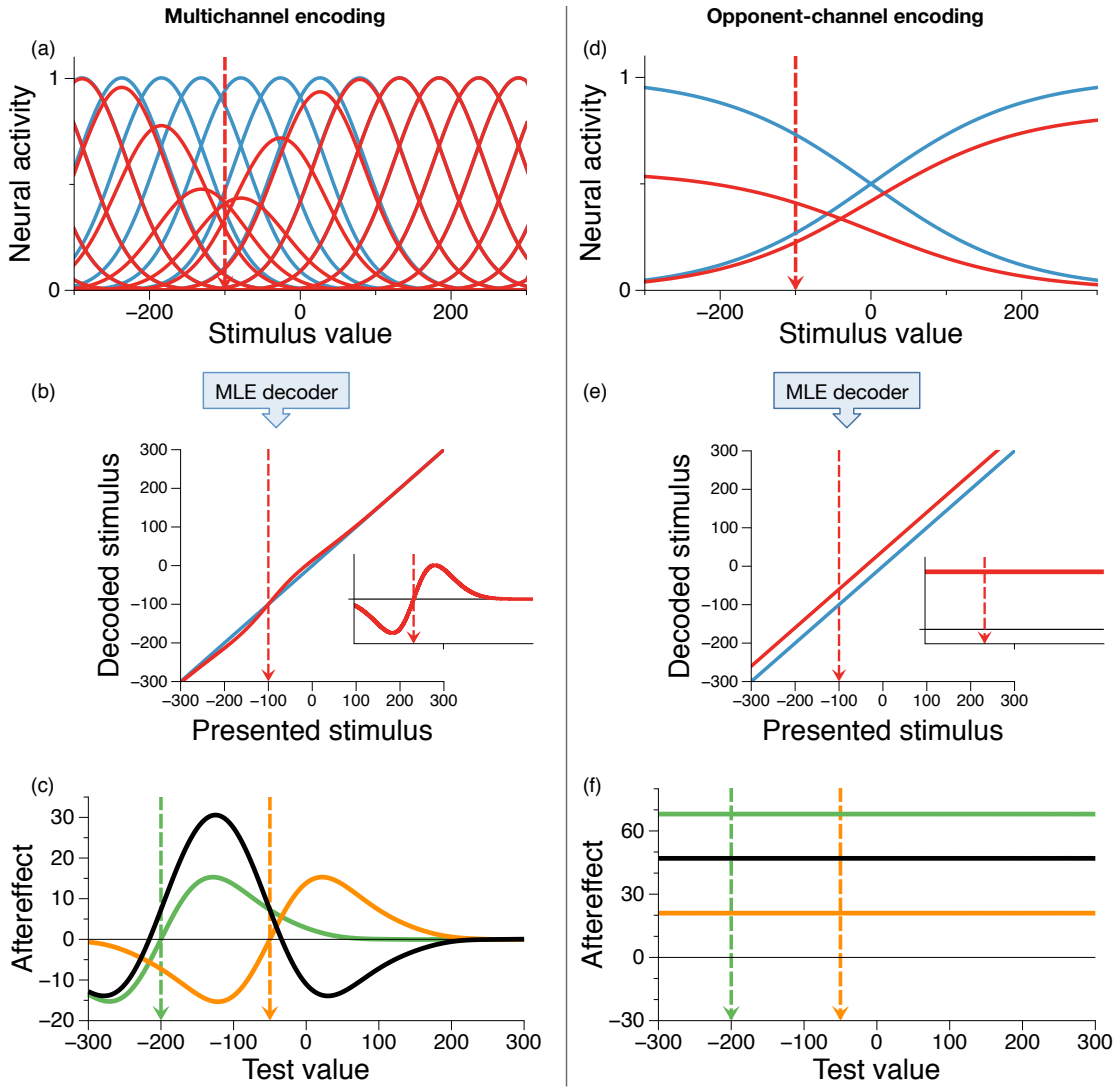


Figure 8. *Simulations of channel structure and adaptation in (left) a multichannel, and (right) an opponent code.* ***(a)*** *The response of each channel in a multichannel code to each stimulus value, before (blue) and after (red) adapting to a stimulus value of -100 (arbitrary units), shown by vertical dashed red line. Maximum-likelihood estimation (MLE) can be used to decode* ***(b)*** *the most probable value of each stimulus, given the channel activity it elicits, before (blue) and after (red) adaptation. The aftereffect (b, inset) is calculated as the post-adaptation decoded stimulus minus the pre-adaptation decoded stimulus. The aftereffect manifests as a local repulsion of test stimuli away from the adapted value, with no change in the appearance of the adapted value.* ***(c)*** *The relative aftereffect (black) predicted at each test value is given by the difference between the two location-specific aftereffects (green curve minus orange curve).* ***(d)*** *Channel responses in an opponent code before and after adaptation.* ***(e)*** *MLE performed on the upper and lower channel activity veridically decodes stimulus values before adaptation (blue), and predicts a constant bias at all test values after adaptation (red).* ***(e, inset)*** *The aftereffect manifests as a uniform renormalisation, in which the appearance of all test stimuli is biased in the same direction by the same amount.* ***(f)*** *The predicted relative aftereffect between two differently-adapted locations (black) is also uniform across a wide range of test values.*

non-uniform changes, with reduced aftereffects for extreme test values near the limits of the encoded range (see Appendix 1 for an example of a model which predicts this pattern of non-uniform renormalisation). We do not think this possibility could account for our findings for two reasons. First, Upper and Lower Standard values were not more 'extreme'

than Central Standards in our experiments. In both the facial identity and gender experiments, adapting and test stimuli were centred about a point offset from the computational average along the stimulus dimension. In the facial identity experiment, the Upper Standard stimulus was closest to the computational average, yet aftereffects were smaller for this standard than for the Central Standard (see Figure 4b).

Second, in our two facial gender experiments we centred adapting and test stimuli about two different points along the same stimulus continuum, neither of which was the computationally average androgynous face (see Figure 4c and 4d). In both cases we found that aftereffects were greatest for the test value lying midway between the two adapted values (although this pattern only produced a significant quadratic trend for female-face-centred data). This suggests that aftereffect magnitudes were tied to the positions of test stimuli *relative* to adapting stimuli, rather than being determined by the absolute positioning of test stimuli along the stimulus dimension. This is more consistent with a locally-repulsive aftereffect than with a renormalising aftereffect that is reduced for extreme test stimuli.

The locally-repulsive pattern of aftereffects we observe is consistent with a model in which face aftereffects arise from local interactions, such as those that govern the tilt aftereffect (Dickinson et al., 2010; Dickinson & Badcock, 2013; Dickinson et al., 2012). However, as yet it is unclear how this model can account for the transfer of aftereffects between differently-sized adapting and test stimuli. An alternate possibility is that our data are driven by adaptation within a multi-channel-structured representation of facial attributes.

Rather than committing to a particular instantiation of a neural encoding scheme (any form of which would almost certainly be an oversimplification), the key message of our data is that facial identity and gender aftereffects are better described as locally-repulsive than as renormalising aftereffects. If face adaptation involves a re-centring of all stimuli around a new neutral point, as has often been proposed (Burton et al., 2015; Jeffery et al., 2010, 2011; Leopold et al., 2001; McKone et al., 2014; Pond et al., 2013; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins et al., 2007; Susilo et al., 2010; Webster & Maclin, 1999), then an aftereffect measured over a wide range of test values should be uni-directional and approximately equal in magnitude. If this were the case, we should have found approximately equal relative aftereffects for different Standard test values in our spatial-comparison task. This is not what we found. Instead, aftereffects were greatest when tests differed from, and therefore could be influenced by, both adaptors. Aftereffects were smaller when tests were identical to one of the two adaptors and, according to the local repulsion hypothesis, should therefore not be affected by that adaptor.

6.6.1 *Isn't there lots of other evidence in favour of opponent coding?*

Although there are several lines of evidence hinting that face aftereffects involve renormalisation, all of these are currently equivocal. In the Introduction, we described the 'adaptor strength' paradigm (Jeffery et al., 2010, 2011; McKone et al., 2014; Pond et al., 2013; Zhao et al., 2011), and explained why the results of this paradigm are inconclusive. We also described a paradigm previously proposed by us (Storrs & Arnold, 2012), in which participants classify a range of stimuli in a three-category classification task. When tested in this paradigm, the facial gender aftereffect matched the predictions of a locally-repulsive aftereffect, as in the present report. Facial distortion aftereffects, however, better matched a re-normalising pattern (see below for further discussion).

Two other paradigms designed to dissociate locally-repulsive from renormalising aftereffects deserve mention. In the first, identity aftereffects are compared after adapting and testing along facial identity trajectories that pass through the centre of a hypothetical face space, as compared to facial identity trajectories that do not (Anderson & Wilson, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006). Aftereffects were found to be larger along the former than the latter type of trajectory, which was thought to demonstrate that the norm plays a special role in facial encoding. However, there is some recent evidence that both multichannel and norm-based models may predict this pattern of results (Ross et al., 2013).

In a second paradigm, changes in a three-category classification task are compared after adapting either to a neutral stimulus value, or to alternating positive and negative stimulus values. For instance, participants might classify faces as "looking left," "looking straight ahead," or "looking right" before and after adapting either to faces that look straight ahead, or to alternating faces that look left and right (Calder et al., 2008). A multichannel code can predict that, after consistent neutral-adaptation, the range of stimuli placed in the central category should narrow, and that after alternating negative and positive adaptation, the range of central stimuli should broaden. A simple opponent-channel scheme predicts no change in response pattern. Eye gaze direction (Calder et al., 2008) and head direction aftereffects (Lawson et al., 2011) both followed the pattern of changes predicted by a multichannel code when tested in this paradigm. Facial expression aftereffects, however, produced a narrowing of the central range in both conditions (Burton et al., 2015), which was argued to be more consistent with an opponent than multichannel code.

There are important caveats to conclusions based on data from the above paradigm. First, the predictions of an opponent-channel scheme are unclear, as additional assumptions must be made (e.g. that the response functions of adapted channels steepen or flatten) in order to predict a change in categorisation decisions after either neutral or alternating adaptation (Burton et al., 2015; Calder et al., 2008; Lawson et al., 2011). Second, repeated exposures to an 'adapting' stimulus might alter the range of stimuli placed in the middle category, not because such exposure has changed the appearance of test stimuli, but because participants

adopt the repeatedly-seen adaptors as exemplars against which other inputs can be judged when applying the middle category label. Both factors combine to dictate that any result from this paradigm cannot be regarded as conclusive evidence either of multichannel or opponent-channel encoding.

Perhaps the strongest evidence for renormalisation in face aftereffects comes from experiments involving geometrically distorted images. Webster and Maclin (1999) had participants remember a distorted face, and then adjust a test to match this remembered face after adapting to either a distorted or to an undistorted face. They found that adapting to an undistorted face had no effect on participants' matches to distorted faces. They inferred from this that the undistorted face played a special role in perception. In general support of this conclusion is the observation, given anecdotally by Webster and Maclin (1999) and quantified by us and others since (Rhodes et al., 2003; Robbins et al., 2007; Storrs & Arnold, 2012), that during adaptation a distorted face may come to appear more normal. We, however, found that this did not generalise to facial gender adaptation (Storrs & Arnold, 2012). One possible interpretation of this is that distortion adaptation does not tap facial mechanisms specifically, but instead is akin to adapting to a geometric distortion, such as during prism adaptation (Redding, Rossetti, & Wallace, 2005; von Helmholtz, 1909).

6.6.2 *Caveats*

While our data speak to the computational processes underlying face adaptation, they do not provide insight into where these computations might take place.

It is worth emphasising that our spatial-comparison paradigm measures *differences* in perception between two differently-adapted regions. It would therefore be insensitive to any adaptation that uniformly affected all locations in the visual field. Unfortunately psychophysics lacks methods to measure global biases that do not confound perceptual with decision-level effects (Morgan, 2014; Morgan et al., 2013; Storrs, 2015a). On the basis of our data, we cannot therefore preclude the possibility that there is a retinally global face adaptation effect that impacts all face perceptions equally, no matter where in the visual field a face might be presented. However, since facial identity and gender aftereffects as measured by a standard binary-classification task are retinotopically localised (Afraz & Cavanagh, 2009; Afraz & Cavanagh, 2008), we believe our data speak to the same effects as reported in previous face adaptation studies, and that our data and these previous reports are driven by a common computational process.

We would like to stress one final caveat in relation to our testing paradigm. Stimuli presented in the first test interval might have induced adaptation, affecting the appearance of stimuli presented in the second, as facial aftereffects can be induced by 'adapting' for as little as 200ms (Fang & He, 2005). This influence should, however, be smaller than that

induced by our more prolonged adaptors, as facial identity aftereffects reportedly increase logarithmically with adaptation duration (Leopold et al., 2005). Moreover, any interactions between sequential test stimuli would also have been present during baseline trials, and so would not systematically impact our aftereffect data. In future, variance arising due to interactions between test stimuli could be mitigated by introducing a second adaptation period on each trial in-between the two test intervals.

6.6.3 *Conclusion*

Our data add to the growing body of evidence (Ross et al., 2013; Storrs & Arnold, 2012; Zhao et al., 2011) that an opponent-channel model predicting renormalising aftereffects does not well describe the perceptual shifts following face adaptation. Instead, facial identity and gender aftereffects are better described as locally-repulsive aftereffects, in which differences between adapting and test stimuli are exaggerated in all directions, with little or no change in the appearance of the adaptor. This is consistent with an encoding model in which multiple channels are tuned for preferred values, with no value having a special role. Face aftereffects therefore behave similarly to other figural aftereffects (Badcock et al., 2014; Blakemore et al., 1970; Köhler & Wallach, 1944; Mitchell & Muir, 1976), perhaps suggesting common processes underlie the encoding of spatial patterns throughout the visual hierarchy.

7

CHAPTER 7: DISCUSSION AND CONCLUSIONS

Experiments presented in the preceding chapters used visual aftereffects to explore how the brain represents simple, intermediate, and complex spatial forms — edges, shapes, and faces. In each of these domains, I tested the proposal that spatial forms are represented relative to normative values by evaluating whether adaptation induces perceptual renormalisation. In Chapter 2, I replicated previous evidence for renormalisation in orientation perception, but showed that the same pattern of data is also found for very brief 'adaptation' durations, and is more parsimoniously explained by well-known anisotropies in orientation discrimination sensitivity. In Chapter 3 I established that shape aspect ratio aftereffects involve relatively late-stage shape representations, after shape constancy calculations have been completed. In Chapter 4 I went on to show that this high-level aspect ratio adaptation exaggerates differences between adapting and test aspect ratios, rather than re-centering perception around a new normative aspect ratio. Using two novel psychophysical protocols, Chapters 5 and 6 showed that facial gender and identity aftereffects follow a similar pattern, exaggerating differences between successive faces, and are therefore not consistent with a norm-based encoding of these attributes.

Taken together, these results suggest a preponderance of locally-repulsive adaptation aftereffects in spatial vision. I found no evidence for renormalisation except in the case of facial distortion (Chapter 6; discussed further in Section 2.3.1 below). This is surprising, given the long history and growing popularity of theories favouring norm-based representation in spatial vision (Gibson, 1933; Leopold et al., 2006, 2001; Regan & Hamstra, 1992; Rhodes et al., 2005; Webster & Maclin, 1999). In this final chapter I review the theoretical implications and methodological contributions of work in this thesis, before highlighting some of the outstanding questions regarding spatial form representation and adaptation.

7.1 THEORETICAL IMPLICATIONS OF THE PRESENT WORK

7.1.1 *Aftereffects provide little evidence for norm-based representations in spatial vision*

The idea that aftereffects in spatial vision arise from the recalibration of neutral normative values, relative to which the brain represents current input, dates back at least to Gibson (1933). Norm-based representations have been proposed to underlie the perception of orientations (Gibson & Radner, 1937; Regan & Beverley, 1985; Vaitkevicius et al., 2009), curvatures (Gibson, 1933; Poirier & Wilson, 2006), shapes (Kayaert et al., 2005; Regan & Hamstra, 1992; Suzuki, 2005), and faces (Anderson & Wilson, 2005; Leopold et al., 2001; McKone et al., 2014; Pond et al., 2013; Rhodes et al., 2005; Susilo et al., 2010; Valentine, 1991; Webster & Maclin, 1999). Such proposals have sometimes been formalised by encoding models in which opposing pairs of channels respond monotonically to increasingly non-neutral stimulus values (Giese & Leopold, 2005; Regan & Hamstra, 1992; Rhodes et al., 2005; Susilo et al., 2010). A key prediction of opponent-channel schemes is that adaptation should cause the adapted value to appear more neutral after adaptation, with approximately uniform concomitant shifts of other stimulus values (see Webster and MacLeod (2011), and modelling in Appendix 1). Data in this thesis demonstrate that this prediction is not supported in the domains of orientation, shape or face perception, and that norm-based opponent-channel encoding is therefore unlikely to represent a predominant encoding strategy in spatial vision.

Locally-repulsive aftereffects measured in this thesis for shape aspect ratio, facial gender, and facial identity were consistent with these domains being encoded by multiple channels relatively narrowly-tuned for specific instances within relevant sensory domains. Sensory cells in many domains are found to exhibit bandpass tunings for particular preferred values. Bandpass tuning has been reported in channels encoding orientation and spatial frequency (reviewed in De Valois and De Valois (1980)), motion speed (Lagae, Raiguel, & Orban, 1993), motion direction (Georgopoulos et al., 1986), auditory frequency (Phillips & Irvine, 1981) and odour (Mori, Nagao, & Yoshihara, 1999). It is therefore reasonable to expect that complex feature detectors might also exhibit relatively narrow non-monotonic tunings over their input spaces (Riesenhuber & Poggio, 2002), as is suggested by data in this thesis.

Previous evidence for renormalisation underlying spatial aftereffects has been inconclusive. In the case of orientation perception, a long history of research has led to an understanding of the tilt aftereffect as predominantly a local repulsion away from the adapted orientations (Clifford et al., 2000; Howard, 1982; Mitchell & Muir, 1976), perhaps with a small renormalising component (Held, 1963; Prentice & Beardslee, 1950; Templeton, 1972). In Chapter 2, the most compelling recent evidence for renormalisation (Müller, Schillinger, et al., 2009) was re-interpreted as arising from a methodological artifact involving the oblique effect, leading to the conclusion that there is at present little reason to think that the tilt aftereffect

involves any renormalising. Experiments in curvature perception have likewise concluded that curvature aftereffects involve local repulsion rather than renormalisation (Gheorghiu & Kingdom, 2007, 2008). For more complex shape properties, such as aspect ratio, the prediction of renormalisation has been tested only twice, and both studies have so far only been reported via conference abstracts. The first study (Suzuki & Rivest, 1998) found evidence for renormalisation, but used extremely brief 'adaptation' durations (150ms). The second used more conventional adaptation times and found evidence supporting local repulsion of aspect ratio (Badcock et al., 2014), consistent with the results of Chapter 4. The weight of evidence across orientation, curvature, and aspect ratio is therefore in favour of locally repulsive aftereffects with no evidence for any influence of renormalisation.

Evidence for renormalisation of face perception is more substantial, and has been reviewed in the Introduction, and in Chapters 5 and 6. There I argued that even this evidence is inconclusive. In the face aftereffect literature, spatial comparison tasks have not been widely used, presumably because face aftereffects are not tightly retinotopically localised (Afraz & Cavanagh, 2008; Zhao & Chubb, 2001). This is unfortunate, since spatial comparison methods allow the experimenter to measure aftereffects for diagnostic test values, including the adaptor and putatively more 'extreme' values. Instead, several alternate methods have been used to try differentiate local repulsion from renormalisation using method-of-single-stimulus classification tasks. For several of these paradigms, renormalisation and local repulsion predict qualitatively similar data. These include methods in which aftereffect magnitudes are compared following adaptation to increasingly 'extreme' adaptors (Jeffery et al., 2010, 2011; McKone et al., 2014; Pond et al., 2013), or to faces drawn from trajectories that pass through a putative norm vs trajectories that do not (see Ross et al. (2013)). A ternary-classification method, in which changes in category boundaries are compared after adapting to a neutral value vs adapting to alternating opposite extreme values, has provided evidence for local repulsion of facial viewpoint (Lawson et al., 2011) and gaze direction (Calder et al., 2008). Data from facial expression adaptation (Burton et al., 2015) did not clearly fit predictions of either proposal. Data in this thesis suggest facial gender and identity aftereffects are predominantly locally repulsive in nature.

Perhaps the most compelling evidence for renormalisation comes from experiments using multiple-point rating scales to probe observers about the appearance of test faces, before and after adaptation (O'Neil et al., 2014; Rhodes et al., 2003; Robbins et al., 2007). Multi-point rating scales afford higher resolution than more common binary classification tasks, and are therefore potentially sensitive to changes in the appearance of any test value in which the experimenter is interested. Upon reanalysis, data from O'Neil et al. (2014) appear to lend slightly more support to a locally-repulsive model of facial age aftereffects than a renormalising model (see Appendix 2), although this depends on how one chooses to constrain the model parameters (O'Neil et al., 2015). Neither model explains more than a couple of percent of the aftereffect variance in this particular dataset. However, studies

of facial distortion adaptation have provided evidence for renormalisation. After adapting to a face in which eye height (Robbins et al., 2007) or overall configuration (Rhodes et al., 2003) has been manipulated, the adapting faces (as well as more extremely distorted versions) were rated as appearing less distorted than when viewed before adaptation. This is particularly interesting in light of the mixed results in Chapter 5, wherein facial gender aftereffects appeared to involve local repulsion while facial distortion aftereffects were consistent with an influence of renormalisation. The possibility that different facial attribute dimensions might be encoded by different strategies is considered below in Section 7.3.1.

7.1.2 *Face aftereffects support exemplar-based theories of facial representation*

A key theoretical debate in the face perception literature over the past few decades is between exemplar-based and norm-based models of facial recognition (Burton & Vokey, 1998; Byatt & Rhodes, 1998; Jiang, Blanz, & O'Toole, 2006; Lewis, 2004; Loffler et al., 2005; Macke & Wichmann, 2010; Valentine, 1991; Wallis, Siebeck, Swann, Blanz, & Bulthoff, 2008). In the norm-based proposal, units underlying facial representation are conceptualised as vectors, pointing away from the origin of a perceptual face space, encoding the ways in which any given input differs from a norm (Valentine, 1991; Valentine et al., 2015). The norm is the central tendency of all faces in face space, and corresponds to the average experienced face. The perceived identity of a face is determined by the vector's direction, while its magnitude gives the strength of evidence for that identity (Lewis, 2004; Rhodes, Brennan, & Carey, 1987; Valentine, 1991). In the exemplar-based alternative, units underlying facial representation are thought of as fuzzy templates, encoding the similarity of any input relative to multiple points within face space — with the multiple points regarded as 'exemplar' faces (Lewis, 2004; Valentine, 1991). The perceived identity of an input is determined by the best-matching exemplar (various definitions of 'best match' have been suggested; see Lewis (2004) and Wallis (2013)).

Exemplar-based and norm-based theories of facial representation have their roots in an ongoing debate in the categorisation and memory literature, between those who propose that new instances (of a potential category member, for example) are compared against individual previously-experienced instances (Medin & Schaffer, 1978; Nosofsky, 1986, 1988), and those who propose that new instances are compared against prototypes abstracted from many individual experiences (Goldstein & Chance, 1980; Hintzman, 1986; Knowlton & Squire, 1993; Palmer, 1975; Reed, 1972). The norm-based model of facial representation is inspired by prototype models, in that all faces are represented in terms of their deviation from an abstracted facial prototype. It differs from classical prototype models in that not only the magnitude, but also the *direction*, by which the input differs from the prototype is encoded (Valentine, 1991).

Many findings that were presented as evidence for a norm-based representation of faces were later identified as also being consistent with exemplar-based theories. In the caricature effect, exaggerating the distinctive features of a familiar face can render that face easier to recognise (Perkins (1975) and Benson and Perrett (1991); although see Kaufmann and Schweinberger (2008) for a failure to replicate this effect). The caricature effect follows intuitively from the norm-based theory, because exaggerating a face's distinctive features is equivalent to increasing the length of the identity vector, and therefore the evidence for that identity. However, computational simulations (Lewis, 2004; Wallis, 2013) have shown that an exemplar-based model can also predict recognition advantages for caricatures, because these exist in more sparsely-occupied regions of exemplar space and are therefore less easily confused with other identities. The other-ethnicity effect (Shepherd, Deregowski, & Ellis, 1974) refers to the greater discrimination sensitivity people tend to have for faces drawn from their own ethnic group than from another group. The other-ethnicity effect is intuitively consistent with norm-based representation, in which all faces of an unfamiliar identity differ from the own-ethnicity norm in similar ways, making them difficult to differentiate (Valentine, 1991). However, the effect is also consistent with exemplar-based representations, in which discrimination sensitivity is related to the density of stored exemplars within a region of face space, and faces of unfamiliar ethnicities occupy a sparsely-sampled region (Wallis, 2013).

In recent years the debate between norm- and exemplar-based theories of facial representation has been re-invigorated by evidence from face aftereffects (Leopold et al., 2001; Rhodes et al., 2005; Valentine et al., 2015; Webster & MacLeod, 2011). By reframing the theories in terms of opponent-channel and multi-channel neural population codes, researchers have hoped to gain traction on the understanding of how faces are represented (McKone et al., 2014; Pond et al., 2013; Rhodes et al., 2005; Susilo et al., 2010; Webster & MacLeod, 2011). The models I have used to generate experimental predictions (see modelling Appendix 1) follow those described in this recent incarnation of norm- and exemplar-based theories. To the extent that face aftereffects reflect adaptation within face-selective channels (see Section 2.3.3 below), data in this thesis support exemplar-based theories of facial representation.

As well as theoretical implications, the present findings are relevant to the use of face aftereffects to study developmental and clinical questions. Reduced face aftereffects in children with autism spectrum disorder have been taken as evidence for deficiencies in norm-based face encoding in such groups (Ewing, Leach, Pellicano, Jeffery, & Rhodes, 2013; Pellicano, Jeffery, Burr, & Rhodes, 2007; Rhodes, Jeffery, Taylor, & Ewing, 2013; Rhodes, Maloney, et al., 2007; Walsh, Maurer, Vida, Rhodes, & Jeffery, 2015). Similarly, the early development of face aftereffects in children has been taken as evidence for the early maturation of norm-based facial encoding (Burton et al., 2013; Jeffery et al., 2010, 2011; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; Pimperton, Pellicano, Jeffery, & Rhodes, 2009). If face

aftereffects do not arise from an inherently norm-based representational system, but from an exemplar-based representation, these findings will need to be re-interpreted.

7.2 METHODOLOGICAL CONTRIBUTIONS OF THE PRESENT WORK

7.2.1 *New psychophysical methods to differentiate renormalisation from local repulsion*

Face aftereffect experiments have predominantly used tasks in which a single stimulus is classified as belonging to one of two categories, before and after adaptation. This method has two problems, the first and primary one being that it is sensitive only to changes in perception near a neutral category boundary, where local repulsion and renormalisation make qualitatively identical predictions. I have proposed two widely applicable psychophysical protocols to test for local repulsion vs renormalisation. In Chapter 5, I developed a novel ternary classification task, combined with adaptation to one of the category boundaries, to measure changes in the appearance of the adaptor. In Chapter 4 I used a standard spatial comparison task to compare perceived aspect ratio between adapted and unadapted locations. In Chapter 6 I modified the spatial comparison task to accommodate the weaker spatial localisation of face aftereffects, by comparing perception between two *differently-adapted* locations. All three methods facilitated measuring perception at diagnostic test values: the adaptor itself (Chapters 5 and 6) and at points putatively more 'extreme' than the adaptor (Chapters 4 and 6), thereby deriving qualitatively distinct predictions from renormalisation and local repulsion.

Such methods will be instrumental in evaluating the evidence for norm-based representation in many domains not considered in this thesis. In recent reports, aftereffects have been proposed as evidence for norm-based representation of natural scenes (Greene & Oliva, 2010), body identity (Rhodes, Jeffery, Boeing, & Calder, 2013), and, in audition, vocal gender (Latinus, McAleer, Bestelmeyer, & Belin, 2013). These aftereffects have so far been demonstrated only as biases in the perception of putatively 'neutral' stimuli, and so do not yet provide evidence for either renormalisation or local repulsion. Before a conclusion can be drawn in this regard, it will be necessary to apply more diagnostic tests such as those demonstrated in the preceding chapters.

7.2.2 *Bias-minimising methods applied to high-level aftereffects*

A second problem with methods involving classification of a single stimulus is that the location of the category boundary is determined both by the participant's sensory evidence and by their criteria for applying each of the response labels to that evidence (Farell & Pelli, 1999; Green & Swets, 1966; Kingdom & Prins, 2010). A change in perception and a change

in decision criterion can produce exactly the same pattern of response shifts, which has led many psychophysicists to discourage the use of method-of-single-stimulus classification tasks (Gescheider et al., 1970; Green & Swets, 1966; Morgan et al., 2013; Morgan et al., 2012; Storrs, 2015a; Yarrow et al., 2011). This is an important and under-addressed issue: if high-level aftereffects are decisional biases rather than perceptual biases, they may all have a similar origin within amodal cognitive processes and tell us little about the initial encoding of any particular stimulus dimension. This concern applies also to experiments presented in Chapter 3 (which employed a binary classification task to measure perceived shape aspect ratio) and Chapter 5 (which employed ternary classification tasks to measure perceived colour, facial gender, and facial distortion).

One way to dissociate response biases from perceptual biases is to use a two-alternative forced-choice (2AFC) task, provided there exists an unadapted or differently adapted location in the visual field. Isolating perceptual bias is still not straightforward. For example, if one were to show two shapes with identical aspect ratios in adapted and unadapted retinal locations, and ask "which is more elongated?" (a simple 2AFC), a strategy of picking the stimulus in the adapted location if unsure could masquerade as a perceptual bias (Jogan & Stocker, 2014; Morgan, 2014; Morgan et al., 2013; Schneider & Komlos, 2008). Such problems can be alleviated by elaborations to the forced-choice task, such as varying the standard stimulus from trial to trial so that a perceptual bias predicts different PSE shifts for different standard stimuli (Morgan, 2014; Morgan et al., 2013), or by presenting two standard stimuli, both in unadapted locations, from which the participant selects the one most similar to a test shown in the adapted location (Jogan & Stocker, 2014).

The 2AFC task elaboration adopted in Chapters 4 and 6 was to use a forced-choice double-pair task (e.g. "in which interval was the identical pair of shapes presented?"). Because stimuli are presented in *both* spatial locations in *both* intervals, the observer can no longer favour stimuli on the basis of spatial location. This double-pair task has the additional advantage that the dimensions along which stimuli are varied need not be easily expressed in language. For example, in the facial identity experiment in Chapter 6, participants could identify which interval contained *more similar* faces, without having to first learn the identities of particular faces to answer a question like "which interval contained the face most like Bob's?"

Although an 'objective' measure of subjective appearance is impossible, the forced-choice double-pair tasks with multiple interleaved standard stimuli in Chapters 4 and 6 de-confounding changes in appearance from irrelevant response biases, such as choosing stimuli in the adapted location if in doubt. One obvious downside of these methods is that they are constrained to measuring spatially-localised adaptation. They would therefore fail to detect any global component of a perceptual aftereffect. Although we have no reason to believe the facial gender and identity aftereffects reported in Chapter 6 differ qualitatively from those measured by other methods, it remains possible that there is an additional global com-

ponent that went undetected in this experiment. The method proposed in Chapter 5 is not limited to measuring spatially localised components, and could be applied to measuring cross-modal or non-visual aftereffects. However, as with other method-of-single-stimulus approaches, caution must be taken in interpreting data, as shifts in category boundaries due to perceptual biases cannot be differentiated from those involving criterion shifts.

7.3 OUTSTANDING QUESTIONS

7.3.1 *Are different facial dimensions encoded in different ways?*

It is important to note that this thesis did not reveal unanimous evidence for locally-repulsive aftereffects. The exception was facial distortion adaptation (Chapter 5) which, when measured via a ternary classification task, appeared more consistent with renormalisation than local repulsion. Interestingly, some of the strongest previous evidence for facial renormalisation also comes from distortion adaptation (Rhodes et al., 2003; Robbins et al., 2007; Webster & Maclin, 1999). These data are therefore consistent with the visual system using different schemes to encode different properties, but are these properties necessarily facial?

Distortion adaptation might not tap facial mechanisms specifically, but rather be akin to a global visual geometric distortion, as experienced during prism adaptation (Redding et al., 2005; von Helmholtz, 1909). The distortions used in Chapter 5 and by Rhodes et al. (2003) and by Webster and Maclin (1999) mimic the effect of viewing an image through a concave or convex lenses. Distortion aftereffects are not specific to faces, but can also be induced between differently-distorted photographs of various natural scenes and textures (Maclin & Webster, 2001). The ability of the visual system to compensate for certain geometric distortions might involve a form of renormalisation, a possibility that warrants further exploration.

A second possibility is that there is an interaction between low-level properties of adapting and test images. Adaptation to a globally distorted image can be thought of as adaptation to a systematically biased orientation field (Dickinson et al., 2010, 2012). If the adaptor and test were presented at the same size and retinal location, one would expect this field of tilt aftereffects to manifest as a local repulsion along the facial-distortion dimension (Webster & MacLeod, 2011). As tests were smaller than adaptors in the relevant experiments, interactions between orientations in the two images would have manifested in a more complicated fashion.

Finally, facial distortion dimensions might be more susceptible to criterion shifts than dimensions involving face-specific properties, such as gender or identity. Evidence for renormalisation following distortion adaptation so far comes only from paradigms in which

changes in perception and shifts in criteria can both contribute to the measured aftereffect (i.e. the ternary classification task in Chapter 5, and the multi-point rating scales in Robbins et al. (2007) and Rhodes et al. (2003)). After repeated exposure to a slightly distorted face, participants might change their initial criterion for what they are willing to classify as 'normal.' To test this possibility, it would be fruitful to apply the bias-minimising spatial-comparison methods of Chapters 4 and 6 to facial distortion aftereffects.

7.3.2 *Would renormalisation occur under different experimental circumstances?*

The experiments presented here were intended to test the prediction that prolonged exposure to a single non-neutral value would induce renormalisation of the relevant perceptual dimension. This prediction is made by theories in which encoding mechanisms explicitly signal the manner in which an input deviates from a normative value (McKone et al., 2014; Pond et al., 2013; Regan & Hamstra, 1992; Rhodes et al., 2005; Susilo et al., 2010; Webster & MacLeod, 2011; Zhao et al., 2011). Our results do not preclude the possibility that spatial vision might involve 'norms' in a less explicit fashion, and that these values might update during natural experience. For example, vertical and horizontal orientations have unique psychological and linguistic salience, unique behavioural meaning (due to gravity), and are accompanied by uniquely fine perceptual sensitivity (Appelle, 1972). The over-representation of the cardinal axes might well be learned from the statistics of our environments (Girshick et al., 2011), and alter under a different environment.

Many more complex dimensions also contain values that appear psychologically unique and can be thought of as non-explicit norms. When viewing photographs, there is a particular distribution of spatial frequencies that renders the image "in focus" to an observer. After viewing images with distributions that deviate from this, renormalisation can occur, such that the adapted image appears better-focused than it had initially (Elliott et al., 2011). Adaptation to a *single* spatial frequency, on the other hand, tends to induce a locally-repulsive aftereffect on single-frequency tests (e.g. Blakemore et al. (1970)). Elliott and colleagues (2011) showed that a standard multichannel model of spatial-frequency encoding can account for both results. They proposed that after adapting to natural images, the gains in spatial frequency channels adjust so that a natural image elicits approximately equal activation across frequency-tuned channels. This balanced activation corresponds to a subjectively "in focus" image. Adaptation to an image with a different frequency distribution readjusts the gains of channels so that the adapted image elicits more balanced activation after adaptation than it had initially. Multichannel codes can therefore predict locally repulsive aftereffects after narrowband adaptation, and 'recalibration' after broadband adaptation. Perception of an "average" face might correspond to a learned pattern of activation across face-selective channels, which can be updated by experience.

7.3.3 *How does adaptation in multiple substrates contribute to 'high-level' aftereffects?*

An image of a shape or face is a complex spatial pattern, and adaptation to the component local contrasts, spatial frequencies, and orientations likely contributes to changes in how the shape or face is perceived. Direct evidence for the inheritance of adaptation across the visual hierarchy comes from electrophysiology (Dhruv & Carandini, 2014) and from demonstrations that adapting to simple stimuli can alter the appearance of more complex stimuli. For example, adapting to an upwards-curved line can make a neutrally-expressioned face appear sad (Xu et al., 2008); adapting to a "T" shape with a long stem can cause participants to judge the height of the eyes on a face to be unusually low (Susilo et al., 2010); and adapting to a field of oriented lines can alter the apparent emotional expression of a face (Dickinson et al., 2012).

Nevertheless, 'high level' aftereffects cannot be attributed solely to low-level adaptation. Whereas tilt aftereffects are tightly retinotopically localised (Gibson, 1937; Knapen et al., 2010), shape aftereffects can survive retinal location differences of several degrees of visual angle between adapting and test stimuli, as illustrated in Chapter 4 where spatially-jittering adaptors induced robust aspect ratio adaptation (see also Regan and Hamstra (1992), Suzuki (2005) and He, Kersten, and Fang (2012)). Results in Chapter 3 demonstrate that the shape aspect ratio aftereffect is tuned not simply for the retinal shape, but is also affected by perceived shape after the completion of shape constancy operations, suggesting a reasonably late locus of adaptation. Face aftereffects can survive substantial differences between adapting and test images, not only in terms of position (Afraz & Cavanagh, 2008), but also in terms of size (Leopold et al., 2001; Webster et al., 2004; Yamashita et al., 2005; Zhao & Chubb, 2001), orientation (Watson & Clifford, 2003), facial expression (Fox, Oruc, & Barton, 2008) and identity (Fox & Barton, 2007; Lai, Oruç, & Barton, 2011; Walton, 2012). In experiments in Chapters 5 and 6, retinotopically-local adaptation was mitigated by introducing a size change between adapting and test stimuli. This is a standard precaution in the face aftereffect literature, although it is worth noting that it by no means isolates face-specific adaptation effects. Local adaptation to colour, contrast, orientation and spatial frequency might have complex systematic effects on a differently-sized or -positioned test (Blakemore & Over, 1974; Dickinson et al., 2010). Further, rescaling does not eliminate translation-tolerant aftereffects, including those induced by adaptation to curvature (Gheorghiu & Kingdom, 2008), shape (Regan & Hamstra, 1992), and texture (Motoyoshi, 2012). Stricter controls are therefore required in order to evaluate to what extent 'face' aftereffects measure face-specific adaptation.

A re-examination of this question is particularly warranted given findings in this thesis. The claim that face aftereffects manifest as renormalisation, rather than local repulsion, has previously been cited as evidence that face aftereffects cannot be explained entirely by adaptation to simpler image components (Webster & MacLeod, 2011). For example,

adaptation to the edge energy in the adapting image will induce a collection of orientation and spatial frequency aftereffects, which are known to be predominantly locally-repulsive. If a test face is presented at the same size and in the same location, then this accumulation of locally-repulsive local aftereffects should serve simply to exaggerate image differences between the adapting and test faces. This would manifest as a local repulsion within whatever face dimension the experimenter used. The vast majority of data in this thesis, spanning various forms of spatial adaptation, followed a locally repulsive pattern (as do tilt and shape aftereffects). This undermines one of the arguments against face aftereffects being explicable by an aggregation of local aftereffects.

7.3.4 *What function, if any, does adaptation serve in high-level vision?*

The data collected here do not speak to what, if any, *function* is served by high-level adaptation. Face adaptation has been described as "fitting the mind to the world" via renormalisation (Rhodes et al., 2005). Hypothetically this could ensure a person is best able to differentiate faces typical of recent perceptual history, akin to the function light adaptation serves by bringing the dynamic ranges of photoreceptors into useful bounds for the prevailing environment (Rieke & Rudd, 2009; Shapley & Enroth-Cugell, 1984; Webster, Werner, & Field, 2005).

The functional role of locally repulsive aftereffects is less apparent. Indeed, such perceptual biases might seem maladaptive, perhaps arising when an upstream decoder is 'unaware' of adaptation-induced changes at sensory encoding stages a phenomenon referred to as a "coding catastrophe" (Schwartz, Hsu, & Dayan, 2007; Seriès et al., 2009). One function of locally repulsive aftereffects might be to sharpen discrimination sensitivity about the adapted value. Improved discrimination have been reported in some studies (Clifford et al., 2001; Oruç & Barton, 2011; Regan & Beverley, 1985) but not others (Barlow et al., 1976; Rhodes, Maloney, et al., 2007; Westheimer & Gee, 2002). Another interesting and non-exclusive possibility is that locally repulsive aftereffects, by temporarily warping perceptual similarity, might facilitate attentional capture by novel scenes, objects, faces, and facial expressions (McDermott et al., 2010; Ranganath & Rainer, 2003). The functional role of adaptation in high-level vision remains an open question.

7.3.5 *How should the present findings be reconciled with evidence for norm-based coding from neuroimaging and electrophysiology?*

Functional magnetic resonance imaging (fMRI) and single-unit electrophysiological data are also cited as evidence for a norm-based encoding of faces. Loffler et al. (2005) presented schematic faces drawn from a multidimensional artificial face space, while using fMRI to

record blood oxygenation level dependent (BOLD) signal in the fusiform face area (FFA). The BOLD signal was *lowest* for faces near the centre of the artificial space, and scaled with facial distinctiveness. This is consistent with a norm-based representation of faces, in which neurons in FFA are minimally activated by an average face, and can be increasingly activated by faces that deviate from average.

There is complementary single-unit electrophysiological data. Leopold et al. (2006) presented macaques with images of realistic computer-generated human faces drawn from an artificial face space. Face-selective neurons tended to respond minimally to the face at the centre of the space, and increasingly as the face moved toward one or more distinctive identities. In simulations, these data could be more closely fit by a norm-based population model than by an exemplar-based model (Giese & Leopold, 2005). Freiwald et al. (2009) created a space of cartoon faces with nineteen independent feature dimensions (eye height, hair length, etc), and presented thousands of faces drawn from this space while recording from the macaque middle face patch. Cells that were sensitive to variations along individual feature dimensions tended to respond monotonically to increasing or decreasing values of that feature, rather than responding maximally to an average or intermediate feature value. These results again appear consistent with cells being tuned to deviations from a normative face.

Analogous results have been reported in the domain of shape perception. Neurons in monkey inferior temporal cortex have been reported that fire at an increasing rate as simple geometric shapes become increasingly tapered or curved (Kayaert et al., 2005), or as novel two-dimensional shapes move further from the average shape within a stimulus set (De Baene, Premereur, & Vogels, 2007). The evidence is equivocal, though, as other investigations have found relatively narrow tuning for particular preferred shapes (Pasupathy & Connor, 2002; Tanaka, 2003) or for particular facial view and gaze directions (see Barraclough and Perrett (2011)).

Intuitively, monotonic tuning for non-average values suggests a norm-based encoding scheme. However, there are complexities that render many of these results difficult to interpret. Short-term adaptation presents a potential confound in both fMRI and electrophysiology studies (Davidenko, Remus, & Grill-Spector, 2011; Jiang, Rosen, et al., 2006; Kahn & Aguirre, 2012). In fMRI (Grill-Spector & Malach, 2001) and single-unit recording experiments (e.g. Barlow and Hill (1963) and Miller, Gochin, and Gross (1991)), repeated presentation of a stimulus tends to elicit a response smaller than the initial presentation. Adaptation effects can therefore predict the pattern of responses reported — little response to average stimuli, which have the highest overlap of features with preceding other stimuli, but maintained responses to extreme stimuli, which have the lowest feature overlap with other stimuli. A compelling demonstration of this confound was provided by Davidenko et al. (2011), in an fMRI experiment using face silhouette images drawn from an artificial face space. When faces of a fixed distinctiveness level were randomly selected within each

experimental block, blocks containing more distinctive faces elicited greater activation. But when faces of a fixed distinctiveness level were selected in a way that controlled the amount of image variability within each block, highest activation was instead found for the most *prototypical* faces. It is not yet clear to what extent neuroimaging and electrophysiological evidence for norm-based neural tuning will survive careful controlling for adaptation confounds (Kahn & Aguirre, 2012).

7.4 TOWARD A FULLY-EXPLICIT MODEL OF HIGH-LEVEL REPRESENTATION AND ADAPTATION

Throughout this thesis, I have derived experimental predictions from neural population code models. To their credit, these models formalise ideas about encoding, adaptation, and decoding. For the sake of simplicity (and consistency with the descriptions of models in the literature), the models have elided many of the details required of a full model of spatial form representation and adaptation. I have concentrated on single dimensions within perceptual shape or face spaces, and have assumed these map monotonically onto constructed stimulus dimensions. I have also ignored the hierarchy of adaptable mechanisms intervening between the retinal image and the representation of complex attributes, such as 'gender.' Throughout, I have simulated adaptation as a simple fatigue, although there is good evidence for more complex effects on tuning curves and population interactions (Clifford et al., 2007, 2000; Kohn & Movshon, 2004; Solomon & Kohn, 2014).

To begin to fully understand spatial representations, and how they adapt with experience, we will need to move beyond these simplifications. A complete model of face perception, for example, should fully explicate successive processing stages intervening between image encoding and response behaviour (such as categorising the gender of the face). It should be flexible, in that the encoding stage admits the substitution of several different decoding stages corresponding to different behavioural tasks. Such a model should predict psychophysical aftereffect patterns for arbitrary combinations of adaptor and test images, as well as predicting other aspects of face perception — such as detection and classification speed and accuracy, discrimination sensitivity, and the various quirks of human face perception such as inversion effects (Yin, 1969), configural effects (Rossion, 2013; Young, Hellawell, & Hay, 1987) and other-ethnicity effects (Shepherd et al., 1974).

Image-computable hierarchical models of some aspects of face perception already exist (Farzmahdi, Rajaei, Ghodrati, Ebrahimpour, & Khaligh-Razavi, 2015; Hu et al., 2015; Jiang, Rosen, et al., 2006; Wallis, 2013), and are likely to advance rapidly in coming years. Object recognition by artificial systems has proven difficult, but in the past three years deep neural networks have been built which rival human performance in constrained object categorisation tasks (Krizhevsky, Sutskever, & Hinton, 2012; Russakovsky et al., 2014). These

models are engineering solutions, although comparisons of representations within deep neural networks and human and monkey high-level visual cortex suggest some overlap of computational strategies (Cadieu et al., 2014; Guçlu & van Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014). Some neural network models (i.e. radial basis function networks such as HMAX; Riesenhuber and Poggio (1999)) correspond mathematically to exemplar-based models in psychology (Jäkel, Schölkopf, & Wichmann, 2009). Others, such as those that are state-of-the-art in machine vision (Krizhevsky et al., 2012; Ren, He, Girshick, & Sun, 2015), do not correspond precisely to exemplar-based theory, but follow similar principles of detecting similarity to stored features, rather than deviation from an abstracted norm. Simulating adaptation and probing representations within such fully-explicit models might grant a clearer understanding of computations in biological visual systems.

7.5 CONCLUSIONS

Over the past century, the idea that the visual system represents spatial properties relative to normative values has recurred in various domains, including orientation (Gibson, 1937; Gibson & Radner, 1937), curvature (Gibson, 1933; Poirier & Wilson, 2006), aspect ratio (Regan & Hamstra, 1992) and face perception (Rhodes et al., 2005; Webster & Maclin, 1999). In each of these domains, 'opposite' aftereffects, such as the rightward tilt of a vertical line after viewing a left-tilted adaptor, have been taken as evidence of an inherent opponency underlying relevant neural representations. However, such distortions are equally consistent with repulsive aftereffects that exaggerate differences between successive stimuli, and which can arise in exemplar-based models with no explicit norm. More diagnostic tests, such as measures of how the adaptor and more putatively 'extreme' stimuli appear after adaptation, are therefore needed, but this has been under-appreciated.

This thesis addressed these issues by evaluating evidence for renormalisation of orientation, aspect ratio, and facial attributes, using diagnostic psychophysical methods, inspired by hypotheses based on neural population code modeling. In all domains (except facial distortion), I found evidence for local repulsion, rather than renormalisation. These data support models of encoding in which channels are relatively narrowly tuned to sub-regions within the domains they encode. A great deal remains to be understood about how spatial forms are represented in the brain, but it is encouraging to think similar computational principles might operate throughout spatial vision.

II

REFERENCES

REFERENCES

Adams, R. (1834). An account of a peculiar optical phenomenon seen after having looked at a moving body, etc. *London and Edinburgh Philosophical Magazine and Journal of Science*, *5*, 373–374.

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), 1–17.

Afraz, S.-R. & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, *48*(1), 42–54.

Afraz, S.-R., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology*, *20*(23), 2112–2116.

Allan, L. G. (1975). The relationship between judgments of successiveness. *Attention, Perception, & Psychophysics*, *18*(1), 29–36.

Anderson, N. D. & Wilson, H. R. (2005). The nature of synthetic face adaptation. *Vision Research*, *45*(14), 1815–1828.

Andrews, D. P. (1965). Perception of contours in the central fovea. *Nature*, *205*, 1218–1220.

Andrews, D. P. (1967). Perception of contour orientation in the central fovea Part I: Short lines. *Vision Research*, *7*(11), 975–997.

Anstis, S. M. & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. *Colour vision: Physiology and psychophysics*, 155–166.

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences*, *2*(3), 111–117.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278.

Arnold, D. H., Birt, A., & Wallis, T. S. A. (2008). Perceived size and spatial coding. *Journal of Neuroscience*, *28*(23), 5954–5958.

Badcock, D. R., Morgan, S., & Dickinson, J. E. (2014). Evidence for aspect-ratio processing independent of linear dimensions: A channel-based system. *Journal of Vision*, *14*(10), 1181–1181.

Baker, C. L. & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Progress in Brain Research*, *134*, 1–21.

Baker, D. H. & Meese, T. S. (2012). Size adaptation effects are independent of spatial frequency aftereffects. *Perception*, *41*, 33–33.

Barlow, H. B. & Hill, R. M. (1963). Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *200*, 1345–1347.

Barlow, H. B., Macleod, D. I. A., & Van Meeteren, A. (1976). Adaptation to gratings: no compensatory advantages found. *Vision Research*, *16*(10), 1043–1045.

Barraclough, N. E. & Perrett, D. I. (2011). From single cells to social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1739–1752.

Beale, J. M. & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239.

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

Bell, J., Dickinson, J. E., & Badcock, D. R. (2008). Radial frequency adaptation suggests polar-based coding of local shape cues. *Vision Research*, *48*(21), 2293–2301.

Bennett, P. J. & Cortese, F. (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research*, *36*(2), 233–238.

Benson, P. J. & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, *3*(1), 105–135.

Benton, C. P. & Burgess, E. C. (2008). The direction of measured face aftereffects. *Journal of Vision*, *8*(15), 1–8.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, *31*(10), 1–21.

Blakemore, C. T. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, *203*(1), 237–260.

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. & Over, R. (1974). Curvature detectors in human vision. *Perception*, *3*(1), 3–7.

Blakemore, C. & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science*, *166*(3902), 245–247.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Burbeck, C. A. (1987). Locus of spatial-frequency discrimination. *Journal of the Optical Society of America A*, *4*(9), 1807–1813.

Burton, M. & Vokey, J. R. (1998). The face-space typicality paradox: Understanding the face-space metaphor. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 475–483.

Burton, N., Jeffery, L., Calder, A. J., & Rhodes, G. (2015). How is facial expression coded? *Journal of Vision*, *15*(1), 1–13.

Burton, N., Jeffery, L., Skinner, A. L., Benton, C. P., & Rhodes, G. (2013). Nine-year-old children use norm-based coding to visually represent facial expression. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1261–1269.

Byatt, G. & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*, 2455–2468.

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., … DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12), e1003963.

Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a non-opponent multichannel system. *Journal of Experimental Psychology: General*, *137*(2), 244–261.

Campbell, F. W. & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *The Journal of Physiology*, *187*, 437–445.

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

Clifford, C. W. G. (2002). Perceptual adaptation: motion parallels orientation. *Trends in Cognitive Sciences*, *6*(3), 136–143.

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision Research*, *47*(25), 3125–3131.

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.

Clifford, C. W. G., Wyatt, A. M., Arnold, D. H., Smith, D. T., & Wenderoth, P. (2001). Orthogonal adaptation improves orientation discrimination. *Vision Research*, *41*(2), 151–159.

Coltheart, M. (1971). Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, *78*(2), 114–121.

Coren, S. & Festiniger, L. (1967). An alternative view of the "Gibson normalization effect". *Perception & Psychophysics*, *2*(12), 621–626.

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, *75*(3), 485–491.

Daelli, V. (2011). High-level adaptation aftereffects for novel objects: The role of pre-existing representations. *Neuropsychologia*, *49*(7), 1923–1927.

Daelli, V., van Rijsbergen, N. J., & Treves, A. (2010). How recent experience affects the perception of ambiguous objects. *Brain Research*, *1322*(100), 81–91.

Davidenko, N., Remus, D. A., & Grill-Spector, K. (2011). Face-likeness and image variability drive responses in human face-selective ventral regions. *Human Brain Mapping*, *33*(10), 2334–2349.

Day, R. H. (1962a). Excitatory and inhibitory processes as the basis of contour shift and negative after-effect. *Psychologia*, *5*, 183–193.

Day, R. H. (1962b). Excitatory and inhibitory processes as the basis of contour shift and negative after-effect. *Psychologia*, *5*, 185–193.

Day, R. H. & Wade, N. J. (1969). The reference for visual normalization. *The American Journal of Psychology*, *82*(2), 191–197.

De Baene, W., Premereur, E., & Vogels, R. (2007). Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *Journal of Neurophysiology*, *97*(4), 2900–2916.

De Valois, R. L. & De Valois, K. K. (1980). Spatial Vision. *Annual Review of Psychology*, *31*, 309–341.

Dennett, H. W., Edwards, M., & McKone, E. (2012). Global face distortion aftereffects tap face-specific and shape-generic processes. *Journal of Vision*, *12*(11), 1–20.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, *357*(1), 241–265.

Dhruv, N. T. & Carandini, M. (2014). Cascaded effects of spatial adaptation in the early visual system. *Neuron*, *81*(3), 529–535.

Dickinson, J. E., Almeida, R. A., Bell, J., & Badcock, D. R. (2010). Global shape aftereffects have a local substrate: A tilt aftereffect field. *Journal of Vision*, *10*(13), 1–12.

Dickinson, J. E. & Badcock, D. R. (2013). On the hierarchical inheritance of aftereffects in the visual system. *Frontiers in Psychology*, *4*, 1–15.

Dickinson, J. E., Mighall, H. K., Almeida, R. A., Bell, J., & Badcock, D. R. (2012). Rapidly acquired shape and face aftereffects are retinotopic and local in origin. *Vision Research*, *65*, 1–11.

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, *28*(1), 287–298.

Elliott, S. L., Georgeson, M. A., & Webster, M. A. (2011). Response normalization and blur adaptation: Data and multi-scale model. *Journal of Vision*, *11*(2), 1–18.

Epstein, W. & Park, J. N. (1963). Shape constancy: Functional relationships and theoretical formulations. *Psychological Bulletin*, *60*(3), 265–288.

Ewing, L., Leach, K., Pellicano, E., Jeffery, L., & Rhodes, G. (2013). Reduced face aftereffects in autism are not due to poor attention. *PLoS ONE*, *8*(11), e81353.

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

Farell, B. & Pelli, D. G. (1999). Psychophysical methods, or how to measure a threshold and why. In R. Carpenter & J. Robson (Eds.), *Vision research: a practical guide to laboratory methods*.

Farzmahdi, A., Rajaei, K., Ghodrati, M., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2015). A specialized face-processing network consistent with the representational geometry of monkey face patches. *arXiv.org*.

Fox, C. J. & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, *1127*, 80–89.

Fox, C. J., Oruc, I., & Barton, J. J. S. (2008). It doesn't matter how you feel: The facial identity aftereffect is invariant to changes in facial expression. *Journal of Vision*, *8*(3), 1–13.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*(9), 1187–1196.

Georgopoulos, A., Schwartz, A., & Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416–1419.

Gescheider, G. A., Herman, D. D., & Phillips, J. N. (1970). Criterion shifts in the measurement of tactile masking. *Perception & Psychophysics*, *8*(6), 433–436.

Gheorghiu, E. & Kingdom, F. A. A. (2007). The spatial feature underlying the shape-frequency and shape-amplitude after-effects. *Vision Research*, *47*(6), 834–844.

Gheorghiu, E. & Kingdom, F. A. A. (2008). Spatial properties of curvature-encoding mechanisms revealed through the shape-frequency and shape-amplitude after-effects. *Vision Research*, *48*(9), 1107–1124.

Gheorghiu, E., Kingdom, F. A. A., Bell, J., & Gurnsey, R. (2011). Why do shape aftereffects increase with eccentricity? *Journal of Vision*, *11*(14), 1–21.

Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, *16*(1), 1–31.

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*, 553–569.

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), 453–467.

Giese, M. & Leopold, D. A. (2005). Physiologically inspired neural model for the encoding of face spaces. *Neurocomputing*, *65-66*, 93–101.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Publishing Group*, *14*(7), 926–932.

Goldstein, A. G. & Chance, J. E. (1980). Memory for faces and schema theory. *The Journal of Psychology*, *105*(1), 47–59.

Goldstone, R. L. & Hendrickson, A. T. (2009). Categorical perception. *WIREs Cognitive Science*, *1*(1), 69–78.

Goris, R. L. T., Putzeys, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review*, *120*(3), 472–496.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Peninsula Pub.

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.

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

Guçlu, U. & van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014.

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), 1044–1060.

He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high- and low-Level visual aftereffects by perceptual grouping. *Current Biology*, *22*(11), 1040–1045.

Heinemann, E. G. & Marill, T. (1954). Tilt adaptation and figural after-effects. *Journal of Experimental Psychology*, *48*(6), 468–472.

Held, R. (1963). Localized normalization of tilted lines. *The American Journal of Psychology*, *76*(1), 146–148.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *93*.

Howard, I. P. (1982). *Human visual orientation*. John Wiley & Sons.

Hsu, S. M. & Young, A. (2010). Adaptation effects in facial expression recognition. *Visual Cognition*, *11*(7), 871–899.

Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., & Li, S. Z. (2015). When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition. *arXiv.org*, 1–8.

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*(1), 106–156.

Hurvich, L. M. & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, *64*(6), 383–404.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2009). Does cognitive science need kernels? *Trends in Cognitive Sciences*, *13*(9), 381–388.

Jakel, F. & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naive observers. *Journal of Vision*, *6*(11), 1307–1322.

Janssen, P., Vogels, R., & Orban, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences*, *96*(14), 8217–8222.

Jazayeri, M. & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, *9*(5), 690–696.

Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, *10*(5), 1–19.

Jeffery, L., Rhodes, G., & Busey, T. (2007). Broadly tuned, view-specific coding of face shape: Opposing figural aftereffects can be induced in different views. *Vision Research*, *47*(24), 3070–3077.

Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1824–1840.

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.

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.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron*, *50*(1), 159–172.

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.

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), 1–18.

Kahn, D. A. & Aguirre, G. K. (2012). Confounding of norm-based and adaptation effects in brain responses. *NeuroImage*, *60*(4), 2294–2299.

Kaplan, H. L., Macmillan, N. A., & Creelman, C. D. (1978). Tables of d' for variable-standard discrimination paradigms. *Behavior Research Methods & Instrumentation*, *10*(6), 796–813.

Kaufmann, J. M. & Schweinberger, S. R. (2008). Distortions in the brain? ERP effects of caricaturing familiar and unfamiliar faces. *Brain Research*, *1228*, 177–188.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, *22*(1), 212–224.

Kayaert, G., Op de Beeck, H. P., & Wagemans, J. (2011). Dynamic prototypicality effects in visual search. *Journal of Experimental Psychology: General*, *140*(3), 506–519.

Khaligh-Razavi, S.-M. & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 1–46.

Kingdom, F. A. A. & Prins, N. (2010). *Psychophysics: A practical introduction*. Academic Press.

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision*, *10*(1), 1–13.

Knowlton, B. J. & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*(5140), 1747–1749.

Köhler, W. & Wallach, H. (1944). Figural after-effects: an investigation of visual processes. *Proceedings of the American Philosophical Society*, *88*(4), 269–357.

Kohn, A. & Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nature Neuroscience*, *7*(7), 764–772.

Kompaniez, E., Abbey, C. K., Boone, J. M., & Webster, M. A. (2013). Adaptation Aftereffects in the Perception of Radiological Images. *PLoS ONE*, *8*(10), e76175.

Kompaniez-Dunigan, E., Abbey, C. K., Boone, J. M., & Webster, M. A. (2015). Adaptation and visual search in mammographic images. *Attention, Perception, & Psychophysics*.

Kourtzi, Z. & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *TRENDS in Neurosciences*, *29*(5), 250–256.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kunnapas, T. M. (1955). An analysis of the "vertical-horizontal illusion". *Journal of Experimental Psychology*, *49*(2), 134–140.

Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. *Journal of Neurophysiology*, *69*(1), 19–39.

Lai, M., Oruç, I., & Barton, J. J. S. (2011). Facial age after-effects show partial identity invariance and transfer from hands to faces. *Cortex*, 1–10.

Larsson, J., Landy, M. S., & Heeger, D. J. (2006). Orientation-selective adaptation to first-and second-order patterns in human visual cortex. *Journal of Neurophysiology*, *95*, 862–881.

Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-Based coding of voice identity in human auditory cortex. *Current Biology*, *23*(12), 1075–1080.

Lawson, R. P., Clifford, C. W. G., & Calder, A. J. (2011). A real head turner: Horizontal and vertical head directions are multichannel coded. *Journal of Vision*, *11*(9), 1–17.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575.

Leopold, D. A., O'Toole, A., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience*, *26*(2), 40–46.

Leopold, D. A., Rhodes, G., Muller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 897–904.

Levitt, H. C. C. H. (1970). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467–477.

Lewis, M. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition*, *11*(1), 29–69.

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology*, *90*(1), 204–217.

Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008). Category contingent aftereffects for faces of different races, ages and species. *Cognition*, *106*(3), 1537–1547.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, *8*(10), 1386–1391.

Ma, W. J. & Pouget, A. (2009). Population Codes: theoretic aspects. In L. Squire (Ed.), *Encyclopedia of neuroscience* (Vol. 7, pp. 749–755).

Macke, J. H. & Wichmann, F. A. (2010). Estimating predictive stimulus features from psychophysical data: The decision image technique applied to human faces. *Journal of Vision*, *10*(5), 1–24.

Maclin, O. H. & Webster, M. A. (2001). Influence of adaptation on the perception of distortions in natural images. *Journal of Electronic Imaging*, *10*(1), 100–109.

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

Mather, G. (1980). The movement aftereffect and a distribution-shift model for coding the direction of visual movement. *Perception*, *9*(4), 379–392.

Mathôt, S. & Theeuwes, J. (2013). A reinvestigation of the reference frame of the tilt-adaptation aftereffect. *Scientific Reports*, *3*, 1–7.

McDermott, K. C., Malkoc, G., Mulligan, J. B., & Webster, M. A. (2010). Adaptation and visual salience. *Journal of Vision*, *10*(13), 1–17.

McKone, E., Jeffery, L., Boeing, A., & Clifford, C. W. G. (2014). Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces. *Vision Research*, *108*, 119–132.

Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207.

Miller, E. K., Gochin, P. M., & Gross, C. G. (1991). Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. *Visual Neuroscience*, *7*(4), 357–362.

Mitchell, D. E. & Muir, D. W. (1976). Does the tilt after-effect occur in the oblique meridian? *Vision Research*, *16*(6), 609–613.

Mollon, J. D. (1977). Neural analysis. In K. von Fieandt & I. K. Moustgaard (Eds.), *The perceptual world* (pp. 71–97). Academic Press London.

Morant, R. B. & Harris, J. R. (1965). Two different after-effects of exposure to visual tilts. *The American Journal of Psychology*, 218–226.

Morant, R. B. & Mistovich, M. (1960). Tilt after-effects between the vertical and horizontal axes. *Perceptual and Motor Skills*, *10*(2), 75–81.

Morgan, M. J. (2014). A bias-free measure of retinotopic tilt adaptation. *Journal of Vision*, *14*(1), 1–9.

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience*, *30*(5), 197–206.

Morgan, M., Dillenburger, B., Raphael, S., & Solomon, J. A. (2012). Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception, & Psychophysics*, *74*(1), 185–193.

Mori, K., Nagao, H., & Yoshihara, Y. (1999). The olfactory bulb: Coding and processing of odor molecule information. *Science*, *286*(5440), 711–715.

Motoyoshi, I. (2012). Broad spatial tunings of the object aftereffect: Evidence for global statistical representations of 3D shape and material. *Perception*, *41*, 17–17.

Movshon, J. A. & Blakemore, C. (1973). Orientation specificity and spatial selectivity in human vision. *Perception*, *2*(1), 53–60.

Movshon, J. A., Chambers, B. E. I., & Blakemore, C. (1972). Interocular transfer in normal humans and those who lack stereopsis. *Perception*, *1*, 483–490.

Muir, D. & Over, R. (1970). Tilt aftereffects in central and peripheral vision. *Journal of Experimental Psychology*, *85*(2), 165–170.

Müller, K.-M., Schillinger, F., Do, D. H., & Leopold, D. A. (2009). Dissociable perceptual effects of visual adaptation. *PLoS ONE*, *4*(7), e6183.

Müller, K.-M., Wilke, M., & Leopold, D. A. (2009). Visual adaptation to convexity in macaque area V4. *Neuroscience*, *161*(2), 655–662.

Nachmias, J. (2011). Shape and size discrimination compared. *Vision Research*, *51*(4), 400–407.

Nemes, V. A., Whitaker, D., Heron, J., & McKeefry, D. J. (2011). Multiple spatial frequency channels in human visual perceptual memory. *Vision Research*, *51*(23-24), 2331–2339.

Newell, N. N. & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*(2), 113–143.

Nishimura, M., Maurer, D., Jeffery, L., Pellicano, E., & Rhodes, G. (2008). Fitting the child's mind to the world: Adaptive norm-based coding of facial identity in 8-year-olds. *Developmental Science*, *11*(4), 620–627.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708.

O'Neil, S. F., Mac, A., Rhodes, G., & Webster, M. A. (2014). Adding Years to Your Life (or at Least Looking Like It): A Simple Normalization Underlies Adaptation to Facial Age. *PLoS ONE*, *9*(12), e116105.

O'Neil, S. F., Mac, A., Rhodes, G., & Webster, M. A. (2015). Model fitting versus curve fitting: A model of renormalization provides a better account of age aftereffects than a model of local repulsion. *i-Perception*, *6*(6), 1–7.

Orban, G. A. & Kennedy, H. (1981). The influence of eccentricity on receptive field types and orientation selectivity in areas 17 and 18 of the cat. *Brain Research*, *208*(1), 203–208.

Oruç, I. & Barton, J. J. S. (2011). Adaptation improves discrimination of face identity. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1718), 2591–2597.

Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. *Explorations in Cognition*, 279–307.

Panis, S., Wagemans, J., & Op de Beeck, H. P. (2010). Dynamic norm-based encoding for unfamiliar shapes in human visual cortex. *Journal of Cognitive Neuroscience*, *23*(7), 1829–1843.

Paradiso, M. A., Shimojo, S., & Nakayama, K. (1989). Subjective contours, tilt aftereffects, and visual cortical organization. *Vision Research*, *29*(9), 1205–1213.

Parker, A. (1981). Shifts in perceived periodicity induced by temporal modulation and their influence on the spatial frequency tuning of two aftereffects. *Vision Research*, *21*(12), 1739–1747.

Pasupathy, A. & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338.

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D face model for pose and illumination invariant face recognition, 296–301.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology*, *17*(17), 1508–1512.

Perkins, D. (1975). A definition of caricature and caricature and recognition. *Studies in the Anthropology of Visual Communications*, *2*(1), 1–24.

Phillips, D. P. & Irvine, D. R. (1981). Responses of single neurons in physiologically defined primary auditory cortex (A1) of the cat: frequency tuning and responses to intensity. *Journal of Neurophysiology*, *45*(1), 48–58.

Pimperton, H., Pellicano, E., Jeffery, L., & Rhodes, G. (2009). The role of higher level adaptive coding mechanisms in the development of face recognition. *Journal of Experimental Child Psychology*, *104*(2), 229–238.

Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research*, *34*(12), 1637–1658.

Poirier, F. J. A. M. & Wilson, H. R. (2006). A biologically plausible model of human radial frequency perception. *Vision Research*, *46*(15), 2443–2455.

Pond, S., Kloth, N., McKone, E., Jeffery, L., Irons, J., & Rhodes, G. (2013). Aftereffects support opponent coding of face gender. *Journal of Vision*, *13*(14), 1–19.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*(2), 125–132.

Prentice, W. C. H. & Beardslee, D. C. (1950). Visual "normalization" near the vertical and horizontal. *Journal of Experimental Psychology*, *40*(3), 355–364.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, *25*, 111–164.

Ranganath, C. & Rainer, G. (2003). Cognitive neuroscience: Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, *4*(3), 193–202.

Redding, G. M., Rossetti, Y., & Wallace, B. (2005). Applications of prism adaptation: a tutorial in theory and method. *Neuroscience & Biobehavioral Reviews*, *29*, 431–444.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407.

Regan, D. & Beverley, K. I. (1985). Post-adaptation orientation discrimination. *Journal of the Optical Society of America A*, *2*(2), 147–155.

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

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, *19*(4), 473–497.

Rhodes, G. & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*(18), 2977–2987.

Rhodes, G., Jeffery, L., Boeing, A., & Calder, A. J. (2013). Visual coding of human bodies: Perceptual aftereffects reveal norm-nased, opponent coding of body identity. *Journal of Experimental Psychology: Human Perception and Performance*.

Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Research*, *47*(17), 2291–2296.

Rhodes, G., Jeffery, L., Evangelista, E., Ewing, L., & Peters, M. (2011). Enhanced attention amplifies face adaptation. *Vision Research*, 1811–1819.

Rhodes, G., Jeffery, L., Taylor, L., & Ewing, L. (2013). Autistic traits are linked to reduced adaptive coding of face identity and selectively poorer face recognition in men but not women. *Neuropsychologia*, *51*(13), 2702–2708.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science*, *14*(6), 558–566.

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.

Rhodes, G. & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *The oxford handbook of face perception* (pp. 263–286).

Rhodes, G., Maloney, L. T., Turner, J., & Ewing, L. (2007). Adaptive face coding and discrimination around the average face. *Vision Research*, *47*(7), 974–989.

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C. W. G. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: adaptation and after-effects in high-level vision* (pp. 213–240).

Rieke, F. & Rudd, M. E. (2009). The challenges natural images pose for visual adaptation. *Neuron*, *64*(5), 605–616.

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Riesenhuber, M. & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*(2), 162–168.

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.

Rolls, E. T. (2011). Face neurons. In A. J. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), (pp. 51–75).

Rose, D. & Blakemore, C. (1974). An analysis of orientation selectivity in the cat's visual cortex. *Experimental Brain Research*, *20*(1), 1–17.

Rosenberger, W. F. & Grill, S. E. (1997). A sequential design for psychophysical experiments: an application to estimating timing of sensory events. *Statistics in Medicine*, *16*(19), 2245–2260.

Ross, D. A., Deroche, M., & Palmeri, T. J. (2013). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*, *21*(1), 47–70.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, 1–115.

Rousseau, B. & Ennis, D. M. (2001). A Thurstonian model for the dual pair (4IAX) discrimination method. *Perception & Psychophysics*, *63*(6), 1083–1090.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Bernstein, M. (2014). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 1–42.

Schindel, R. & Arnold, D. H. (2010). Visual sensitivity can scale with illusory size changes. *Current Biology*, *20*(9), 841–844.

Schneider, K. A. & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision*, *8*(15), 1–10.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*(7), 522–535.

Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, *9*(4), 1–19.

Schweinberger, S. R., Zäske, R., Walther, C., Golle, J., Kovács, G., & Wiese, H. (2010). Young without plastic surgery: Perceptual adaptation to the age of female and male faces. *Vision Research*, *50*(23), 2570–2576.

Seriès, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the homunculus "aware" of sensory adaptation? *Neural Computation*, *21*(12), 3271–3304.

Seyama, J. & Nagayama, R. S. (2006). Eye direction aftereffect. *Psychological Research*, *70*(1), 59–67.

Seyama, J. & Nagayama, R. S. (2010). Photorealism aftereffect. *Psychological Research*, *75*(3), 179–187.

Shapley, R. & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 263–346.

Shepherd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology*, *9*(3), 205–212.

Skinner, A. L. & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, *21*(9), 1248–1253.

Skinner, A. L. & Benton, C. P. (2012). The expressions of strangers: Our identity-independent representation of facial expression. *Journal of Vision*, *12*(2), 1–13.

Solomon, S. G. & Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. *Current Biology*.

Spence, M., Storrs, K. R., & Arnold, D. H. (2014). Why the long face? the critical role of vertical configural relations in face 'barcodes' for recognition. *Journal of Vision*, *14*(8), 1–12.

Spiess, A. N. & Neumeyer, N. (2010). An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacology*, *10*(6), 2–11.

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 913–932.

Stevenson, S. B., Cormack, L. K., Schor, C. M., & Tyler, C. W. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research*, *32*(9), 1685–1694.

Storrs, K. R. (2015a). Are high-level aftereffects perceptual? *Frontiers in Psychology*, *6*(157), 1–4.

Storrs, K. R. (2015b). Facial age aftereffects provide some evidence for local repulsion (but none for re-normalisation). *i-Perception*, *6*(2), 100–103.

Storrs, K. R. & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research*, *64*, 7–16.

Storrs, K. R. & Arnold, D. H. (2013). Shape aftereffects reflect shape constancy operations: appearance matters. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 616–622.

Storrs, K. R. & Arnold, D. H. (2015a). Evidence for tilt normalization can be explained by anisotropic orientation sensitivity. *Journal of Vision*, *15*(1), 1–11.

Storrs, K. R. & Arnold, D. H. (2015b). Face aftereffects involve local repulsion, not renormalization. *Journal of Vision*, *15*(8), 1–18.

Storrs, K. R. & Arnold, D. H. (2016). Shape adaptation exaggerates shape differences. *Journal of Experimental Psychology: Human Perception and Performance*, *in press*.

Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50*(3), 300–314.

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

Suzuki, S. (2003). Attentional selection of overlapped shapes: A study using brief shape aftereffects. *Vision Research*, *43*(5), 549–561.

Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: adaptation and after-effects in high-level vision* (pp. 135–172).

Suzuki, S. & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1315–1341.

Suzuki, S. & Rivest, J. (1998). Interactions among "aspect-ratio channels". *Investigative Ophthalmology & Visual Science*, *39*(4), S855.

Swindale, N. V. (1998). Orientation tuning curves: Empirical description and estimation of parameters. *Biological cybernetics*, *78*(1), 45–56.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci. 19*(1), 109–139.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, *13*(1), 90–99.

Templeton, W. B. (1972). Visual tilt normalization: The method of kinesthetic matching. *Perception & Psychophysics*, *12*(5), 422–424.

Templeton, W. B., Howard, I. P., & Easting, G. (1965). Satiation and the tilt after-effect. *The American Journal of Psychology*, 656–659.

Thompson, P. & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11–R14.

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

Tsao, D. Y. & Freiwald, W. A. (2006). What's so special about the average face? *Trends in Cognitive Sciences*, *10*(9), 391–393.

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Attention, Perception, & Psychophysics*, *42*(3), 224–239.

Vaitkevicius, H., Viliunas, V., Bliumas, R., Stanikunas, R., Svegzda, A., Dzekeviciute, A., & Kulikowski, J. J. (2009). Influences of prolonged viewing of tilted lines on perceived line orientation: the normalization and tilt after-effect. *Journal of the Optical Society of America A*, *26*(7), 1553–1563.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204.

Valentine, T., Lewis, M. B., & Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 1–24.

Vernon, M. D. (1934). The perception of inclined lines. *British Journal of Psychology: General Section*, *25*(2), 186–196.

Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, *8*(10), 1401–1410.

von Helmholtz, H. (1909). *Treatise on physiological optics* (J. Southall, Ed.). Courier Corporation.

Wallis, G. (2013). Toward a unified model of face and object recognition in the human visual system. *Frontiers in Psychology*.

Wallis, G., Siebeck, U. E., Swann, K., Blanz, V., & Bulthoff, H. H. (2008). The prototype effect revisited: Evidence for an abstract feature model of face recognition. *Journal of Vision*, *8*(3), 1–15.

Walsh, J. A., Maurer, D., Vida, M. D., Rhodes, G., & Jeffery, L. (2015). Norm-based coding of facial identity in adults with autism spectrum disorder. *Vision Research*.

Walton, B. R. P. (2012). Face distortion aftereffects in personally familiar, famous, and unfamiliar faces. *Frontiers in Psychology*, 1–8.

Ware, C. & Mitchell, D. E. (1973). The spatial selectivity of the tilt aftereffect. *Vision Research*, *14*, 735–737.

Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, *17*(4), 423–429.

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

Webster, M. A. (1996). Human colour perception and its adaptation. *Network: Computation in Neural Systems*, *7*(4), 587–634.

Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, *11*(5), 1–23.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557–561.

Webster, M. A. & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America A*, *25*(11), 2817–2825.

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.

Webster, M. A. & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, *6*(4), 647–653.

Webster, M. A. & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature*, *373*, 694–696.

Webster, M. A., Werner, J. S., & Field, D. J. (2005). Adaptation and the phenomenology of perception. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: adaptation and after-effects in high-level vision*.

Westheimer, G. (1990). Simultaneous orientation contrast for lines in the human fovea. *Vision Research*, *30*(11), 1913–1921.

Westheimer, G. & Gee, A. (2002). Orthogonal adaptation and orientation discrimination. *Vision Research*, *42*, 2339–2343.

Whitaker, D., McGraw, P. V., & Levi, D. M. (1997). The influence of adaptation on perceived visual location. *Vision Research*, *37*(16), 2207–2216.

Wichmann, F. A. & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods & Instrumentation*, *42*(3), 671–684.

Wissig, S. C., Patterson, C. A., & Kohn, A. (2013). Adaptation improves performance on a visual search task. *Journal of Vision*, *13*(2), 1–15.

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.

Yamashita, J. A., Hardy, J. L., De Valois, K. K., & Webster, M. A. (2005). Stimulus selectivity of figural aftereffects for faces. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 420–437.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.

Yarrow, K., Jahn, N., Durant, S., & Arnold, D. H. (2011). Shifts of criteria or neural timing? The assumptions underlying timing perception studies. *Consciousness and Cognition*, 1–16.

Yarrow, K., Sverdrup-Stueland, I., Roseboom, W., & Arnold, D. H. (2013). Sensorimotor temporal recalibration within and across limbs. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1678–1689.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.

Zhao, C., Seriès, P., Hancock, P. J. B., & Bednar, J. A. (2011). Similar neural adaptation mechanisms underlying face gender and tilt aftereffects. *Vision Research*, *51*(18), 2021–2030.

Zhao, L. & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision Research*, *41*(23), 2979–2994.

III

APPENDICES

1

APPENDIX 1: DETAILS OF MODELS PRESENTED IN CHAPTER 6

1.1 ENCODING AND ADAPTATION

1.1.1 *Multichannel model*

Noise-free neural populations were simulated in Matlab corresponding to simple versions of the multichannel and opponent-channel hypotheses. The multichannel model comprised 20 simulated channels, with each channel's response profile described by a Gaussian function:

$$f_i(s) = \frac{G_0}{\sigma\sqrt{2\pi}} e^{\frac{-(s-\mu_i)^2}{2\sigma^2}} \tag{1}$$

The peaks of each Gaussian, μ_i, were uniformly spaced along the interval -500 to $+500$ (arbitrary units). All channels had a standard deviation $\sigma = 60$, and an initial gain $G_0 = \sigma/2.5$, which produced a maximum response of 1 when a channel was presented with a stimulus matching its peak tuning. The value of $f_i(s)$ can therefore be thought of as the normalised average activity elicited in channel i by stimulus s.
After adaptation, each channel followed the same response function, but with its maximum response reduced proportional to its baseline response to the adapted stimulus. The maximum response of each channel post-adaptation, $G_{i(adapt)}$ was determined by the equation (McKone et al., 2014; Seriès et al., 2009; Zhao et al., 2011):

$$G_{i(adapt)} = G_0\left(1 - \alpha e^{\frac{-(\mu_i - s_{adapt})^2}{2\sigma^2_{adapt}}}\right) \tag{2}$$

Here, α determines the strength of the maximum suppression after adaptation (set to 0.6 of the unadapted response), s_{adapt} indicates the stimulus value of the adaptor, and σ_{adapt}

determines how broadly or narrowly adaptation affects channels selective for neighbouring stimulus values (set to 60). Pre- and post-adaptation channel responses are shown in Figure 8a of Chapter 6.

1.1.2 *Opponent-channel model*

The opponent-channel model comprised two monotonic channels intersecting at stimulus value zero, each described by a logistic function:

$$f_i(s) = \frac{1}{1 + e^{ks}} \tag{3}$$

The single free parameter, k, controls the slope of the sigmoid, and was set to 0.01 for the lower channel and -0.01 for the upper channel. After adaptation, each channel was suppressed proportional to its baseline response to the adapting stimulus, so that the adapted response function $f_{i(adapt)}(s)$ is given by:

$$f_{i(adapt)}(s) = f_i(s) \times (1 - \alpha f_i(s_{adapt})) \tag{4}$$

where α was again set to 0.6. Pre- and post-adaptation channel responses are shown in Figure 8d of Chapter 6.

1.2 DECODING

Several methods can be used to estimate, from a pattern of neural activity, which stimulus value was presented (see Ma and Pouget (2009)). Some of these methods ignore the overall shape of each channel's response profile, and simply consider the activity in a channel to be a 'vote' for the stimulus at the peak of the channel's response function, with the strength of activity determining the strength of the vote (in a winner-take-all or weighted-average fashion). Such methods are not ideal for decoding the activity in opponent channels. First, they produce biased estimates when channels do not uniformly tile the stimulus dimension, and second, they are ill-defined for channels which do not have a single peak 'preferred stimulus'. Instead, I opted to use maximum-likelihood estimation (MLE), which estimates the stimulus most likely to have produced a given pattern of activity across channels, based on full knowledge of the response function of each channel (assuming that all stimuli are

equally likely *a priori*; Pouget et al. (2000)). MLE is the optimal decoding method for many encoding schemes, in the senses that it recovers an unbiased estimate of the presented stimulus, and when neural responses are noisy, its estimates have the smallest variance (Pouget et al., 2000).

If the i_{th} channel in a population of n channels has a response $f_i(x)$ in response to some unknown stimulus x, then the log-likelihood of each possible stimulus s is (Jazayeri & Movshon, 2006):

$$log(L(s)) = \sum_{i=1}^{n} f_i(x) log(f_i(s)) - \sum_{i=1}^{n} f_i(s) \tag{5}$$

The best guess as to the presented stimulus, $\hat{x}$, is the stimulus value for which the log-likelihood is highest:

$$\hat{x} = argmax_s(log(L(s))) \tag{6}$$

In both models, it was assumed that aftereffects occur because the decoding mechanism is 'unaware' of the changes in the channels' response functions that occur after adaptation (Seriès et al., 2009). After adaptation, the presented stimulus value was therefore estimated by taking:

$$\hat{x} = argmax_s\left(\sum_{i=1}^{n} f_{i(adapt)}(x) log(f_i(s)) - \sum_{i=1}^{n} f_i(s) \right) \tag{7}$$

where $f_{i(adapt)}$ is the i_{th} channel's post-adaptation response function and f_i is its unadapted response function.

1.2.1 *Multichannel model*

The predicted aftereffect at each test stimulus value was calculated as the decoded value of that stimulus before adaptation, subtracted from its decoded value after adaptation (see Chapter 6, Figure 8b inset). The multichannel model predicts a classic locally-repulsive aftereffect, in which the adapting stimulus is decoded veridically both before and after adaptation, but the values of nearby stimuli are 'repelled' away from the adapted value

after adaptation. The same pattern of aftereffects is produced following adaptation to any value within the encoded range.

1.2.2 *Opponent-channel model*

The opponent-channel model predicts a uniform 'renormalisation,' in which the adaptor appears more neutral after adaptation, and the appearance of all other stimuli is altered in the same direction by the same amount (see Chapter 6, Figure 8e inset). The magnitude and direction of this uniform shift depends on the adapting value, with adaptors further from the norm (zero) producing larger shifts, and adaptation to the norm uniquely producing no aftereffect.

I also considered a ratio-based version of the opponent-channel model, as it has been proposed that 'opponency' may arise because a readout mechanism takes the ratio between an upper and a lower response channel (Robbins et al., 2007; Susilo et al., 2010). For the ratio-based model, the value of the $\frac{upper}{lower}$ response functions was taken, and used as the input to an MLE decoding stage. This produced identical results to performing MLE directly on the channel responses and so is not further considered.

1.3 RELATIVE AFTEREFFECTS BETWEEN TWO DIFFERENTLY-ADAPTED LOCATIONS

To derive predictions for the relative aftereffect induced between two differently-adapted locations, it was assumed that stimuli in each of two locations are encoded by identical, independent populations. This is likely not realistic; receptive field sizes of face-selective cells are highly variable, but some are likely large enough to include both stimulus locations in our experiments (Rolls, 2011). However, it is an acceptable simplification, because adaptation in neurons equally responsive to both adapting locations should affect both test stimuli equally, and so any differences in adapted state between the two locations must be due to neurons solely or disproportionally responsive to one location.

First, the value of each test stimulus was separately decoded in two identical but independent location-specific populations, one before and after adapting to a stimulus value of -200 (arbitrary units) and the other before and after adapting to a value of -50. Then, the relative aftereffect was calculated as the aftereffect in the -200-adapted population minus that in the -50-adapted population, for each possible test stimulus value.

In the multichannel model, the relative aftereffect has an undulating pattern, being greatest for test values midway between the two adapted locations on the stimulus dimension, becoming smaller or opposite in direction at and just beyond the two adaptors, and diminishing to zero far from the adapted values (black curve in Figure 8c of Chapter 6). The magnitude and shape of the relative aftereffect varies depending on the width of the un-

derlying channels, the spread of adaptation through the channels, and the position of the two adaptors, but across a wide range of simulations, these qualitative properties remained true. The opponent-channel model predicted a qualitatively different pattern of results, in which the relative aftereffect is of the same magnitude and direction for all test stimuli (black line in Figure 8f of Chapter 6).

1.4 ROBUSTNESS OF MODEL PREDICTIONS

1.4.1 *Effect of using a weighted-average decoder instead of MLE*

Maximum-likelihood estimation determines the stimulus that was most likely presented, given full knowledge about the unadapted response functions of each channel, and provides an unbiased readout of both models before adaptation. Other decoding methods may predict different post-adaptation perceptual biases. To consider the impact of decoder choice, I repeated the simulations above using a weighted-average decoder, as used in previous models of face adaptation presented in McKone et al. (2014) and Chapter 5 of the present thesis (Georgopoulos et al. (1986); also called a population-vector decoder when applied to circular dimensions).

The weighted average decoder considers activity in each channel as a 'vote' for its preferred stimulus (weighted by the strength of that activity), and so is ill-defined for models in which some or all channels increase monotonically, with no single preferred stimulus. Following McKone et al. (2014), a 'dummy' preference was assigned to each opponent channel, of $\mu = -300$ for the lower channel, and $\mu = +300$ for the upper channel. These values led to the smallest decoding errors over the visualised range (± 300); other values predicted qualitatively similar patterns of aftereffects, but created larger biases in the unadapted readout. For each presented stimulus, x, the decoded value, $\hat{x}$, was given by the weighted 'votes' of all channels:

$$\hat{x} = \frac{\sum_{i=1}^{n} f_i(x)\mu_i}{\sum_{i=1}^{n} f_i(x)} \tag{8}$$

For the multichannel model, this decoding method produces a veridical readout before adaptation (see Figure 1b; biases occur only near the ends of the channels, outside of the visualised region). The pattern of aftereffects after adapting to a single stimulus value (see Figure 1c, red and green curves) and the pattern of relative aftereffects between two

differently-adapted locations (Figure 1c, black curve) are identical to those predicted using MLE decoding (compare to Figure 8c in Chapter 6).

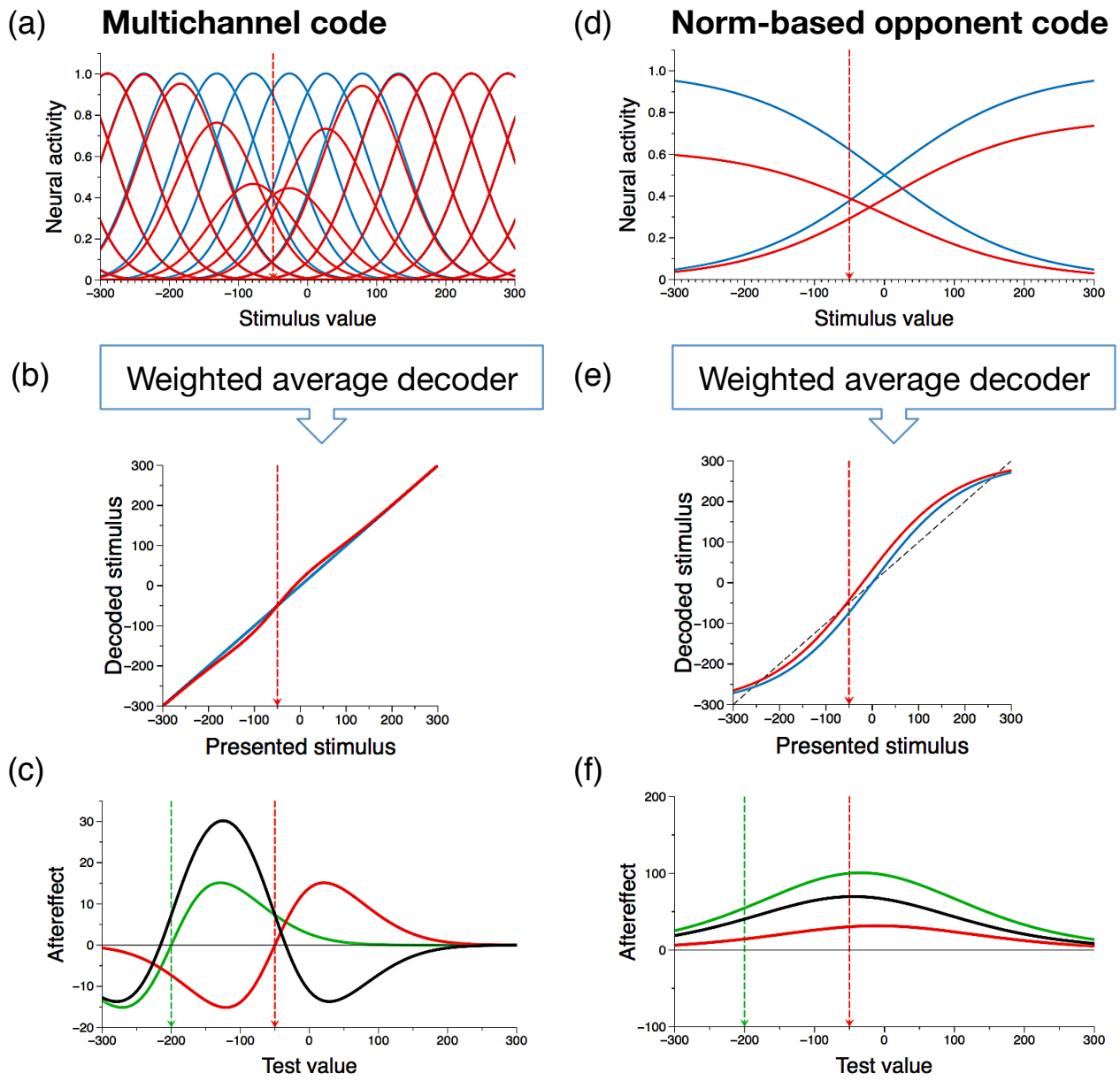


Figure 1. *Adaptation simulations based on weighted-average decoding rather than MLE.* ***(a)*** *Response functions of each channel in a multichannel code, and* ***(d)*** *in an opponent-channel code, before (blue) and after (red) adapting to a stimulus at -50 (arbitrary units).* ***(b)*** *A weighted-average readout of each stimulus before (blue) and after (red) adaptation in a multichannel code, and* ***(e)*** *an opponent-channel code; dashed black line shows the veridical value.* ***(c)*** *Predicted aftereffect (post-adaptation decoded value minus unadapted decoded value) after adapting to -200 (green) and -50 (red) in a multichannel code, and* ***(f)*** *in an opponent code. Black curves show the predicted relative aftereffect (i.e. green curve minus red curve).*

For the opponent-channel model, the weighted average method produces decoding errors — note that in Figure 1e the unadapted readout (blue curve) deviates from the veridical value (dashed black line). Predicted aftereffects are somewhat different from those predicted using an MLE decoder. A weighted-average readout predicts a non-uniform re-

normalisation, in which the renormalisation is greatest around previously neutral stimuli, regardless of the adapted value (Figure 1f, red and green curves). The relative aftereffect follows the same pattern (Figure 1f, black curve). This could in principle produce the undulating relative aftereffect pattern evident in our data, if adaptors were far apart on the encoded dimension, and symmetrically spanned the neutral stimulus. However, we do not think this provides a good account of our results, for two reasons. First, our adapting stimuli lay predominantly to one side of the central stimulus along each continuum, rather than spanning the most identity- or gender-neutral face. Second, in the facial gender experiment, a similar pattern of aftereffects ensued when adaptors and tests were centred either around a male or around a female face, suggesting that aftereffect magnitudes were determined by the position of test faces relative to adapting faces, rather than by their absolute position along the stimulus dimension (this point is further considered in the Discussion of Chapter 6).

1.4.2 *Effect of different adapted locations on the relative aftereffect*

Finally, I explored the robustness of the predicted relative aftereffect pattern for different combinations of adapted values. The relative aftereffect was simulated after adaptation to all combinations of two adaptors selected from the set {-300, -200, -100, 0, +100, +200, +300}. I used MLE decoding, and the parameters described above, for both models. The qualitative pattern of relative aftereffects was identical within each encoding scheme for all combinations of adapted values. In the multichannel code, relative aftereffects were largest for a test point midway between the two adapted values, and in the opponent-channel code, relative aftereffects were of equal magnitude at all test points.

2

APPENDIX 2: RE-ANALYSIS OF DATA IN O'NEIL ET AL. (2014)

A version of this appendix originally appeared as:

Storrs, K. R. (2015b). Facial age aftereffects provide some evidence for local repulsion (but none for re-normalisation). *i-Perception*, *6*(2), 100–103.

2.1 ABSTRACT

Face aftereffects can help adjudicate between theories of how facial attributes are encoded. O'Neil et al. (2014) compared age estimates for faces before and after adapting to young, middle-aged, or old faces. They concluded that age aftereffects are best described as a simple renormalisation — e.g. after adapting to old faces, all faces look younger than they did initially. Here I argue that this conclusion is not substantiated by the reported data. The authors fit only a linear regression model, which captures the predictions of renormalisation, but not alternative hypotheses such as local repulsion away from the adapted age. A second concern is that the authors analysed absolute age estimates after adaptation, as a function of baseline estimates, so goodness-of-fit measures primarily reflect the physical ages of test faces, rather than the impact of adaptation. When data are re-expressed as aftereffects and fit with a nonlinear 'locally repulsive' model, this model performs equal to or better than a linear model in all adaptation conditions. Data in O'Neil et al. (2014) do not provide strong evidence for either renormalisation or local repulsion in facial age aftereffects, but are more consistent with local repulsion (and exemplar-based encoding of facial age), contrary to the original report.

2.2 FACIAL AGE AFTEREFFECTS PROVIDE SOME EVIDENCE FOR LOCAL REPULSION (BUT NONE FOR RENORMALISATION)

The properties of a face can seem to change depending on what other faces you have seen recently. For example, after looking at an old face, participants tend to rate middle-aged faces as being younger than they had initially (Schweinberger et al., 2010). Face aftereffects may result from neural adaptation of channels encoding facial attributes and, if so, studying the patterns of biases induced by adaptation might help reveal the number and selectivity of such channels (Webster & MacLeod, 2011). There are two prominent theories concerning how facial attributes are encoded. According to exemplar-based proposals, multiple channels encode each attribute, with each channel selective for a particular value of that attribute, and no value playing a special role (Lewis, 2004; Valentine, 1991). According to norm-based proposals, channels are selective for the ways in which a face differs from a perceptual 'norm' (e.g. the average face), which is constantly updated according to recent experience (Rhodes et al., 2005; Valentine, 1991). In simulations, these two encoding schemes predict markedly different patterns of adaptation-induced aftereffect (Storrs & Arnold, 2012, 2015b; Webster & MacLeod, 2011). Exemplar-based encoding can predict a locally-repulsive pattern of biases after adaptation (see Figure 1, left), like that found after adapting to simple spatial properties such as orientation and spatial frequency (Blakemore et al., 1970; Mitchell & Muir, 1976; Nemes, Whitaker, Heron, & McKeefry, 2011; Seriès et al., 2009). Norm-based encoding can instead predict renormalisation, involving a uniform bias at all points along the test continuum (see Figure 1, right).

Despite the qualitative differences between local repulsion and renormalisation, it is not clear which best describes face aftereffects. The two proposals can only be clearly dissociated by measuring changes in perception at the adapting face and for more 'extreme' test faces, and this is difficult to do using common methods such as binary classification tasks (see Chapters 5 and 6). In a recent paper, O'Neil et al. (2014) attempted to overcome this problem by measuring aftereffect patterns along a natural facial dimension for which people can easily report the appearance of any test face, using a reasonably precise and reliable numerical value: age. In O'Neil et al. (2014), age estimates for 80 test faces were collected before adaptation, and again after adapting to a sequence of young, middle-aged, or old faces. This provides a rich dataset, from which the perceptual change induced at any point along the age continuum, after each type of adaptation, can in principle be measured. The authors argued that the changes in age ratings after adaptation were best described as a uniform shift in perceived age across all test faces, and therefore that these data support norm-based theories of facial encoding.

However, there are two issues relating to this conclusion. First, and most importantly, the authors presented and analysed fits to their data from a regression model containing only a linear term. This straight-line model well captures the predictions of a uniform renor-

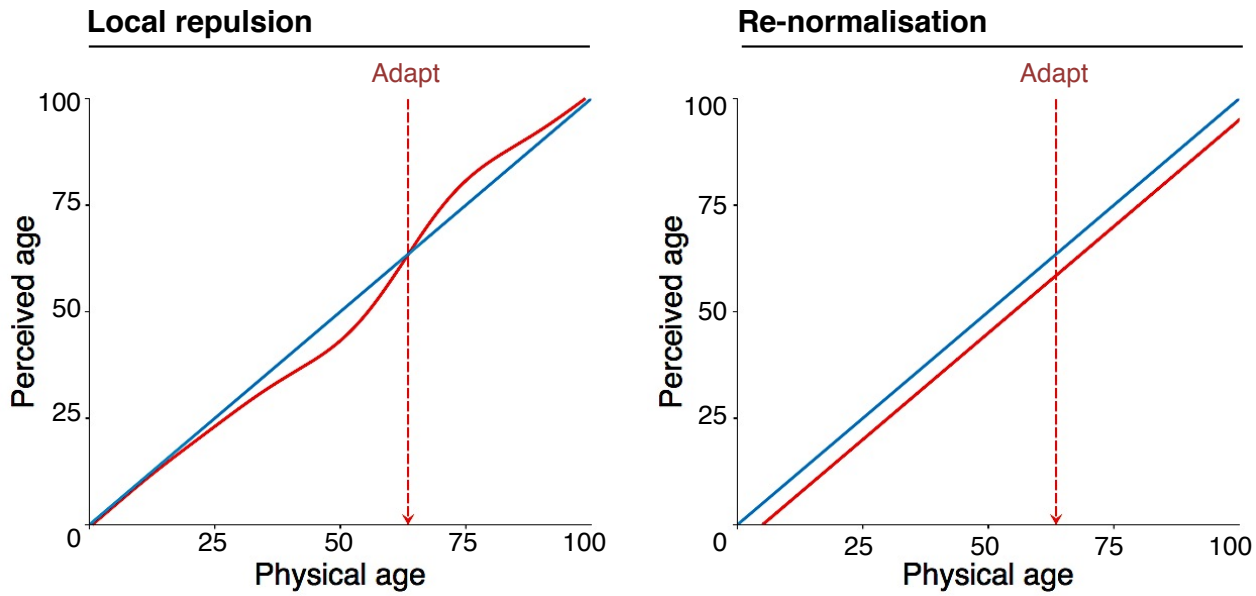


Figure 1. *A schematic illustration of the predicted changes in perceived age after adapting to a moderately old face, if facial age aftereffects follow* ***(left)*** *a locally-repulsive or* ***(right)*** *a renormalising pattern. Before adaptation (blue curves) perceived age may match physical age. After adaptation (red curves), local repulsion predicts that the apparent age of the adapting face will not change, but younger faces will look exaggeratedly young, and older faces will look exaggeratedly old. Renormalisation predicts that the adapting face and all other faces will appear younger by approximately the same amount.*

malisation (a shift in the intercept with no change in slope), but it doesn't well capture the predictions of a local repulsion. The authors proposed that local repulsion should manifest as a change in both the intercept and the slope of the best fitting line after adaptation, relative to baseline. This might reasonably describe shifts predicted by a very broadly-tuned local repulsion, but since the tuning of any hypothetical local repulsion as a function of physical age is unknown, a linear model cannot comprehensively capture the local repulsion hypothesis. The authors note that including higher-order terms in the linear regression model accounted for negligible additional variance. However, models with such terms (e.g. quadratic or cubic functions) also fail to neatly capture the local-repulsion hypothesis. Below I compare the linear fits in the original report to an alternative model — a first derivative of a Gaussian function — which well describes the locally repulsive aftereffect pattern found following adaptation to simpler spatial patterns, such as tilt and spatial frequency (Blakemore et al., 1970; Mitchell & Muir, 1976; Nemes et al., 2011; Seriès et al., 2009).

Second, it is unclear in the original report how well model fits capture *aftereffect* patterns, as the authors performed regression on *absolute* age estimates for each test face after adaptation, as a function of absolute age estimates for each test face before adaptation. Unsurprisingly, post-adaptation age ratings were highly correlated with pre-adaptation ratings. The authors specify that physical age accounts for ∼94% of the variance in all regressions involving post-adaptation age ratings, so only minimal variance was left to be accounted for via the effects of adaptation. The linear model reported in O'Neil et al. (2014) explains ∼98-

99% of the variance in post-adaptation age ratings, but it is important also to know how well the model explains the relatively minor shifts in age ratings induced by adaptation.

Since we are interested in the aftereffect pattern, it is more appropriate to analyse *differences* in estimated age for each test face before vs after adaptation. In Figure 1 I have re-expressed these data (available in the Supplementary data file of O'Neil et al. (2014)) in terms of the shift in average age estimate for each test face, as a function of the pre-adaptation average age estimate for that face. I fit each dataset independently with a linear model (blue lines) and with a nonlinear Gaussian-derivative model (red curves), in Matlab.

Across all three datasets, neither model provided a good fit to aftereffect data (see Figure 2). For aftereffects induced by adapting to Young faces, linear regression $R^2 = 0.008$, explaining less than 1% of the variance in aftereffect data. Root mean squared error (RMSE) of both the linear and nonlinear models[1] = 2.16 years. For Middle-Aged adaptation, model fits are again poor and similar: linear model $R^2 = 0.019$ (RMSE = 2.92 years) and nonlinear model RMSE = 2.86 years. For Old adaptation, linear model $R^2 = 0.005$ (RMSE = 2.82 years) and nonlinear model RMSE = 2.55 years. Since the two models are not nested, it is not appropriate to compare them on the basis of an F-ratio. An alternative metric is the Bayesian Information Criterion (BIC) which provides unbiased estimates of goodness-of-fit for nonlinear models (Spiess & Neumeyer, 2010) and contains a term to penalise more complex models. The lowest BIC value indicates the model that fits the data best with the fewest parameters, and the magnitude of the difference in BIC values between two models indicates the strength of evidence in favour of one over the other. For the Young adaptation data, linear model BIC = 355 and nonlinear model BIC = 357. According to the conventions suggested by Raftery (1995) a difference in BIC of 2 or less should not be considered evidence in favour of either model. For Middle-Aged adaptation, linear model BIC = 403 and nonlinear model = 402 (again a neglible difference). For Old adaptation, linear model BIC = 398, and nonlinear model BIC = 384. A BIC difference of >10 constitutes "very strong" evidence (Raftery, 1995), here in favour of the nonlinear local-repulsion model.

This new analysis makes two points. First and foremost, the shifts in age ratings for each test face are highly variable, and neither model provides a good fit to the data. The best linear fit explains less than 2% of the variance in aftereffect data for each test face. Second, if we were to attempt to compare the two models, two of the three datasets provide no evidence in favour of either model, and in the third, a Gaussian-derivative model (capturing a locally-repulsive pattern of biases) provides a better fit to the aftereffect data than does a linear model (and is judged as superior using a metric which penalises this model for having one more parameter than the linear model).

The source of the variability in the age aftereffect across test faces is unknown. One possibility is that different test faces interacted differently with the series of adapting faces; for example because some test faces share many features with members of the adapting

1 R^2 does not have a clear interpretation for nonlinear models (see Spiess and Neumeyer (2010)).

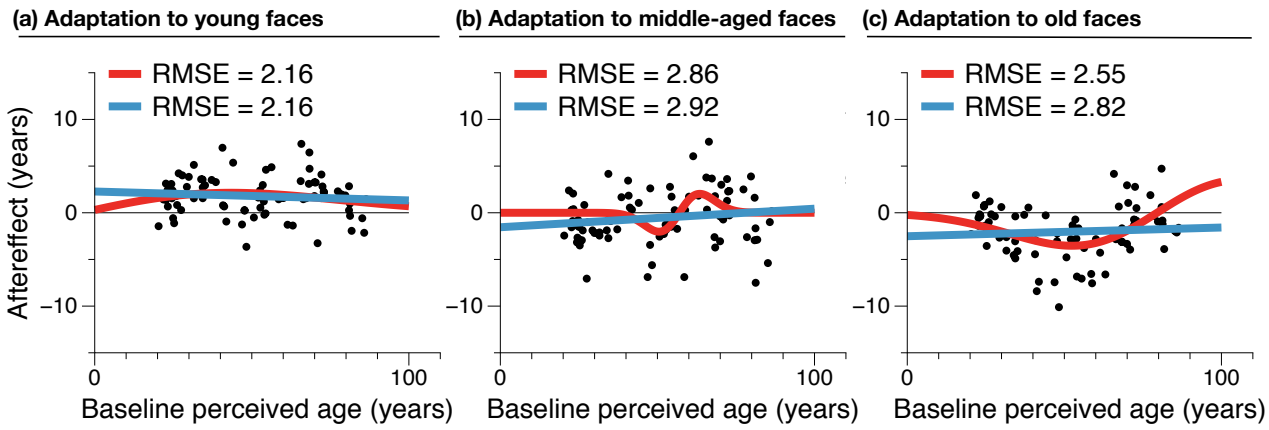


Figure 2. *Average shift in estimated age for each test face, relative to its estimated age at baseline, after* ***(a)*** *adapting to young faces,* ***(b)*** *middle-aged faces, or* ***(c)*** *old faces. Aftereffect data were fitted with a linear model (blue lines) and a first derivative of a Gaussian function (red curves). Root mean squared error (RMSE) is shown for each fit, and additional details are provided in the main text. All graphs and analyses are based on mean age ratings as given in the Supplementary data file of O'Neil, Mac, Rhodes, and Webster (2014).*

series while other test faces share few features with the adaptors. Aftereffects along one facial dimension are known to be largest when test and adapting stimuli are similar along other dimensions (Little, DeBruine, Jones, & Waitt, 2008). Such exemplar-specific effects may arise either from interactions between local properties of the adapting and test images (Dickinson et al., 2010, 2012), or from the overlap between test and adapting faces in higher-level representations.

In conclusion, the pattern of perceptual changes induced by age adaptation is far from clear in these data. Where there is evidence for either hypothesis, they more strongly support the local-repulsion hypothesis (and exemplar-based encoding), than renormalisation (and norm-based encoding), contrary to the title claim of O'Neil et al. (2014).

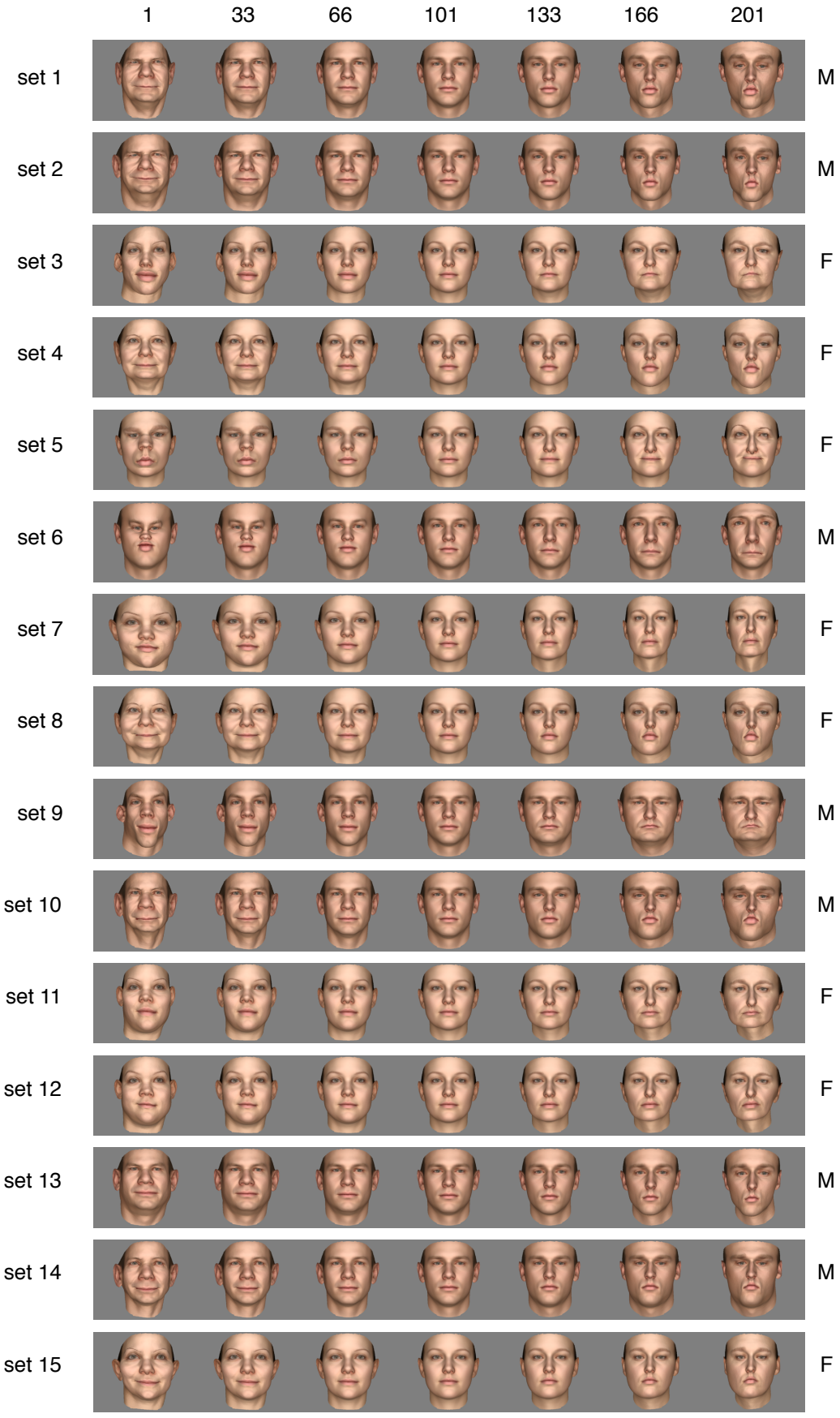
3

APPENDIX 3: SUPPLEMENTARY MATERIAL FOR CHAPTER 6

3.1 SUPPLEMENTARY FIGURE

Figure 1 (following page). *Selected face images from each of the stimulus sets used in the facial identity experiment. Each depicts a continuum between two different same-gender identities; letters to the right of each row indicate whether the face continuum was centred on a male or female average face (shown as face 101). Each of the 15 participants saw a different stimulus set. For all participants, the Central Standard was face image 66.*

Figure 1


1
33
66
101
133
166
201
set 1
M
set 2
M
set 3
F
set 4
F
set 5
F
set 6
M
set 7
F
set 8
F
set 9
M
set 10
M
set 11
F
set 12
F
set 13
M
set 14
M
set 15
F

COLOPHON

This thesis was typset in LaTeX.